Review of "Improved Method for Merging Particle Filter Using Random Merging Coefficients"

Manuscript: egusphere-2026-4073, by N. Hiramoto, G. Ueno, and D. Murakami.

AI disclosure: I take full ownership and responsibility for the review, but note that an AI/LLM "agent" assisted in all aspects of the review.

Summary

The manuscript proposes the randomized merging particle filter (RMPF), which modifies the merging particle filter (MPF) of Nakano et al. (2007) by (i) drawing the merging coefficients at random, uniformly on the set $\sum_j \alpha_j = 1$, $\sum_j \alpha_j^2 = 1$, and (ii) computing the (diagnostic) state estimate from the weighted ensemble before resampling and merging. The authors derive the density of a merged particle, interpret the choice of the merging size n as a bias–variance trade-off, propose selecting n by maximising the predictive likelihood, and report experiments with Lorenz-63 and Lorenz-96.

Generalizing the fixed coefficients is a reasonable goal, and random coefficients do remove a genuine weakness of MPF: with fixed coefficients, a recurring group of parents yields identical merged particles, and groups recur often under weight degeneracy. However, I cannot recommend publication. First, the method is not placed in the context of closely related prior work: the coefficient sampler is known (Sakov and Oke, 2008), and RMPF is a kernel-shrinkage particle filter whose bandwidth does not vanish as N grows and which tends to the Gaussian particle filter as n grows (comment 1). Recognising this would both shorten and correct the theory (comments 2 and 6). Second, the experiments do not support the conclusions. The baseline particle filter has collapsed in both experiments, and no ensemble Kalman filter (EnKF) or regularised particle filter is included. In my own runs of the stated configurations, such baselines perform comparably to or better than RMPF, with far fewer members (comments 3 and 5). The Lorenz-96 results are also inconsistent with those of Nakano et al. (2007), and the setups are not described in enough detail to resolve this (comment 7). Third, the claim that RMPF resolves the RMSE saturation of MPF is not supported (comment 4). Addressing these issues requires new experiments and is likely to change the conclusions. I therefore recommend **rejection**.

Major comments

1. Relation to existing methods

(a) Each column of the mean-preserving random orthonormal matrices of Sakov and Oke (2008, Eqs. 17–19) satisfies both constraints of Eq. (3), and is distributed exactly as the coefficients of Sect. 3.3. Related constructions appear in the second-order exact sampling of the SEIK filter (Pham, 2001). These works should be cited, and Sect. 3.3 shortened. Also, merging each group of n parents into n particles, using the n columns of one such matrix, would preserve each group's sample mean and covariance exactly (cf. Appendix A), at no extra cost.

(b) Writing a merged particle as $\mathbf{x}_{\text{new}} = \bar{\mathbf{x}}_I + \sqrt{\frac{n-1}{n}} \sum_j v_j (\hat{\mathbf{x}}_j - \bar{\mathbf{x}}_I)$ shows that the centres $\bar{\mathbf{x}}_I$ have covariance $\approx \Sigma/n$, and the kernel around them $\approx (1 - 1/n)\,\Sigma$. This is the kernel shrinkage of Liu and West (2001) with $h^2 = 1 - 1/n$, except that the centres are group averages. Consequently:

- Since $n \geq 3$, $h^2 \geq 2/3$ does not vanish as $N \to \infty$, so RMPF is not consistent for any fixed n (as the authors' own Eq. 41 indicates: its bias term is independent of N).

- As $n \to \infty$, RMPF tends to Gaussian resampling, i.e. the Gaussian particle filter of Kotecha and Djurić (2003), which is precisely what Nakano et al. (2007) positioned MPF against (and not the Gaussian *sum* particle filter cited in Sect. 9.2). The best Lorenz-96 results, at $n = 40$–100 ($h^2 \geq 0.975$), thus come from near-Gaussian resampling, which makes an EnKF baseline essential (comment 3).
- The observations enter only through the weights, and merging follows the weighting: particles remain in the affine span of the resampled particles and are never moved towards the observations, as they are by Kalman-type updates. Merging can thus remove duplicates, but does not obviously alleviate weight degeneracy.

