Reviewer # 1

We are grateful to the reviewer for the thoughtful comments and suggestions to our manuscript. We have compiled a revised version and in the following provide a point-by-point reply to all issues raised.

Comment # 1.1

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.

Reply:

We would like to thank the reviewer for their positive and constructive comments. The specific comments related to the PIES scheme are addressed below in response to Comment #1.5.

Comment # 1.2

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.*

Reply:

We are indeed aware of this emerging field of ensemble Kalman inversion and are excited to see a broader unification of data assimilation and inverse modeling techniques under the common umbrella of Bayesian inference. In particular, we are adopting a broad definition of the term data assimilation from the Bayesian perspective outlined in (1). Under this definition, data assimilation is also concerned with parameter estimation in batch smoothing problems that arise in areas as diverse as reservoir history matching (2), snow reanalysis (3), and the present study. As the reviewer points out, this problem turns out to be identical to the Bayesian inverse problem formulation outlined by (4) that we cite in our manuscript. To make this connection clearer we have now also included more of the related studies (such as (5; 6)), as well as the study by Iglesias and Yang (2021) (7), to the reference list and moved these to earlier portions of the text. In particular, our broad definition of DA and the overlap with Bayesian inverse problems is now explicitly discussed towards the end of the Introduction and the link to ensemble Kalman inversion has been added to the section introducing the ES-MDA in Section 2.4:

Changes:

2.1.4 Introduction

. . .

This view implies that a mathematically optimal technique for consistent data-model fusion can be formulated as a kind of Bayesian inference problem ~~(8; 9)~~(8; 10; 9; 11) , which is typically referred to as data assimilation (DA) or inverse modeling in the geosciences (12; 1). Herein, we adopt a broad Bayesian definition of the field of DA in line with (13). In addition to the conventional DA problem of state estimation, this definition also encompasses the problem of parameter estimation. The latter is often referred to as an inverse problem (4) rather than a DA problem. Since the flux estimation problem at hand is precisely such a parameter estimation or inverse problem we are also leaning on developments in this field (5; 6). In this study, we do not make any distinction between DA and inversion and take a unifying approach through the lens of Bayesian inference following (9). Such a unified view is especially helpful as the methods used herein can be applied in a hierarchical framework that jointly solve both state and parameter estimation problems (14).

. . .

2.1.4 Data assimilation schemes

...

At the root of these iterative schemes we find the idea of tempered transitions, which is a technique that is widely used in challenging Bayesian inference ~~problems (15; 16)~~ tasks (15; 16; 7). This tempering, in combination with their derivative-free implementation, has placed iterative ensemble Kalman methods at the frontier of ongoing research in Bayesian inverse problems (4; 5; 6) which is helping to both formalize, improve, and generalize this family of methods (17; 7; 18; 19). The equations and workflow for the ES-MDA scheme used herein are presented in Appendix B.

COMMENT # 1.3

It would be illustrative to unwrap the classic equation (1). For example the authors state L116 "$\mathsf{G}(\cdot)$ 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 "$\mathsf{H} \circ \mathsf{F} \circ \mathsf{T}$= Observation_operator∘LES∘Transform" in (1) would illuminate this.

Reply:

We have followed the reviewer's suggestion of unwrapping Equation (1) to make the various operations more explicit as outlined below.

Changes:

2.1 Data assimilation framework

...

The aim here is to infer surface fluxes of sensible and latent heat using sparse and uncertain drone measurements of meteorological variables in the atmospheric boundary layer. Solving this inverse problem requires a forward (or data generating) model that maps the parameters, namely the surface fluxes of interest and other uncertain boundary conditions, to the drone observations through

$$\mathbf{y} = \mathcal{G}(\mathbf{x}) + \boldsymbol{\epsilon}\,, \tag{1}$$

where $\mathbf{y} \in \mathbb{R}^d$ is the observation vector, $\mathcal{G}(\cdot)$ is the forward model~~(e.g. RANS or LES)~~, $\mathbf{x} \in \mathbb{R}^m$ is the target parameter vector, and $\boldsymbol{\epsilon} \in \mathbb{R}^d$ is the observation error. ~~For~~ In practice, $\mathcal{G}(\cdot)$ is a composition of multiple operations (c.f. 1)

$$\mathcal{G}(\mathbf{x}) = \mathcal{H}\left(\mathcal{M}\left(\mathcal{T}(\mathbf{x})\right)\right)\,. \tag{2}$$

