



## Estimating and including observation-error correlations in data assimilation

Takemasa Miyoshi , Eugenia Kalnay & Hong Li

To cite this article: Takemasa Miyoshi , Eugenia Kalnay & Hong Li (2013) Estimating and including observation-error correlations in data assimilation, Inverse Problems in Science and Engineering, 21:3, 387-398, DOI: [10.1080/17415977.2012.712527](https://doi.org/10.1080/17415977.2012.712527)

To link to this article: <https://doi.org/10.1080/17415977.2012.712527>



Published online: 31 Aug 2012.



Submit your article to this journal [↗](#)



Article views: 259



Citing articles: 26 View citing articles [↗](#)

## Estimating and including observation-error correlations in data assimilation

Takemasa Miyoshi<sup>a\*</sup>, Eugenia Kalnay<sup>a</sup> and Hong Li<sup>b</sup>

<sup>a</sup>*Department of Atmospheric and Oceanic Science, University of Maryland, College Park, MD 20742, USA;* <sup>b</sup>*Shanghai Typhoon Institute, Shanghai, China*

(Received 16 June 2012; final version received 19 June 2012)

Usually in data assimilation with geophysical systems, the observation-error covariance matrix  $\mathbf{R}$  is assumed to be diagonal for simplicity and computational efficiency, although there are studies indicating that several types of satellite observations contain significantly correlated errors. This study brings to light the impact of the off-diagonal terms of  $\mathbf{R}$  in data assimilation. The adaptive estimation method of Li *et al.*, which allows online estimation of the observation-error variance using innovation statistics, is extended to include off-diagonal terms of  $\mathbf{R}$ . The extended method performs well with the 40-variable Lorenz model in estimating non-diagonal observation-error covariances. Interestingly, the analysis accuracy is improved when the observation errors are correlated, but only if the observation-error correlations are explicitly considered in data assimilation. Further theoretical considerations relate the impact of observing systems (characterized by both  $\mathbf{R}$  and an observation operator  $\mathbf{H}$ ) on analysis accuracy. This analysis points out the importance of distinguishing between observation-error correlations (i.e. non-diagonal  $\mathbf{R}$ ) and correlated observations (i.e. non-orthogonal  $\mathbf{H}$ ). In general, observations with a non-diagonal  $\mathbf{R}$  carry more information, whereas observations with a non-orthogonal  $\mathbf{H}$  carry less information, but it turns out that the combination of  $\mathbf{R}$  and  $\mathbf{H}$  is essential: more information is available from positively (negatively) correlated observations with negatively (positively) correlated errors, resulting in a more accurate analysis.

**Keywords:** data assimilation; observation-error covariances; adaptive estimation; observation operator; Shannon entropy

### 1. Introduction

Since observation-error correlations are generally difficult to estimate and to consider in data assimilation, it is a common practice in numerical weather prediction (NWP) and other geophysical applications to assume that the observation-error covariance matrix  $\mathbf{R}$  is diagonal. The use of a diagonal  $\mathbf{R}$  simplifies the implementation of data assimilation equations and improves computational efficiency, but it may not be accurate if the actual (unknown) observation errors are correlated and  $\mathbf{R}$  is non-diagonal [1,2].

Although many geophysical observations are made independently and their errors can be assumed to be uncorrelated, some observations such as satellite radiances and retrieved

---

\*Corresponding author. Email: [miyoshi@atmos.umd.edu](mailto:miyoshi@atmos.umd.edu)

temperature profiles, atmospheric motion vectors, sea-surface winds by satellite scatterometers and radar data, which play important roles in NWP, are known to have significantly correlated errors (e.g. [3–7]). It is not surprising that radiosonde observations have vertically correlated errors since a single instrument observes an atmospheric profile at all vertical levels. However, the impact of observation-error correlations in data assimilation is a relatively new direction of research, with just a few studies published thus far [1,2]. This study aims to corroborate and further extend the previous studies.

A typical approach in satellite data assimilation to account for the correlated observation errors is to use a manually inflated diagonal  $\mathbf{R}$ . Garand *et al.* [5] found that considering the inter-channel error correlations of the Atmospheric Infrared Sounder (AIRS) lowered the weight given to the observations in the one-dimensional variational retrievals. Inflating the diagonal  $\mathbf{R}$  would also lower the weight to the observations. Since observations with correlated errors are usually assumed to have less information, inflating a diagonal  $\mathbf{R}$  could be an approximation of considering a full  $\mathbf{R}$  matrix with significant off-diagonal terms. This approach is not equivalent to considering the full  $\mathbf{R}$  matrix explicitly: one of the main goals of this study is to investigate the impact of inflating the diagonal of  $\mathbf{R}$ .

