RMetS
Royal Meteorological Society

An adaptive covariance relaxation method for ensemble data assimilation

Yue Ying and Fuqing Zhang*
Department of Meteorology, The Pennsylvania State University, University Park, USA

*Correspondence to: F. Zhang, Department of Meteorology, The Pennsylvania State University, 503 Walker Building, University Park, PA 16802, USA. E-mail: fzhang@psu.edu

For ensemble filters, accounting for unrepresented errors by inflating the ensemble perturbations can help improve filter performance. However, tuning the inflation factor can be costly, thus demanding adaptive covariance inflation (ACI) algorithms that give an online estimate of a temporally varying inflation factor. Additionally, a spatially varying inflation factor should be used to account for an irregular observation network. Anderson's adaptive inflation method offers a spatially and temporally varying inflation factor estimated from innovation statistics using a hierarchical Bayesian approach. In this study, we propose an alternative adaptive covariance relaxation (ACR) method that estimates a relaxation parameter online. Instead of treating inflation parameters as spatially varying random variables as in Anderson's method, the relaxation-to-prior-spread method provides an ensemble spread reduction term that serves as a spatial mask to account for an irregular observation network. We demonstrate with a set of experiments using the 40-variable Lorenz model that the ACR method is able to improve filter performance with the presence of sampling/model errors over a range of severity. Its reliability and ease of implementation suggest potential for future applications with atmospheric models.

Key Words: ensemble Kalman filter; covariance inflation; covariance relaxation; filter divergence; sampling and model error

Received 26 November 2014; Revised 18 April 2015; Accepted 30 April 2015; Published online in Wiley Online Library

1. Introduction

The ensemble Kalman filter (EnKF) combines information from a prior state estimate and its associated uncertainty (from an ensemble of model realizations) with observations and observation uncertainties to get an improved posterior state estimate and an updated uncertainty (Evensen, 1994). The accuracy of prior error covariance is one of the necessary criteria for optimal filter performance. In the presence of sampling and model error, the ensemble may underestimate the true uncertainty in the prior causing the filter to place too much weight on the prior mean state. Over time, the filter will begin to ignore the observed information, resulting in filter divergence. There are a number of methods for handling unrepresented error sources, among which covariance inflation methods are widely used. Empirical covariance inflation methods include multiplicative inflation (Anderson and Anderson, 1999), which increases the ensemble perturbations by a specified factor, and additive inflation (Mitchell and Houtekamer, 2000), which adds a random perturbation drawn from a specified error distribution to each member. Whitaker and Hamill (2012) hypothesize that multiplicative inflation methods can account for observation network related sampling error, while additive inflation is more suitable for treating model error.

There are also methods specifically designed to handle sampling error due to limited ensemble size. For example, covariance localization (Hamill *et al.*, 2001) treats sampling errors by tapering the Kalman gain far from the observation location, which removes some of the spurious correlations. Houtekamer and Mitchell (1998) proposed a double-ensemble EnKF in which ensemble-estimated covariance is used to update the other ensemble, therefore avoids inbreeding. More recently, a new formulation of the EnKF is proposed to account for the sampling bias due to the use of a limited-size ensemble (Bocquet, 2011; Bocquet and Sakov, 2012). For model error, treatments related to the forecast model can help maintain ensemble spread and ensure filter performance; such treatments include the use of a multi-physics or multi-model ensemble (e.g. Meng and Zhang, 2007), stochastic kinetic energy backscatter (Shutts, 2005; Berner *et al.*, 2009), and stochastically perturbed physics tendencies (Buizza *et al.*, 1999).

Among the aforementioned methods, covariance inflation is favoured due to its simplicity. However, given the large dimension of the dynamical systems of interest in atmospheric science, the process of tuning the inflation factor to suit a particular application is often costly. A variety of adaptive covariance inflation (ACI) methods are proposed to estimate inflation factors from innovation (observation-minus-forecast) statistics. The innovation statistics was first used to estimate parameters in forecast and observation error covariance matrices using a maximum-likelihood approach (Dee, 1995; Dee and da Silva, 1999; Dee *et al.*, 1999). For example, Wang and Bishop (2003)

formulated an online inflation estimation algorithm within the ensemble transform Kalman filter (ETKF) framework. Li *et al.* (2009) further extended the algorithm to simultaneously estimate covariance inflation and the observation error, using a set of observation-minus-forecast diagnostics (Desroziers *et al.*, 2005). They applied temporal smoothing to reduce the sampling bias due to a small innovation sample at each time step. Anderson (2007) provides an alternative algorithm by treating each innovation as a random variable and updates the inflation factor using a hierarchical Bayesian approach. These methods assume observations are uncorrelated. Zheng (2009) and Liang *et al.* (2012) relaxed this assumption and estimated an inflation factor from the innovation vector at each time step using a maximum-likelihood approach. They found that their algorithm is only efficient when estimating a scalar inflation factor constant in space.

Another important issue for any effective adaptive inflation method is how the method handles a spatially irregular observation network. For areas where no observations are available, applying an inflation factor constant in space may cause the variance of these unobserved variables to keep growing, sometimes even exceeding climatological variance. To alleviate this problem, Anderson (2009) extends the Bayesian approach to estimate a spatially and temporally varying inflation parameter, and assume that the spatial correlation factor of the inflation parameters is the same as the state variables. Miyoshi (2011) also introduces an algorithm that estimates inflation parameters that are spatially varying. He advances Li *et al.* 's method by including variance of the estimated inflation parameter from the Central Limit Theorem, making it a Gaussian approximation of Anderson's method.

In practice, the inflation factor is usually not physically constrained, which will often give rise to imbalance issues for complicated dynamic models. Zhang *et al.* (2004) developed a method that relaxes the posterior ensemble perturbations to the prior perturbations that effectively inflates the posterior ensemble yet preserves physical balances of the ensemble perturbations. Following this concept, Whitaker and Hamill (2012) proposed a relaxation-to-prior-spread (RTPS) method that is equivalent to applying a spatially varying multiplicative inflation. As a property of ensemble filters, the posterior ensemble spread should be smaller than prior spread after assimilating observations. Spatially, the reduction in ensemble spread should only be found where observations are available. Therefore, the reduction of spread can serve naturally as a spatial mask for inflation in case of an irregular observation network. The scalar relaxation parameter can effectively be an inflation factor given that it is allowed to be larger than 1.

In this study, we introduce an adaptive covariance relaxation (ACR) method that estimates the relaxation parameter of RTPS online according to innovation statistics. We will compare this new method to the non-adaptive RTPS as well as Anderson's ACI method. The advantage of RTPS is its simplicity and its ability to provide a spatially varying inflation as in Anderson's method. In the next section, we will provide the mathematical formulation of the proposed ACR method, along with Anderson's method described with the same notation for comparison. Experimental designs are presented in section 3 and the results are given in section 4. Section 5 gives the concluding remarks.