(c) Sect. 9.2 contrasts RMPF with the regularised PF (RPF) on grounds that do not hold. RMPF's performance also "hinges on bandwidth", namely n (to which Sect. 8 is devoted). And the RPF does not "construct" a KDE: it samples from it by jittering, which, using the ensemble anomalies, is itself a random linear combination of particles, at a cost comparable to merging. Sect. 9.2 should be rewritten to position RMPF with respect to the above, as well as to PF–EnKF hybrids and localised particle filters.

2. Sections 4 and 5 can be greatly condensed and corrected

The density of a merged particle (Sect. 4, Appendix B) follows from classical results: $\mathbf{x}_{\text{merged}} - \bar{\mathbf{x}}$ is a linear image of a uniform point on S^{n-2}, hence elliptically contoured (e.g. Cambanis et al., 1981), with the radial profile of a marginal of the uniform distribution on a sphere, as in Eq. (24). Citing this, dropping the unused normalisation constant, and stating the intuition (the shape is governed by the number of projected-out dimensions, from an ellipsoid surface, via uniform, to centrally concentrated and Gaussian) would replace several pages. Sect. 5 largely repeats Sect. 4 (and both are restated in Sect. 9.1); they should be merged.

More importantly, the relevant dimension is not d, but the rank $k = \min(n_{\text{distinct}} - 1,\ d)$ of the parents' anomalies. The analysis assumes n distinct parents, but the parents are drawn by resampling, with replacement, so duplicates are common precisely when the weights are degenerate. Duplicates collapse the kernel onto a lower-dimensional subspace (down to a copy of the parent), and rule out the "surface" case. For example, with two particles of weights 0.8 and 0.2 and $n = 3$, half of the merged particles are exact copies. In general, a group is duplicate-free with probability $\approx \exp(-\frac{n(n-1)}{2N_{\text{eff}}})$, which for $n = 40$ requires $N_{\text{eff}} \gg 780$, unlikely in the Lorenz-96 experiments. The classification of Table 2, the "continuous and smooth" mixture (Sect. 4.3.2), and the "transition" around $n \approx d$ (Sect. 7.2.1) should be revised accordingly. Similarly, the experiment of Sect. 5.2 uses distinct, equally spaced parents with $n = N$, which is exactly the case that resampling does not produce.

Finally, the advantage of random over fixed coefficients (Figs. 1, 3) is not demonstrated in the filter: fixed-coefficient MPF performs comparably for moderate N (Table 3), indistinguishably for $n = 3$ on Lorenz-96 (Table 5), and much better at small N ($N = 128$: 0.265 vs 7.57 for RMPF with $n = 3$). The authors should report the fraction of duplicate merged particles under both schemes, e.g. in an experiment without model noise.

3. Lorenz-96: inconsistent with Nakano et al. (2007), and no EnKF baseline

The configuration of Sect. 7.1 appears to be that of Nakano et al. (2007, Sect. 4.2), where the PF attains an RMSE of 0.77 at $N = 262\,144$, MPF about 0.80, and the EnKF about 0.87. Here (Table 5) the PF attains 2.70, fixed-coefficient MPF 2.6–4.6, and the best RMPF 2.09, and many entries

are close to the climatological level (≈ 3.6). The headline 22% improvement over the PF is thus a comparison between filters that do not track the truth.