Another important issue addressed in this study is whether it is possible to estimate the observation-error correlations. Li *et al.* [8] ('LKM09' hereafter) proposed using the innovation statistics derived by Daley [9] and Desroziers *et al.* [10] for simultaneous online estimation of both the diagonal  $\mathbf{R}$  and the covariance inflation parameter of an ensemble Kalman filter (EnKF, e.g. [11,12]). Covariance inflation deals with a typical problem of the underestimated forecast error variance in EnKF. A common inflation method multiplies by a factor slightly larger than one the forecast ensemble perturbations [13]; the inflation factor is a tuning parameter for optimization of the state estimate and its error variance estimate. Since the innovation statistics include both  $\mathbf{R}$  and forecast error covariance, LKM09 proposed a method to estimate not only the diagonal  $\mathbf{R}$  but also the inflation parameter to optimize the forecast error variance. In this study, the LKM09 method is extended to include off-diagonal terms of  $\mathbf{R}$  in a straightforward manner.

Experiments are performed using an EnKF with a low-order model known as the Lorenz-96 model with 40 variables [14,15]. Using synthetic observations with correlated errors, 'identical twin' experiments are performed for purely focusing on the potential impact of observation-error correlations in data assimilation. In this study, we start by applying the LKM09 method for adaptive estimation of both the EnKF covariance inflation and the diagonal  $\mathbf{R}$  in order to assess how well it works when the observations have correlated errors. Then, the LKM09 method is extended to estimate the observation-error correlations explicitly. The comparison between the two experiments addresses the impact of the explicit consideration of full  $\mathbf{R}$  matrix in EnKF data assimilation.

Section 2 describes the numerical experiments, and the results are presented in Section 3. The discussion in Section 4 includes further consideration of the relations between the analysis accuracy and observing systems characterized by both  $\mathbf{R}$  and an observation operator  $\mathbf{H}$  using simple conceptual models. Finally, Section 5 provides conclusions.

## 2. Numerical experiments

The experimental system in this study is composed of the Lorenz-96 model with 40 variables [14,15] and the ensemble square root filter (EnSRF, [16]) but without the serial

treatment of observations. The ensemble size is fixed at 10 throughout this study, considerably smaller than the model's number of degrees of freedom.

The EnSRF analysis equations are given by

$$\bar{\mathbf{x}}^a = \bar{\mathbf{x}}^f + \mathbf{K}(\mathbf{y} - \mathbf{H}\bar{\mathbf{x}}^f), \quad (1)$$

$$\delta\mathbf{X}^a = [\mathbf{I} - \tilde{\mathbf{K}}\mathbf{H}]\delta\mathbf{X}^f, \quad (2)$$

where  $\mathbf{x}$  denotes the 40-dimensional state vector and  $\delta\mathbf{X}$  denotes the  $40 \times 10$  matrix whose  $i$ th column is the ensemble perturbation of the  $i$ th member. The overbar indicates the ensemble mean and the superscript indicates analysis ( $a$ ) or forecast ( $f$ ).  $\mathbf{y}$  denotes a vector composed of observed values and  $\mathbf{H}$  indicates the observation operator that maps a model variable to observation space. Only a linear observation operator is considered in this article, but the tangent-linear extension to deal with a nonlinear observation operator is straightforward (e.g. [17]); a detailed description is beyond the scope of this article.

The Kalman gain matrix  $\mathbf{K}$  in Equation (1) is given by

$$\mathbf{K} = (\rho \circ \Delta\mathbf{P}^f)\mathbf{H}^T\mathbf{E}^{-1}, \quad (3)$$

where  $\mathbf{P}^f$  is the  $40 \times 40$  forecast error covariance matrix estimated by 10 ensemble forecast perturbations as follows:

$$\mathbf{P}^f = \frac{1}{9}\delta\mathbf{X}^f(\delta\mathbf{X}^f)^T, \quad (4)$$

$\Delta$  is the covariance inflation factor, a scalar slightly larger than one. The matrix  $\mathbf{E}$  is defined by

$$\mathbf{E} = \mathbf{H}(\rho \circ \Delta\mathbf{P}^f)\mathbf{H}^T + \mathbf{R}, \quad (5)$$

$\rho$  is the  $40 \times 40$  covariance localization matrix whose  $(i, j)$  component is defined by the Gaussian function:

$$\rho(i, j) = \exp\left(-\frac{d(i, j)^2}{2\sigma^2}\right), \quad (6)$$

where  $d(i, j)$  denotes the grid distance between the  $i$ th and  $j$ th grid points with considering the cyclic boundary, and the symbol  $\circ$  indicates the Schur or element-wise product. Throughout this study, the localization scale parameter is fixed at  $\sigma = 2.0$ , and the inflation parameter  $\Delta$  is adaptively estimated as in LKM09 using the following innovation statistics at each assimilation step:

$$\Delta = \frac{(\mathbf{H}\bar{\mathbf{x}}^a - \mathbf{H}\bar{\mathbf{x}}^f)^T(\mathbf{y} - \mathbf{H}\bar{\mathbf{x}}^f)}{\text{tr}[\mathbf{H}(\rho \circ \Delta\mathbf{P}^f)\mathbf{H}^T]} \quad (7)$$

(cf. Equation (5) of LKM09). The ensemble perturbations are updated by Equation (2) with

$$\tilde{\mathbf{K}} = (\rho \circ \Delta\mathbf{P}^f)\mathbf{H}^T\mathbf{E}^{-\frac{1}{2}}[\mathbf{E}^{\frac{1}{2}} + \mathbf{R}^{\frac{1}{2}}] \quad (8)$$

as in Equation (10) of [16].

