Re: egusphere-2026-615 (Investigating Information Transfer in CO_2 Flux Inversions: An Analysis of Ensemble Kalman Filter Based on Monte Carlo Simulations).

Dear editors,

We are grateful to the editors and the reviewers for the comments and suggestions. Following is a point-by-point response to the reviewer's comments. Texts in ***Italic Bold*** are the reviewers' comments, and those in normal black are our responses. The blue texts are revised sentences in the revised manuscript. All the line numbers in blue are referred to the change tracking version. We hope that you and the referees will find the changes satisfactory and we are looking forward to hearing from you soon.

Responds to the reviewers' comments

Reviewer #1:

The study addresses a critical knowledge gap in top-down atmospheric inversions: the internal "information transfer" mechanism by which pointwise observations are translated into flux estimates. This paper introduces a paradigm shift in understanding atmospheric CO2 inversions by moving beyond the traditional view of the prior covariance matrix B as a static statistical assumption. While the community has long recognized that B is important, this study is the first to mechanistically dissect how B fundamentally governs the spatial resolution, detection sensitivity, and cross-contamination risks of the entire inversion system, using a perturbation-response strategy with high-resolution 500-member simulations in the Ensemble Kalman Filter (EnKF) framework. It explained how correlated and uncorrelated flux components, respectively, amplify or suppress observational influence. The paper is generally well structured and clearly written. I recommend publication after minor revisions addressing the points below.

1. The study relies on a 500-member ensemble to minimize noise in remote areas. Is 500 members the "convergence point" where spurious correlations become negligible for the 27km resolution used?.

<u>Response:</u>

The choice of a 500-member ensemble is not intended to represent a strict "convergence point" at which spurious correlations vanish, but rather a practical compromise that ensures sufficiently low sampling noise for the purpose of diagnosing correlation structures. Our choice is based on two considerations. First, ensemble sizes of similar magnitude have been adopted in previous studies for related diagnostic analyses (e.g., Super et al., 2020). Second, while larger ensembles would further reduce sampling noise, they would also lead to substantially increased computational cost, particularly given the high-resolution (27 km) transport simulations used here.

To demonstrate that the 500-member ensemble is adequate for our analysis, we provide additional sensitivity tests (Figs. R1 and R2). Figure R1 shows the spatial patterns of concentration–flux correlations at the WLG site for different ensemble sizes. When the ensemble size is small (e.g., 20 members), strong spurious correlations appear in remote regions and obscure the physically meaningful correlation patterns near the observation site. As the ensemble size increases, these spurious correlations are progressively reduced, and the underlying spatial structure becomes clearer.

Figure R2 provides a more quantitative assessment by showing the correlation coefficients at grid cells located ~2000 km away from the site, where true correlations are expected to be weak. The results show that the magnitude and variability of these remote correlations decrease with increasing ensemble size and approach a stable level. This behavior indicates that sampling noise is substantially reduced at larger ensemble sizes, although it does not completely vanish.

Based on these results, we conclude that an ensemble size of 500 is sufficient to suppress spurious correlations to a level that does not obscure the physically meaningful patterns relevant to this study.

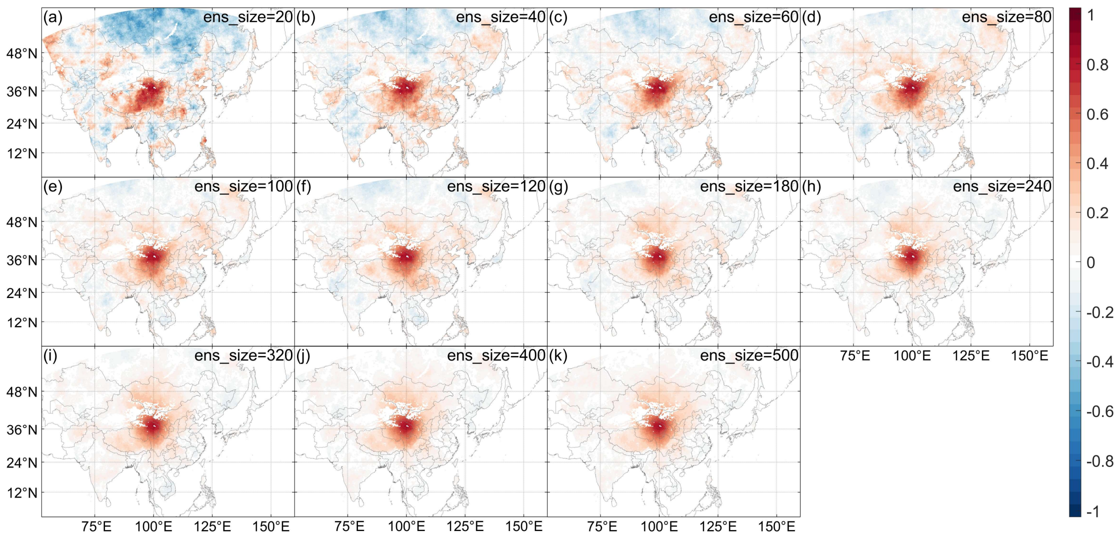

Fig. R1. Correlation patterns at the WLG site. Different panels have different ensemble sizes labeled at the right upper corner of the panels.

