6B.1A TIME-EXPANDED SAMPLING FOR ENSEMBLE-BASED FILTER WITH COVARIANCE LOCALIZATION: ASSIMILATION EXPERIMENTS WITH A SHALLOW-WATER EQUATION MODEL

Qin Xu[1]*, Li Wei[2], Huijuan Lu[2], Qingyun Zhao[3] and Chongjian Qiu[4]

[1]NOAA/National Severe Storms Laboratory, Norman, Oklahoma
[2]Cooperative Institute for Mesoscale Meteorological Studies, University of Oklahoma
[3]Naval Research Laboratory, Monterey, California
[4]College of Atmospheric Sciences, Lanzhou University, China

1. INTRODUCTION

When the probability density function (pdf) of the model state is represented by an ensemble of state vectors in an ensemble-based filter, such as the ensemble square root filter (EnSRF) – a variant of the EnKF (Whitaker and Hamill 2002; Tippett et al. 2003), the mean and covariance are directly estimated from and updated through the ensemble. This makes the filter much easier to implement than the four-dimensional variational assimilation (4DVar). Computationally, however, an ensemble-based filter is still very expensive for operational applications due to large ensemble sizes required by the Monte Carlo method that the filter is based on. Theoretically, the ensemble size should be sufficiently large to adequately represent the pdf. Practically, however, an ensemble-based filter has to use a limited or even small ensemble to reduce the computational cost. How to deal with problems caused by limited or small ensemble is thus a primary issue for ensemble-based filters. Another primary issue for ensemble-based filters (also for 4DVar) concerns how to solve or alleviate problems caused by unknown model errors. In this paper, a time-expanded sampling is proposed to treat or alleviate problems concerned above in a simple and effective way. The potential merits of the proposed approach are demonstrated by assimilation experiments with a shallow-water equation model.

2. DESCRIPTION OF THE METHOD

It is well recognized that a weather system predicted by a numerical model often develops and/or propagates either faster or slower than the truth and thus often contains amplitude and/or phase errors. If the predicted system develops or propagates faster (or slower) than the true system, the predicted field at a time level before (or after) the analysis time will very likely better represent the true field than the predicted field at the analysis time, at least over the local area covered by the weather system. Thus, the difference between the predicted field sampled before (or after) the analysis time and the one at the analysis time may represent the prediction error associated with the aforementioned amplitude and/or phase errors. This implies that the mean and localized covariance may be computed from an ensemble of predicted fields sampled before, at and after the analysis time within a certain time window.

Based on the above considerations, we propose a time-expanded sampling approach for ensemble-based filters. For an ensemble of size N, the conventional approach with an ensemble-based filter updates N ensemble members at each analysis step to initialize N prediction runs and then propagates the ensemble to the next analysis time. At each analysis step, only one perturbed state vector is sampled at the analysis time from each prediction run. In our proposed approach, however, a series of perturbed state vectors are sampled from each prediction run not only at the analysis time but also at other time levels properly selected in the vicinity of the analysis time. All the sampled state vectors are used to compute the background covariance for the analysis so that the ensemble size is increased without increasing the number of prediction runs. For a given ensemble size of N, the number of prediction runs are thus reduced from N to N/S and so is the computational cost, where S is the number of sampling time levels.

The EnSRF serial algorithm (Whitaker and Hamill 2002) is used in this paper to test the proposed time-expanded sampling approach in comparison with the conventional approach. This EnSRF algorithm uses a compactly-supported smooth correlation function to localize the background covariance computed from the ensemble and thus to reduce the adverse impact of noise and spurious long-range correlation caused by limited ensemble size (Houtekamer and Mitchell 2001). The localization is necessary for an ensemble-based filter when the ensemble size is not sufficiently large. The EnSRF serial algorithm is particularly convenient for covariance localization and thus is used with the proposed time-expanded sampling in this study.