The inner operation, $\mathcal{T}(\cdot)$, is a transformation step that maps the parameters from an unbounded space to a bounded physical space. This step helps satisfy the Gaussian assumption of the ensemble Kalman methods while avoiding unphysical values (Section 2.1.2), although it adds an extra layer of non-linearity to the forward model~~, we employ~~. The subsequent middle operation, $\mathcal{M}(\cdot)$, is the dynamical model used to simulate the state of the boundary layer given the boundary conditions specified by the parameters. The outer operation, $\mathcal{H}(\cdot)$, is the observation operator that maps the states of the model to the corresponding predicted observations by extracting the flight paths of drones and (when necessary) performing temporal aggregation (see Section 2.1.3). By employing a turbulence-resolving LES ~~that is~~ as opposed to a RANS model for the dynamics $\mathcal{M}(\cdot)$ in our forward model $\mathcal{G}(\cdot)$, we are able to generate ~~this mapping if~~ the surface flux to drone observation mapping since the LES is run at an appropriate level of spatio-temporal detail. ~~Note that, even~~

Even in the absence of observation error,...

Comment # 1.4

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)?

Reply:

We have followed the reviewer's suggestion and presented clear equations for the observational error covariance matrices, including how we scaled these to account for the number of samples S (what the reviewer calls T) in the averaging operations (factor $1/\sqrt{S}$ for error standard deviations, $1/S$ for error variances) and local mean gradients (introducing a factor $\sqrt{2}$ for error standard deviations, so 2 for error variances). The use of Gaussian observation errors in this study is an assumption and, given CLT, it is more likely to hold for longer averaging periods (i.e. larger S). We did not test the appropriateness of the CLT approximation for shorter timescales per se, instead our experimental design was based on our prior expectations of the turbulence spectra and practical constraints related to the drones (battery time etc...). This could be an important topic to explore further in future work. We nonetheless added changes to Section 2.1.3 to clarify some of these concerns as outlined below

Changes:

2.1.3 Drone measurements, observations and errors

...

Systematic errors that occur for error distributions that are asymmetrically distributed with respect to zero, are assumed to be negligible. This leads to the following definition for the diagonal observation error covariance matrix $\mathbf{R} \in \mathbb{R}^{d \times d}$ employed in this study

$$\mathbf{R} = \operatorname{diag}\left(\boldsymbol{\tau} \odot \boldsymbol{\sigma}^2\right), \tag{5}$$

where $\operatorname{diag}(\cdot)$ is the diagonal operator that converts a vector to a diagonal matrix, $\boldsymbol{\tau} \in \mathbb{R}^d$ is a scaling vector, $\odot$ denotes the element-wise product, $\boldsymbol{\sigma} \in \mathbb{R}^d$ contains the measurement error standard deviation for each observation. The elements of the scaling vector are defined as follows

$$\tau_i = \begin{cases} 1/S & \text{if mean,} \\ 2/S & \text{if local mean gradient,} \end{cases} \tag{6}$$

where S is the number of measurement samples that are averaged to form an observation. As elaborated in Section 2.2, we test two types of flight plans. The first type involves step-wise vertical profiles while the drones hover in place for a 2 minute averaging period with a 10 s sampling interval such that $S = 12$. The second type involves random exploration where no averaging is performed such that $S = 1$. In summary, following independent Gaussian error propagation, this observation error covariance matrix implies that observation errors are uncorrelated, decrease with number of samples S in an averaging period, and are larger for local mean gradients than for means.

The elements of σ are determined by the measurement error standard deviation of the respective sensors. For temperature measurements on drones,...

COMMENT # 1.5

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)".