2. Methodology

2.1. Ensemble Kalman filter

Generally speaking, the inflation methods described here should be applicable to most if not all variations of ensemble Kalman filters. In this study, we will use an ensemble square-root filter (EnSRF) introduced by Whitaker and Hamill (2002) for testing. The EnSRF algorithm is described as follows. Let $\mathbf{x}$ be an n-by-1 column vector that holds all state variables. We introduce an ensemble of N members, or state vectors. The ensemble mean and ith ensemble perturbation are denoted by $\bar{\mathbf{x}}$ and $\mathbf{x}'_i = \mathbf{x}_i - \bar{\mathbf{x}}$, respectively. The background error covariance matrix is calculated from this ensemble using $P = \frac{1}{N-1}\sum_{i=1}^{N} \mathbf{x}'_i \mathbf{x}'^{\mathrm{T}}_i$. Let $\mathbf{y}^{\mathrm{o}}$ be a p-by-1 column vector that contains all observations at the current time. The observations are drawn from a normal distribution $\mathcal{N}(H\mathbf{x}^{\mathrm{t}}, R)$, where $\mathbf{x}^{\mathrm{t}}$ is the truth state vector, H is a p-by-n linear operator that maps the state vector to observation space and R is the observation error covariance matrix that is assumed to be diagonal. We denote the jth row of H_j by H_j. The EnSRF assimilates observations serially to avoid large matrix inversion. The following equations are applied for $j = 1, 2, \ldots, p$ to update $\bar{\mathbf{x}}$ and $\mathbf{x}'_i$ from the prior ($\bar{\mathbf{x}}^{\mathrm{b}}$ and $(\mathbf{x}^{\mathrm{b}}_i)'$) to the posterior ($\bar{\mathbf{x}}^{\mathrm{a}}$ and $(\mathbf{x}^{\mathrm{a}}_i)'$).

$$\bar{\mathbf{x}}^{\mathrm{new}} = \bar{\mathbf{x}} + \rho_j \circ K_j(y^{\mathrm{o}}_j - H_j\bar{\mathbf{x}}), \tag{1}$$

$$(\mathbf{x}^{\mathrm{new}}_i)' = \mathbf{x}'_i + \epsilon_j\rho_j \circ K_j(0 - H_j\mathbf{x}'_i), \text{for } i = 1, 2, \ldots, N, \tag{2}$$

where $K_j = \mathrm{cov}(H_j\mathbf{x}, \mathbf{x})/\{(\sigma^{\mathrm{o}}_{y,j})^2 + (\sigma^{\mathrm{b}}_{y,j})^2\}$ is the Kalman gain, ρ_j is a localization function, the circle denote an element-wise production and $\epsilon_j = \left(1 + \sqrt{\frac{(\sigma^{\mathrm{o}}_{y,j})^2}{(\sigma^{\mathrm{o}}_{y,j})^2 + (\sigma^{\mathrm{b}}_{y,j})^2}}\right)^{-1}$ is a square-root modification term to account for the use of unperturbed observations (see Eq. 13 of Whitaker and Hamill (2002), denoted as α in their article). Here $(\sigma^{\mathrm{b}}_{y,j})^2$ and $(\sigma^{\mathrm{o}}_{y,j})^2$ represent the background and observation error variances in observation space; they are the jth diagonal terms of HPH^{T} and R, respectively. The model integrates the posterior estimates forward in time, providing the prior estimates for the next assimilation cycle.

2.2. Calculation of inflation factor according to innovation statistics

The first component of an adaptive inflation algorithm is the calculation of the inflation factor according to the spread deficiency indicated by innovation statistics. The innovation vector associated with the p observations is $\mathbf{d}^{\mathrm{o-b}} = \mathbf{y}^{\mathrm{o}} - H\bar{\mathbf{x}}^{\mathrm{b}}$. Dee (1995) derived the following expression for the prior innovation statistics:

$$\mathrm{E}\{\mathbf{d}^{\mathrm{o-b}}(\mathbf{d}^{\mathrm{o-b}})^{\mathrm{T}}\} = HP^{\mathrm{b}}H^{\mathrm{T}} + R. \tag{3}$$

The expected value for $\mathbf{d}^{\mathrm{o-b}}(\mathbf{d}^{\mathrm{o-b}})^{\mathrm{T}}$ needs multiple realizations to estimate, which typically comes from observations taken at different times. For adaptive inflation, we need an online estimate of the innovation statistics for each time, so a reduced statistic using only the p observations at one time is used. Given that p is large, the following relationship will hold:

$$\mathrm{tr}\{\mathbf{d}^{\mathrm{o-b}}(\mathbf{d}^{\mathrm{o-b}})^{\mathrm{T}}\} = (\mathbf{d}^{\mathrm{o-b}})^{\mathrm{T}}\mathbf{d}^{\mathrm{o-b}} = \mathrm{tr}(HP^{\mathrm{b}}H^{\mathrm{T}}) + \mathrm{tr}(R). \tag{4}$$

The deficiency in prior error variance can be expressed in terms of an inflation factor,

$$\lambda_{\mathrm{b}} = \sqrt{\{(\mathbf{d}^{\mathrm{o-b}})^{\mathrm{T}}\mathbf{d}^{\mathrm{o-b}} - \mathrm{tr}(R)\}/\mathrm{tr}(HP^{\mathrm{b}}H^{\mathrm{T}})}. \tag{5}$$

Note that, if $(\mathbf{d}^{\mathrm{o-b}})^{\mathrm{T}}\mathbf{d}^{\mathrm{o-b}} < \mathrm{tr}(R)$, the square root cannot be evaluated. In this case, λ_{b} is set to 1.

Similarly, covariance inflation can be applied to the posterior ensemble according to the posterior innovation statistics (Desroziers *et al.*, 2005):

$$\mathrm{E}\{\mathbf{d}^{\mathrm{a-b}}(\mathbf{d}^{\mathrm{o-a}})^{\mathrm{T}}\} = HP^{\mathrm{a}}H^{\mathrm{T}}, \tag{6}$$

where $\mathbf{d}^{\mathrm{o-a}} = \mathbf{y}^{\mathrm{o}} - H\bar{\mathbf{x}}^{\mathrm{a}}$ and $\mathbf{d}^{\mathrm{a-b}} = H\bar{\mathbf{x}}^{\mathrm{a}} - H\bar{\mathbf{x}}^{\mathrm{b}}$. And the inflation factor can be calculated as

$$\lambda = \sqrt{(\mathbf{d}^{\mathrm{a-b}})^{\mathrm{T}}\mathbf{d}^{\mathrm{o-a}}/\mathrm{tr}(HP^{\mathrm{a}}H^{\mathrm{T}})}. \tag{7}$$

Note that the inflation factor in both Eqs (5) and (7) is estimated in observation space, thus additional steps are needed for applying this inflation in state space.

2.3. Anderson's adaptive covariance inflation (ACI) method

Anderson (2009) proposes a Bayesian approach for estimating a temporally and spatially varying inflation parameter from the innovation statistics. More specifically, the inflation parameters λ_k are treated as random variables with normal distributions, $\Pr(\lambda_k) = \mathcal{N}(\bar{\lambda}_k, \sigma^2_{\lambda,k})$. This method uses Bayes' theorem to update the distribution of λ_k by processing each innovation $d_j^{\mathrm{o-b}}$ serially. For $j = 1, 2, \ldots, p$, the following procedure is performed for each state vector element k:

$$\Pr(\lambda_k | d_j^{\mathrm{o-b}}) \propto \Pr(d_j^{\mathrm{o-b}} | \lambda_k) \Pr(\lambda_k), \tag{8}$$

where $d_j^{\mathrm{o-b}}$ is the jth innovation $y_j^{\mathrm{o}} - H_j \bar{\mathbf{x}}^{\mathrm{b}}$. Each d_j is assumed to be drawn from a zero-mean normal distribution with variance $\theta^2 = (\lambda_k^{\mathrm{o}} \sigma_{y,j}^{\mathrm{b}})^2 + (\sigma_{y,j}^{\mathrm{o}})^2$, where λ_k^{o} is the expected inflation in the observation space given the prior inflation value at k. The method assumes that the inflation parameters and state variables have the same spatial correlation structure,

$$\mathrm{corr}(H_j \mathbf{x}, x_k) = \mathrm{corr}(\lambda^{\mathrm{o}}, \lambda_k). \tag{9}$$

Given the prior value $\bar{\lambda}_k$, the expected inflation λ_k^{o} can be calculated as

$$\lambda_k^{\mathrm{o}} = 1 + \rho_k \mathrm{corr}(\lambda^{\mathrm{o}}, \lambda_k)(\bar{\lambda}_k - 1), \tag{10}$$

where ρ_k is the kth component of the localization function used in Eq. (1). The observation likelihood in Eq. (6) can be expressed as:

$$\Pr(d_j^{\mathrm{o-b}} | \lambda_k) = (\sqrt{2\pi}\theta)^{-1} \exp\{-(d_j^{\mathrm{o-b}})^2 / 2\theta^2\}. \tag{11}$$

The updated $\bar{\lambda}_k$ is found when $\Pr(\lambda_k | d_j)$ reaches its maximum, thus $\bar{\lambda}_k$ can be solved by taking the derivative of Eq. (6) with respect to λ_k and setting it equal to 0. See Appendix A in Anderson (2009) for a detailed numerical method for finding the maximum of $\Pr(\lambda_k | d_j^{\mathrm{o-b}})$. Also note that in Anderson's derivation he uses $\sqrt{\lambda}$ as the λ used in this article.