* Corresponding author address: Qin Xu, National Severe Storms Laboratory, 120 David L. Boren Blvd., Norman, OK 73072-7326; E-mail: Qin.Xu@noaa.gov

3. MODEL SYSTEM AND DESIGN OF EXPERIMENTS

3.1 *Model Equations and Simulated Observations*

To test the proposed approach, assimilation experiments are designed with a shallow-water equation model. The shallow water equations are formulated in the *f*-plane by

$$\partial_t u = -u\partial_x u - v\partial_y u + fv - g\partial_x h + \mu(\partial_x^2 + \partial_y^2)u,$$
$$\partial_t v = -u\partial_x v - v\partial_y v - fu - g\partial_y h + \mu(\partial_x^2 + \partial_y^2)v,$$
$$\partial_t h = -u\partial_x h - v\partial_y h - (H + h)(\partial_x u + \partial_y v) + \mu(\partial_x^2 + \partial_y^2)h.$$

Here, (u, v) denote the velocities, h is the perturbation height, $H = 3000$ m is the basic-state depth; $f = 10^{-4}$ s^{-1} is the Coriolis parameter at 45^o N, $g = 9.8$ m s^{-2} is the acceleration of gravity, and $\mu = 10^5$ m^2 s^{-1} is the coefficient of eddy diffusivity. The model domain is a square with periodic boundary conditions at $x, y = 0$ and D, where $D = 13200$ km is the length of one side of the model domain. The spatial derivatives are discretized by the first-order central finite difference scheme. The local time derivatives are discretized by using the two-step backward difference scheme of Matsuno (1966).

The "true" state is produced by integrating the above model on an 89×89 grid with $D = 88d$, where $d = \Delta x = \Delta y = 150$ km is the grid spacing. The time step is set to $\Delta t = 6$ min to ensure the computational stability. The integration is initialized 48 hours before the starting time of data assimilation with the following geostrophically balanced initial conditions:

$$h = H\{1 + (y'/\pi)\exp(-2y'^2)[1 + 0.1\sin(4\pi x/D)]\}^{-1},$$
$$u = -f^{-1}g\partial_y h \text{ and } v = f^{-1}g\partial_x h \text{ at } t = -48 \text{ hours}, \quad (1)$$

where $y' = 2\pi y/D - \pi$ for $0 \le y \le D$ over the model domain. This initial state is a zonal jet flow superimposed with a small-amplitude wave, which is similar to the initial state used in section 5.2 of Frank and Reich (2002) for a barotropic instability experiment. The wave grows to moderate amplitude after 48 hours of integration, and then develops further to large-amplitude wave perturbations and nonlinear eddies in the subsequent 5.5 days over the time period of data assimilation.

By adding samples of simulated observational error to the above model-produced "true" fields on a coarse grid spaced every 900 km in the *x*- and *y*-directions, "observations" are generated every 12 hours during the time period from $t = 12$ to 132 hour. The observation errors are uncorrelated between different variables and different points in both space and time. The observation error standard deviations are 12 m for h and 1.2 m s^{-1} for u and v. The above generated observations can be sorted into three types: (i) height h observations only (type-1), (ii) velocity (u, v) observations only (type-2), and (iii) both height and velocity observations (type-3). To examine the robustness of the proposed approach with respect to incomplete observations, we will consider mainly the type-1 observations.

3.2 *Experiment Set-up*

In all the experiments, the prediction model uses a doubly coarsened grid (with d increased from 150 to 300 km) so the model is not perfect. The initial background state (at $t = 0$) is also imperfect and is produced by averaging "true" fields over 4 days centered at $t = 0$. The background root-mean-square (RMS) errors (estimated by spatially averaged RMS differences between the background state and the "true" state) are 28 m in height h and 3.4 m s^{-1} in velocity $\mathbf{v} = (u, v)$. We denote by Nr the number of perturbed predictions runs, so the ensemble size is $N = N\text{r}S$ where S is the number of sampling time levels as introduced earlier. The Nr prediction runs are initialized at $t = 0$ (the starting time of data assimilation) by adding Nr geostrophically-balanced pseudo-random fields to the above background state. The pseudo-random height fields are generated by using the spectral method in Appendix E of Evensen (2003) with the de-correlation length set to $3d$ (= 900 km) and the standard deviation set to 22 m (which is close to the above estimated background RMS error in h). The pseudo-random velocity fields are given by the geostrophic balance conditions as in Eq. (1) and their standard deviation for $N\text{r} = 30$ (or 10) is very close to (or slightly larger than) the above estimated background RMS error in $\mathbf{v}$.