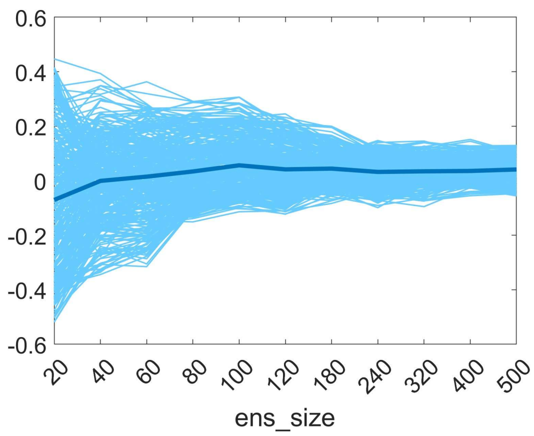

Fig. R2. Changes of correlation coefficients with ensemble size at the grids ~2000km away from the WLG sites. The bold line indicates the average.

2. Table 1: In the "Perturbation variance" column, "40% of mean" is used. Ensure it is clear whether this refers to the standard deviation or the variance itself.

Response:

We have corrected the terminology in Table 1 by replacing "perturbation variance" with "perturbation standard deviation".

3. In Section 3.1.1 (Line 222), it is mentioned that negative correlations occasionally arise and may be attributed to "negative diffusivity". Please discuss briefly the effects on the Kalman gain calculation.

Response:

We thank the reviewer for this insightful comment. Following the suggestion of Reviewer 2 (specific comment #9), we revised the corresponding figures in the main text, and the panels that previously showed occasional negative correlations are no longer included. Accordingly, to avoid distracting from the main message of the manuscript, we have removed the related discussion from the revised text (Line 280-287).

From the perspective of the EnKF framework, such negative correlations mainly lead to a sign reversal in the corresponding Kalman gain. In practice, this does not necessarily introduce an issue, because the sign of the observation increment is expected to change consistently with the concentration response to the underlying flux perturbation. As a result, the posterior adjustment remains dynamically consistent, with the update direction simply reversed relative to the positively correlated case.

4. It is suggested that following observation-based short correlation lengths (e.g., <100 km) is not recommended, for sparse observation networks. Please clarify if the "600 km" recommendation is specific to the East Asian domain or a general rule of thumb for any region with similar station density? Furthermore, some emphasis is

needed to avoid interpreting this as an observationally derived or physically “true” correlation length.

Response:

We thank the reviewer for highlighting this important point. The 600 km recommendation should be interpreted as a region- and network-specific practical choice, rather than an observationally derived or physically “true” flux correlation length. In our study, this value reflects the particular balance between station density, transport influence, and the need to efficiently propagate observational information across the East Asian domain. Therefore, it should not be taken as a universal rule of thumb applicable unchanged to other regions.

That said, the underlying trade-off is general across regional inversion systems. Specifically, there is a common balance between observation availability and the physical realism of prior flux autocorrelation structures, regardless of the study region. To clarify this point, we have added the following discussion in the revised manuscript (Line 413-422):

More broadly, similar to the choice of a longer correlation length, this trade-off reflects a balance between physical realism and inversion efficiency. Ideally, prior correlation functions (including their shape, correlation length, and related properties) should be informed by process understanding or model–data mismatch statistics, which may favor shorter or more heterogeneous structures. However, under sparse observational coverage, such physically realistic choices may fail to sufficiently propagate observational information to the target fluxes. Conversely, stronger and smoother correlation structures may improve the mathematical efficiency of the inversion by allowing sparse observations to constrain wider regions, albeit at the cost of reduced physical specificity. The results presented here should therefore be interpreted not as prescribing a universally optimal prior structure, but rather as illustrating the practical consequences of this balance under current observational limitations. Although the resulting balance may be region-specific, the methodology used to identify an appropriate configuration is straightforward and can be readily adapted to regions beyond our study domain.

See also related revision in Conclusion (Line 691-697).

5. The comparison between EnKF and 4D-Var in the discussion is entirely theoretical. Without running an actual 4D-Var experiment, claiming “essential equivalence” is a stretch.

Response:

We agree that the original wording was too strong given that no explicit 4D-Var experiment was conducted in this study. Our intention was not to claim empirical equivalence based solely on the present analysis, but rather to place our findings within the context of established theoretical and applied studies. The basis for this comparison is twofold. First, under the standard linear-Gaussian assumptions, the

Kalman filter update can be derived as the analytic solution of the corresponding 4D-Var cost function. Second, previous inversion studies have shown that the two methods often yield similar posterior estimates in regions well constrained by observations. In this sense, our analysis should be viewed as an intermediate diagnostic step, linking the theoretical equivalence at the equation level to the practical similarity observed in inversion applications. To better reflect this scope, we revised the conclusion sentence from:

This demonstrates the essential equivalence of EnKF and 4D-Var in their information-transfer kernel under linear-normality assumptions.

to:

Consistent with the theoretical equivalence between the Kalman filter update and the 4D-Var cost-function solution under linear–Gaussian assumptions (Evensen et al., 2022), as well as previous inversion studies showing similar posterior estimates from the two approaches in observation-constrained regions (Liu et al., 2016), our analysis suggests a corresponding equivalence in their information-transfer kernel under these assumptions. (Line 709-712)

The related sentences in Sect. 3.3 are also revised (Line 615-619).

