Referee Comment (RC2)

Manuscript: egusphere-2026-5032, *New and existing covariance estimation techniques for ensemble data assimilation*, by S. Gilpin, M. Morzfeld and K. K. Lin
Recommendation: Major revisions

General comments

The manuscript compares covariance estimators for small-ensemble DA (GC and GenGC localization, hybrid estimators with an "average of the day" background, Ledoit–Wolf shrinkage, soft thresholding and NICE) on a static Gaussian problem and in cycling stochastic EnKF experiments with two inhomogeneous Lorenz-96 variants, and introduces GenGC localization with spatially varying cut-off lengths. The experiments are carefully executed and the discussion of PSD violations and filter divergence under thresholding is useful. The paper is well written.

My main concern is the scope. The paper is framed as a bridge between covariance estimation in statistics and in geophysical DA, and its central conclusion is that "spatially-agnostic" statistical estimators are not competitive with localization. However, the statistical methods tested are the weakest representatives of that literature (linear shrinkage towards the identity, element-wise thresholding), while the family of *structured precision-matrix estimators*, which is the main line of high-dimensional covariance estimation in statistics and has an established EnKF implementation, is absent. This family is directly relevant: it resolves the two problems the authors themselves document (loss of positive definiteness, and the lack of a natural notion of distance), and its omission makes both the review and the headline conclusion premature. The second concern is that the "new" GenGC localization is constructed from prior knowledge of the model's inhomogeneity (the forcing or damping profile) rather than estimated from data, so its practical applicability is unclear. Details follow.

Specific comments

1. **Structured precision estimation is missing from the review and from the experiments.** The authors cite Bickel and Levina (2008b) and Friedman et al. (2008) in the introduction but test neither. Banded (modified) Cholesky estimation of the precision matrix (Pourahmadi, 2007; Bickel and Levina, 2008b), graphical lasso, and their EnKF implementation (Niño-Ruiz, Sandu and Deng, 2018, *SIAM J. Sci. Comput.*; see also Niño-Ruiz et al., 2018, *J. Comput. Appl. Math.*) belong to the "statistics" side of the paper's own dichotomy, yet they use locality through conditional independence rather than through tapering of marginal correlations, and are therefore precisely the intermediate class the paper is looking for. Three consequences for the manuscript:
   (a) Lines 486–501 explain at length that thresholding yields non-PSD estimates and filter divergence. A banded Cholesky precision is SPD by construction, and this is the standard statistical remedy for the problem reported. The paper should at least say so, and preferably test it.
   (b) Section 2.3 states that spatially-agnostic methods "have no natural notion of distance". The modified Cholesky estimator incorporates distance through the ordering and the regression neighbourhood without imposing a taper, which weakens the dichotomy on which the paper's discussion rests.
   (c) The conclusion (lines 544–548) that statistical estimators are "not competitive with localization or hybrid estimators in all experiments" cannot be drawn from Ledoit–Wolf and thresholding alone. Please either add a precision-space estimator to Figures 5–8 (the modified-Cholesky EnKF is a drop-in replacement for $\mathbf{P}^f$ in information form) or restrict the conclusion to the estimators actually tested.

   Niño-Ruiz et al. (2021) is cited only as having "showed little promise" (line 183); please characterize that result more precisely, since it is used to justify not exploring the statistical side further. Since the ordering of localization versus structured estimation is known to depend on observation density and ensemble size, the conclusion depends strongly on which estimator is chosen to represent the statistical family.
2. **GenGC localization requires knowledge the user does not have.** In Eqs. (18), (21) and (23) the cut-off function $c(x)$ is prescribed to mirror the damping or forcing of the model, and c^* is then tuned against the truth. This is an oracle construction. Without a procedure to estimate $c(x)$ from the ensemble or from climatology (e.g. via the optimal localization of Eq. 5 applied to a large ensemble, or via the correlation-length diagnostics of Ménétrier et al. 2015), GenGC is not a method a practitioner can apply, and the paper should say so explicitly. The gain over GC is also marginal (Figures 7–8: overlapping quantiles), which should be reflected in the abstract and conclusions.
3. **Tuning protocol favours methods with more hyperparameters.** All hyperparameters are tuned jointly to minimize analysis RMSE against the truth (line 437). Hybrid and GenGC estimators have several tunable quantities; Ledoit–Wolf and NICE have none or one. Please report the number of tuned parameters per method, and ideally tune on a separate set of experiments from those used for evaluation.
4. **Single ensemble size.** All cycling experiments use $n_e = 20$. Since the question is estimation from few samples, a sweep over n_e (e.g. 10, 20, 40, 80) is needed; the ranking of shrinkage versus localization is known

to depend on n_e/n_x.

5. **Terminology.** "AoD shrinkage" (Eq. 7 with $\mathbf{B} = \mathbf{B}_{\mathrm{AoD}}$) is a hybrid with a cycle-derived climatology, not shrinkage in the statistical sense; the genuinely statistical variants (nonlinear shrinkage, Ledoit and Wolf 2018, 2020; factor-based targets) are not tested. Please either rename or add one such estimator.
6. **Missing experimental details.** Inflation factors, localization radii, c^* values, the NICE powers and the thresholds actually selected are not reported. Please add a table. The code is only promised "should the paper be accepted"; EGUsphere policy favours availability at the discussion stage.
7. **Observation operator.** Only the stochastic EnKF with linear $\mathbf{H}$ is considered (lines 100–106). Since the paper argues that its conclusions are relevant to NWP, a short discussion of which estimators extend to square-root filters and observation-space localization is needed.

Technical corrections

- Line 136: "(Morzfeld and Hodyss, 2023, their Eq. 12)" — Eq. (5) with n_e in the numerator and denominator should be checked against the cited form.
- Figure 5(i) and 6(i): state the ensemble size and the number of draws in the caption.
- Line 267: "the last 3500 time steps" versus line 249 "the last 500 time steps" — please make the climatology estimation consistent or explain the difference.
- References: Lorenc (2003a,b) are duplicates; Necker et al. 2022 and 2023 appear to be the same work.
- "thesholding" (line 231); "effected" → "affected" (line 210).