Three control experiments, denoted by E5, E10 and E30, are designed with the conventional approach in which $S = 1$ and $N = N\text{r} = 5$, 10 and 30, respectively. To optimize the performance of these experiments, (especially the E10), the covariance computed from the ensemble forecasts in each assimilation cycle is localized by using a compactly-supported smooth correlation function (Gaspari and Cohn 1999) with the localization cut-off radius tuned to $12d = 3600$ km (which is optimal for the control experiment E30). The time length of each data assimilation cycle is the same as the observation time interval, which is $T = 12$ hours.

With the time-expanded sampling, the ensemble consists of both the standard members and the time-expanded members. The Nr standard members are sampled at the analysis time from Nr perturbed runs. The time-expanded members are sampled at equally separated time levels before and after the analysis time. In particular, the sampling time levels centered at the jth analysis time are $t = t_j + m\tau$ for $m = 0, \pm1, \pm2, \ldots \pm M$, where $t = t_j$ denotes the jth analysis time ($j = 1, 2, \ldots 11$) and τ is the sampling time interval. For each analysis, the number of sampling time levels is $S = 2M + 1$ and the sampling time window is $2M\tau$. The ensemble size is thus increased by $S = 2M + 1$ times, that is, from $N = N\text{r}$ to $N\text{r}S$.

To examine the effectiveness of the time-expanded sampling in comparisons with the conventional approach, test experiments are designed by setting Nr to be the same as in E10 but with $S = 2M + 1 > 1$. For the experiment with Nr = 10 and S = 3, the initial 10 standard ensemble members are generated by random perturbations in the same way as in E10, but the total ensemble size for each analysis is expanded to $N = N\text{r}S$ = 10×3 by the time-expanded sampling. We denote this experiment by E10×3τ3 (or E10×3τ5, …) for τ = 3 (or 5, …) hours. Similarly, E5×3τ3 (or E5×3τ5, …) represents the experiment with Nr = 5 and S = 3 for τ = 3 (or 5, …) hours.

As we will see later, when M is increased from 0 to 1 and thus S is increased from 1 to 3, the analyses are improved significantly compared with those obtained from the conventional approach with S = 1. However, when M is further increased from 1 to 2 and S from 3 to 5, the time-expanded members are sampled more frequently and become more similar to each other (as the sampling time window is confined between the previous analysis time and next one, that is, $2M\tau \le 2T$). In this case, the analyses can hardly be further improved but the computational costs will increase. Because of this, the test experiments are performed with S = 3 only and compared with the control experiments with S = 1.

4. RESULTS AND DISCUSSIONS

4.1 *Optimal Sampling Time Interval*

The above designed control and test experiments are conducted with different settings of Nr (= 30 10, 5), S (= 1, 3, 5) and τ (from 3 to 9 hours). The performances of the filter in each experiment can be evaluated by spatially-averaged RMS errors of the assimilated ensemble mean height $\overline{h}$ and velocity $\overline{\mathbf{v}} = \overline{(u,v)}$. These two RMS errors are defined by

$$\sigma_h \equiv \{\{(\overline{h} - h^t)^2\}\}^{1/2} \text{ and } \sigma_v \equiv \{\{|\overline{\mathbf{v}} - \mathbf{v}^t|^2\}\}^{1/2}, \quad (2)$$