Anderson's algorithm updates the inflation parameter at k according to the jth innovation and the spatial correlation between the observation and the kth state variable. If the correlation is large (e.g. for observations near k), λ_k^{o} is close to $\bar{\lambda}_k$. When the correlation is small, λ_k^{o} approaches one. The amount of adjustment on $\bar{\lambda}_k$ brought by $d_j^{\mathrm{o-b}}$ depends on the selection of $\sigma^2_{\lambda,k}$. A larger $\sigma^2_{\lambda,k}$ indicates that the prior distribution of $\bar{\lambda}_k$ has larger uncertainty, and should be adjusted more toward the value suggested by the innovation. Due to the nature of Bayesian inference, $\sigma^2_{\lambda,k}$ should decrease after each update, limiting the innovation impact on $\bar{\lambda}_k$ over time. In this study, we implement the method using the posterior $\bar{\lambda}_k$ as the prior value for the next assimilation cycle, and $\sigma^2_{\lambda,k}$ is fixed in time for all k.

2.4. Adaptive covariance relaxation (ACR)

Covariance relaxation can be considered as another approach for covariance inflation. Zhang *et al.* (2004) first proposed to relax the posterior ensemble perturbations to the prior:

$$(\mathbf{x}_i^{\mathrm{a}})'_{\mathrm{inf}} = (1 - \alpha)(\mathbf{x}_i^{\mathrm{a}})' + \alpha(\mathbf{x}_i^{\mathrm{b}})', \tag{12}$$

so that the posterior ensemble spread is artificially increased and the posterior retains a certain degree of physical balance from the prior ensemble perturbations. Whitaker and Hamill (2012) suggested relaxing the ensemble spread instead of the perturbations, and formulated the RTPS method. For the kth state variable, its ensemble spread is rescaled to a mix of prior and posterior spread values controlled by α:

$$(x_{k,i}^{\mathrm{a}})'_{\mathrm{inf}} = (x_{k,i}^{\mathrm{a}})' \left(\alpha \frac{\sigma_k^{\mathrm{b}} - \sigma_k^{\mathrm{a}}}{\sigma_k^{\mathrm{a}}} + 1 \right), \tag{13}$$

where $\sigma_k = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} x'^2_{k,i}}$ is the ensemble spread of the state variable x_k, and the superscripts 'b' and 'a' denote prior and posterior values, respectively. Note that $\alpha = 0$ implies keeping the original posterior spread, while $\alpha = 1$ implies increasing the posterior spread back to the prior spread for each state variable. The rescaling term $\alpha(\sigma_k^{\mathrm{b}} - \sigma_k^{\mathrm{a}})/\sigma_k^{\mathrm{a}} + 1$ works effectively as a multiplicative inflation factor for the ensemble perturbations of that state variable. Recall that, after assimilating observations, the posterior ensemble spread σ^{a} becomes smaller than the prior spread σ^{b}. This reduction in spread depends on the Kalman gain, which in turn depends on the localization distance and the observation location. When localization is applied, the reduction in spread should be larger in densely observed regions and taper to nearly 0 in unobserved regions, assuming all else in the Kalman gain (i.e. observation errors, etc.) are equal. Therefore, in RTPS, the relative reduction in ensemble spread, $(\sigma^{\mathrm{b}} - \sigma^{\mathrm{a}})/\sigma^{\mathrm{a}}$, naturally serves as a spatial mask and a constant α controls the magnitude of inflation.

The ACR method we propose takes advantage of the relative spread reduction term from RTPS, and reduces the problem to estimating a scalar parameter α online. To use innovation statistics, we need to calculate the overall relative reduction in ensemble spread in observation space. The overall relative spread reduction can be expressed as $(\bar{\sigma}_y^{\mathrm{b}} - \bar{\sigma}_y^{\mathrm{a}})/\bar{\sigma}_y^{\mathrm{a}}$, where $\bar{\sigma}_y = \sqrt{\mathrm{tr}(HPH^{\mathrm{T}})/p}$ and the over-bar indicates an average over the p observations. We match the expected inflation in observation space with the inflation factor suggested by posterior innovation statistics Eq. (7), and solve for α:

$$\alpha \frac{\bar{\sigma}_y^{\mathrm{b}} - \bar{\sigma}_y^{\mathrm{a}}}{\bar{\sigma}_y^{\mathrm{a}}} + 1 = \lambda. \tag{14}$$

Note that the calculated α value does not necessarily fall in the range of (0, 1). If the observation sample size p is small for each cycle, the inflation factor could become very noisy; in this case we apply temporal smoothing to λ in Eq. (14) to resolve this issue:

$$\lambda_{t,\mathrm{smooth}} = \lambda_{t-1} + (\lambda_t - \lambda_{t-1})/\tau, \tag{15}$$

where λ_t is the λ calculated for the current cycle and τ determines how fast α responds to changes suggested by the innovation statistics. The use of a larger τ is equivalent to using the innovations from a longer period of time to calculate the observed inflation factor. The estimated α are applied in Eq. (13) in state space to relax the ensemble spread.

3. Numerical experiment design

To test the methods described in the previous section, we run trials of cycling data assimilation using the Lorenz 40-variable model (Lorenz, 1996). The forecast model equations are

$$\frac{\mathrm{d}x_k}{\mathrm{d}t} = a(x_{k+1} - x_{k-2})x_{k-1} - dx_k + F, \text{for } k = 1, 2, \ldots, n, \tag{16}$$

where a, d and F are model parameters for the nonlinear advection, damping and forcing terms and $n = 40$. The model is defined on a one-dimensional domain with a periodic boundary. The true model sets $a = 1$, $d = 1$ and $F = 8$. We set the model time step to $\Delta t = 0.05$, which corresponds to 6 h. At each time step, the EnSRF update equations described in section 2.1 are applied.

The radius of influence is set to ten grid points for trials when covariance localization is applied.

To create synthetic observations, we add random noise to state variables generated from a simulation designated as the truth run. We set the observation noise σ_y^{o} to 1.0,* commonly used in previous studies (e.g. Anderson, 2009; Miyoshi, 2011). Two different types of observation networks are tested: fully-observed and half-observed. The fully-observed network has observations located on all state variables (i.e. H is the identity matrix), while the half-observed network has observations only for the first 20 'land' variables (the remaining 20 'ocean' variables are unobserved) following the set-up in Lorenz and Emanuel (1998).

For the RTPS, ACI, and ACR inflation methods, we run ten trials, each trial consisting of 5000 time steps (corresponding to ∼3.4 years). To evaluate the filter performance of each method, we calculate the analysis root mean square error (RMSE) and the consistency ratio (CR) for the final 1000 time steps of all ten trials, mitigating issues that may arise during the spin-up period of data assimilation. The RMSE is calculated by averaging the squared analysis error $(\bar{x}^{\mathrm{a}} - x^{\mathrm{t}})^2$ over the n state variables for the final 1000 time steps of all ten trials, then taking the square root. For the fully-observed cases, all $n = 40$ state variables are used to calculate RMSE, while for the half-observed cases the 'land' ('ocean') variables located at $k = 1, \ldots, 20 (k = 21, \ldots, 40)$ are used to calculate RMSE_land (RMSE_ocean). The CR is calculated by averaging the ratio $\sqrt{\mathrm{tr}(HP^{\mathrm{b}}H^{\mathrm{T}} + R)/(\mathbf{d}^{\mathrm{o-b}})^{\mathrm{T}}\mathbf{d}^{\mathrm{o-b}}}$ over these time steps. If CR < 1, the background error variance is underestimated and the ensemble spread is too small, and vice versa. We specify the true R in our filter so that an erroneous P^{b} is the only reason why the statistical relation in Eq. (4) does not hold and CR deviates from one. We define 'filter divergence' as a significant drop in the CR below 1 (ensemble spread is too small) and a growth in the RMSE to a value above the observation noise.