Reply:

The PIES scheme did not turn out to be particularly effective in this study and suffered from the same degeneracy as the simpler PBS scheme. We suspect that this could be improved by running more iterations of the ES-MDA and using the final (rather than pen-ultimate) iteration as the posterior, but that would come at a considerable increase in computational cost. Following the reviewer's suggestion, we have thus removed the discussion about the KLD of the PIES scheme on L413. We have nonetheless retained the discussion of the RMSE of the particle methods since this is a point metric based on the ensemble mean (rather than the entire ensemble) and we have in the same section (Section 3.2) made it clear that these schemes were degenerate with effective sample sizes of 1. A discussion of the motivation for testing the PIES method in this paper has also been added. It is worth emphasizing that all the non-iterative schemes (ES, PBS, PIES) can essentially be run for free (other than the cost of the assimilation step) while running the ES-MDA iterations. In particular, they do not require any additional LES runs. As such, including these schemes as tests or benchmarks for the performance of the ES-MDA does not add any noticeable computational burden to our experiments.

Changes:

2.1.4 Data assimilation schemes

...

The motivation for pursuing the PIES scheme is that the ES-MDA produces a biased approximation of the posterior for non-linear forward models (16). Although this bias is typically less severe than that of non-iterative ensemble Kalman methods (2), it would nonetheless be advantageous to find efficient methods to reduce it. PIES is a straightforward translation of the scheme of (20) to iterative ensemble smoothers such as the ES-MDA. As such, PIES can be viewed as a simple extension of the ES-MDA that does not necessarily impose any noticeable computational burden and might improve performance. As with all particle methods, the effective sample size can be used to diagnose degeneracy in the ensemble of particles (21). A low ($\ll N_e$) effective sample size indicates degeneracy due to the fact that the proposal is too far from the target posterior. ...

3.1 Synthetic experiments

...

The ES-MDA ~~and PIES schemes give~~ scheme gives the largest information gain from the prior to posterior, as indicated by ~~their~~ its KLD.

COMMENT # 1.6

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.

Reply:

We agree that structural model errors should be clarified when moving the algorithm from synthetic to field data, so we suggest to add the text below to Section 2.1. Overall, we believe that the main structural errors for field experiments are due to topography and spatio-temporal flux variability. Assuming flat terrain, as well as homogeneous and stationary surface fluxes are simplifications of reality, which should be improved in future studies.

As requested by the reviewer, Figure R1 below shows an example of the marginal distributions for a field data experiment (the same flight as used in Figure 5 of the main article). Note the important difference to the synthetic experiments that true parameter values are not known (vertical dashed lines for H and LE only show independent EC estimates). We do not see indications of significant structural model errors from these distributions, indicating that the choice of LES model and observation operator give an appropriate representation of reality.

Changes:

2.1 Data assimilation framework

...

In the real experiments, where we compare with independent EC data, some of the mismatch between the EC estimates and drone-based inferences will undoubtedly be due to the strong assumptions made in the respective approaches. Given the level of realism in LES, these structural model errors introduced when moving the algorithm application from synthetic to field data are likely dominated by simplifications of topography and spatio-temporal flux variability. The Bayesian approach to inference

...

COMMENT # 1.7

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?

Reply:

As the two methods estimate the surface fluxes over slightly different footprint areas, we wish to see that they agree in the estimated order of magnitude of fluxes and the relative flux variability. We agree with the reviewer that this understanding of "plausibility check" should be clarified.

Changes:

Introduction

...

To be clear, given the differences in footprint and underlying assumptions, we do not argue that this comparison offers a validation per se – only a plausibility check of the estimated order of magnitude of fluxes and their relative variability.