where the overbar denotes the ensemble average, $\{\{\ \}\}$ the spatial average, and $(\ )^t$ the "true" value of $(\ )$. The velocity RMS errors computed from E5, E10, E30, E10×3τ4, E10×3τ5, and E10×3τ6 with the type-1 (height only) observations are plotted as functions of cycle number over the entire data assimilation period (from t = 0 to 132 hour) in Fig. 1. As shown, except for E5, the filter reduces the RMS error at each analysis time and an overall convergence has been reached through the assimilation in each experiment in spite of the vigorous growth of the forecast RMS error between every two adjacent analyses (from one cycle to the next) caused by barotropic instability and unknown model errors. Measured by the RMS error σ_v plotted in Fig. 1, the three test experiments E10×3τ4, E10×3τ5 and E10×3τ6 perform much better than E10. For the last four assimilation cycles, these test experiments even slightly outperform E30. Among the test experiments, E10×3τ5 performs best. These results suggest that (i) the performance of an ensemble-based filter can be improved by time-expanded sampling, and (ii) the improvement can be optimized by properly selecting the sampling time interval.

The RMS errors σ_h and σ_v for the ensemble mean forecast (from t = 120 to 132 hour) and analysis (at t = 132 hour) in the last assimilation cycle are listed in the first four columns of Table 1, where nine experiments are listed and they are all performed with the type-1 (height only) observations. Measured by the RMS errors listed in Table 1, the four test experiments E10×3τ3 E10×3τ4, E10×3τ5 and E10×3τ6 perform significantly better than E10 and slightly better than E30. Again, among the four test experiments, E10×3τ5 performs best. As Nr reduces to 5, E5 fails to converge (as shown in Fig. 1), but E5×3τ5 converges well, and the convergence can be further improved when τ is further increased up to τ = 9 hours as shown by E5×3τ9 in Table 1. Similar experiments have been performed with the type-2 (velocity only) and type-3 (both height and velocity) observations, and the results (not shown) are qualitatively the same as those in Table 1. All these results support the aforementioned two points, and their implications are further examined in the next subsection.

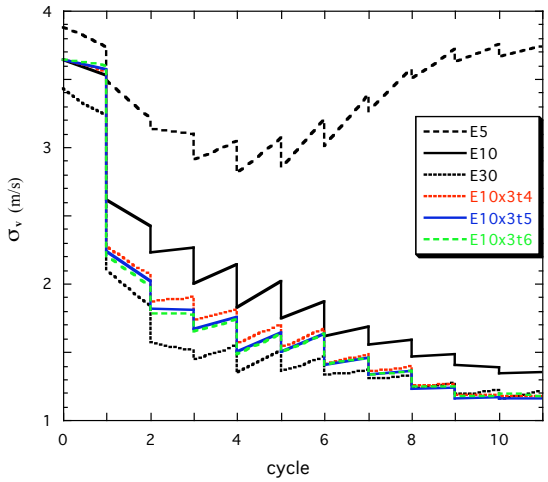

Fig. 1. Velocity RMS errors σ_v plotted as functions of cycle number over the entire assimilation period (from t = 0 to 132 hour) for three control experiments (E5, E10 and E30) and three test experiments (E10×3τ4, E10×3τ5 and E10×3τ6) performed with the type-1 (height only) observations. The drop of the error curve at each analysis time (every 12 hours) corresponds the error reduction made by the analysis. See Eq. (2) for the definition of σ_v.

Table 1. RMS errors [σ_h and $\sigma_\mathbf{v}$ defined in Eq. (2)] and normalized consistency ratios [r_h and $r_\mathbf{v}$ defined in Eq. (3)] for the ensemble forecasts (denoted by superscript f) and analyses (denoted by superscript a) produced in the last assimilation cycle. Nine experiments are listed and they are all performed with the type-1 (height only) observations.