Identical twin experiments are performed. The model settings are chosen to be identical to those of Lorenz and Emanuel [15], except for the time step of 0.01 instead of 0.05. With the settings, 0.05 time unit corresponds to 6 h if we consider error doubling time analogous to synoptic weather [15]. A 1-year ‘nature’ run is generated by running the Lorenz-96 model after spinning-up for 5 years from a state composed of 40 random numbers drawn from  $N(0, 1)$ . Here,  $N(m, s)$  indicates the normal distribution of mean  $m$  and standard deviation  $s$ . 10-member ensemble initial conditions to initiate the EnKF are generated similarly by 5-year spin-up from states of 10 different realizations of 40 random numbers drawn from  $N(0, 1)$ .

The observing system is chosen to be 20 observations, located at every other grid point, so that the observation operator  $\mathbf{H}$  is a  $20 \times 40$  matrix composed of either 1 or 0. The synthetic observations are simulated every 6 h, i.e. every five model time steps, by applying  $\mathbf{H}$  to the nature state and adding to the true values 20 random numbers that satisfy the error covariance matrix  $\mathbf{R}$ . A simple choice of a non-diagonal  $\mathbf{R}$  is

$$(\mathbf{R}_{ij}) = \frac{v}{1 + \frac{d(i,j)}{2L}}, \quad (9)$$

where the observation-error variance is fixed at  $v=1.0$ .  $L$  is the correlation strength parameter, so that larger values of  $L$  correspond to stronger observation-error correlations. In the limit of  $L=0$ ,  $\mathbf{R}$  is diagonal. The  $\mathbf{R}$  matrix is shown in Figure 1. Although we only show results for this case (Equation 9), other choices for  $\mathbf{R}$  including negative correlations were tested, and they all yielded similar results.

Correlated random numbers are generated using uncorrelated random numbers and the Cholesky decomposition of  $\mathbf{R}$ . Let  $\mathbf{n}(0, 1)$  denote a vector composed of 20 normally distributed uncorrelated random numbers drawn from  $N(0, 1)$ . The vector of correlated random observation errors  $\varepsilon^o$  is given by

$$\varepsilon^o = \mathbf{C}\mathbf{n}(0, 1), \quad (10)$$

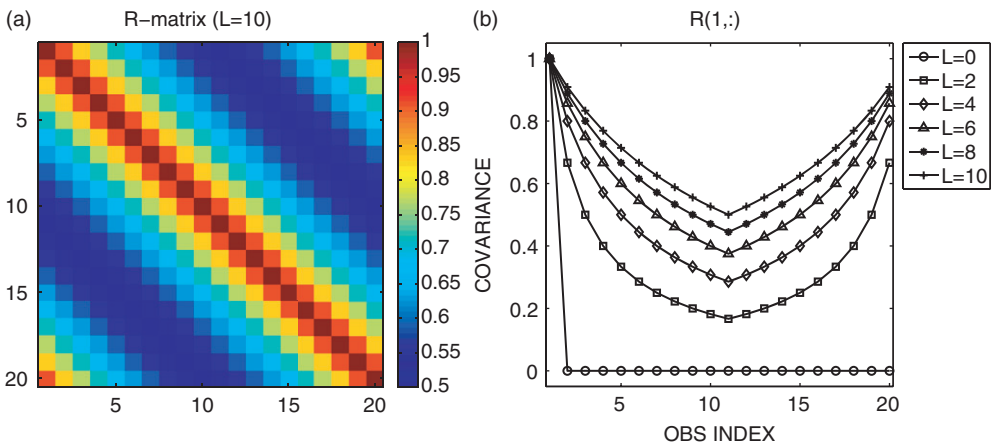


Figure 1. Observation-error covariance  $\mathbf{R}$  defined by Equation (9): (a) the full  $\mathbf{R}$  matrix when  $L=10$ ; (b) the first column (or row) of the  $\mathbf{R}$  matrix with various choices of  $L$ .

where  $\mathbf{R} = \mathbf{C}\mathbf{C}^T$  is the Cholesky decomposition of  $\mathbf{R}$ . The observations are then simulated by

$$\mathbf{y} = \mathbf{H}\mathbf{x}^T + \varepsilon^o, \quad (11)$$

where  $\mathbf{x}^T$  is the nature state or ‘truth’. The analysis accuracy is measured by the root mean-square difference from the nature run, which we call root mean-square errors (RMSEs). The RMSE is averaged over the last 6 months of 1-year EnKF experiments to exclude the initial spin-up.