COMMENT # 1.8

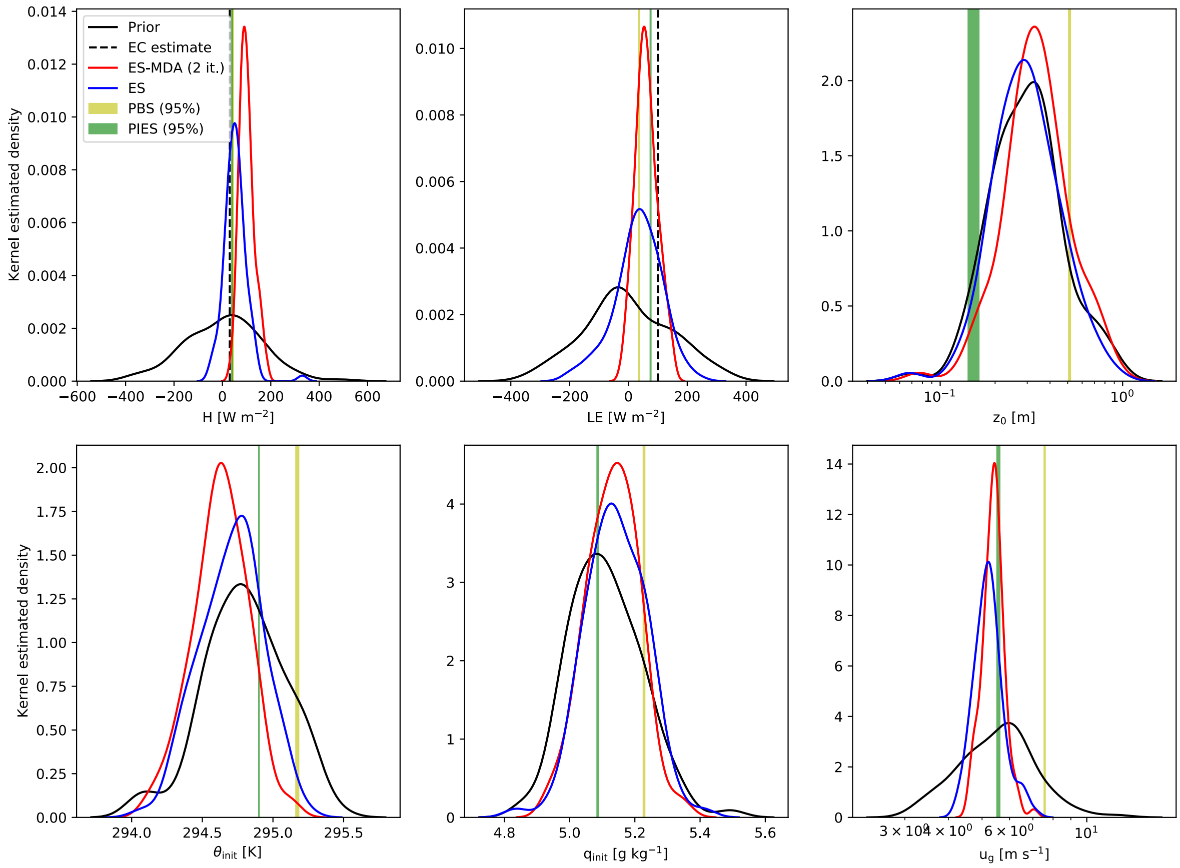

Figure R1: Marginal parameter distributions for the prior (black) as well as the ES-MDA (with $N_a = 2$ iterations, red), ES (blue), PBS (yellow shading shows the central 95^{th}-percentile range), and PIES (green shading shows the central 95^{th}-percentile range) posterior estimates along with the location of the EC flux estimates (black dashed vertical line) for flight 1 of the Iškoras campaign, a step profile on 2020-07-27 with takeoff at 15:20 local time.

Throughout, single or double quotes appear backward before quotations, typically from using character ' and not ' in latex

Reply:

Fixed, thanks for spotting this.

COMMENT # 1.9

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

Reply:

Indeed, our interest is in H and LE, but the other parameters are learned from the data and then 'integrated out' by focusing on the marginal distribution of of H and LE. We were of the impression that our use of the word nuisance was in line with the norm in Bayesian statistics following (10) and (22) as well as (11) which is a standard reference in the field. This procedure may not be the norm in conventional state estimation-based DA, but we have now made it clear (see response to Comment # 1.2) that we are adopting a broader Bayesian definition of the term that includes parameter estimation (and marginalization). We have clarified this understanding of 'nuisance' parameters in the following change of the manuscript.

Changes:

2.1.1 LES model and parameters

...

Of these six parameters, the primary interest is in H and LE while the remaining four parameters can be regarded as ~~'~~'nuisance' parameters ~~(10)~~(22; 10; 11). The nuisance parameters are still inferred from the data, but are then implicitly 'integrated out' as we primarily focus on the marginal posterior distributions of H and LE....