Exp#	σ_h^f	σ_h^a	$\sigma_\mathbf{v}^f$	$\sigma_\mathbf{v}^a$	r_h^f	r_h^a	$r_\mathbf{v}^f$	$r_\mathbf{v}^a$
E5	28.981	26.231	3.748	3.630	0.109	0.110	0.140	0.135
E5×3τ5	12.316	9.677	1.822	1.795	0.546	0.418	0.581	0.407
E5×3τ9	9.315	7.511	1.473	1.437	1.192	0.574	1.133	0.535
E10	8.336	7.224	1.358	1.297	0.474	0.485	0.485	0.475
E10×3τ3	6.959	5.779	1.203	1.150	0.737	0.681	0.688	0.606
E10×3τ4	6.637	5.624	1.173	1.132	0.903	0.730	0.810	0.646
E10×3τ5	6.375	5.560	1.150	1.127	1.070	0.757	0.941	0.674
E10×3τ6	6.529	5.792	1.167	1.149	1.158	0.736	1.049	0.688
E30	7.050	5.872	1.216	1.155	0.633	0.665	0.620	0.614

Table 2. As in Table 1 but for three control experiments with covariance inflation. The inflation factor c is optimally tuned in each experiment, and the tuned value of c is listed for each experiment.

Exp#	σ_h^f	σ_h^a	$\sigma_\mathbf{v}^f$	$\sigma_\mathbf{v}^a$	r_h^f	r_h^a	$r_\mathbf{v}^f$	$r_\mathbf{v}^a$
E5, c =1.5	13.196	10.769	2.334	2.312	0.585	0.667	0.655	0.686
E10, c = 1.1	6.602	5.726	1.226	1.191	0.842	0.877	0.814	0.827
E30, c =1.05	6.418	5.419	1.177	1.126	0.847	0.889	0.826	0.838

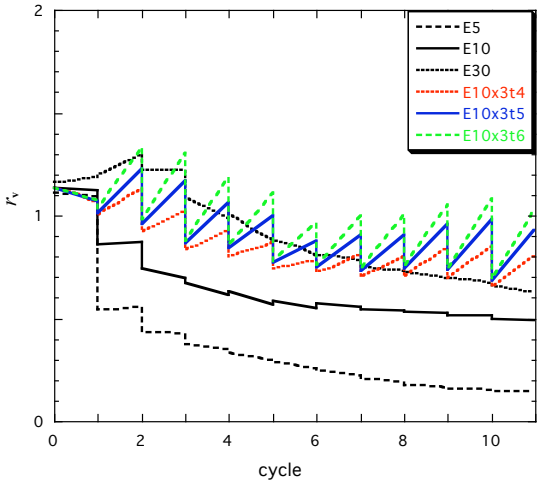

Fig. 2. As in Fig. 1 but for normalized consistency ratio $r_\mathbf{v}$. See Eq. (3) for the definition of $r_\mathbf{v}$.

4.2 *Ensemble Spread and Consistency Ratio*

The spatially averaged ensemble spreads in h and $\mathbf{v} = (u, v)$ can be defined by

$$s_h \equiv \{\{\overline{(h - \bar{h})^2}\}\}^{1/2} \text{ and } \quad s_\mathbf{v} \equiv \{\{\overline{|\mathbf{v} - \bar{\mathbf{v}}|^2}\}\}^{1/2}.$$

Ideally, the ensemble spread in each variable should satisfy a consistency relationship with the ensemble-mean RMS error. This consistency relation requires the ratio between the ensemble spread and the ensemble-mean RMS error to be statistically equal to $[N/(N+1)]^{1/2}$ (Murphy 1988). We call this ratio the consistency ratio. The normalized consistency ratios can be defined by

$$r_h \equiv (s_h/\sigma_h)[(N+1)/N]^{1/2} \quad \text{and}$$
$$r_\mathbf{v} \equiv (s_\mathbf{v}/\sigma_\mathbf{v})[(N+1)/N]^{1/2} \qquad (3)$$

for h and $\mathbf{v}$, respectively. If the uncertainties of the ensemble mean are well quantified by the ensemble spreads, then r_h and $r_\mathbf{v}$ should be statistically equal to 1 according to the consistency relationship. As mentioned in section 3.2, the initial Nr ensemble members are generated at the beginning of the assimilation by adding pseudo-random height fields and their associated geostrophic velocity fields to the background height and velocity fields, respectively. Since the initial ensemble spreads are close to the initial background RMS errors in h and $\mathbf{v}$, respectively, the normalized consistency ratios r_h and $r_\mathbf{v}$ are close to 1 initially in all the experiments. These ratios, however, undergo different variations with time over the assimilation period in different experiments.