LKM09 developed an adaptive method to estimate observation-error variances and the covariance inflation parameter simultaneously. It uses the innovation statistics derived by Desroziers *et al.* [10]

$$\mathbf{R} = \left\langle (\mathbf{y} - \mathbf{H}\bar{\mathbf{x}}^a)(\mathbf{y} - \mathbf{H}\bar{\mathbf{x}}^a)^T \right\rangle. \quad (12)$$

The bracket  $\langle \cdot \rangle$  indicates statistical expectation, usually obtained by averaging over many cases. LKM09 used this relationship and estimated observation-error variances successfully by assuming observations in each type of observation, e.g. wind and temperature, share the same error variance but without error correlations. In this study, we simply use the same Equation (12) to estimate the non-diagonal elements of  $\mathbf{R}$ . Although we can apply the method to estimate each  $\mathbf{R}$  for each group of observations, we assume the 20 observations in the present experiments are in the same group. In addition, we further assume that the error covariance is a function of the distance between observations in order to reduce the degrees of freedom of the  $\mathbf{R}$  matrix. Namely, the 20 diagonal elements of the  $\mathbf{R}$  matrix correspond to the zero distance, which is a single unknown. Similarly, the 20 elements adjacent to the diagonal correspond to the distance of two grid points, another single unknown, since the 20 observations are located at every other grid points. Overall, there are only 20 unknown elements needed to construct the full  $20 \times 20$   $\mathbf{R}$  matrix. We have 20 observations or statistical samples to estimate the 20 unknowns at each assimilation step. The estimates are temporally smoothed by a simple weighted average with the prior estimate from the previous analysis cycle:

$$\mathbf{R}_t = (1 - \alpha)\mathbf{R}_{t-1} + \alpha\mathbf{R}_t^{\text{est}}. \quad (13)$$

Here, the subscripts indicate the assimilation time steps and  $\mathbf{R}_t^{\text{est}}$  is the estimate obtained from Equation (12) at time  $t$ . We use  $\alpha = 0.03$ , i.e., we give 97% of the weight to the prior estimate. The initial (wrong) guess of the  $\mathbf{R}$  matrix is a diagonal with a variance twice as large as the variance in the true  $\mathbf{R}$ .

### 3. Results

Using a diagonal estimation of  $\mathbf{R}$  from the original LKM09 method, the EnKF performs stably with observations with correlated errors (Figure 2, square marks). With various magnitudes of observation-error correlations varying from  $L = 0$  (no correlation) to  $L = 10$  (strong correlation), the RMSEs are below 0.4, less than a half of the observation-error standard deviation, and the RMSEs tend to decrease slightly with larger  $L$ . The estimated observation-error variance (i.e. diagonal values) is very close to 1.0, the true observation-error variance. As a result, using a fixed diagonal  $\mathbf{R}$  with the true observation-error variance gives essentially the same analysis accuracy (not shown, but indistinguishable

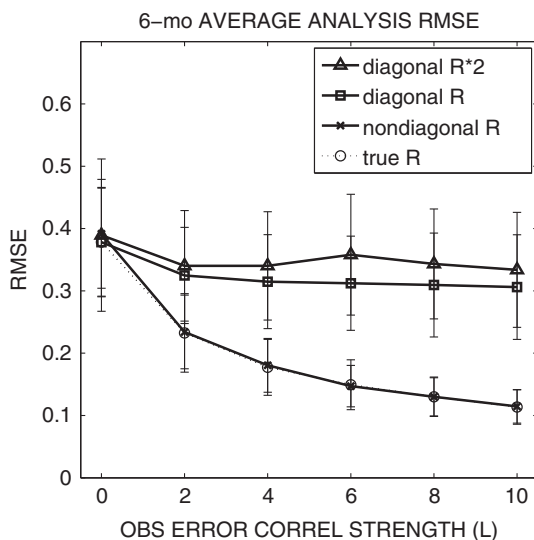


Figure 2. A 6-month average analysis RMSE for the cases of using true  $\mathbf{R}$  (circle with thin dashed line), adaptively estimated non-diagonal  $\mathbf{R}$  (cross), adaptively estimated diagonal  $\mathbf{R}$  (square) and fixed diagonal  $\mathbf{R}$  with twice the true variance (triangle). The abscissa indicates the observation-error correlation strength parameter  $L$ . The error bars indicate the standard deviation of temporal fluctuations. The two curves of true  $\mathbf{R}$  and adaptively estimated non-diagonal  $\mathbf{R}$  are nearly identical and on top of each other.

from the curve with square marks). When the diagonal  $\mathbf{R}$  is inflated, for example to double the true observation-error variance, the analysis accuracy is degraded when the observation errors are correlated (Figure 2, triangle marks). The approach of using an inflated diagonal  $\mathbf{R}$  is not appropriate to deal with observation-error correlations.