COMMENT # 1.10

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

Reply:

We agree that this change of order can clarify the manuscript and are happy to follow the reviewer's suggestion. To emphasize that the table presents average statistics over a number of synthetic experiments, we also suggest to add the following sentence to the paragraph describing the table.

Changes:

3.1 Synthetic experiments

...

Varying the sampling strategies (step profile vs random exploration), flight time (12 ~~to~~ vs 24 minutes), number of drones (1 vs 5), uncertainty in initial conditions (narrow vs broad), and the geostrophic wind speed (1.5 vs 6.0 m s^{-1}) led to a total of 16 synthetic experiments.

COMMENT # 1.11

Figure 4 mention the spread of the violin plots (95%) in the caption

Reply:

The violins are plotted with Matplotlib Violinplot, where the caps mark the extrema of the ensemble. We propose to add the following sentence to the caption of Figure 4 to clarify this.

Changes:

Figure 4 caption:
The caps of the violins mark the extrema of the ensemble and the dots the mean values.

COMMENT # 1.12

L490 typo "constrains"

Reply:

Fixed, thanks for spotting this typo.

COMMENT # 1.13

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").

Reply:

We have now referred to the appropriate part of the manuscript, namely Section 4.2. Moreover, we have added a sentence to provide an example of what kind of external information we are referring to.

Changes:

4.2 Possible improvements

...

A complementary approach could also be to directly incorporate land cover information, e.g. from satellite retrievals (23), into the design of flux maps in the turbulence simulation as was done in (24).

COMMENT # 1.14

Were more than two inflation steps tried with ES-MDA?

Reply:

In the pilot phase of this study, we did try a few experiment with more iterations, but quickly realized, that given our allocation of computational resources we could not afford to test this systematically across many different experiments in our study. We chose to prioritize more experiments (including field data) with fewer iterations. So we can only hypothesize that more iteration in the ES-MDA scheme would give an improved performance (which should be addressed in future studies). In this context, we should also explore reducing the ensemble size and increasing the number iterations (while keeping the number of simulations fixed), to identify the optimal ratio of this computational trade-off. The paragraph we added to Section 4.3 (in relation to Comment #2.2 by Reviewer 2) summarizes these considerations.

COMMENT # 1.15

Figure 5 - mention these are posterior draws of ES-MDA on the caption.

Reply:

We have followed the reviewer's suggestion as shown below.

Changes:

Figure 5 caption:

Drone observations and posterior ensemble predictions from the ES-MDA for flight 1 of the Iškoras campaign, a step profile on 2020-07-27 with takeoff at 15:20 local time. The upper panels show the successive 2-min mean values, whereas the lower panels show the local mean gradients. The line colors of the vertical profiles for the ~~100~~ $N_e = 100$ posterior ensemble members from the ES-MDA correspond to their log-likelihood with more likely values in yellow and less likely values in blue. The prior predictions are not shown, because their range is so wide that one could not see any details in the posterior profiles.

COMMENT # 1.16

L455 "...compared to the less calibrated uncertainty estimates of the EC technique" - I'm not sure what this means here. Please rephrase

Reply:

We agree that this can be clarified and suggest the following changes.

Changes:

2.3 Field experiments

...

Along with the EC fluxes, EddyPro also ~~estimate~~ estimates their random error ~~based on integral turbulence statistics~~(the variance of the flux covariance) due to sampling errors that arise from the small number of large eddies that dominate the flux during typical sampling periods following (25).

4.1 Potential and limitations of drone data assimilation

Using typical sensor configuration and flight times of small drones, we find a relatively large posterior spread of the surface fluxes, compared to the ~~less calibrated~~ uncertainty estimates of the EC technique that are solely accounting for sampling errors arising from the small number of large eddies captured in the 30-min flux interval...

COMMENT # 1.17

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?

