

Referee report

Title: Estimating cross-correlation parameters between forecast and observation errors within the ensemble transform Kalman filter with cross correlation (ETKFCC)

Manuscript: egusphere-2026-2653

Authors: Y. Kobayashi, S. Ohishi, T. Miyoshi

14th October 2026

Recommendation: Major revisions

Summary of the manuscript

The standard Kalman filter and ensemble Kalman filter are formulated under the assumption that forecast and observation errors are uncorrelated. This breaks down in systems that assimilate analysis products, such as optimally interpolated ocean analyses or atmospheric reanalyses, where the observation errors inherit a fraction of the forecast errors themselves. Building on the authors' previous work (Ohishi et al., 2025, "OKM25"), which derived the KF and ETKF with cross-correlation (KFCC and ETKFCC) under the assumption that the cross-correlation parameters are known, this manuscript proposes an offline method to estimate those parameters using innovation statistics from Desroziers et al. (2005). Using perfect-model twin experiments with the Lorenz-96 model, the authors show that the combination of $\langle d^{(a-b)}(d^{(o-b)})^T \rangle$ [their Eq. (29)] with $\langle d^{(o-b)}(d^{(o-b)})^T \rangle$ or $\langle d^{(o-a)}(d^{(o-b)})^T \rangle$ [their Eq. (34)] recovers the cross-correlation parameters well for positive correlations, and that the ETKFCC with estimated parameters significantly outperforms the standard ETKF for cross-correlation coefficients between 0.2 and 0.7, and remains robust even when the underlying ETKF tuning is suboptimal.

The mathematical development is rigorous and consistent across the appendices. The extension of the Desroziers diagnostic framework to correlated background and observation errors is a genuine and worthwhile methodological contribution. The manuscript in its current form though technically correct, lacks a developed scientific narrative. Additionally, several claims and scope limitations require attention before the work is suitable for publication.

Major comments

1. Insufficient introduction and motivation. The introduction spans barely fifteen lines and jumps almost immediately to the authors' previous study. The research question (estimating previously assumed-known cross-correlation parameters) is stated, but the operational urgency is not established. Why does this problem matter now? Which operational or quasi-operational systems (e.g., JCOPE-FGO, NAVGEM with ECMWF analyses, cited in passing) suffer from this correlation, and with what estimated magnitude? A paragraph, ideally with a motivating example quantifying the degradation caused by ignoring the correlation, would substantially strengthen the paper. As is, the manuscript reads as the second instalment of a technical series rather than a self-contained scientific argument.
2. Excessive reliance on OKM25 without scaffolding. The paper assumes the reader is intimately familiar with OKM25. Equation (1) (the error generation mechanism $\vec{\epsilon}^o = \vec{A}H(\vec{\epsilon}^f) + \vec{\eta}$) is introduced without physical interpretation, and the KFCC formulation (Eqs. 3–7) is quoted rather than explained. A conceptual figure illustrating how assimilating optimal-interpolation or reanalysis products injects forecast error into the observation error would resolve this for readers outside the

immediate group, and would also serve to justify the diagonal structure of $\vec{A}$. Without this, the paper is inaccessible to much of its target audience.

3. Spatially uniform parameter assumption. The entire experimental design and the primary estimation equations rest on $A = aI$ and $R^u c = r^u cI$. This is acknowledged in the concluding section, but the implications are not discussed anywhere in the body, and the general non-uniform equations (27), (28), and (31)–(33) are derived and then never exercised. In realistic geophysical systems, observation error characteristics are strongly heterogeneous in space, time, and observation type. The authors should discuss the robustness of the estimation when the uniformity assumption is violated, and ideally provide a limited sensitivity experiment with spatially varying a_i . The non-uniform estimators should be tested in the Lorenz-96 framework.
4. Figure strategy does not support the argument. Figures 1–6 report spatio-temporally averaged RMSEs and parameter estimates. This is sufficient to confirm that the method works in aggregate, but it does not demonstrate why it works or when it might fail. I recommend: (i) a schematic of the correlated-error mechanism (see comment 2); (ii) a time-series or histogram showing the convergence and scatter of a^{est} against a^{true} over the assimilation window, rather than only the 10-year average, so the reader can judge the estimator variance and any temporal drift; and (iii) for the key result (Fig. 4), confidence intervals that are visible on the plot itself rather than only implied by the bootstrap procedure described in the text.
5. The $a^{\text{true}} = 0.9$ degeneracy deserves more than one sentence. The filter divergence at $a^{\text{true}} = 0.9$, attributed to $a^{\text{est}} \approx 1$ rendering $K' \approx O$, is arguably the most practically important caveat in the paper: it identifies a regime where the proposed estimator produces a catastrophically wrong gain rather than merely a suboptimal one. The authors should analyse this degeneracy more carefully. Is there an a priori diagnostic (e.g., a diagnostic of the estimated innovation covariance consistency) that would flag when the estimator is operating near this regime?