The variations of $r_\mathbf{v}$ are shown in Fig. 2 for the same six experiments as in Fig. 1. As shown, $r_\mathbf{v}$ is reduced sharply by the first analysis (at t = 12 hour) in

E5 and then decreases continuously in subsequent cycles. The same feature is seen for r_h (not shown) in E5. Since r_h and r_v decrease rapidly in E5 and become increasingly far below the ideal value of 1 during the assimilation, the ensemble spread is too small to cover the true state. This explains why E5 fails to converge toward the true state. In E10, r_v and r_h (not shown but similar to r_v in Fig. 2) are reduced significantly by the first analysis and then decrease slowly and become nearly constant (around 0.5) toward the end of the assimilation. The ensemble spread in E10 is thus still not large enough to adequately cover the true state, and this explains the poor performance of E10. In E30, r_v and r_h increase (due to the decreases in σ_v and σ_h) after the first analysis and then decrease gradually in the subsequent. Clearly, the normalized consistency ratios in E30 are closer to the ideal value of 1 than those in E10. This explains why E30 performs significantly better than E10. In E10×3τ4, E10×3τ5 and E10×3τ6, r_v and r_h (not shown but similar to r_v in Fig. 2) are larger than those in E30, and they go up and down through each assimilation cycle. The sharp increases in r_v and r_h in each forecast step are caused mainly by the time-expanded sampling. Consequently, r_v and r_h can be kept close to the ideal value of 1 in these test experiments more persistently than in E30, especially in the last four assimilation cycles. This explains why these test experiments slightly outperform E30 in the last four cycles (see Fig. 1).

The normalized consistency ratios r_h and r_v for the ensemble forecasts and analyses in the last assimilation cycle are listed in the last four columns of Table 1. Judged by the closeness of the listed ratios to the ideal value of 1, the ratios produced in E10×3τ3 E10×3τ4, E10×3τ5 and E10×3τ6 are clearly better than those in E10 and slightly better than those in E30. The ratios in E5×3τ5 and E5×3τ9 are much better than those in E5. For Nr = 10, E10×3τ5 produces the best ratios. When Nr is reduced to 5, E5×3τ9 produces the best ratios. Clearly, the consistency ratios listed in the last four columns of Table 1 further support and explain the relative performances of the nine experiments evaluated in section 4.1 based on the RMS errors listed in the first four columns of Table 1.

By including the time-expanded members (in addition to the standard members as described in section 3.2), the time-expanded sampling can not only enlarge the ensemble size but also improve the ensemble spread. As explained in section 2, the time-expanded members represent possible amplitude and/or phase errors in the predicted fields. The inclusion of these members enhances the ensemble spread effectively in regions where the predicted fields undergo significant variations (such as local growths and/or propagations) over the sampling time period. In this sense, the time-expanded sampling can also enrich the spread structure and improve the spread representation of background forecast uncertainties.

As mentioned in section 3.2, the covariance computed from the ensemble forecasts is localized with the localization cut-off radius tuned to $12d$ = 3600 km so that the correlation is reduced to zero as the separation goes beyond the cut-off radius. Within this cut-off radial range, the localized covariance in E10×3τ5 exhibits a significant level of geostrophic balance and has roughly the same gross structure as those in E10 and E30 (not shown). As the structure differences between these localized covariances in E10×3τ5, E10 and E30 are relatively small in comparison with their gross structures, the localized covariance structures do not seem to be significantly improved or changed by the time-expanded sampling in E10×3τ5. Besides, because the model is imperfect and the unknown model errors are neglected, it is difficult to identify or evaluate possible structure improvements in the localized covariance caused by the time-expanded sampling. This problem deserves further investigation.