Reply:

We thank the reviewer for this interesting observation. Drone-DA uncertainty estimates are largely a result of our experimental design (flight time, sensor noise, etc.) and the prior distributions we used. These settings were kept constant in our field experiments shown in Figure 6, which would explain why the uncertainty estimates are largely constant. For EC, the used method by (25) to estimate the relative error of the flux through the variance of the flux covariance, i.e. the sampling uncertainty. As shown in the original paper of the method (25), the relative error is largely constant (around 10-30%) over a range of flux magnitudes, wind speeds, and even ecosystem types (i.e. forested vs agricultural surfaces). It is therefore indeed expected that the absolute random error increases linearly with flux magnitude, as also seen in Figure 6 of our study. In sum, the error bars shown in Figure 6 measure fundamentally different uncertainties (epistemic for the drone DA, vs aleatoric for EC). We propose to add the following sentence describing this observation in the discussion.

Changes:

3.2 Field experiments

As the drone-DA uncertainty estimates are largely a result of our experimental design (flight time, sensor noise, etc.) and the used prior distributions, all 18 flights show largely the same epistemic uncertainty. For the EC estimates, error bars in Figure 6 only indicate the absolute aleatoric uncertainty due to sampling limitations, which is expected to increase with flux magnitude (25).

COMMENT # 1.18

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.

Reply:

Done.

Changes:

2.1.4 Data assimilation scheme

...

~~In practice, it may~~ Importance sampling is more effective the closer the proposal is to the target posterior distribution (8). So in theory it would be better to use the posterior estimate from the final (rather than penultimate) iteration of the ES-MDA for the proposal in PIES, but this would come at a high computational cost of requiring an additional round of runs of the LES ensemble.

COMMENT # 1.19

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.

Reply:

These are nice suggestions. We propose to incorporate them with the following changes in the sentence.

Changes:

4.2 Possible improvements
~~Narrowing the posterior spread~~ Improved surface flux inferences can be achieved in a number of different ways, including technical improvements during data collection, such as using higher quality sensors, more drones, and better quantifying initial and boundary conditions, as well as modifying the data assimilation framework by using a larger ensemble size to improve the Monte Carlo approximation, more assimilation cycles ~~, emulators~~ to better account for nonlinearity, emulators to increase the smoothness of the likelihood function (18), and a higher spatial resolution of the LES model to reduce structural model errors.

COMMENT # 1.20

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?

Reply:

The legal framework for drone applications is very country-specific and cannot be readily forecast. Typically, the legislation allows for the common use cases for drones (below 120 m above ground level and within visual line of sight) and describes possibilities to acquire permits for more advanced use cases. We understand that a reference to some legal guidelines is needed to clarify this aspect, so we propose to add a reference to an article focusing on a discussion of the European airspace regulations.

Changes:

4.1 Potential and limitations of drone data assimilation
...
Most applications of drones are currently still restricted to manual flights with a hu-

man pilot in charge of the system (see, e.g., (26) for a discussion of European airspace regulations).

References

[1] G. Evensen, F. C. Vossepoel, and P. J. van Leeuwen, *Data Assimilation Fundamentals: A Unified Formulation of the State and Parameter Estimation Problem*. Springer Textbooks in Earth Sciences, Geography and Environment, Cham: Springer International Publishing, 2022.

[2] A. A. Emerick and A. C. Reynolds, "Ensemble smoother with multiple data assimilation," *Computers & Geosciences*, vol. 55, pp. 3–15, June 2013.

[3] K. Aalstad, S. Westermann, T. V. Schuler, J. Boike, and L. Bertino, "Ensemble-based assimilation of fractional snow-covered area satellite retrievals to estimate the snow distribution at Arctic sites," *The Cryosphere*, vol. 12, p. 247–270, 2018.

[4] A. M. Stuart, "Inverse problems: A Bayesian perspective," *Acta Numerica*, vol. 19, pp. 451–559, May 2010.

[5] M. A. Iglesias, J. H. Law, and A. M. Stuart, "Ensemble Kalman methods for inverse problems," *Inverse Problems*, vol. 29, no. 4, p. 045001, 2013.

[6] C. Schillings and A. M. Stuart, "Analysis of the Ensemble Kalman Filter for Inverse Problems," *SIAM Journal on Numerical Analysis*, vol. 55, no. 3, pp. 1264–1290, 2017.

[7] M. Iglesias and Y. Yang, "Adaptive regularisation for ensemble kalman inversion," *Inverse Problems*, vol. 37, no. 2, p. 025008, 2021.

