Chapter 9. Ensemble based Data assimilation

Youmin Tang

In previous chapters, we introduced modeling techniques based on either observed data (statistical modeling) or numerical models (numerical modeling). In this chapter, we will talk the modeling and simulation using both approaches, that is, the observations will be incorporated into numerical models for optimal modeling and simulation. In statistics, this is called state-space estimation. In the earth science, it is called data assimilation. For example, a strict definition of data assimilation in atmospheric and oceanic sciences is "the process to estimate the state of a dynamic system such as atmospheric and oceanic flow by combining the observational and model forecast data (Talagrand 1997)".

Intuitionally, one might think that an optimal simulation scheme is to directly replace model variables by observations during numerical integrations. Such a direct replacement is usually not correct since observations are not perfect and contain errors. A simple replacement will introduce observation errors into models, and ignore possible impact of observation errors on model behaviors, easily resulting in imbalance of model dynamics and physics. Thus, the application of observations into numerical models must consider both model and observation errors, which play a critical role in the assimilation process.

We will start to display the assimilation concept by a simple example.

For an unknown true state value, denoted by T_t, there are two samples, denoted by T_1 (e.g. model simulation) and T_2 (observation), which have the errors ε_1 and ε_2, respectively. Thus, we have:

$$T_1 = T_t + \varepsilon_1$$
$$T_2 = T_t + \varepsilon_2$$

If the measurement or observation is unbiased, and the variances of errors are known, i.e.,

$$E(\varepsilon_1)=E(\varepsilon_2)=0;\ Var(\varepsilon_1)=\sigma_1;\ Var(\varepsilon_2)=\sigma_2. \qquad (1)$$

The question here is to seek an optimal estimate, denoted by T_a (called analysis in the assimilation field), for T_t using T_1 and T_2. This optimal estimate is the central issue of data assimilation.

There are several methods for this solution, as demonstrated below.

7.1 The general framework of several assimilation approaches

7.1.1 Least square method

Denote $T_a = a1 * T_1 + a2 * T_2$. T_a should be unbiased. i.e., E(T_a)=E(T_t), so

$a1 * E(T_1) + a2 * E(T_2) = E(T_t)$, i.e.,

$a1 + a2 = 1$, where (1) is invoked.

The best (optimal) estimate should satisfy the below condition:

$\sigma_a^2 = E(T_a - T_t)^2$ ➔ min

$$\sigma_a^2 = E(a1 * T_1 + a2 * T_2 - (a1 + a2) * T_t) = E[a1 * (T_1 - T_t) + a2 * (T_2 - T_t)]^2$$
$$= E(a1^2 * \varepsilon_1^2 + a2^2 * \varepsilon_2^2 + 2 * a1 * a2 * \varepsilon_1 * \varepsilon_2) = a1^2 * \sigma_1^2 + a2^2 * \sigma_2^2 = a1^2 * \sigma_1^2 + (1 - a1)^2 \sigma_2^2$$

Thus,

$\frac{\partial \sigma_a^2}{\partial a1} = 0$ ➔ $a1 = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}$; Namely;

$$T_a = a1 * T_1 + (1 - a1) * T_2 = T_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}(T_2 - T_1)$$

7.1.2 Variational approach

In general, assimilation methods can be classified into two categories: variational and sequential. Variational methods such as three-dimensional variational (3D-VAR) method and four-dimensional variational (4D-VAR) method (Dimet and Talagrand 1986, Courtier et al. 1998) are batch methods, whereas sequential methods such as Kalman filter proposed by Kalman (1961) belong to the estimation theory. They both have had great success. The European Centre forMedium-RangeWeather Forecasts (ECMWF) introduced the first 4D-VAR method into the operational global analysis system in November 1997 (Rabier et al. 2000, Mahfouf and Rabier 2000, Klinker et al. 2000). EnKF was first introduced into the operational ensemble prediction system by Canadian Meteorological Centre (CMC) in January 2005 (Houtekamer et al. 2005).