We introduce sampling error by reducing ensemble size (N) from 80 to 5. To introduce model error, we change the default parameters a, d and F in the forecast model. The forcing parameter F is gradually reduced from the true value 8 to 5, incrementally producing a less chaotic model. We change parameter a from 1 to 0.8 (d from 1 to 1.2) to introduce error in the advection (damping) process in the forecast model. The inflation methods are also tested with combinations of errors in a, d and/or F to examine if they can handle model errors from a mixture of processes.

For each error regime, we test the performance of Anderson's method and the proposed ACR method. For simplicity, we will denote Anderson's method as 'ACI σ_λ^2', where σ_λ^2 is the selected value for the variance of inflation parameter. We have tested ACI with $\sigma_\lambda^2 = 0.1$, 1 and 10, and show that results are sensitive to this parameter. For ACR, we have tested with the temporal smoothing parameter $\tau = 1$, 10 and 100, and only show results for $\tau = 100$. Although the estimated inflation factor depends on the choice of τ, we found that the analysis RMSEs, on average, are not sensitive to this choice for the Lorenz 40-variable toy model. We also show results using the RTPS method with a range of α values (0–1) to serve as a benchmark. Note that $\alpha = 0$ corresponds to the case when no inflation is applied. For both sampling and model errors, we test the inflation methods either with or without applying covariance localization.

4. Results

4.1. Performance with sampling error due to limited ensemble size

Sampling error may occur when using a limited-size ensemble to estimate the background error covariance. Figure 1 shows the results from the perfect-model fully-observed cases where we gradually increase sampling error by reducing sample size N. Figure 1(a)and (c) show the analysis RMSE as a function of sample size for the case without and with localization, respectively, and Figure 1(b) and (d) show the corresponding CR. RMSE values in Figure 1(a) are also documented in Table 1. Generally, as N decreases, the filter diverges (RMSE jumps above observation noise while CR drops below 1) at some point. We can define three ranges of N values: (i) stable range in which the filter is stable even without inflation; (ii) unstable range in which the filter diverges even for the best-tuned inflation; and (iii) the transition range in between the unstable and stable ranges. We are interested in the transition range, because within this range a properly tuned inflation method can help restore filter stability. For the case without localization (Figure 1(a) and (b)), the transition range is (15, 40). Results from non-adaptive RTPS show that, within this range, a larger amount of inflation (larger α for RTPS) is needed to restore filter stability for larger sampling error (smaller N). At $N = 30$, $\alpha = 0.1$ is enough, while at $N = 17$, $\alpha = 0.3$ is needed. In the stable range, no inflation is needed and $\alpha = 0$ results in the lowest RMSE; increasing inflation (larger α) will increase RMSE. In the unstable range, we see that using a larger α can help reduce RMSE slightly, but cannot prevent filter divergence. The ACR method is able to restore filter stability through the transition region, although the resulting RMSE is not as small as the best-tuned RTPS. For the ACI method, we found that relatively large $\sigma_\lambda^2 = 1$ is needed to bring a similar improvement as ACR. Recall that the value of σ_λ^2 tunes how much an innovation impacts the final inflation field, thus a large σ_λ^2 will lead to a final inflation that is close to the value suggested by innovation statistics. Note that Anderson (2009) sets $\sigma_\lambda^2 = 0.01$ in his study. We found that this value only works well in the stable range. With localization (Figure 1(c) and (d)), the filter's tolerance to sampling error is greatly increased. The transition range shifts to (4, 10). The ACR method can also restore filter stability in this case, and the ACI still needs at least $\sigma_\lambda^2 = 1$ to bring such improvement.

For the half-observed case, we only show results when localization is applied (Figure 2), because the model becomes unstable when the unobserved variables are also inflated; localization helps limit the inflation within the observed region in RTPS. In this case, we calculated analysis RMSE separately for the observed variables (RMSE_land, Figure 2(a)) and unobserved variables (RMSE_ocean, Figure 2(b)), and Figure 2(c) shows CR in observation space. Comparing RMSE_land to RMSE in Figure 1(c), we can see that the analysis RMSE is smaller for the land variables in the half-observed case compared with the fully-observed case. The ocean variables maintain a relatively large ensemble spread that propagates downstream, effectively inflating the ensemble covariance over land. We found that the filter does not diverge for the land variables until N reaches 5, while the ocean variables maintain a noise level higher than the observation noise that remains within the climatological bound. Similar to the fully-observed case, the ACR method also restores filter stability for $N < 6$. The ACI still needs a larger σ_λ^2 value to exhibit the same improvement. The RMSEs for ocean variables are also improved for $N < 6$ when adaptive inflation is applied to the land variables, which may result from the more accurate state estimates over land that propagates downstream to the ocean.

4.2. Performance with an imperfect forecast model

In this subsection, we test the inflation methods under the condition where an imperfect forecast model is used. Figure 3 shows results from the fully-observed cases with $N = 40$ and no localization. RMSE values in Figure 3(a) are also documented in Table 1. In Figure 3(a) and (b), the F parameter is the only model error source. In this case, we see that only a tiny model error ($F = 7.9$) will lead to filter divergence without inflation. The RTPS method shows that, for larger model error, larger α values are needed to restore filter stability. For example, $\alpha = 0.2$ is enough for $F = 7.9$, while $\alpha = 0.9$ is needed for

*We also tested most of the cases with $(\sigma_y^{\mathrm{o}})^2 = 0.1$ (i.e. $\sigma_y^{\mathrm{o}} \sim 0.316$). The results (not shown) are qualitatively the same as presented in this study.