[8] D. J. C. MacKay, *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 2003.

[9] S. Särkkä, *Bayesian Filtering and Smoothing*. Cambridge University Press, 2013.

[10] E. Jaynes, *Probability Theory: The Logic of Science*. Cambridge University Press, 2003. doi:10.1017/CBO9780511790423.

[11] A. Gelman, J. Carlin, H. Stern, D. Dunson, A. Vehtari, and D. Rubin, *Bayesian Data Analysis*. Chapman and Hall/CRC, 3 ed., 2013.

[12] A. Carrassi, M. Bocquet, L. Bertino, and G. Evensen, "Data assimilation in the geosciences: An overview of methods, issues, and perspectives," *WIREs Climate Change*, vol. 9, Sept. 2018.

[13] G. Evensen, F. C. Vossepoel, and P. J. van Leeuwen, *Data Assimilation Fundamentals*. Springer, 2022.

[14] M. Katzfuss, R. S. Stroud, and C. K. Wikle, "Ensemble Kalman Methods for High-Dimensional Hierarchical Dynamic Space-Time Models," *Journal of the American Statistical Association*, vol. 115, no. 530, pp. 866–885, 2020.

[15] R. M. Neal, "Sampling from multimodal distributions using tempered transitions," *Statistics and Computing*, vol. 6, pp. 353–366, Dec. 1996.

[16] A. S. Stordal and A. H. Elsheikh, "Iterative ensemble smoothers in the annealed importance sampling framework," *Advances in Water Resources*, vol. 86, pp. 231–239, Dec. 2015.

[17] A. Garbuno-Inigo, F. Hoffmann, W. Li, and A. M. Stuart, "Interacting langevin diffusions: Gradient structure and ensemble kalman sampler," *SIAM Journal on Applied Dynamical Systems*, vol. 19, pp. 412–441, Jan. 2020.

[18] E. Cleary, A. Garbuno-Inigo, S. Lan, T. Schneider, and A. M. Stuart, "Calibrate, emulate, sample," *Journal of Computational Physics*, vol. 424, p. 109716, Jan. 2021.

[19] O. Dunbar, A. Duncan, A. Stuart, and M.-T. Wolfram, "Ensemble Inference Methods for Models With Noisy and Expensive Likelihoods," *SIAM Journal on Applied Dynamical Systems*, vol. 21, no. 2, pp. 1539–1572, 2022.

[20] N. Papadakis, E. Mémin, A. Cuzol, and N. Gengembre, "Data assimilation with the weighted ensemble Kalman filter," *Tellus A: Dynamic Meteorology and Oceanography*, vol. 62, pp. 673–697, Jan. 2010.

[21] N. Chopin and O. Papaspiliopoulos, *An Introduction to Sequential Monte Carlo*. Springer, 2020.

[22] G. Bretthorst, *Bayesian Spectrum Analysis and Parameter Estimation*. Springer, 1988.

[23] K. Aalstad, S. Westermann, and L. Bertino, "Evaluating satellite retrieved fractional snow-covered area at a high-Arctic site using terrestrial photography," *Remote Sensing of Environment*, vol. 239, p. 111618, 2020.

[24] L. D. van der Valk, A. J. Teuling, L. Girod, N. Pirk, R. Stoffer, and C. C. van Heerwaarden, "Understanding wind-driven melt of patchy snow cover," *The Cryosphere*, 2022.

[25] P. L. Finkelstein and P. F. Sims, "Sampling error in eddy correlation flux measurements," *Journal of Geophysical Research: Atmospheres*, vol. 106, pp. 3503–3509, Feb. 2001.

[26] E. Bassi, "From Here to 2023: Civil Drones Operations and the Setting of New Legal Rules for the European Single Sky," *Journal of Intelligent & Robotic Systems*, vol. 100, pp. 493–503, Nov. 2020.

[27] R. T. Palomaki, N. T. Rose, M. van den Bossche, T. J. Sherman, and S. F. J. De Wekker, "Wind Estimation in the Lower Atmosphere Using Multirotor Aircraft," *Journal of Atmospheric and Oceanic Technology*, vol. 34, pp. 1183–1191, May 2017.