I ran the configuration of Sect. 7.1 with a standard DA benchmarking framework (perfect model, since no system noise is specified; one realisation; analysis RMSE):

Method	N	RMSE
Climatology	–	3.63
Square-root EnKF, no localisation (inflation 1.02)	40	0.45
Bootstrap PF, no regularisation	1024, 8192	5.1 (diverged)
Regularised PF (coarsely tuned jittering)	8192	1.47

The EnKF is about five times more accurate than the best RMPF, with about 6000 times fewer members, and a simple regularised PF with $N = 8192$ beats RMPF with $N = 262\,144$. The setups may differ (comment 7), but Table 5 evidently lacks competitive baselines. The authors should fully specify the setup, explain the discrepancy with Nakano et al. (2007), for which the system noise ($0.25\,\mathbf{I}$) and the inflated likelihood ($\sigma = 3$) used there may be responsible, and include an EnKF baseline.

4. The claim that RMPF "resolves" the saturation of MPF

The saturation of MPF is attributed to the merging step distorting the distribution, but RMPF retains this mechanism, being asymptotically biased for any fixed n (comment 1b). Much of the gain comes instead from computing the estimate before merging, which improves *all* merging methods by 7–10% (Tables 4, 7), including fixed-coefficient MPF. That estimates are best computed before resampling is standard (e.g. Chopin, 2004, Sect. 2.4), so this should not be presented as a methodological contribution. The effects of the two modifications should be separated, and the claims toned down.

5. Small-N behaviour and the Lorenz-63 baseline

At small N, RMPF is often much worse than fixed-coefficient MPF (Table 3); this deserves investigation, since this is the regime relevant in geophysics. Moreover, $N \leq 128$ is called "extremely small" (Sect. 6.2.1), but only relative to an unregularised bootstrap PF. Running the configuration of Sect. 6.1 as in comment 3:

Method	N	RMSE
Climatology	–	7.59
Square-root EnKF (inflation 1.02)	10	0.33
Bootstrap PF, no regularisation	32–1024	10–11 (diverged)
Regularised PF (jittering)	32 / 128 / 1024	0.28 / 0.21 / 0.17

A lightly regularised PF tracks the truth with 32 particles, and keeps improving with N, rather than saturating; a 10-member EnKF is comparable to the MPF and RMPF results of Table 3 for $N \leq 128$. (Exact comparison requires the definition of the average RMSE, see comment 7.) The authors should include a regularised PF (Musso et al., 2001) and an EnKF as baselines.

6. Selection of the merging size (Sect. 8)

Sect. 8.1 is not an analysis: $\lambda_{n,r}$ is undefined, Eq. (37) is heuristic, a cross term is dropped, and Fig. 7 is qualitative. The variance term also appears to be mis-derived: the mixture of Eq. (32) sums over all index tuples and is thus deterministic given the particles, so the fluctuations of a single group (Eqs. 38–39) do not yield the $C_t(N)/n$ scaling of Eq. (40). The correct argument is the standard one of kernel density estimation, with n as bandwidth (Silverman, 1986; comment 1b).

Sect. 8.2: likelihood-based tuning is sensible but not new (e.g. Kitagawa, 1996; Carrassi et al., 2017), and should be presented as such, replacing the unusual references (fault diagnostics, lecture notes, Douc et al., 2004) by Gneiting and Raftery (2007). The evidence is weak: Figs. 8 and 9 use one ensemble size and apparently one realisation; for Lorenz-63 the RMSEs span only 0.235–0.252, likely within sampling error; and for Lorenz-96 all RMSEs (3.6–4.0) are at or above climatology, where a correlation between likelihood and RMSE is unsurprising and uninformative. *Sect. 8.3* offers no concrete guidelines.