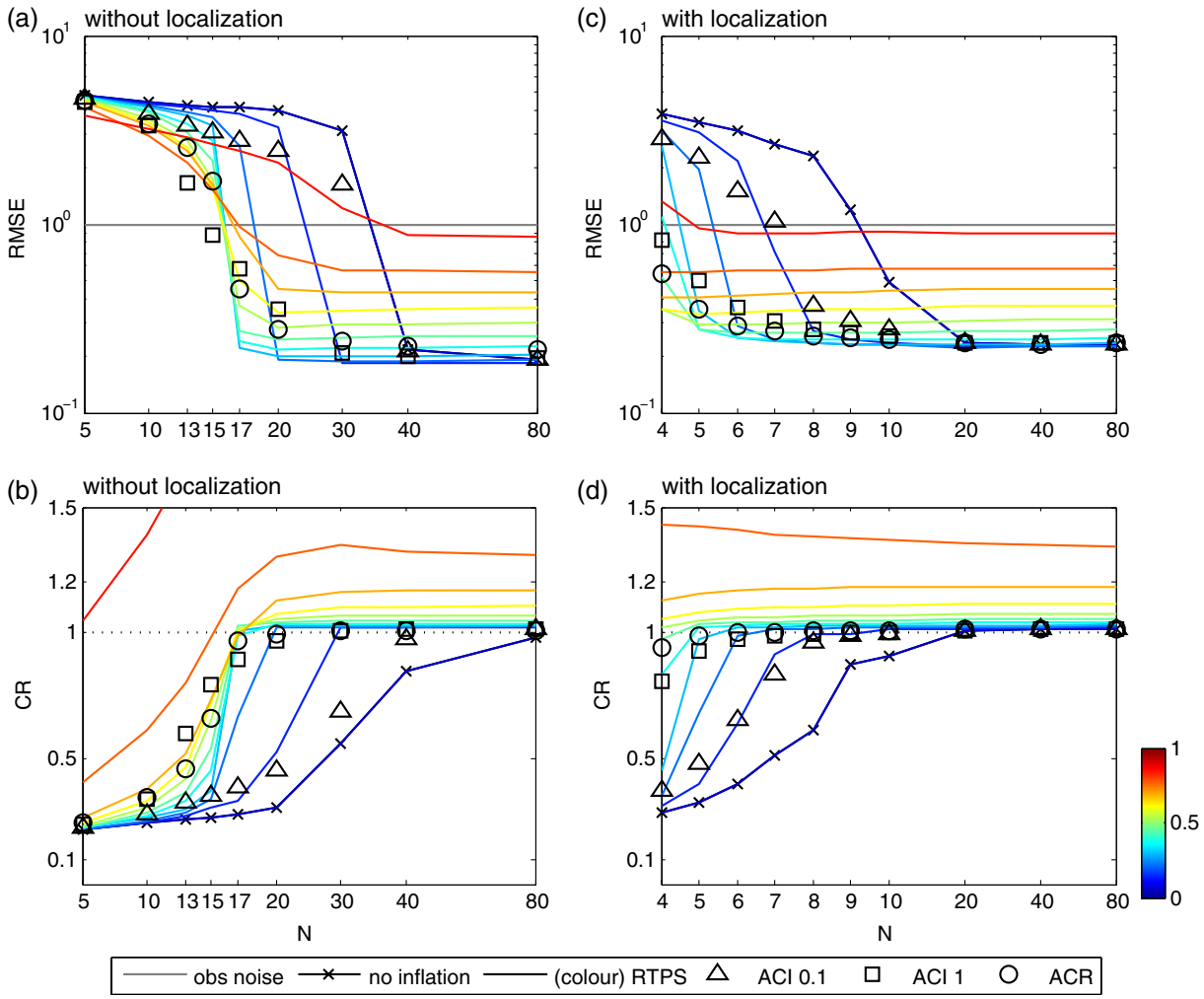

Figure 1. (a,c) Analysis RMSE and (b,d) consistency ratio (CR), for the fully-observed cases with different degrees of sampling errors introduced by changing the ensemble size (N). Panels (a,b) and (c,d) shows result without and with localization (radius of influence = 10), respectively. The grey line shows the observation noise level. The coloured lines show inflation parameters for RTPS with α ranging from 0 to 1; the cross markers (no inflation) are equivalent to $\alpha = 0$. Triangle and square markers correspond to the ACI methods with $\sigma_\lambda^2 = 0.1$ and 1, respectively. Circle markers correspond to the ACR method. Note that we show more data points in the transition region ranging from the filter being stable to diverging. The transition region is different for the cases without localization ($N = 15$–40) and with localization ($N = 4$–10).

Table 1. Analysis RMSEs using no inflation, ACR, best-tuned RTPS and ACI methods (columns) for different error sources (rows) including sampling error N and model error F. The fully-observed case without localization is shown (corresponding to figures 1(a) and 3(a)).

		No inflation	ACR	RTPS (optimal α)	ACI 0.1	ACI 1
N	80	0.1920	0.2163	0.1851 (0.1)	0.1925	0.1955
	40	0.2181	0.2275	0.1821 (0.1)	0.2106	0.1977
	20	4.0032	0.2766	0.1926 (0.2)	2.4608	0.3541
	17	4.1459	0.4561	0.2198 (0.3)	2.7773	0.5846
	15	4.2028	1.6785	1.5101 (0.9)	3.0480	0.8755
	10	4.4331	3.3941	2.9290 (0.9)	3.8840	3.3389
	5	4.7771	4.5219	3.7310 (1.0)	4.6173	4.4831
F	8	0.2154	0.2378	0.1880 (0.1)	0.1979	0.2012
	7.9	3.9566	0.2918	0.2221 (0.3)	1.9667	0.3285
	7.5	4.0564	0.4435	0.3424 (0.6)	2.0047	0.6668
	7	4.0107	0.5835	0.4231 (0.7)	2.2781	0.8577
	6	4.0423	0.7783	0.5234 (0.8)	2.7079	1.0630
	5	4.1770	0.9044	0.5939 (0.9)	2.9864	1.1891

$F = 5$. Similar to the sampling error cases, we found the ACR method able to estimate an α close to the optimal value and reduce the RMSEs significantly. As for ACI, $\sigma_\lambda^2 = 0.1$ only brings limited improvement, while $\sigma_\lambda^2 = 1$ brings improvement that is comparable to the ACR method. To show the performance under a combination of model error sources, we also increase damping by 20% ($d = 1.2$, Figure 3(c) and (d)) and decrease nonlinear advection by 20% ($a = 0.8$, Figure 3(e) and (f)). In both cases, the optimal value for α is much larger ($\sim$0.8) with a smaller reduction in RMSE, with both cases' RMSE still under the observation noise. We found that the ACR method handles all types of model error well; a similar improvement can be achieved by ACI only when using a relatively large σ_λ^2.

Figure 4 shows results from half-observed cases for $N = 10$ with localization. In this case, similar to sampling error cases in Figure 2, the filter tolerates model error more when localization is applied. The filter does not diverge without inflation until F is as small as 6 (Figure 4(a)). We also show a combination of error sources from F, a and d as in Figure 3, and found that the

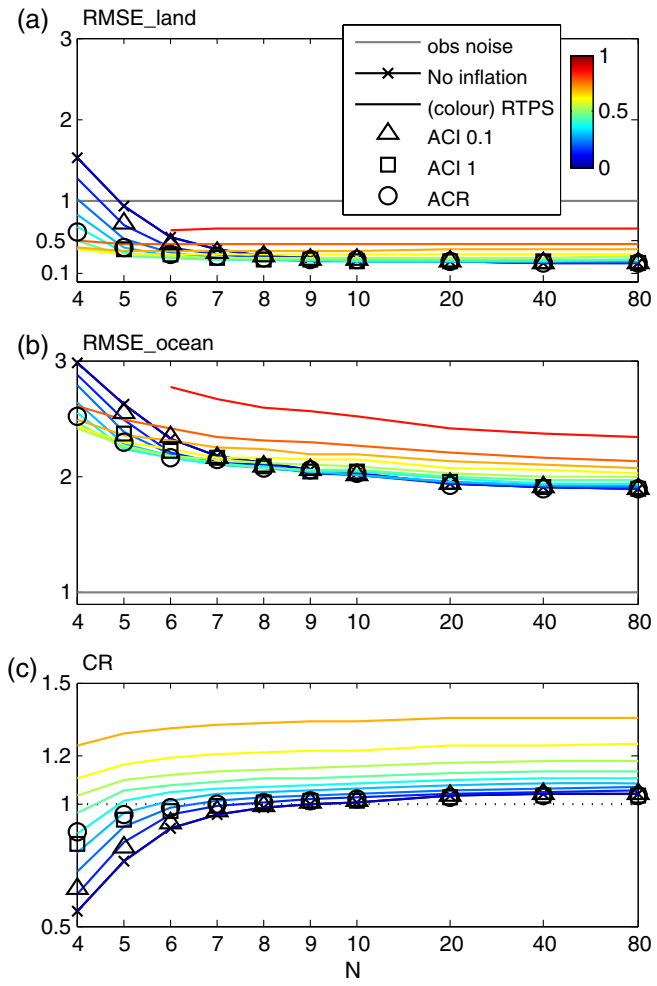

Figure 2. Analysis RMSE and consistency ratio (CR) for the half-observed cases ($N = 10$ with localization). (a) RMSE_land ((b) RMSE_ocean) is calculated using the observed (unobserved) variables along with the truth, while the (c) consistency ratio is calculated in observation space (observed variables only). The legend is the same as for Figure 1.

