Major comments:

1.a. Thanks for the equation. Your readers would benefit if the equation were integrated into the existing discussion, which has otherwise not been modified at all.

Additional discussion has been added throughout section 2.3 to better explain the implications of this equation (20). Page 9, lines 12-14, lines 22-26; Page 10, lines 1-7.

The new equation also introduces new notation for the “resampling index” a(ni,ki). Equation (13) does not explain this.

An explanation has been added for this analogous gridpoint-dependent index. Page 6, lines 17-20.

I suggest also stating that the index a is defined by
$X^a_{mi,ki} = X^b_{mi,a(mi,ki)}$.

We have included the additional equation suggested above by the reviewer into the text following equation (13). Page 6, line 20, eqn (14).

(I hope that is true, else I am completely confused.)

This is true, and it is also reflected in the pseudo-code that had been added in the appendix. Page 18, line 19.

b. I still believe that the authors should elaborate on the length scale for the weight smoothing and its role in the algorithm (namely, to smooth--spatially--the discontinuous weights produced by resampling). They have not modified the text. I don’t think it’s fair to say (as in their response) that there is no length scale.

We have added discussion to address this point at the end of section 2.3. Pages 9-10.

We rely on the observation-space localization to generate a zeroth-order smoothing of the weights, with a further smoothing correction formed by a point-wise convex combination computed in the ensemble-space. In the latter correction, there is no explicit length scale in the sense that in the model/physical space the ‘length scale’ of the smoothing is 0. We have elaborated in the text to make this point more clear at the end of section 2.3. Pages 9-10.

As far as I can tell from the new (unnumbered) equation, in the case of one spatial dimension and at a discontinuity introduced by resampling, the weights will vary linearly from 0 to 1 (or vice versa) across the set of neighborhood points.

This has been interpreted incorrectly. We clarify - regardless of the number of dimensions, the weights vary from 0 to 1 in a convex combination (i.e. sum to 1) only at the single grid point. The neighbor points are not explicitly affected by this calculation. We now clarify this point in the text. Page 9, lines 22-26. Page 10, lines 1-7.

As an example, Figure 1 illustrates this point for a 2-dimensional space.

Thus, the smoothness of the spatial transition is definitely set by the radius chosen for the neighborhood.

The total smoothing is achieved via a combination of localization (the selection of observations local to this grid point, which are highly overlapping across contiguous points - this is equivalent to LETKF) and smoothing of weights locally in the ensemble space at one specific grid point. Using a larger radius for the selection of neighbor points does not necessarily increase the smoothness of the results, as additional sampling noise may be introduced. Examples were demonstrated in the previous response to the reviewers.

This is now summarized on Page 9, lines 22-26. Page 10, lines 1-7.

The choice of neighbor points can only impact the proportion of each member used at a given grid point in the construction of the convex combination at that grid point. Now noted on Page 9, lines 12-14.

Moreover, the fact that radii larger than a single grid interval don’t seem to help in L96 is hardly proof that the radius of neighborhood points will not matter in more realistic applications.

We did not intend to make any claims about what the impact might be to a more realistic model until evaluating the approach with such a model.

2a. The authors’ response--that the results for fewer than 20 observations merely reflect filter divergence--simply raises the question of why the LPF should diverge as the number of observations decreases.

Control theory concepts regarding observability indicates that with too few observations any filter will diverge. In the linear theory for autonomous systems the conditions are straightforward. For chaotic systems, the observability of a system is difficult to identify analytically. For the particular experiment parameters given here, these results indicate that minimum number of required observations is approximately 20 observations per model equivalent of 60-hour update intervals.

Abarbanel et al. (2009) have made analogous conclusions regarding observability with L96 using synchronization methods, and Whartenby et al. (2013) for shallow water flows.

Abarbanel, H.D.I., D.R. Creveling, R. Farsian, and M. Kostuk, 2009: Dynamical State and Parameter Estimation. *SIAM J. Appl. Dyn. Syst.*, 8(4), 1341–1381. DOI:10.1137/090749761

Whartenby, WG, Quinn JC, Abarbanel HDI. 2013: The number of required observations in data assimilation for a shallow-water flow. Monthly Weather Review. 141:2502-2518.

For the LETKF, it is easy to see how having fewer observations (and thus larger prior errors and more nonlinearity) can lead to divergence; it will increasingly underestimate the posterior spread since the mean update will be steadily degraded by nonlinearity and non-Gaussianity, and it will also suffer from the sensitivity of its updates to non-Gaussian effects (Lawson and Hansen 2004 MWR; Lei et al. 2010 MWR). The LPF, on the other hand, is designed to handle arbitrarily nonlinear/non-Gaussian situations.

We do not claim the LPF is designed to handle arbitrarily nonlinear situations. It is designed to emulate the qualities of a standard particle filter that requires exponentially larger ensemble sizes to function. We do not claim that the LPF is insensitive to system nonlinearities when under-constrained. The LPF still suffers from small sample sizes, but we have demonstrated that it does not suffer to the same degree as the traditional non-localized particle filter and has potential for use in more realistic applications.

Why is it diverging? (And, if divergence is the problem, why don't both the errors of both the LETKF and the LPF approach the same value, twice the climatological standard deviation?)
c. Stating that the LPF diverges again does not address why the LPF should behave poorly as the number of observations decreases.

The results show the average error over the entire experiment for many (1600) different parameter settings. The difference in value indicates that with insufficient observations (<20 with model-equivalent 60-hr forecasts) the LPF diverges more often than LETKF. We clarify this point on Page 13, line 30.

There is an important difference between sequential DA methods and particle filters. The former has a forcing term built into the algorithm, thus no matter how far the members may stray from the true state in an EnKF, there is a mechanism to drive the system towards the observations. In a particle filter this is not the case - a diverging PF relies on random chance to generate members that are close enough to the observations to survive the next resampling step and produce quality samples for the next round of the procedure.

We now address this point at the end of section 2.4. Page 10, lines 27-32.

We also refer to this discussion in presenting the results. Page 14, lines 1-3.

New, minor comment:

i. Equations (12-14) are confusing as they stand. I suggest removing set notation from (12) and (14): b = [1 2 3 … k], e_j = [0 … 0 1_j 0 … 0].

To accommodate the reviewer’s request for improved clarity while maintaining our preferred notation, we have included additional descriptive text to benefit a general audience. Page 6, lines 15-22.

In (13), replace = with \in (i.e. “is an element of”).

Thanks, we’ve made this correction to eqn (13).