7. Reproducibility

The setups are not described in enough detail for the results to be reproduced, and, as comments 3 and 5 show, the conclusions depend on the missing details. The manuscript should be self-contained, particularly as the replication package is under restricted access. Missing or ambiguous are: (1) the system noise, both in the truth and in the forecast (Nakano et al., 2007, use $0.01\,\mathbf{I}$ for Lorenz-63 and $0.25\,\mathbf{I}$ for Lorenz-96); (2) the observation-error standard deviation used in the likelihood; (3) any jittering or inflation, for each method; (4) the resampling scheme, when resampling is triggered, and how particles are grouped (minor comment 2); (5) whether coefficients are re-used (Sect. 3.2), and the justification for the three fixed-coefficient sets of Table 3; (6) the initial conditions; (7) the integration scheme for Lorenz-96; (8) the definition of the "average RMSE" (analysis times or all time steps; average of RMSE or root of average squared error); (9) the number of realisations behind each table entry.

Minor comments

1. Algorithms 1–3 are nearly identical, and Algorithm 1 is unnecessary. A short prose description, with the merging step as an equation, would be clearer.
2. The covariance argument (Appendix A) requires independent parents. Systematic and residual resampling return sorted indices, so merging the i-th members of n resampled sets combines (near-)identical particles. Please state the resampling scheme, and whether particles are permuted before grouping.
3. Sect. 4.2: the merged particles lie on an $(n-2)$-dimensional, not $(n-1)$-dimensional, ellipsoid surface (cf. Fig. 1, where $n = 3$ gives a circle). Also, $(\mathbf{X}\mathbf{X}^\top)^+$ (Sect. 4.2) and $(\mathbf{X}\mathbf{P}\mathbf{X}^\top)^+$ (Sect. 4.3) should be unified.
4. Sect. 6.1 states 12 levels of N, but Table 3 shows 9.
5. **Abstract.** Half of it describes MPF's problems, and the results are not quantified. "However, these coefficients are chosen empirically" (L. 7) repeats L. 4–5. Terminology: calling the coefficients "weight parameters" (L. 5–6) invites confusion with importance weights; "merging dimension" (L. 9) is undefined and misleading (n is a count, not a dimension; Sect. 3.2 calls it "merging size"); use "bootstrap PF" consistently instead of "basic"/"standard" PF. "Good RMSE" (L. 3) should be "low RMSE". The bias–variance "theoretical analysis" (L. 9–10) and "highly accurate estimation" (L. 11) overstate the results.

6. **Introduction.** It is only three paragraphs, while the literature review is deferred to Sect. 2. Merging the two would allow positioning RMPF properly (comment 1). Several citations do not support the claims they accompany: L. 14–15 cites no meteorological or oceanographic work for those applications; L. 19–20 cites Nakano et al. (2007) for general trade-offs of other methods; the relevance of Shen et al. (2017) is not explained. L. 16: in the geosciences, EnKFs and variational methods are far more common than particle filters. L. 18: "weight bias" is not standard terminology. L. 18–19: Snyder et al. (2008) show that the required ensemble size grows exponentially with the variance of the log-likelihood, which is not the same as the effective sample size "decreasing exponentially". L. 84 ends with "as follows:", but Algorithm 2 appears on the next page. Several works are cited twice in one sentence (e.g. L. 70–76, and Muller (1959) in Sect. 3.3). The hat denotes both the estimate and the resampled particles (Table 1).
7. **Discussion and conclusions.** Sects. 9–10 restate the density (Eq. 43 repeats Eq. 24) and Table 2 again, and should be merged and shortened. Sect. 9.1: Fig. 1 is not the Lorenz-63 experiment; the pre- vs post-merging difference is resampling noise rather than "information loss" (comment 4); "Davis et al. (2011)" is a reprint of Rosenblatt (1956). Sect. 10.2: "$n < r + 2$ … edge concentration" is inconsistent with Table 2. Sects. 10.1, 10.3 and 10.5 overstate the results (e.g. "high-dimensional applicability … from 3 to 40 dimensions"; particle savings "in both Lorenz63 and Lorenz96 models", shown only for Lorenz-96; "greatly enhance the practicality of particle filters"). An explicit statement of limitations is needed: poor small-N performance, asymptotic bias for fixed n, no localisation, and experiments restricted to linear, Gaussian observations.