ACR method also handles all types of model error well, while ACI needs large σ_λ^2 to achieve this improvement.

4.3. *Spatial and temporal behaviour of inflation parameters*

In this subsection, we examine the spatial and temporal behaviour of the inflation parameters estimated by the ACR method, namely the $\alpha(\sigma_k^{\mathrm{b}} - \sigma_k^{\mathrm{a}})/\sigma_k^{\mathrm{a}} + 1$ term in Eq. (13). It is desirable that the adaptive methods automatically adjust the inflation parameters both in space and time to account for the inhomogeneity. Figure 5 compares the spatial pattern of inflation parameters averaged over time between different inflation methods for two half-observed cases with large error sources. The inflation from RTPS can be thought of as a scalar α value ranging from 0 to 1 that tunes a spatial pattern (the relative ensemble spread reduction term). We set the localization distance to 10, so that observations over land ($k = 1, \ldots, 20$) do not impact the centre of the ocean ($k = 30$). The average inflation is larger over land than over the ocean, with a maximum near $k = 1$ and tapering to 0 from $k = 20$ to $k = 30$. Since the domain is cyclic (periodic boundary condition), the maximum is at the downstream side of the ocean, where large uncertainties propagate toward the land. For the ACR method, the resulting inflation pattern has the same properties as the RTPS method. On the other hand, the ACI method uses the σ_λ^2 value to tune the overall amount of inflation. In the presence of large model error (Figure 5(b)), $\sigma_\lambda^2 = 0.1$ leads to overall inflation of only $\sim$1.01 over land. Increasing σ_λ^2 to 1 increases the inflation to $\sim$1.05, but the maximum near $k = 1$ is not well captured. Increasing σ_λ^2 to 10 produces the inflation pattern similar to ACR, but it appears to be noisier.

Figure 6 shows the time series of λ, the inflation factor estimated in observation space from ACR, for the cases with no error sources (Figure 6(a)) and with large model error (Figure 6(b)). The temporal smoothing τ value is set to 1, 10 and 100. As expected, a larger τ value leads to a more smoothed λ in time. For the case with no error sources, the λ value is close to 1 throughout the time period, while for the case with large model error, λ is elevated to $\sim$1.2. In our experiments, although λ has different temporal variability depending on the τ value used, the resulting temporally averaged λ and analysis RMSE are not particularly sensitive to the choice of τ (not shown). One caveat arises if τ is set too large, e.g. as large as the model integration time period, so the algorithm adjusts the inflation too slowly causing the filter to diverge before inflation reaches the required magnitude. Therefore, the optimal τ should not be too large so that the adjustment of the inflation magnitude is faster than the accumulation rate of unrepresented errors, but not too small so that inflation becomes too noisy in time.

The ACI method treats each innovation as one observation and uses it to update the inflation field. The updated inflation will be a weighted average of its prior value and the 'observed' value suggested by the innovation, with the weight dependent on the selection of σ_λ^2 values. The ACR method will be equivalent to ACI if we use one innovation at a time in Eq. (7) to get an 'observed' inflation factor and also update the prior inflation factor by the weighted average. We can reformulate the temporal smoothing equation (15) as an update equation that weights the observed and prior inflation factors, similar to the reformulation of the temporal smoothing equation in Li *et al.* (2009) by Miyoshi (2011) as a Gaussian approximation to the ACI method. In doing so, we should expect both methods to have similar sensitivity to the choice of σ_λ^2 or τ values. However, in our experiments, we used all p innovations at each time step to derive the 'observed' inflation factor deterministically in Eq. (7), rather than treating each innovation as a random variable. Even if $\tau = 1$, the inflation factor is still derived from innovation statistics using p observations, which is equivalent to using a large σ_λ^2 in the ACI method. In consequence, the analysis RMSE is insensitive to the choice of τ from 1 to 100. However, increasing τ reduces the noise and variability in the estimated inflation factor. From Figure 6 we see that the inflation factor has larger variance and frequently exceeds 1.5 when τ is set to 1. Though the noisy inflation factor works for the Lorenz 40-variable model, further research is needed to show whether it still works for other more complex models and/or observation networks. It is possible that an inflation factor that has too large a variance in time may be a potential issue for less autonomous models.

5. Concluding remarks

Covariance inflation can improve filter performance by inflating the ensemble to account for the unrepresented sampling/model errors. Tuning the inflation parameter can be costly, therefore adaptive algorithms that estimate the inflation online according to innovation statistics have been introduced in previous studies (Wang and Bishop, 2003; Anderson, 2007, 2009; Li *et al.*, 2009; Miyoshi, 2011). The online estimated inflation parameter adjusts to the value suggested by innovation statistics. When a spatially irregular observation network is used leaving some state variables unobserved, a spatially varying inflation parameter is likely needed. Inflating the unobserved variables may cause their ensemble spread to keep increasing, and therefore likely leads to filter/model instability. To introduce spatially varying covariance inflation, the ACI method (Anderson, 2009) treats the inflation parameters as individual random variables and updates them using 'observed' inflation from innovations. In this study, an alternative ACR method is introduced. We start with the RTPS method (Whitaker and Hamill, 2012), in which a