4.3 *Covariance Inflation*

With a limited ensemble size, the conventional EnKF tends to underestimate the analysis uncertainty especially when the model is imperfect. Because of this, matrix composed of the state vectors computed from the ensemble often needs to be inflated, and the inflation can be done simply by multiplying an empirically-tuned inflation factor, $c > 1$, to the computed covariance (Anderson 2001). In a sense, according to the results presented in section 4.2, the time-expanded sampling can be viewed as a way of inflation. What it does, however, is not just simply inflate the ensemble variance. It also enriches the spread structure and enlarges the space spanned by the ensemble. The time-expanded sampling is thus expected to be more effective than the conventional covariance inflation.

To compare the time-expanded sampling with the conventional covariance inflation, the control experiments E5, E10 and E30 are re-performed with the conventional covariance inflation. In each experiment, the inflation factor c is optimally tuned to minimize the RMS errors σ_h and σ_v of the ensemble mean analysis at the end of the assimilation (t = 132 hour). The velocity RMS errors are plotted as functions of cycle number in Fig. 3 for three optimally tuned control experiments (E5 with c = 1.5, E10 with c = 1.1 and E30 with c = 1.05) against two optimally tuned test experiments (E5×3τ9 and E10×3τ5) performed with the type-1 (height only) observations. The RMS errors σ_h and σ_v and normalized consistency ratios r_h and r_v obtained in the last assimilation cycle from each experiment are listed in Table 2, where the optimally tuned value of c is also listed for each experiment. The improvement made by the conventional covariance inflation in each experiment in Table 2 can be evaluated by the reduced values of

σ_h and σ_v and the increased values of r_h and r_v toward 1 compared to those from the same experiment without inflation (c = 1) listed in Table 1. By comparing Fig. 3 with Fig. 1 and comparing Table 2 with Table 1, it is clear that E5 is improved significantly by the conventional covariance inflation (with c tuned to 1.5) but it still far underperforms E5×3τ9 with c = 1 (as shown in Fig. 3 and listed in Table 1). E10 is also improved by the conventional covariance inflation (with c = 1.1) but still underperforms E10×3τ5 with c = 1. E30 is slightly improved by the conventional covariance inflation with c = 1.05 and it performs just as well as E10×3τ5. Thus, for given Nr = 5 (or 10), the filter performance is improved more by the time-expanded sampling than by the conventional covariance inflation.

Finally, it is necessary to point out that the localization cut-off radius used in this paper is fixed at 12d = 3600 km. This localization radius is optimally tuned for the control experiment E30 (as mentioned in section 3.2) but is not necessarily optimal for the test experiments and certainly not optimal for other two control experiments E10 and E5. Thus, without optimally tuning the localization radius for each experiment, the comparison presented in this paper is not complete. In view of this, additional experiments are performed to tune the localization radius for each experiment. The preliminary results indicate that the best test experiment with Nr = 5 (or 10) still outperforms the associated control experiment E5 (or E10) although the difference between each pair of the best test and control experiments is not as significant as presented above.

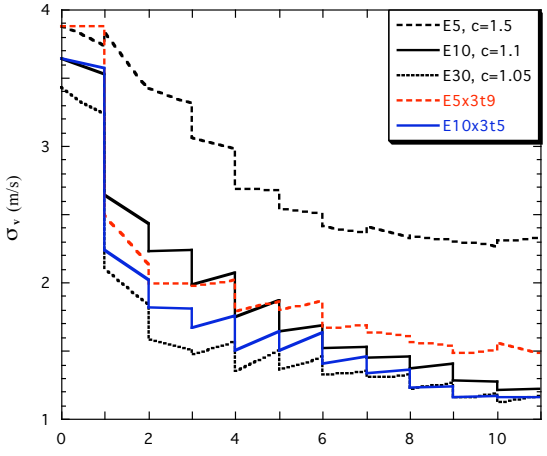

Fig. 3. As in Fig. 1 but for three optimally tuned control experiments (E5 with c = 1.5, E10 with c = 1.1 and E30 with c = 1.05) and two test experiments (E5×3τ9 and E10×3τ5) performed with the type-1 (height only) observations.