Below we will demonstrate the idea of variaonal assimilation by the above example. First, a cost function should be defined for variational assimilation approach. For this simple example, we define the cost function as below:

$$J(T) = \frac{1}{2}[\frac{(T - T_1)^2}{\sigma_1^2} + \frac{(T - T_2)^2}{\sigma_1^2}] \quad (2)$$

$$T = a1 * T`_1 + a2 * T_2 \quad (3)$$

The solution is to seek an analysis T_a, determined by a1 and a2, leading to the cost function minimum, i.e., $J(T_a)$ ➔ minimum. Obviously, we have

$$\frac{\partial J(T)}{\partial a1} = 0; \qquad \frac{\partial J(T)}{\partial a2} = 0$$

$$\frac{\partial J(T)}{\partial a1} = \frac{T - T_1}{\sigma_1^2}\frac{\partial T}{\partial a1} + \frac{T - T_2}{\sigma_2^2}\frac{\partial T}{\partial a1} = 0; \tag{4}$$

Eq. (3) leads to $\frac{\partial T}{\partial a1} = T_1$. Thus, the solution of (4), denoted by T_a, satisfies

$$T_a = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}T_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}T_2 \tag{5}$$

The above is a simple example of variaonal assimilation approach, where we only consider the analysis error (cost function) for a time point. However, in many cases, we need to consider the error growth during a period, i.e., the sum of errors during the period, in the cost function (2). For example, the cost function is defined as below

$$J(T) = \frac{1}{2}\sum_{t=1}^{N} [\frac{(T(t) - T_1(t))^2}{\sigma_1^2} + \frac{(T(t) - T_2(t))^2}{\sigma_1^2}] \tag{6}$$

Meanwhile $T(t)$ follows a dynamical model, saying $T(t) = \int_{t_0}^{N} F(T(t))dt = M(T(t_0))$, where *F* is a nonlinear dynamical model, M is the integral operator, and t_o is the initial time. Thus, the cost function value of (6) is only determined by T_0. Namely, the objective here is to seek optimal initial condition T_0 that enables (6) minimum, i.e., minimizing (6) subject to dynamical model *F*. This is a standard conditional extreme problem which can be solved by Lagrange Multiplier approach. However the complexity of dynamical model excludes the possibility to get the analytical solution. We have to solve the minimum problem by aid of numerical methods, e.g., Newton conjugate gradient method. All of numerical methods require the gradient value ($\frac{\partial J}{\partial T_0}$) for solution. Again, it is almost impossible for obtaining analytical solution of $\frac{\partial J}{\partial T_0}$ due to the complexity of *F*. Usually researchers get the gradient value numerically using an approach of "tangent-linear and adjoint model". The details on tangent-linear and adjoint model can be found in relevant references as cited above. It should be noticed that it is

very difficult, even intractable sometimes, to construct tangent-linear and adjoint model in some cases. Thus, more and more researchers have started to apply sequential assimilation methods instead of 4-D variational assimilation methods in recent years. Next we will introduce the concept of the sequential assimilation method using the above example.

7.1.3 Bayesian approach

Assume T_1 is model prediction, which implies *a prior* probability distribution of truth *T,*

$$p(T) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(T_1-T)^2}{2\sigma_1^2}}$$

Given another observation T_2, the posterior distribution of the truth can be expressed by Bayesian formula:

$$p(T \mid T_2) = \frac{p(T_2 \mid T) * P(T)}{P(T_2)} \propto \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{(T_2-T)^2}{2\sigma_2^2}} * \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(T_1-T)^2}{2\sigma_1^2}} \quad (7)$$

$p(T_2)$ was ignored in (7) since it is independent of *T*. The $p(T_2 \mid T)$ is assumed to be Gaussian $\propto N(T, \sigma_2^2)$. The objective here is to estimate the truth by maximizing the *a posteriori* probability $p(T \mid T_2)$ (namely, we ask the truth to occur as much as possible --- maximum probability). Maximizing $p(T \mid T_2)$ is equivalent to maximizing the logarithm of the right item of (7), i.e.,

$$\log(p(T \mid T_2)) = \log(\frac{1}{\sqrt{2\pi}\sigma_2}) - \frac{(T_2 - T)^2}{2\sigma_2^2} + \log(\frac{1}{\sqrt{2\pi}\sigma_1}) - \frac{(T_1 - T)^2}{2\sigma_1^2} = const - \frac{1}{2}[\frac{(T - T_2)^2}{\sigma_2^2} + \frac{(T - T_1)^2}{\sigma_1^2}]$$

➔ maximum (8)

Obviously, maximum of $p(T \mid T_2)$ occurs at the minimum of the second item of right side hand of (8), i.e., minimum of the cost function *J* of (2). Thus, under the assumption of Gaussian distribution, maximizing *a posterior* probability (Bayesian approach) is equivalence to minimizing cost function (varitonal assimilation approach). Further, if the model *F* is linear and the probability distribution is Gaussian, it can be proved that the Kalman filter is equivalent to 4-D Var adjoint assimilation method.