6.Some terminology (e.g., "information transfer", "resonance") is physically intuitive but metaphorical. Consider briefly restating these concepts in strictly statistical terms when first introduced.

Response:

We thank the reviewer for this helpful suggestion. We agree that terms such as "information transfer" and "resonance" are physically intuitive and useful for conceptual interpretation, but may be overly metaphorical if not statistically defined when first introduced. In the revised manuscript, we therefore added brief clarifications at their first occurrence to restate these concepts in strictly statistical terms.

Specifically, "information transfer" is now explicitly linked to the propagation of observational increments to state variables through the prior covariance structure and Kalman gain (Line 86-88), while "resonance" is described in terms of the enhanced alignment between the assumed prior correlation structure and the effective transport-induced observation sensitivity pattern, which leads to stronger posterior constraint (Line 342-345). These additions are intended to preserve the intuitive interpretation while ensuring methodological precision.

Reviewer #2:

Fan and Li presented a study on investigating information transfer in CO2 flux inversions using the approach of Ensemble Kalman Filter based on Monte Carlo. The authors examined the information transfer pathways between CO2 concentrations and fluxes through analysing their spatial and temporal correlations across 500 ensemble members varying by fluxes and transported using WRF-Chem transport model. They introduced a comprehensive analysis highlighting the effect of applying different correlation length scales across several selected sites representing both surface station and remote sensing networks at different seasons. The authors suggested to use longer spatial and temporal length scales where a resonance can be achieved to maximize gains in the inversions, particularly when using limited observations from sparse sites. The study is well written and worthy of publication, as it provides useful insights for improving CO2 inversion estimates. I recommend publication after the authors address minor issues listed below.

General comments:

1. The introduction needs to be more comprehensive to align with the scope of atmospheric inversions. Specifically, there is still a gap in literature review that needs to be bridged to help introduce the scientific objectives.

Response:

We thank the reviewer for this important and constructive comment. We agree that the original Introduction was too compressed in its opening flow and did not sufficiently position the study within the broader context of greenhouse gas flux estimation and atmospheric inversion literature.

In response, we substantially revised the Introduction in three ways (considering also the specific comments #1-#3).

First, we added a broader overview of the major complementary approaches used to estimate greenhouse gas fluxes, including bottom-up inventories, process-based ecosystem models, remote sensing products, and top-down atmospheric inversions (Line 35-37). This allows atmospheric inversion to be introduced more naturally.

Second, we revised the discussion of the Global Carbon Budget to focus less on the number of participating inversion systems and more on the unique information atmospheric inversions contribute, namely their large-scale atmospheric constraint that complements bottom-up carbon budget estimates (Line 37-44).

Third, we expanded the literature review bridging the uncertainty discussion and the study objectives, particularly by adding context on how prior covariance matrices and their spatiotemporal autocorrelation assumptions are commonly constructed in inversion studies and how the autocorrelation structures may influence the inversion results (Line 54-64). This helps motivate why diagnosing the covariance-controlled information pathway is both scientifically relevant and methodologically necessary.

In addition, we inserted a clearer transition paragraph to explicitly distinguish our study from previous covariance-sensitivity studies by emphasizing that our focus is the mechanistic statistical pathway linking covariance assumptions to observational influence propagation, rather than only the resulting posterior flux differences (Line 65-69). We also revise the general objectives (Line 78-82).

These revisions improve the logical flow of the Introduction and better align the scientific objectives of the study with the broader literature on atmospheric inversions.

2. Some elements in Table 1 need further clarifications, e.g., aim of using offline and online perturbations, the selection of 40% of the variance (and make sure you do not mix it with standard deviation as is through Section 2.3).

Response:

We thank the reviewer for this helpful comment. We have revised Table 1 and the corresponding text in Sect. 2 to provide clearer explanations of the experimental design and parameter choices.

First, we clarify the purpose of the different perturbation strategies. The three cases are designed with two main objectives: (1) To systematically vary the spatial correlation length scales (e.g., 100 km vs. 600 km) in order to examine their influence on the information transfer and inversion behavior; (2) To include an ONLINE case, in which flux perturbations are generated through VPRM parameter perturbations, producing a spatial autocorrelation structure linked to land-cover types. This allows us to further assess the impact of the structure (shape) of the autocorrelation, beyond its characteristic length scale.

We also clarify that perturbations are applied to all three flux components to investigate the interactions among co-existing uncertain fluxes, which is central to the analysis of signal "dilution".

In addition, we have corrected the terminology in Table 1 by replacing "perturbation variance" with "perturbation standard deviation" to avoid ambiguity.

These clarifications have been incorporated into the revised manuscript (Line 179-187).

3. The emphasis in some conclusions on using longer spatial correlation scales is somehow not fully understandable because it is not about increasing correlation length to get fluxes constrained by distant observation sites as long as there is surface influence. Instead, it is about to what extent the errors of fluxes are dependent spatially and temporally in biosphere models compared to reality. I agree, longer correlation lengths allow for more information transfer from observation space to flux space in inversions, but this is not always true or at least differs among biosphere

flux models used as priors in inversions.