Minor comments

1. Negative correlations are dismissed too lightly. The rationale (“observations negatively correlated with the forecast errors would not exist in practice”) may well be correct for the specific contamination mechanism considered, but the paper includes the negative- a experiments, and their failure should be explained rather than waved away. Moreover, the asymmetry between estimation success for positive a and failure for negative a (attributed in the text to spread biases, Fig. 2c) deserves a cleaner theoretical treatment. If the estimator has a systematic bias whose sign depends on the spread error $\delta\sigma_f^2$, this should be stated as a general property, not just an empirical observation about this setup.
2. “Negligible additional computational cost” needs qualification. The cost per analysis step is indeed negligible, but the proposed workflow requires a full prior ETKF run to accumulate innovation statistics, followed by a second ETKFCC run. The claim is true of the update step and misleading of the overall workflow. Please clarify.
3. The two-parameter tuning of the baseline ETKF (ρ and R). The comparison in Fig. 4 gives the standard ETKF a generously large tuning space (R up to 10 times its true value). This is commendably fair to the baseline, but the text should state explicitly whether the tuned- R baseline is meant to represent the situation where a practitioner compensates for the unmodelled correlation by inflating R , and what the comparison implies about that common practice.
4. Line 260, “significantly and 5% more accurate”. The phrasing conflates statistical significance and effect size. The bootstrap test establishes significance at the 99% level; the 5% is the effect size. Please rephrase.
5. Descriptions of filter divergence thresholds. The definition of filter divergence (analysis RMSE exceeding observational RMSE) is buried in Section 3. Consider moving it to Section 2 or defining it at first use in the results.
6. Eq. (37) notation. The forcing $F = 8.0$ and Δt conventions follow Lorenz and Emanuel (1998); the mapping of $\Delta t = 0.05$ to six hours is stated correctly, but the manuscript uses “model timestep”

loosely ($\Delta t = 0.01$ for RK4 integration, 0.05 for observation interval). One clarifying sentence would help.

7. References. The Hirose et al. (2019) citation appears to concern tidal forcing of interplate earthquakes, which seems unrelated to optimal interpolation in ocean DA. Please verify this citation is intended, and check Kido et al. (2022) is cited in the appropriate context. In general, more citations at sensible points in the introduction and methods would strengthen the argument.
8. Presentation of Eqs. (31)–(36). Six near-identical estimation formulas are presented before the reader learns which combination is ultimately adopted (only in Section 4.1). Consider presenting the recommended combination ((29) with (34)) prominently in Section 2 and relegating the alternatives to the appendix, which would sharpen the methodological message.

Assessment of novelty and significance

The extension of Desroziers-style innovation diagnostics to the correlated-error setting is a sound and useful contribution, and the demonstration that the offline estimation remains effective even from sub-optimally tuned ETKF runs (Fig. 6) addresses a genuinely practical concern. The restriction to the Lorenz-96 perfect-model framework is acceptable for a methodological paper of this type, provided the authors more candidly frame the distance between this setting and their stated target applications.

Recommendation

Major revisions. The core mathematics is correct and the contribution is worth publishing, but the manuscript requires substantial revision: an expanded and properly motivated introduction, self-contained exposition of the OKM25 framework (ideally with a schematic), engagement with the spatial-uniformity limitation, careful analysis of the $a \approx 1$ degeneracy, and a figure strategy that demonstrates rather than merely tabulates the estimator’s behaviour. None of these corrections require new conceptual development beyond the authors’ evident capability, and I would expect a revised version to be suitable for publication. I would be pleased to review this subsequent version.