When the non-diagonal  $\mathbf{R}$  is explicitly considered in data assimilation, the analysis accuracy is significantly improved as the observation-error correlations become more significant (Figure 2, cross and circle marks). Non-zero error correlations imply that the errors compared to the truth have similar patterns. Since we used the same number of observations to estimate such correlated errors, we were able to achieve a better analysis. In other words, the error correlation implies more information, and we can potentially extract more from observations by considering non-diagonal  $\mathbf{R}$  explicitly in data assimilation.

The straightforward extension of the LKM09 method for the non-diagonal  $\mathbf{R}$  matrix performs properly and gives an analysis accuracy similar to that obtained using the true  $\mathbf{R}$  (cross and circle marks in Figure 2). Several other choices for the true  $\mathbf{R}$  matrix including negative correlations were tested, and they all gave essentially equivalent results, namely improving the analysis when estimating the error correlations. The 6-month average of the adaptively estimated  $\mathbf{R}$  indicates very accurate estimation of the true  $\mathbf{R}$  (cf. Figure 3 with Figure 1b). The time series of the adaptively estimated observation-error covariance indicates a certain level of noise (Figure 4), which largely depends on the choice of the time-smoothing parameter  $\alpha$  in Equation (13). A larger  $\alpha$  gives noisier time series due to sampling noise at each time, as shown by LKM09. Although a smaller  $\alpha$  gives a smoother time series, it spins up more slowly. Using the Kalman filter update for Equation (13) as

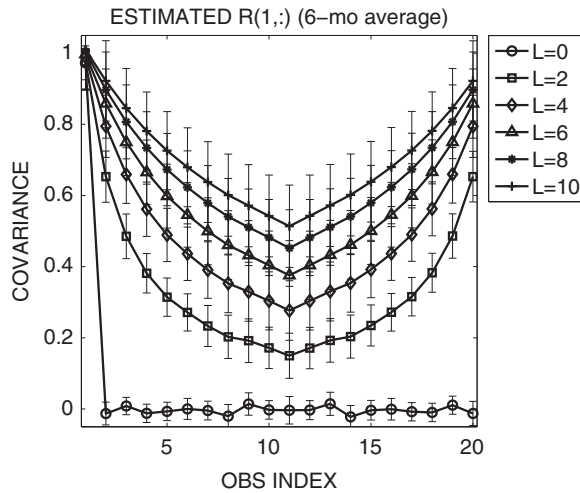


Figure 3. As Figure 1(b), but for the 6-month average of the adaptively estimated  $\mathbf{R}$ . The error bars indicate the standard deviation of temporal fluctuations.

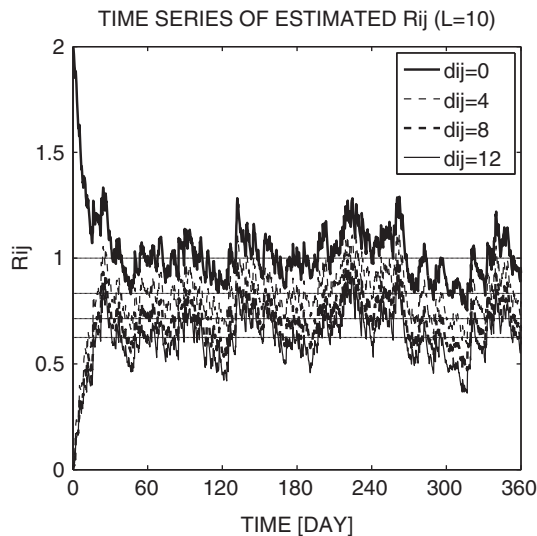


Figure 4. Time series of the adaptively estimated components of the full  $\mathbf{R}$  matrix in the case of  $L = 10$ .

Note: Thin dotted lines indicate the true covariance values. Each line type corresponds to the element  $\mathbf{R}_{ij}$  whose grid distance  $d_{ij}$  between the  $i$ th and  $j$ th observations is as shown in legend.

done by LKM09 provides faster spin-up, but the two approaches are essentially equivalent after the spin-up. The estimate is quite accurate after about 1-month spin-up; the initial guess of the  $\mathbf{R}$  matrix (diagonal with twice as large observation-error variance) quickly approaches the true  $\mathbf{R}$  values. Since the adaptively estimated  $\mathbf{R}$  with some noise gave almost identical analysis accuracy as the true  $\mathbf{R}$  (cross and circle marks in Figure 2), the