Response:

We thank the reviewer for this thoughtful comment. We fully agree that the true spatial correlation scale of biosphere flux errors is likely short (<100 km) and is model-dependent. Given this agreement, the reviewer raises a fair question: can we still recommend a longer length (600 km) without relying on the physical length? To make it clearer, we add the following explanation to the revised manuscript (Line 373-389):

The specification of the spatial correlation length L in the prior covariance matrix $\boldsymbol{B}$ serves two distinct purposes. The first is to represent the assumed spatial structure of flux errors. The second, related to the first and revealed by our results, is to regulate the propagation of observational information. Under a sparse observation network, the latter can become the dominant constraint. Mathematically, in the Kalman gain expression $\boldsymbol{K} = \boldsymbol{B}\boldsymbol{H}^T(\boldsymbol{H}\boldsymbol{B}\boldsymbol{H}^T + \boldsymbol{R})^{-1}$, the term $\boldsymbol{B}\boldsymbol{H}^T$ governs how observational increments propagate to state variables. Even when atmospheric transport ($\boldsymbol{H}$) physically links distant locations, prescribing a short correlation length (e.g., 100 km) forces the prior covariances to near zero for separations beyond approximately 300 km. This mathematical cutoff is independent of transport; the Kalman gain simply has no pathway to connect the observation to that distant flux, regardless of how well the transport model represents the physical connection. Here, it is important to distinguish between two aspects of concentration–flux correlations: the location of the correlation maximum (determined by atmospheric transport) and the spatial extent of the high-correlation region (determined by the prescribed autocorrelation length L). Even with a short L, transport can shift the correlation maximum far from the flux source. However, a short L causes the high-correlation region to be narrow, so that sparse observations can only constrain a small area around the location with strongest correlation. A longer L broadens the high-correlation region, enabling the same sparse network to constrain a much wider domain, as shown in Fig. 3. Thus, the recommended 600 km length is not a claim about the true spatial scale of flux errors, but a mathematical requirement to keep the information pathway open under sparse observations.

As the reviewer rightly implies, the optimal length is not universal. We have clarified the statement in the conclusion (Line 690-697):

This explains why, under sparse observations, using overly short autocorrelation lengths (e.g., <100 km) not only shrinks the spatial influence of information but also systematically reduces the information extraction efficiency of the inversion system, leading to extensive "blank areas" of observational constraint (Fig. 3). It should be emphasized, however, that the choice of longer autocorrelation lengths (e.g., the 600 km recommendation for the current East Asian network) involves a trade-off between physical realism and mathematical efficiency. This choice is conditional on network sparsity and is region-specific. As observation networks become denser in the future, the optimal correlation length will converge toward the shorter, physically realistic scale. Our methodological framework provides a quantitative tool to re-evaluate this optimal length as observational constraints improve.

4. Although the effect of CO2 concentration and fluxes correlations is clear, quantifying the impact in flux estimates would be more useful by implementing inversions using the similar setup as in Table 1. This would give some insights on how much spatial and temporal variations in fluxes can be explained by tuning correlation scales. I know this might be beyond the setup of this study as such an inversion should be conducted for a longer period of time than one week.

Response:

We thank the reviewer for this important suggestion. We agree that implementing actual inversion experiments would provide a more direct link between the diagnosed information-transfer mechanisms and inversion performance.

While a full, systematic inversion study is beyond the scope of the present work, we have conducted a set of simplified OSSE-based inversion experiments using an existing inversion system developed for a follow-up study (CMAQ-VPRM with DART). These experiments are intended as a consistency check, rather than a comprehensive parameter study, and are therefore presented in the Supplementary Information with a brief summary added to the main text (Line 396-403).

In the OSSE framework (two-month simulation, 100-member ensemble), we tested two spatial autocorrelation length scales (100 km vs. 600 km) with and without satellite observations. OFF100surf is the case with 100 km correlation length and surface observations only, while OFF100 uses additionally satellite data. The results (see Fig. R3) show that when surface-only observations are used, with a 100 km correlation length, the inversion produces minimal posterior adjustments (Fig. R3h), and corrections remain localized near observation sites, indicating very limited constraint. Increasing the length scale to 600 km leads to substantially larger and more spatially coherent corrections, and improves agreement with the true flux (Fig. R3i). When satellite data are added, satellite data increase overall constraint, particularly in regions with dense sampling (e.g., South and Southeast Asia). However, even in this case, the 100 km setting still fails to adequately constrain fluxes in regions with strong true signals (e.g., China, Fig. R3j), whereas the 600 km case shows consistent improvement across regions (Fig. R3k).

For a quantitively evaluation, compared to the prior, the 100 km (surface-only) case shows almost no improvement. In contrast, the 600 km case reduces mean bias by ~44% (surface-only) and ~88% (with satellite data), with corresponding RMSE reductions of ~9% and ~32%.

These results are consistent with our diagnostic findings that short autocorrelation lengths suppress the propagation of observational constraints, particularly under sparse observational coverage, and that increasing the correlation length enhances inversion effectiveness.