7.2 Optimal Interpolation (OI) and Kalman filter (KF) and KF-based derivatives

7.2.1 Optimal Interpolation (OI)

The most special case in data assimilation is that the forecast model is linear and the errors are Gaussian. The solution among sequential methods to this case is provided by Kalman filter. Typically the Kalman filter applies to the below state-space model:

$$\begin{aligned} x_{t+1} &= Mx_t + \eta_t \\ y_t &= hx_t + \varsigma_t \end{aligned} \qquad (9)$$

where M and h are linear operator of model and measurement. The x is model state and y is the observation. The η_t and ς_t are the model errors and observed errors, respectively, which have variance: var(η_t)=$<\eta_t, \eta_t^T>$ = Q, var(ς_t)=$<\varsigma_t, \varsigma_t^T>$ =R. The objective here is to estimate model state x using y, making it close to true state (unknown) as much as possible.

Assuming the estimate of model state x_a at a time step is a linear combination of model forecast x_b and observation y_0, i.e., the filter itself is linear, so

$$x_a = x_b + K[y_0 - hx_b] \qquad (10)$$

Eq (10) is the standard expression of Kalman filter. K is called Kalman gain which determines the optimal estimate. Before deriving the K, we denote the variance of the analysis error ε_a by P_a, i.e., $P_a = <\varepsilon_a, \varepsilon_a^T>$ where $\varepsilon_a = x_a - x_{tr}$ and x_{tr} is the true value of model state. Similarly, observed errors and forecast errors are defined by $\varepsilon_o = \varsigma = y_o - hx_{tr}$ and $\varepsilon_b = x_b - x_{tr}$, respectively. It should be noticed that the forecast error ε_b is different from the model error η_t that is a systematic bias. Also, denote

$B = <\varepsilon_b, \varepsilon_b^T>$ ---------- background (forecast) error covariance
$R = <\varepsilon_o, \varepsilon_o^T>$ ---------- Observation error covariance

Assuming the observation error is not related to forecast error, so

$$<\varepsilon_b, \varepsilon_o^T> = <\varepsilon_o, \varepsilon_b^T> = 0$$

Clearly, we ask the K that can lead to P_a minimum. Subtracting x_{tr} on both sides of eq (10) leads to the below eq.

$$x_a - x_{tr} = x_b - x_{tr} + K[y_0 - hx_b + hx_{tr} - hx_{tr}] \qquad (11)$$

Namely,

$$\boldsymbol{\varepsilon}_a = \boldsymbol{\varepsilon}_b + K(\boldsymbol{\varepsilon}_o - h\boldsymbol{\varepsilon}_b) \qquad (12)$$

$$P_a = E[\boldsymbol{\varepsilon}_b + K(\boldsymbol{\varepsilon}_o - h\boldsymbol{\varepsilon}_b)][\boldsymbol{\varepsilon}_b + K(\boldsymbol{\varepsilon}_o - h\boldsymbol{\varepsilon}_b)]^T = E[\boldsymbol{\varepsilon}_b + K(\boldsymbol{\varepsilon}_0 - h\boldsymbol{\varepsilon}_b)][\boldsymbol{\varepsilon}_b{}^T + (\boldsymbol{\varepsilon}_0 - h\boldsymbol{\varepsilon}_b)^T K^T]$$
$$= E[\boldsymbol{\varepsilon}_b\boldsymbol{\varepsilon}_b{}^T + \boldsymbol{\varepsilon}_b(\boldsymbol{\varepsilon}_o - h\boldsymbol{\varepsilon}_b)^T K^T + K(\boldsymbol{\varepsilon}_o - h\boldsymbol{\varepsilon}_b)\boldsymbol{\varepsilon}_b{}^T + K(\boldsymbol{\varepsilon}_o - h\boldsymbol{\varepsilon}_b)(\boldsymbol{\varepsilon}_o - h\boldsymbol{\varepsilon}_b)^T K^T]$$
$$= B - Bh^T K^T - KhB + K(R + hBh^T)K^T = B - Bh^T K^T - K(Bh^T)^{\mathrm{T}} + K(R + hBh^T)K^T \qquad (13)$$