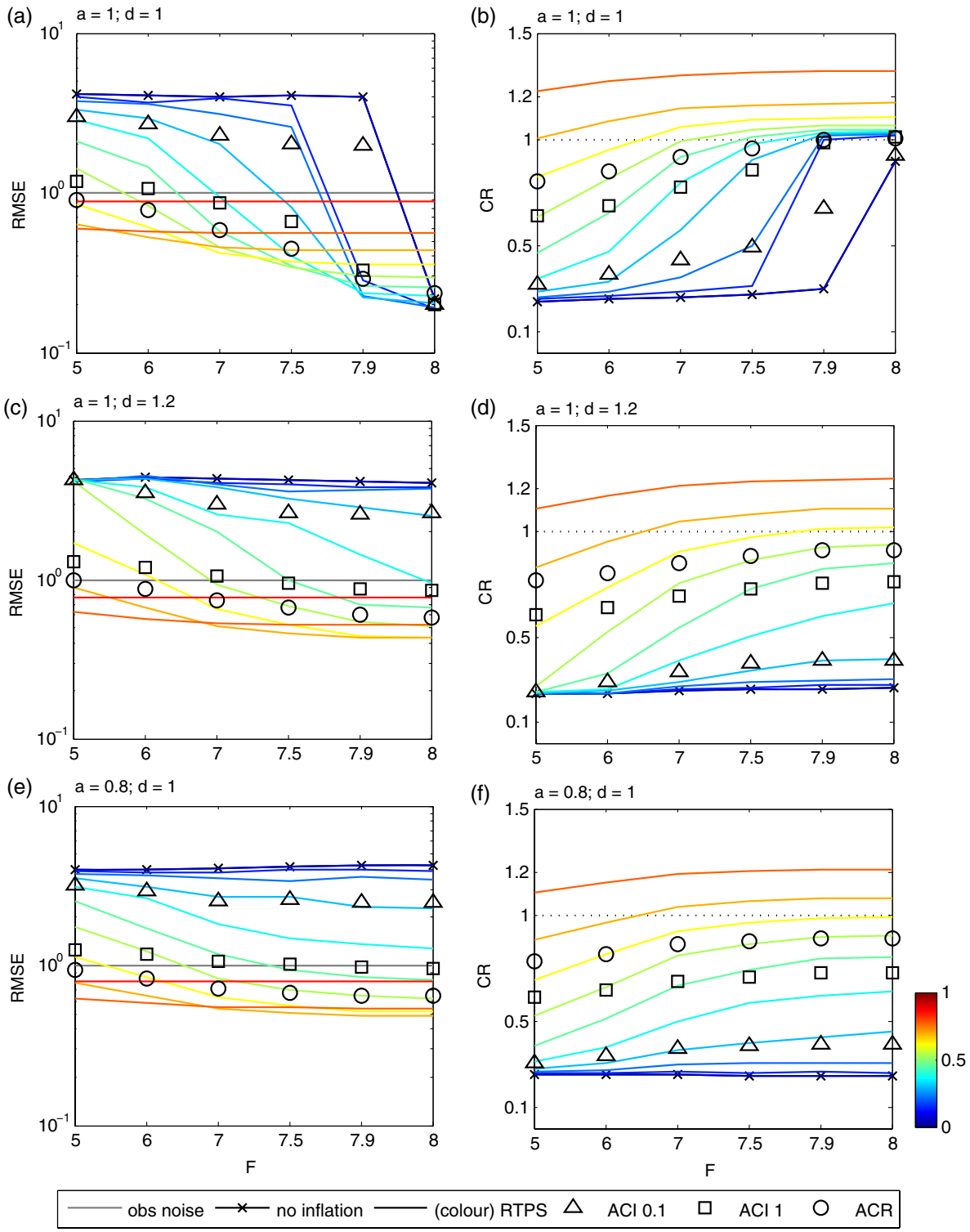

Figure 3. (a,c,e) Analysis RMSE and (b,d,f) consistency ratio (CR), for the fully-observed cases ($N = 40$ without localization) with different degree of model errors introduced by varying parameters F, a and d. Within each panel, F varies from 5 to 8. For each row, a different combination of a and d values are used. The legend is the same as for Figure 1.

spatially varying relative ensemble spread reduction term is used to mask the inflation to only the observed variables. We show that this reduction term is a direct result from assimilating observations whose spatial pattern depends on observation location and localization distance. The ACR method estimates the relaxation parameter in RTPS by matching innovation statistics in observation space, therefore it is equivalent to a spatially and temporally varying inflation. Numerical experiments with the Lorenz 40-variable model show that the ACR method can restore filter stability with the presence of sampling/model error for a range of severities. It is found that the spatial inflation pattern estimated by ACI is noisier compared to ACR and does not capture the location where ensemble spread deficiency is most severe. Also, ACI results are sensitive to the tunable variance of the inflation parameter, while ACR results are less sensitive to the tunable temporal smoothing parameter.

Both the RTPS and the algorithm for online estimation of the relaxation parameter are easy to implement for the proposed ACR method. Therefore it has the potential for real-data atmospheric model applications. The results in this study only apply to the Lorenz 40-variable model, thus there are several potential issues regarding the use of real observations and more complicated models that we want to stress. First, the real observations usually contain variables with different units. In this case, the innovation statistics Eqs (5) and (7) should use normalized observations so that each type of observation has a comparable contribution (see

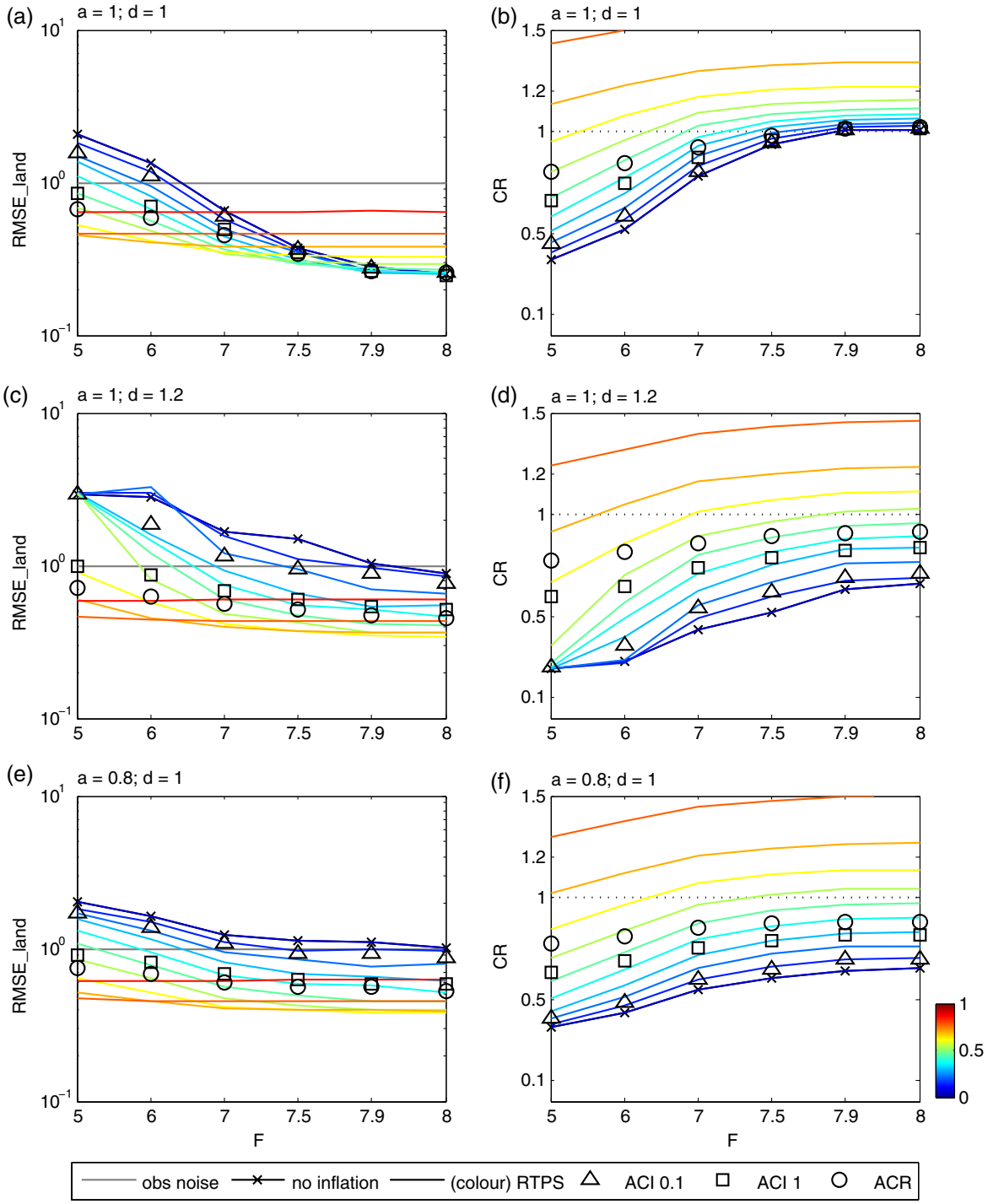

Figure 4. Same as Figure 3, but showing half-observed cases ($N = 10$ with localization).

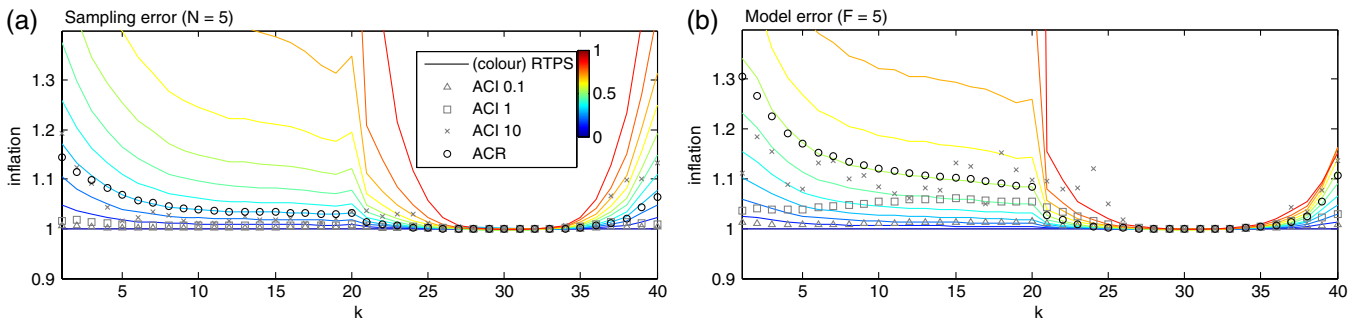

Figure 5. The spatial structure of inflation parameters for the half-observed cases ($N = 10$ with localization) with two types of severe error sources: (a) sampling error with $N = 5$ and (b) model error with $F = 5$. The coloured lines show inflation parameters for RTPS with α ranging from 0 to 1. Grey triangle, square and cross markers correspond to ACI methods with $\sigma_\lambda^2 = 0.1$, 1 and 10, respectively.