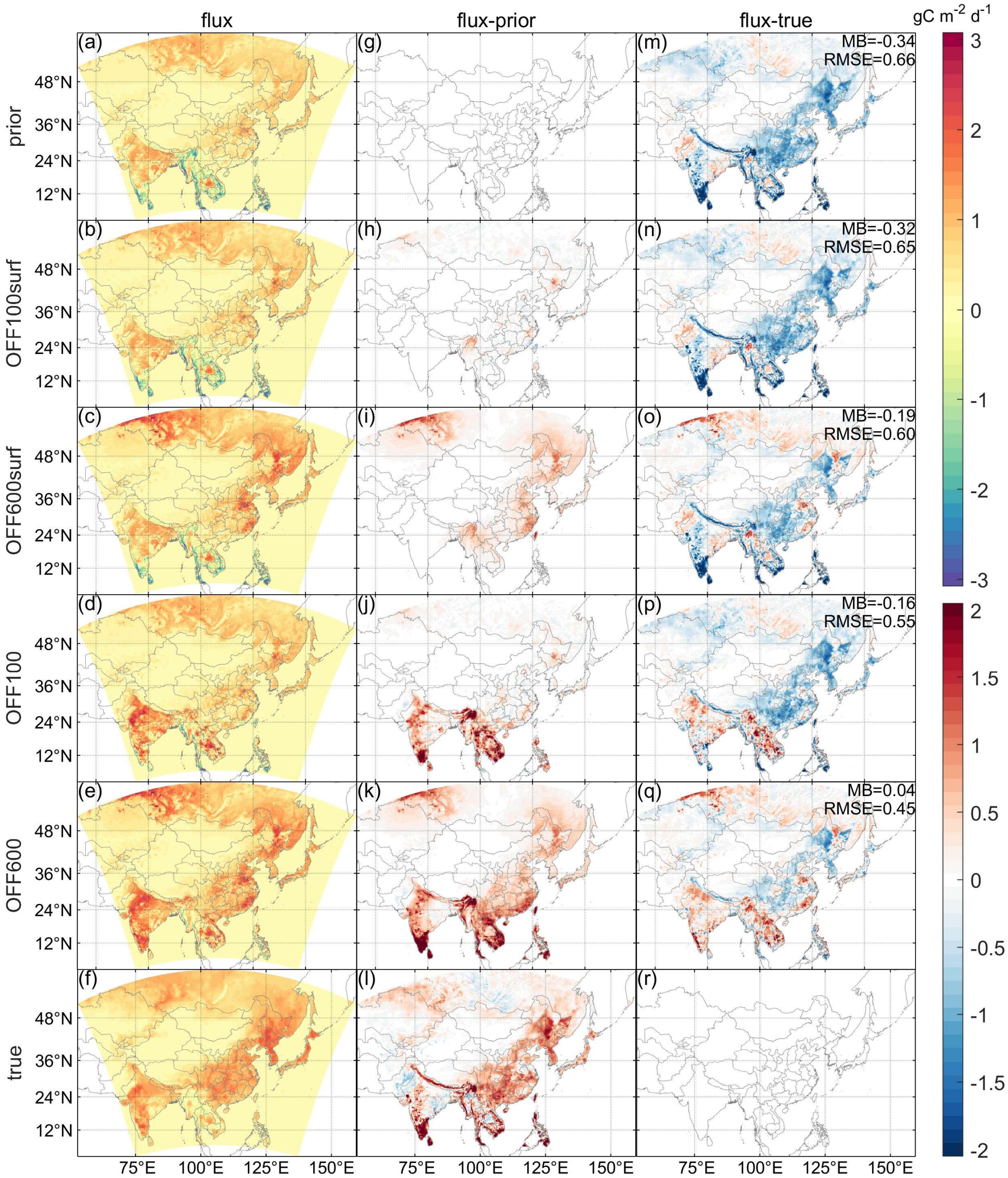

Fig. R3. Monthly biospheric fluxes in the inversions. (a) is the prior flux. (b)-(e) are the posterior fluxes. (f) is the true flux. The second column is the difference between the first column and the prior (a). The third column is the difference between the first column and the true (f). The numbers in the third column panels represents the mean bias (MB) and root mean square error (RMSE) across all the land grids.

Specific comments:

1. L26-29: The first sentence feels quite squeezed in its flow. So, there should be a sentence before to briefly describe what are the various estimation methods, used to quantify GHG fluxes besides atmospheric inversions. Also, consider splitting and rephrasing this sentence to clearly first define atmospheric inversions and then you can indicate their usefulness afterwards.

Response:

Please see general comment #1.

2. L30-31: It does not matter how many inversions participating to GCP but what would be worthy of mentioning here is what information atmospheric inversions add on top, or alongside, of the other methods to estimate GCB, for example.

Response:

Please see general comment #1.

3. L41-46: Here is a big jump in the flow with lack of literature. You need to elaborate the objectives of your study (in the next paragraph) with more details and relevant literature. For instance, there are many studies elucidating their assumptions and approaches for the construction of covariance matrices that consist of autocorrelation information in space and time.

Response:

Please see general comment #1.

4. L87: Indicate clearly what method you adopt here, e.g. nonlinear Kalman filter, as "the latter" is still ambiguous at this stage, or just link the sentence like: "…….(Tanizaki, 1996), which we adopt in this study."

Response:

Revised (Line 124).