Here we used $B = B^T$. The optimal estimate asks the trace of P_a minimum, namely,

$$\frac{\partial[trace(P_a)]}{\partial K} = 0, \text{ so}$$

$$-2Bh^T + 2K(R + hBh^T) = 0$$
$$K = Bh^T(hBh^T + R)^{-1} \qquad (14)$$

$$P_a = B - Bh^T K^T - K(Bh^T)^{\mathrm{T}} + K(R + hBh^T)K^T = B - Bh^T K^T - KhB + Bh^T(hBh^T + R)^{-1} * (R + hBh^T)K^T$$
$$= B - KhB = (I - Kh)B$$

where we invoked the below properties:

$$\frac{\partial Ax}{\partial x} = \frac{\partial x^T A}{\partial x^T} = A;$$
$$\frac{\partial x^T Ax}{\partial x} = x^T(A + A^T);$$
$$\frac{\partial A^T x}{\partial x} = \frac{\partial x^T A}{\partial x} = A^T;$$
$$\frac{\partial(trace[xAx^T])}{\partial x} = 2xA;$$
$$\frac{\partial(trace[xA^T])}{\partial x} = \frac{\partial(trace[Ax^T])}{\partial x} = A$$

Thus we have the optimal estimate filter:

$$x_a = x_b + K[y_0 - hx_b] \qquad (15)$$
$$K = Bh^T(hBh^T + R)^{-1} \qquad (16)$$
$$P_a = (I - Kh)B \qquad (17)$$

In the estimate (15) – (17), if the background error covariance B is prescribed, the estimate is called optimal interpolation (OI). The OI does not involve state equation

(9) and *B* is unchanged during the entire assimilation process.

7.2.2 Kalman filter

Now, we consider *B* in (16) change with the assimilation cycle. This is more realistic since the model prediction errors should be expected to decrease with the assimilation.

From Eq. (9), we have

$$x_{t+1}^{tr} = Mx_t^{tr} + \eta_t \quad (18)$$

$$x_{t+1} = E(Mx_t^a + \eta_t) = Mx_t^a \quad (19)$$

Eq (18) indicates that even the true value is input at a time step, model can't get a true value for next step due to model bias η_t. Eq (19) shows a standard procedure for the model prediction of next step starting from the analysis of previous step.

Subtracting (18) from (19) produces:

$$x_{t+1}^{tr} = Mx_t^{tr} + \eta_t \quad (18)$$

$$x_{t+1} = E(Mx_t^a + \eta_t) = Mx_t^a \quad (19)$$

$$\varepsilon_{t+1}^b = M\varepsilon_t^b - \eta \quad (19a)$$

$$B_{t+1} = E(\varepsilon_{t+1}^b * (\varepsilon_{t+1}^b)^T) = E(M\varepsilon_t^b - \eta)(M\varepsilon_t^b - \eta)^T) = MB_tM^T + Q \quad (20)$$

The above equation considers the evolution of the background (prediction) error covariance with the time, controlled by the dynamical model operator *M*. The equations (15)-(19) plus the above equation constitutes the framework of Kalman filter for eq (9), namely

---

$$x_a = x_b + K[y_0 - hx_b] \quad (21)$$

$$K = Bh^T(hBh^T + R)^{-1} \quad (22)$$

$$P_a = (I - Kh)B \quad (23)$$

$$B_{t+1} = MB_tM^T + Q \quad (24)$$

$$x_{t+1} = Mx^a \quad (25)$$

Sometimes, B is also denoted by P^f. The detailed procedure of Kalman filter (KF) is like below:

(i) Q and R are specified. For convenience, Q and R are usually assumed to be diagonal matrix. An initial value of B is given, e.g., $B_{t=0} = Q$;

(ii) K is calculated using (21);

(iii) The analysis x^a is obtained using (20);

(iv) prediction (background) error covariance B is calculated;

(v) The model (9) is integrated forward using (24);