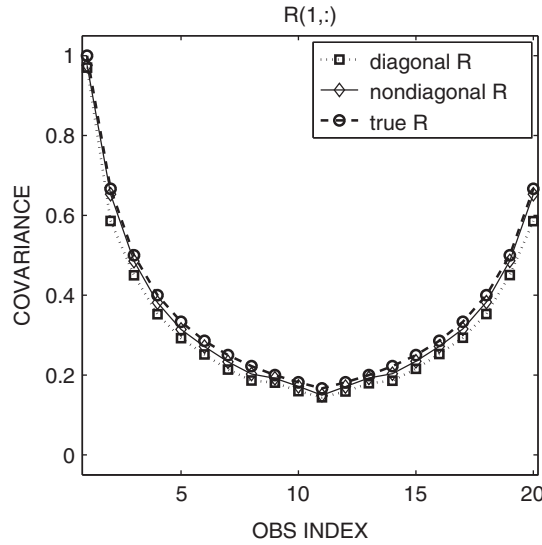


Figure 5. First row (or column) of the true  $\mathbf{R}$  matrix (dashed line with circles), adaptively estimated  $\mathbf{R}$  matrix using non-diagonal  $\mathbf{R}$  (solid line with diamonds) and diagonal  $\mathbf{R}$  (dotted line with squares) in data assimilation.

sensitivity to the choice of  $\alpha$  would not be significant; a smaller  $\alpha$  would give smoother time series of estimated  $\mathbf{R}$ , but the impact to the analysis accuracy would be small.

In real applications, it may be useful to first attempt to estimate the off-diagonal components of  $\mathbf{R}$  without considering it in data assimilation. Namely, while using a diagonal  $\mathbf{R}$  in EnKF assimilation, we can estimate a non-diagonal  $\mathbf{R}$  in order to check whether the real observations have significant error correlations. The statistical relationship Equation (12) would be accurate if the full  $\mathbf{R}$  matrix is considered in data assimilation and if the analysis is optimal. However, even when a diagonal  $\mathbf{R}$  is used, we obtain fairly accurate estimate of non-diagonal  $\mathbf{R}$  after spin-up (Figure 5). The results are encouraging, suggesting that it is possible to estimate a non-diagonal  $\mathbf{R}$  of real observations without changing the current data assimilation system assuming a diagonal  $\mathbf{R}$ .

#### 4. Information content analysis

The results showed that the analysis was more accurate when observations had correlated errors. This implies more information available from observations with correlated errors. In information theory, entropy is a quantitative measure of information (e.g. [18]). Stewart *et al.* [1] used entropy reduction (known as the Shannon information content) for numerically evaluating the information loss due to simplifying observation-error correlations. Entropy  $S$  is defined as a function of probability density function (PDF)  $p$  of probabilistic variable  $x$ :

$$S = - \int p(x) \log p(x) dx. \quad (14)$$

Larger entropy corresponds to less information. In data assimilation, observations are considered to be a probabilistic variable with multi-dimensional Gaussian PDF, i.e. the PDF of  $n$  observations is given by the  $n$ -dimensional normal distribution:

$$p(\mathbf{y}) = \frac{1}{(2\pi)^{\frac{n}{2}}|\mathbf{R}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{y} - \bar{\mathbf{y}})^T \mathbf{R}^{-1}(\mathbf{y} - \bar{\mathbf{y}})\right), \quad (15)$$

where  $\bar{\mathbf{y}}$  indicates most probable values, which are usually observed values. Combining Equations (14) and (15), we obtain that the entropy of the observation PDF is given by

$$S = \frac{1}{2} \log|\mathbf{R}| + \text{const.} \quad (16)$$

$|\mathbf{R}|$  decreases if the diagonal terms are smaller and if the absolute values of the off-diagonal terms are larger. Since the logarithm function is monotonic, this implies that observations have more information if their error variances are smaller and if their error correlations are larger, which agrees with our results.

Although Equation (16) is considered to be the information of observations, it does not necessarily correspond to information *available* to the analysis through the data assimilation process. The following simple conceptual model highlights this point. Assume that we have two observations for a scalar variable, such as two thermometers in a small room. The observation operator  $\mathbf{H}$  is given by

$$\mathbf{H} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}. \quad (17)$$

A general non-diagonal observation-error covariance matrix  $\mathbf{R}$  is

$$\mathbf{R} = \begin{bmatrix} \sigma_1^2 & c\sigma_1\sigma_2 \\ c\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}, \quad (18)$$

where the two observations have the error standard deviations  $\sigma_1$  and  $\sigma_2$ , respectively, and there is an error correlation  $c$ . Assuming  $\sigma_1 = \sigma_2 = \sigma$  for simplicity, we obtain the analysis PDF of the Kalman filter analysis update which ensures the optimal use of observations:

$$\mathbf{P}^a = \left(1 - \frac{2\mathbf{P}^f}{2\mathbf{P}^f + \sigma^2(1+c)}\right)\mathbf{P}^f. \quad (19)$$

With a given background error  $\mathbf{P}^f$ , the analysis error  $\mathbf{P}^a$  is smaller when the two observations have negatively correlated errors ( $c < 0$ ). Conversely, positive error correlations ( $c > 0$ ) yield larger analysis errors than the case with uncorrelated errors. Therefore, even if observations have more information as given by the entropy in Equation (16), it may not ‘be available’ to reduce the analysis error in data assimilation. In summary, the available information of observations depends not only on  $\mathbf{R}$  but also on  $\mathbf{H}$ . Both  $\mathbf{R}$  and  $\mathbf{H}$  define the observing network, and their combination determines the available information in data assimilation.