5. L129-130: You need to mentioned what is meant by ACM2 scheme as it was not defined previously.

Response:

The sentence is revised as The planetary boundary layer scheme for vertical mixing was the Asymmetric Convective Model, version 2 (ACM2, Pleim, 2007). (Line 168-169)

6. L196-197: Although correlations between observations and fluxes affect how much observational information to the state space variables, atmospheric transport and uncertainties assumed in fluxes and observations play a major role in that. So, they are worthy of mentioning in this context to provide a full view of the most impactful components in inversions for the reader.

Through all texts, clearer to use "long or short" instead of "large or small" once referring to spatial and temporal correlation scales.

Response:

We thank the reviewer for this helpful suggestion. We agree that, in addition to concentration–flux correlations, atmospheric transport and the assumed uncertainties in both fluxes and observations are also fundamental components controlling inversion behavior. To provide a more complete context, we have revised the text to explicitly acknowledge that these factors influence the uncertainty of modeled concentrations and therefore affect the effectiveness of observational constraints.

At the same time, atmospheric transport, as well as the assumed uncertainties in both fluxes and observations, also play fundamental roles by shaping the uncertainty of modeled concentrations and the resulting observational influence. In the present study, we specifically focus on the spatial extent of observational constraints on fluxes. (Line 250-253)

In addition, throughout the revised manuscript, we have replaced "large/small" with "long/short" when referring to spatial and temporal correlation scales.

7. L221-222: The negative correlation between CO2 simulations and fluxes in Fig. 1Bg requires further explanation than diffusivity or perhaps diagnostics to find out the main reason if it is something physical property of the transport (in advection and in vertical convection) but also to ensure this is not just an artifact due to transport error at that local scale. I would suggest to displace the location of that site one or a few grids around and see whether this phenomenon still persists.

Response:

Following the reviewer's recommendation, we examined the surrounding grid cells (±4 grid points around the original site location). The negative concentration-flux correlations persist at several neighboring grid cells (Fig. R4a), indicating that this feature is not confined to a single grid cell and is therefore unlikely to result from a purely local numerical artifact.

To further assess whether this behavior is specific to near-surface transport, we additionally analyzed the M_{TOT}-F_{ANT} correlations at other altitudes (Fig. R4b,c). Negative correlations appear in a larger number of grid cells at ~1000m, whereas they are much less pronounced near ~100m. This vertical contrast suggests that the feature is more likely related to the structure of atmospheric transport and vertical mixing pathways than to isolated grid-scale noise. That said, based on the current diagnostic framework, it remains difficult to unambiguously determine whether the underlying mechanism is primarily associated with physical transport processes (e.g., advection or vertical convection) or numerical aspects of the transport simulation. A dedicated transport-process analysis would be required to fully resolve this issue, which is beyond the scope of the present study.

From the perspective of the EnKF framework, such negative correlations mainly reverse the sign of the Kalman gain. Because the sign of the observation increment is also expected to reverse consistently with the concentration response to the flux perturbation, the resulting posterior adjustment is not necessarily problematic in itself.

Importantly, after revising and simplifying the main figures, the panel containing this specific feature is no longer included in the main text. Accordingly, to avoid distracting from the main message, we have removed the related discussion from the manuscript (Line 280-287) and retain these additional tests only in the response for completeness.