5. CONCLUSIONS

In this paper, a time-expanded sampling is proposed for ensemble-based filters with localized covariances in data assimilation. Assimilation experiments have been conducted with the EnSRF serial algorithm (Whitaker and Hamill 2002) applied to an imperfect shallow-water equation model and the proposed approach is demonstrated to have the following merits:

(i) By sampling ensemble members at several time levels in the vicinity of the analysis time rather than only at the analysis time (as the conventional approach does), the number of prediction runs (and hence the CPU cost) is reduced without reducing the ensemble size.

(ii) By adjusting the sampling time interval, the proposed approach can optimally enhance the ensemble spread and thus improve the filter performance even thought the number of prediction runs is greatly reduced.

(iii) As the ensemble spread is optimized (and inflated implicitly), the algorithm can alleviate the filter divergence problem caused by small ensemble in the presence of model errors (to a certain extent).

The assimilation experiments also suggest that the sampling time interval can be optimized approximately by tuning the ensemble spread based on the consistency relationship [see Eq. (3)] provided the RMS error of the predicted background mean field is known or well estimated (Xu et al. 2001). It is also possible to tune the sampling time interval adaptively if the RMS error of the predicted background mean field can be estimated and updated with the most recent ensemble forecasts in each assimilation cycle. This problem is under our investigation.

In this paper the time-expanded sampling is applied with the EnSRF serial algorithm to simulated observations of a developing wave excited owing to barotropic shear instability in a shallow-water equation model. This approach is also applied to simulated and real radar observations of a developing tornadic supercell storm with the EnSRF in a non-hydrostatic model (Xu et al. 2007; Lu 2007). The results are encouraging and will be highlighted at the conference.

Acknowledgements. The research was supported by the ONR Grant N000140410312 to the University of Oklahoma and by FAA contract IA# DTFA03-01-X-9007 to NSSL. The computational resources were provided by the OU Supercomputing Center for Education & Research at the University of Oklahoma. Funding was also provided to CIMMS by NOAA/Office of Oceanic and Atmospheric Research under NOAA-University of Oklahoma Cooperative Agreement #NA17RJ1227, Department of Commerce.

REFERENCES

Anderson, J. L., 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.*, **129**, 2884–2903.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99**, 10 143–10 162.

——, 2003: The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dyn.*, **53**, 343–367.

Frank, J., and S. Reich, 2002: A particle-mesh method for the shallow water equations near geostrophic balance. *J. Comput. Phys.*, **180**, 407–426.

Gaspari, G., and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.*, **125**, 723–757.

Houtekamer, P. L., and H. L. Mitchell, 2001: A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.*, **129**, 123–137.

Lu, H., 2007: Efficient approaches in applying ensemble Kalman filter to storm-scale radar data assimilation. Ph. D. Thesis, Institute of Atmospheric Physics, Chinese Academy of Sciences, 103 pp.

Matsuno, T., 1966: Numerical integration of the primitive equations by a simulated backward difference method. *J. Meteor. Soc. Japan* , **44**, 76-84.

Murphy, J. M., 1988: The impact of ensemble forecasts on predictability. *Quart. J. Roy. Meteor. Soc.*, **114**, 463–493.

Tippett, M. K., J. L. Anderson, C. H. Bishop, T. M. Hamill, and J. S. Whitaker, 2003: Ensemble square root filters. *Mon. Wea. Rev.*, **131**, 1485–1490.

Whitaker, J. S., and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev*, **130**, 1913–1924.

Xu, Q., H. Lu, S. Gao, M. Xue, and M. Tong, 2007: Time-expanded sampling for ensemble Kalman filter: Assimilation experiments with simulated Radar observations. Submitted to *Mon. Wea. Rev*.

Xu, Q., L. Wei, A. VanTuyl and E. H. Barker, 2001: Estimation of three-dimensional error covariances. Part I: Analysis of height innovation vectors. *Mon. Weather Rev*., **129**, 2126–2135.