In order to further elaborate the relation between the observing network and the available information, another simple conceptual model is considered. Let us

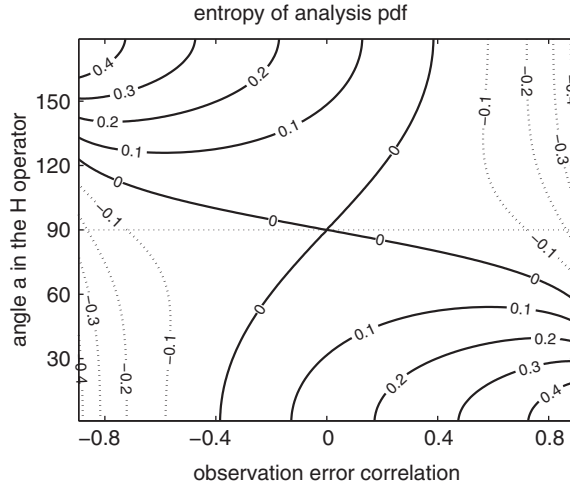


Figure 6. Entropy of analysis PDF, an indicator of available information of observations, as a function of the observation-error correlations (abscissa) and the angle between the two observations (ordinate).

Note: The smaller the entropy, the more information is available from the observations.

assume two observations for a two-dimensional system. Assume the observing system is defined by

$$\mathbf{H} = \begin{bmatrix} \sin \theta & \cos \theta \\ \sin(\theta + a) & \cos(\theta + a) \end{bmatrix}, \quad (20)$$

$$\mathbf{R} = \begin{bmatrix} \sqrt{1+c^2} & c \\ c & \sqrt{1+c^2} \end{bmatrix}. \quad (21)$$

Here, information of observations is fixed at  $|\mathbf{R}| = 1$  in order to purely focus on the *available* information. The observation-error correlations are given by  $\frac{c}{\sqrt{1+c^2}}$  and  $a$  is the angle between the two observations. Since  $\theta$  is arbitrary due to isotropy,  $a$  and  $c$  are the two parameters that define the observing network. Without loss of generality, the background error covariance matrix is assumed to be the identity matrix:  $\mathbf{P}^f = \mathbf{I}$ . Since the analysis PDF of the Kalman filter update is the result of the optimal use of the observations, entropy of the analysis PDF is considered to be an indicator of the available information of the observations in data assimilation (Figure 6). The smaller the entropy, the more information is available from observations. When the observation errors are uncorrelated, the available information is maximum at  $a = 90^\circ$ , i.e. the observation vectors are perpendicular. If the two observations observe a similar structure, the observation operator  $\mathbf{H}$  is non-orthogonal, and less information is available in data assimilation. A typical example is satellite radiance observations whose channels have overlapping weighting function, so that the angles between different channels' observation operators are not perpendicular. If  $a = 90^\circ$ , i.e.  $\mathbf{H}$  is orthogonal, the available information is minimum when the observation errors are uncorrelated. Namely, orthogonal observations carry more information when observation errors are correlated either negatively

or positively. This agrees with the Lorenz model results in the previous section, in which only orthogonal  $\mathbf{H}$  was considered, and both positively and negatively correlated errors improved the analysis accuracy. Uncorrelated conventional observations are a typical example of orthogonal  $\mathbf{H}$ . The entropy of the analysis PDF (Figure 6) shows a general pattern of smaller values in the top-right and bottom-left corners, and of larger values in the top-left and bottom-right corners. The pattern corresponds to the previous one-dimensional case in the sense that the best combination is negatively (positively) correlated observations with positively (negatively) correlated errors. Therefore, the correlations in the observation operator have to be considered in addition to the observation-error correlations.

## 5. Conclusions

Although at least some of the important observations in NWP have significantly correlated errors, the observation-error covariance matrix  $\mathbf{R}$  is commonly assumed to be diagonal. This study aimed to investigate the impact of considering off-diagonal terms of  $\mathbf{R}$  in data assimilation. One of the two important findings is that the error-correlated observations give a better analysis only when the non-diagonal  $\mathbf{R}$  is explicitly considered in data assimilation. The results may appear to be against the intuitive notion that error-correlated observations may have less information. However, information theory tells us that off-diagonal terms of the error covariance matrix correspond to more, not less information. Moreover, the simple conceptual examples considered indicate that observations can provide more information to the analysis when the observation network has a favourable combination of the observation operator  $\mathbf{H}$  and the observation-error covariance matrix  $\mathbf{R}$ . Namely, more information is available from positively (negatively) correlated observations with negatively (positively) correlated errors. This can be further investigated by introducing various non-orthogonal  $\mathbf{H}$  operators with the Lorenz model, which is a subject of future research.