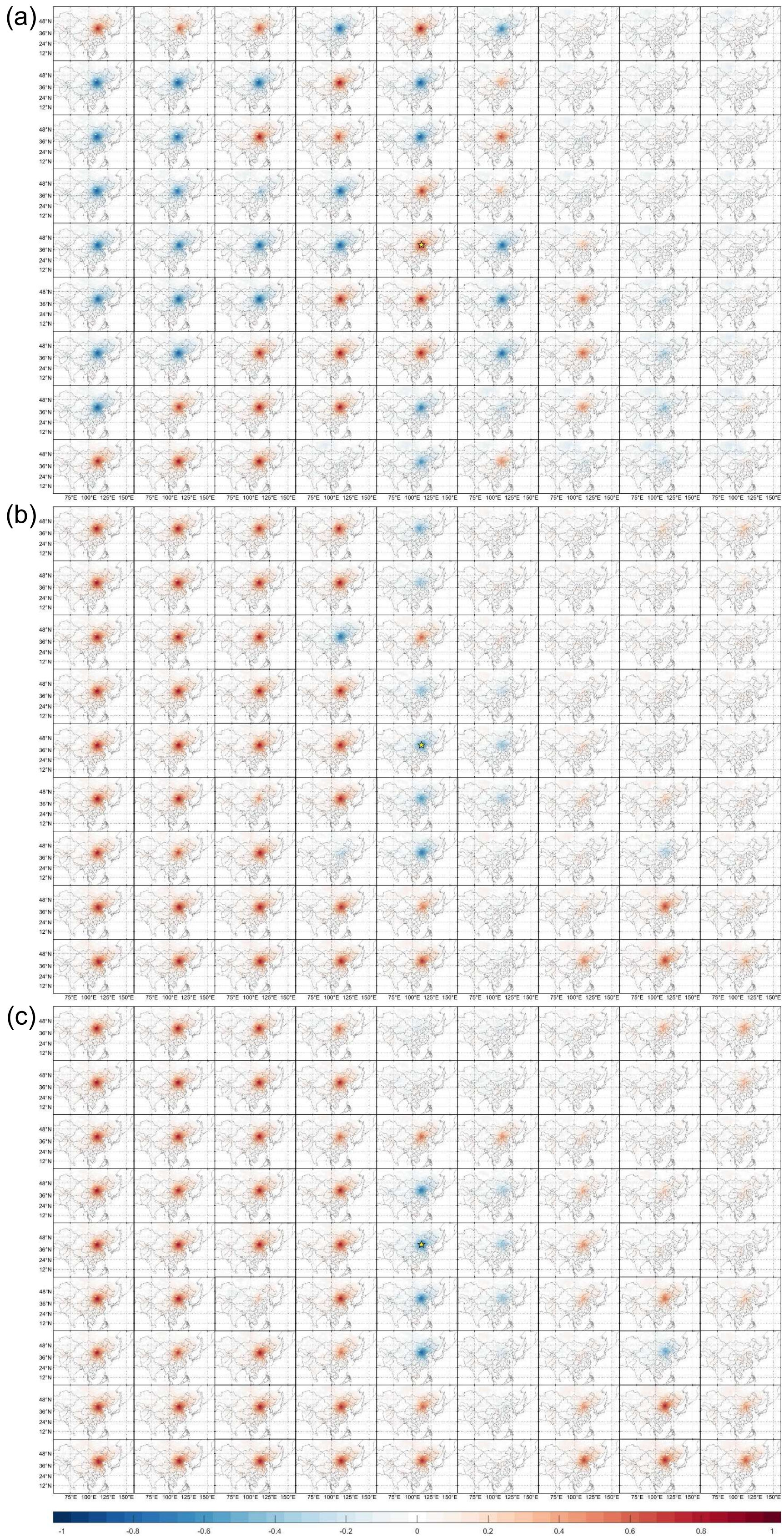

Fig. R4. Correlation patterns at the original site and nearby grids. The original site is at the center panel. (A) XCO_2, (B) M_{TOT} at ~1000m, (C) M_{TOT} at ~100m.

8. L280-283: In indication to Kountouris et al. (2015, 2018), the use of a shorter correlation length than 100 km was coincided with applying a different decaying function (hyperbolic), which has another property with longer tails compared to the exponential function. This also allows for adjusting fluxes based on distant observations and that experiment relied on analysis of model-data mismatch using the same biosphere model VPRM used in this study as well. Of course, you are using here a statistical method to generate the covariances that are associated with what you indicated "resonant effect" that maximises the gain in information transfer but the question remains is how realistic is this in terms of flux estimates compared to the truth, despite the issue of data sparsity.

Response:

We thank the reviewer for this insightful comment regarding the role of the shape of the correlation function. The reviewer is correct that, in principle, a correlation function with a longer tail can allow more remote fluxes to be adjusted by observations. However, our additional sensitivity test suggests that this effect is limited in practical EnKF applications. As shown in Fig. R5, we compared three correlation-function shapes (Gaussian, exponential, and a slowly decaying hyperbolic form, $1/(1+d)$) while keeping the overall correlation length scale comparable. The longer-tailed functions do extend the reach of the prior autocorrelation. However, the resulting M-F correlations at remote distances rapidly decrease to values that are small and, in practice, become indistinguishable from sampling noise in the ensemble framework (Fig. R5b–d). As a result, the effective observational influence is not substantially extended in a meaningful way. At the same time, a longer tail redistributes the total correlation structure and tends to weaken the correlation strength in nearby regions. Therefore, the choice of correlation-function shape introduces a practical trade-off: (1) a longer tail may slightly increase the theoretical range of adjustment, (2) but it can reduce the strength of local correlations and thus weaken the most informative constraints. For this reason, the shape of the correlation function should be selected by balancing these two competing effects. (Line 404-412)

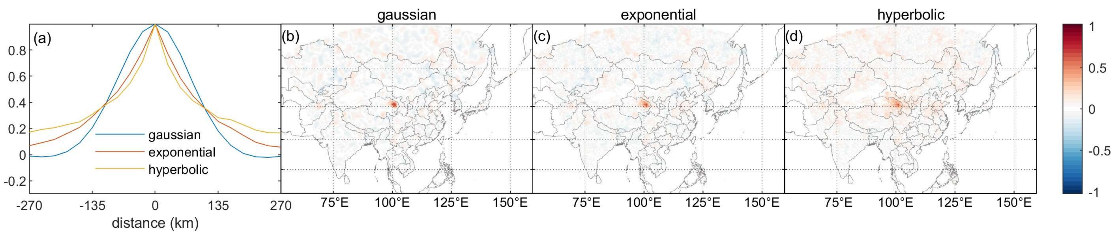

Fig. R5. Influence of correlation function shape. (a) Spatial autocorrelations along latitude centered at the WLG site. Lines are averaged from 21 lines with centers near the WLG site for smoothness. Colors denote correlation functions: blue—Gaussian, orange—exponential, yellow—hyperbolic (1/(1+d)). (b)–(d) Spatial patterns of M_{TOT}–F_{BIO} correlations using (b) Gaussian, (c) exponential, and (d) hyper functions. The experiments are conducted using CMAQ instead of WRF-VPRM, and the temporal coverage is from November 1 to November 8, 2015, without perturbing the initial conditions.