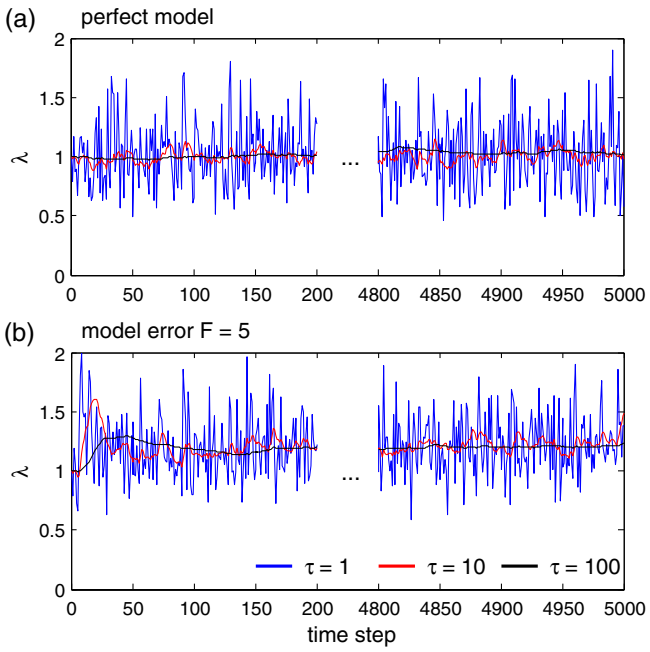

Figure 6. Time series of the inflation factor (λ) from ACR method for half-observed cases ($N = 10$ with localization) with (a) perfect model and (b) model error $F = 5$. Only the first and last 200 steps are shown. Blue, red and black lines correspond to $\tau = 1$, 10 and 100, respectively.

Eq. 18 in Wang and Bishop (2003)). Second, for complicated models, the dynamics can be also inhomogeneous in space. In this case, one needs to verify that the system of interest (e.g. a hurricane) is well covered by observations and the observations outside the system are excluded so the innovation statistics truly reflect the inflation needed for the system of interest. Finally, for real observations, the observation operator may be nonlinear. Since we estimate the relaxation parameter in observation space and apply it in state space, the nonlinear relation between observation and state space may cause the relaxation parameter to be either under- or overestimated. In this case, one may need a tangent linear observation operator to map state space to observation space.

Acknowledgements

This research was supported by NSF grant AGS-1305798. Y. Ying was also sponsored by the China Scholarship Council (CSC) visitor programme. Review comments from Takemasa Miyoshi and another anonymous reviewer on earlier versions of the manuscript were extremely beneficial. Discussions with John Harlim, Tyrus Berry, Steven Greybush, Jeff Anderson, Jeff Whitaker, and Jon Poterjoy were very helpful. Computing is conducted at the Texas Advanced Computing Center (TACC).

References

Anderson JL. 2007. An adaptive covariance inflation error correction algorithm for ensemble filters. *Tellus A* **59**: 210–224.

Anderson JL. 2009. Spatially and temporally varying adaptive covariance inflation for ensemble filters. *Tellus A* **61**: 72–83.

Anderson JL, Anderson SL. 1999. A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Mon. Weather Rev.* **127**: 2741–2758.

Berner J, Shutts GJ, Leutbecher M, Palmer TN. 2009. A spectral stochastic kinetic energy backscatter scheme and its impact on flow-dependent predictability in the ECMWF ensemble prediction system. *J. Atmos. Sci.* **66**: 603–626.

Bocquet M. 2011. Ensemble Kalman filtering without the intrinsic need for inflation. *Nonlinear Processes Geophys.* **18**: 735–750.

Bocquet M, Sakov P. 2012. Combining inflation-free and iterative ensemble Kalman filters for strongly nonlinear systems. *Nonlinear Processes Geophys.* **19**: 383–399.

Buizza R, Miller M, Palmer TN. 1999. Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. *Q. J. R. Meteorol. Soc.* **125**: 2887–2908.

Dee DP. 1995. On-line estimation of error covariance parameters for atmospheric data assimilation. *Mon. Weather Rev.* **123**: 1128–1145.

Dee DP, da Silva AM. 1999. Maximum-likelihood estimation of forecast and observation error covariance parameters. Part I: Methodology. *Mon. Weather Rev.* **127**: 1822–1834.

Dee DP, Gaspari G, Redder C, Rukhovets L, da Silva AM. 1999. Maximum-likelihood estimation of forecast and observation error covariance parameters. Part II: Applications. *Mon. Weather Rev.* **127**: 1835–1849.

Desroziers G, Berre L, Chapnick B, Poli P. 2005. Diagnosis of observation, background and analysis-error statistics in observation space. *Q. J. R. Meteorol. Soc.* **131**: 3385–3396.

Evensen G. 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.* **99**: 10143–10162, doi: 10.1029/94JC00572.

Hamill TM, Whitaker JS, Snyder C. 2001. Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Mon. Weather Rev.* **129**: 2776–2790.

Houtekamer PL, Mitchell HL. 1998. Data assimilation using an ensemble Kalman filter technique. *Mon. Weather Rev.* **126**: 796–811.

Li H, Kalnay E, Miyoshi T. 2009. Simultaneous estimation of covariance inflation and observation errors within an ensemble Kalman filter. *Q. J. R. Meteorol. Soc.* **135**: 523–533.

Liang X, Zheng XG, Zhang SP, Wu GC, Dai YJ, Li Y. 2012. Maximum likelihood estimation of inflation factors on error covariance matrices for ensemble Kalman filter assimilation. *Q. J. R. Meteorol. Soc.* **138**: 263–273.

Lorenz EN. 1996. Predictability: A problem partly solved. *Proceedings of Seminar on Predictability*, Vol. 1 1–18 ECMWF: Reading, UK.

Lorenz EN, Emanuel K. 1998. Optimal sites for supplementary weather observations: simulation with a small model. *J. Atmos. Sci.* **55**: 399–414.

Meng ZY, Zhang FQ. 2007. Tests of an ensemble Kalman filter for mesoscale and regional-scale data assimilation. Part II: Imperfect model experiments. *Mon. Weather Rev.* **135**: 1403–1423.

Mitchell HL, Houtekamer P. 2000. An adaptive ensemble Kalman filter. *Mon. Weather Rev.* **128**: 416–433.

Miyoshi T. 2011. The Gaussian approach to adaptive covariance inflation and its implementation with the local ensemble transform Kalman filter. *Mon. Weather Rev.* **139**: 1519–1535.

Shutts G. 2005. A kinetic energy backscatter algorithm for use in ensemble prediction systems. *Q. J. R. Meteorol. Soc.* **131**: 3079–3102.

Wang XG, Bishop CH. 2003. A comparison of breeding and ensemble transform Kalman filter ensemble forecast schemes. *J. Atmos. Sci.* **60**: 1140–1158.

Whitaker JS, Hamill TM. 2002. Ensemble data assimilation without perturbed observations. *Mon. Weather Rev.* **130**: 1913–1924.

Whitaker JS, Hamill TM. 2012. Evaluating methods to account for system errors in ensemble data assimilation. *Mon. Weather Rev.* **140**: 3078–3089.

Zhang FQ, Snyder C, Sun JZ. 2004. Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble Kalman filter. *Mon. Weather Rev.* **132**: 1238–1253.

Zheng XG. 2009. An adaptive estimation of forecast error covariance parameters for Kalman filtering data assimilation. *Adv. Atmos. Sci.* **26**: 154–160.