The second main finding of this study is the successful and promising performance of the adaptive estimation of a non-diagonal  $\mathbf{R}$  with EnKF. The method is a straightforward extension of the LKM09 method for adaptive estimation of a diagonal  $\mathbf{R}$  to include off-diagonal components. The extended method provided fairly accurate estimate of observation-error correlations even with suboptimal data assimilation system using a diagonal  $\mathbf{R}$ .

This study was carried out with only idealized low-order experiments and two conceptual experiments. It is important to further extend this study to estimating a non-diagonal  $\mathbf{R}$  matrix of real observations, and to considering the non-diagonal  $\mathbf{R}$  matrix in the local ensemble transform Kalman filter [17], with realistic NWP models. In real applications, a number of complications would arise, such as the inevitable presence of bias in both the observations and forecasts. Those biases would be spatially correlated and would naturally interfere with the estimate of the observation-error correlations. These possible complications should be identified and tackled in future studies.

## Acknowledgements

The lead author (Takemasa Miyoshi) is grateful to Haroldo Fraga de Campos Velho of INPE (Brazilian National Institute for Space Research) for his kind invitation to the session IN22A

'Inverse Problems in Geosciences' at the AGU meeting in Brazil. We also thank Mark Buehner of the Environment Canada and Chris Snyder of the National Center for Atmospheric Research for insightful discussions. This study was supported by the Office of Naval Research grant N000141010149 under the National Oceanographic Partnership Program.

## References

- [1] L.M. Stewart, S.L. Dance, and N.K. Nichols, *Correlated observation errors in data assimilation*, Int. J. Numer. Methods Fluids 56 (2008), pp. 1521–1527.
- [2] L.M. Stewart, *Correlated observation errors in data assimilation*, Ph.D. diss., University of Reading, Reading, 2009, p. 195.
- [3] R. Keeler and S. Ellis, *Observational error covariance matrices for radar data assimilation*, Phys. Chem. Earth Part B 25 (2000), pp. 1277–1280.
- [4] N. Bormann, S. Saarinen, G. Kelly, and J.-N. Thepaut, *The spatial structure of observation errors in atmospheric motion vectors from geostationary satellite data*, Mon. Weather Rev. 131 (2003), pp. 706–718.
- [5] L. Garand, S. Heilliette, and M. Buehner, *Interchannel error correlation associated with AIRS radiance observations: Inference and impact in data assimilation*, J. Appl. Meteorol. Climatol. 46 (2007), pp. 714–725.
- [6] N. Bormann and P. Bauer, *Estimates of spatial and interchannel observation-error characteristics for current sounder radiances for numerical weather prediction. I: Methods and application to ATOVS data*, Q. J. R. Meteorolog. Soc. 136 (2010), pp. 1036–1050.
- [7] N. Bormann, A. Collard, and P. Bauer, *Estimates of spatial and interchannel observation-error characteristics for current sounder radiances for numerical weather prediction. II: Application to AIRS and IASI data*, Q. J. R. Meteorolog. Soc. 136 (2010), pp. 1051–1063.
- [8] H. Li, E. Kalnay, and T. Miyoshi, *Simultaneous estimation of covariance inflation and observation errors within an ensemble Kalman filter*, Q. J. R. Meteorolog. Soc. 135 (2009), pp. 523–533.
- [9] R. Daley, *Estimating model-error covariances for application to atmospheric data assimilation*, Mon. Weather Rev. 120 (1992), pp. 1735–1746.
- [10] G. Desroziers, L. Berre, B. Chapnik, and P. Poli, *Diagnosis of observation, background and analysis-error statistics in observation space*, Q. J. R. Meteorolog. Soc. 131 (2005), pp. 3385–3396.
- [11] G. Evensen, *Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics*, J. Geophys. Res. 99(C5) (1994), pp. 10143–10162.
- [12] G. Evensen, *The ensemble Kalman filter: Theoretical formulation and practical implementation*, Ocean Dyn. 53 (2003), pp. 343–367.
- [13] J.L. Anderson and S.L. Anderson, *A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts*, Mon. Weather Rev. 127 (1999), pp. 2741–2758.
- [14] E. Lorenz, *Predictability: A Problem Partly Solved*, Proceedings of the ECMWF Seminar on Predictability, Vol. 1, Reading, UK, 1996.
- [15] E. Lorenz and K. Emanuel, *Optimal sites for supplementary weather observations: Simulation with a small model*, J. Atmos. Sci. 55 (1998), pp. 399–414.
- [16] J.S. Whitaker and T.M. Hamill, *Ensemble data assimilation without perturbed observations*, Mon. Weather Rev. 130 (2002), pp. 1913–1924.
- [17] B.R. Hunt, E.J. Kostelich, and I. Szunyogh, *Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman Filter*, Phys. Nonlinear Phenom. 230 (2007), pp. 112–126.
- [18] C.D. Rodgers, *Inverse Methods for Atmospheric Sounding: Theory and Practice*, Vol. 2, World Scientific, Singapore, 2000, p. 256.