Regarding realism relative to the true fluxes. The reviewer asks a critical question: how realistic are the flux estimates obtained with the 600 km correlation length compared to the true fluxes? We address this directly using the OSSE framework mentioned in our response to general comment #4. In the OSSE experiments, the “true” flux is known by construction, allowing us to evaluate posterior estimates against a known truth. The results (Fig. R3) show that, under the current sparse surface network, the 600 km setting reduces the mean bias relative to the true flux by approximately 44% compared to the prior, whereas the 100 km setting produces almost no improvement (bias reduction < 2%). The RMSE is reduced by about 9% with 600 km, versus less than 1% with 100 km. When satellite observations are added, the 600 km setting achieves a bias reduction of 88% and an RMSE reduction of 32%. These quantitative results demonstrate that the 600 km correlation length not only enhances mathematical information transfer but also yields posterior flux estimates that are substantially closer to the true fluxes. The “resonance” effect is therefore not merely a mathematical artifact; it enables the inversion to extract meaningful signal from sparse observations. Conversely, the physically more “realistic” short correlation length (100 km) fails to constrain the fluxes, producing posterior estimates that remain close to the prior and thus do not reflect the true flux distribution.

We acknowledge that the OSSE framework assumes perfect model physics and known error statistics; the absolute improvements reported here should therefore be interpreted as upper bounds. In real-world applications, the true improvement may be smaller, but the relative advantage of longer correlation lengths under sparse observations is expected to hold.

Our intention is not to claim that the more mathematically efficient choice is universally “best,” but rather to explicitly illustrate the cost of prioritizing physical realism when observational density is limited. We have clarified this point in the revised manuscript (Line 396–403).

9. Cross comparison between Fig. 1 and Fig. 2 is confusing. To simplify it with minimum redundancy, you can choose only one site as best example for each case (surface vs airborne and winter vs summer). This will clearly help compare the resulting effect for each case you specified in Table 1, e.g., ONLINE and OFFLINE perturbations, long and short correlation length scales (e.g., OFF600 and OFF100), etc. The rest can be moved to the supplements. In Fig. 3, correlations cannot be added linearly, at least need to be normalized by the sample size. In addition, captions of figures are not fully clear and need to be well explained even with lengthy text, and better to avoid cross mentions from figure to figure, provided that each one is very busy with too much info.

Response:

Following the reviewer’s recommendation, we have redesigned Figs. 1, 2, 4, and 5 to reduce redundancy and improve readability. In the revised manuscript, we retain only three representative sites in the main text, chosen to illustrate the three major flux types and the key contrasts among the experimental settings. The remaining site examples and extended comparisons have been moved to the SI. We have also substantially revised the figure captions to make each figure more self-contained and easier to interpret without cross-reference to other figures.

Regarding Fig. 3, we agree that correlation coefficients themselves are not linearly additive in a strict statistical sense. Our intention in this figure, however, is not to derive a formal combined correlation metric, but rather to provide a qualitative visualization of the spatial extent of observational influence across multiple observation types and times. Because all correlation fields are derived from the same ensemble size and identical sampling framework, the comparison remains internally consistent for this illustrative purpose.

To avoid possible misunderstanding, we have revised both the figure caption and the main text to explicitly clarify that this figure should be interpreted as a diagnostic composite of correlation extent, rather than a statistically rigorous combined correlation coefficient or a weighted measure of observational effectiveness. Our goal here is to emphasize the spatial propagation pathways of observational constraints, rather than their formally aggregated magnitude.

10. Throughout Section 3.1.3, replace "CO_2"$_{BIO,ANT,OCE}$ with any code else (like C or M) to differentiate between concentration and flux (F), because both relate to CO2.

<u>Response:</u>

We thank the reviewer for the suggestion and we use M to represent concentration across the manuscript when necessary and use subscript TOT to represent normal concentration other than tracers.

11. Fig. 4: please see note above regarding Fig. 1 and 2.

<u>Response:</u>

See #9 above.

12. L414: color lines for F_{BIO} looks much more into cyan other than green!

<u>Response:</u>

Revised.

13. L437-439: Transport influence plays a more important role in designing station network and thus correlation effect is stronger at sites with broader transport influence and can be beneficial for inversion.

<u>Response:</u>

We agree that atmospheric transport plays a major role in station network design. Our intention in the main text was to quantify this transport influence and provide a visualized diagnostic framework that may help guide network design, as we expect that some characteristic transport behaviors can be meaningfully visualized, such as the clear January-July contrast shown in Fig. 6. We appreciate this helpful comment and have revised the corresponding sentence to further emphasize the central role of transport.

The sentence has been changed from:

Due to the rapid and significant variations in atmospheric dynamics, optimal measurement locations may change quickly, meaning our findings are not definitive.

to:

Since atmospheric transport plays a critical role in determining the locations of high correlations, and because atmospheric dynamics can vary rapidly and substantially (see also the January–July contrast in Fig. 6), the optimal measurement locations may change quickly; therefore, our findings should not be regarded as definitive. (Line 565-567)

Technical corrections:

1. L35: Missing object!

<u>Response:</u>

Corrected. (Line 48)

2. Sentence starts in L280 suggesting to be joint with the next one!

<u>Response:</u>

Revised. (Line 354)