(vi) Repeat (ii) to (v) till the end of the assimilation period. So, KF is a recursive algorithm

7.2.3 Extended Kalman filter (EKF)

In deriving eqs. (12) and (20), we assume the state model M and measurement model h are both linear. Further, we also assume the error has Gaussian distribution. Therefore, classic KF only works for linear models and Gaussian distribution. If the dynamical model and measurement model are not linear, we can't directly apply for KF. Instead, linearization must be performed prior to applying for KF. The linearized version of KF is called extened KF (EKF), which solves the below state-space estimate problem:

$$\begin{aligned} x_{t+1} &= F(x_t, \eta_t) \\ y_t &= H(x_t, \varsigma_t) \end{aligned} \tag{26}$$

where F and H are nonlinear models. Assume the error is additive noise, i.e.,

$$\begin{aligned} x_{t+1} &= F(x_t) + \eta_t \\ y_t &= H(x_t) + \varsigma_t \end{aligned} \tag{27}$$

The filter is still assumed to be linear, i.e.,

$$x_a = x_b + K[y_0 - H(x_b)]$$

$$H(x + \delta x) = H(x) + \frac{\partial H}{\partial x}\delta x = H(x) + h\delta x \tag{28}$$

where $h_{i,j} = \dfrac{\partial H_i}{\partial x_j}$

So,

$$y_0 - H(x_b) = y_0 - H(x_{tr} + x_b - x_{tr}) = y_0 - H(x_{tr}) - h(x_b - x_{tr}) = \varepsilon_0 - h\varepsilon_b \tag{29}$$

$$x_a = x_b + K(\varepsilon_0 - h\varepsilon_b) \tag{30}$$

Eq (30) is identical to Eq (10). Similarly, subtracting x_{tr} on both sides of eq (30) leads to the below eq.:

$$\boldsymbol{\varepsilon}_a = \boldsymbol{\varepsilon}_b + K(\boldsymbol{\varepsilon}_o - h\boldsymbol{\varepsilon}_b) \quad (31)$$

which is the same as eq. (12). Following the same derivation as that for eq (12), we can obtain the equations similar to (15) to (17). Therefore, if the measurement model *H* is nonlinear, the KF can be still applied with a linearization of *H*.

Similar to (18) and (19), the state model is as below:

$$x_{t+1}^{tr} = F(x_t^{tr}) + \boldsymbol{\eta}_t \quad (32)$$

$$x_{t+1} = E(F(x_t^a) + \boldsymbol{\eta}_t) = F(x_t^a) \quad (33)$$

Subtracting (32) from (33) produces:

$$\boldsymbol{\varepsilon}_{t+1}^b = F(x_t^a) - F(x_t^{tr}) - \boldsymbol{\eta} = F(x_t^a) - F(x_t^{tr} - x_t^a + x_t^a) - \boldsymbol{\eta} = F(x_t^a) - F(x_t^a - \boldsymbol{\varepsilon}_t^b) - \boldsymbol{\eta} = M\boldsymbol{\varepsilon}_t^b - \boldsymbol{\eta} \quad (34)$$

where $M_{i,j} = \dfrac{\partial F_i}{\partial x_j}$

Comparing (19a) with (34) reveals that (20) still works here. Thus, the EKF can be summarized as below:

$$x_a = x_b + K[y_0 - H(x_b)] \quad (34)$$

$$K = Bh^T(hBh^T + R)^{-1} \quad (35)$$

$$P_a = (I - Kh)B \quad (36)$$

$$B_{t+1} = MB_tM^T + Q \quad (37)$$

$$x_{t+1} = F(x^a) \quad (38)$$

$$M_{i,j} = \frac{\partial F_i}{\partial x_j} \quad (39)$$

$$h_{i,j} = \frac{\partial H_i}{\partial x_j} \quad (40)$$

The procedure to perform EKF is similar to that for KF, as listed above. The disparities and similarities between EKF and KF include i) Kalman gain *K* has the same form for both, especially the linear or linearized measurement model should be used; ii) the update equation of model error covariance has the same form, with linear and linearized state model used; iii) forecast model is different, with KF using linear eq. (25) and EKF using nonlinear model (38); iv) the Kalman filter itself is different, linear measurement model *h* used in KF (eq. (21)) and nonlinear model *H* in EKF (eq. (34)).
It should be noticed that EKF is only an approximate KF for nonlinear state model.

7.2.4 Ensemble Kalman filter (EnKF)

A challenge in EKF is to update background (prediction) error covariance by (37), which requires the linearization of nonlinear model. The linearization of nonlinear model is often difficult technically, and even intractable in some cases, e.g. non-continuous functions existing in models. Another drawback of EKF is to neglect the contributions from higher-order statistical moments in calculating the error covariance.

To avoid the linearization of nonlinear model, the ensemble Kalman filter (EnKF) was introduced by Evensen etc. (Evensen 1992; Houtekamer and Mitchell 1998), in which the prediction (background) error covariances *B* of (37) are estimated approximately using an ensemble of model forecasts. The main concept behind the formulation of the EnKF is that if the dynamical model is expressed as a stochastic differential equation, the prediction error statistics, which are described by the Fokker–Plank equation, can be estimated using ensemble integrations (Evensen 1994, 1997); thus, the error covariance matrix *B* can be calculated by integrating the ensemble of model states. The EnKF can overcome the EKF drawback that neglects the contributions from higher-order statistical moments in calculating the error covariance. The major strengths of the EnKF include the following: (i) there is no need to calculate the tangent linear model or Jacobian of nonlinear models, which is extremely difficult for ocean (or atmosphere) general circulation models (GCMs); (ii) the covariance matrix is propagated in time via fully nonlinear model equations (no linear approximation as in the EKF); and (iii) it is well suited to modern parallel computers (cluster computing) (Keppenne 2000). EnKF has attracted a broad attention and been widely used in atmospheric and oceanic data assimilation.

Simply saying, EnKF used the below formula to replace (37) while other equations are kept the same, i.e,

$$B = \frac{1}{M-1}\sum_{i=1}^{M}(\mathbf{x}_i^b - \overline{\mathbf{x}^b})(\mathbf{x}_i^b - \overline{\mathbf{x}^b})^T \qquad (41)$$

where $\mathbf{x}_t^b$ represents the system state vector at step *t*, and *M* is the ensemble size. The use of (41) avoids processing the *M*, the linearized operator of nonlinear model. In eq. (35), the measurement function h is still linear or linearized, which causes concern. To avoid the linearization of nonlinear measurement function, Houtekamer and Mitchell (2001) and Hamill (2006) wrote Kalman gain (35) by

$$Bh^T = \frac{1}{M}\sum_{i=1}^{M}(\mathbf{x}_i^b - \overline{\mathbf{x}^b})(H(\mathbf{x}_i^b) - \overline{H(\mathbf{x}^b)})^T \qquad (42)$$

$$h\mathbf{P}^b h^T = \frac{1}{M}\sum_{i=1}^{M}(H(\mathbf{x}_i^b) - \overline{H(x^b)})(H(\mathbf{x}_i^b) - \overline{H(\mathbf{x}^b)})^T \qquad (43)$$

where $\overline{H(\mathbf{x}^b)} = \frac{1}{M}\sum_{i=1}^{M} H(\mathbf{x}_i^b)$

(42) and (43) allow direct evaluation of the nonlinear measurement function *h* in calculating Kalman gain. However, (42) and (43) have not been proven mathematically, and only were given intuitionally. Tang and Amabadan (2009) argued that (42) and (43) approximately hold if and only if

$$\overline{H(\mathbf{x}^b)} = H(\overline{\mathbf{x}^b}) \tag{44}$$

$$\mathbf{x}_i^b - \overline{\mathbf{x}^b} = \boldsymbol{\varepsilon}_i, \quad \text{Norm}(\boldsymbol{\varepsilon}_i) \text{ is small for i} = 1, 2..., \text{M} \tag{45}$$

Under the conditions of (44) and (45), Tang and Amabadan (2009) argued (42) and (43) actually linearize the nonlinear measurement functions *H to* h. Therefore, direct application of the nonlinear measurement function in (42) and (43) in fact imposes an implicit linearization process using ensemble members. In next discussions, we will see that (42) and (43) have good statistical foundations and hold without requiring conditions (44) and (45).

Thus, the equations and procedure of EnKF are summarized as below:

$$x_a^i = x_b^i + K[y_0 + \boldsymbol{\varepsilon}_i - H(x_b)] \tag{46}$$

$$B = \frac{1}{M-1}\sum_{i=1}^{M}(\mathbf{x}_i^b - \overline{\mathbf{x}^b})(\mathbf{x}_i^b - \overline{\mathbf{x}^b})^T \tag{47}$$

$$K = Bh^T(hBh^T + R)^{-1} \tag{48}$$

$$Bh^T = \frac{1}{M}\sum_{i=1}^{M}(\mathbf{x}_i^b - \overline{\mathbf{x}^b})(H(\mathbf{x}_i^b) - \overline{H(\mathbf{x}^b)})^T \tag{49}$$

$$h\mathbf{P}^b h^T = \frac{1}{M}\sum_{i=1}^{M}(H(\mathbf{x}_i^b) - \overline{H(x^b)})(H(\mathbf{x}_i^b) - \overline{H(\mathbf{x}^b)})^T \tag{50}$$

$$P_a = (I - Kh)B \tag{51}$$

$$x^i{}_{t+1} = F(x_a^i + \gamma_i) \tag{52}$$

---

i) put perturbations on initial conditions and integrate the model, i.e., $x^i{}_1 = F(x_0 + \gamma_i)$ where i=1,2…M (ensemble size) and x_0 is the initial condition;

ii) using (48), (49) and (50) to calculate Kalman gain K;

iii) calculate analysis using (46) after K is obtained. It should be noted that each ensemble member produces an analysis; the average of all analyses (M) can be obtained;

iv) using (52) to obtain new ensemble members for next round analysis. Here γ_i is perturbation on the *i*th member.

v) repeat ii) – v) until the end of assimilation period.

vi) It should be noted that the observation should be treated as a random variable with the mean equal to 0 and covariance equal to $\boldsymbol{R}=<\varepsilon\varepsilon^T>$. This is why there is ε_i in (46). Simply, ε_i is often drawn from a normal distribution

$$\varepsilon_i \sim N(0, R)$$

From the above procedure, we find that eq (47) is not directly applied here. Instead, we use (49) and (50) to calculate K. This is because i) (48) and (49) avoid the linearization of nonlinear model; ii) avoid the explicit expression of matrix B, which is often very large and can't be written in current computer sources in many realistic problems. This should be noticed that (48) and (49) avoid writing B directly. The measurement function H, projecting model space (dimension) to observation space (dimension), greatly reduce the number of dimension.

7.2.5 Some remarks on EnKF

(1) Initial perturbation

The success of EnKF highly depends on the quality of ensemble members produced by initial perturbations. It is impractical to represent all possible types of errors within the ensemble because of the computational cost, the method of generating initial perturbations must be chosen judiciously. i) The first issue is the amplitude of initial perturbations. Usually the following two factors are considered when selecting the amplitude of initial perturbations: the amplitude of observation error and the amplitude of model errors induced by model parameters and uncertainty in model physics. If a model is perfect, the amplitude of the perturbations should be the same as the amplitude of observation errors. This combined error is used to determine the amplitude of perturbations. ii) When the perturbation amplitude is determined, the practical initial perturbation field generating each ensemble member could be constructed by a normalized pseudorandom field multiplied by the prescribed amplitude. Considering the spatial coherence, the pseudorandom field is red-noise as proposed by Evensen (2003), summarized as below:

i) Calculate the statistical characteristics for the pseudorandom field to meet its variance of 1 and the mean of 0.

$$e^{-1} = \frac{\sum_{l,p} e^{-2(k_l^2+r_p^2)/\sigma^2} \cos(k_l * r_h)}{\sum_{l,p} e^{-2(k_l^2+r_p^2)/\sigma^2}} \qquad (53)$$

$$k_l = \frac{2\pi l}{x_n} = \frac{2\pi l}{N\Delta x};\ r_p = \frac{2\pi p}{y_m} = \frac{2\pi p}{M\Delta y},$$

where N and M are the number of grid points in x-axis (lon.) and y-axis (lat.). For example, if your model domain is 114*42, N=114 and M=42. The *l* and *p* are wave-number, changing from 1 to the maximum value of N/2 and M/2. Δx and Δy are the interval of two adjacent points, often set to 1. r_h is the de-correlation length. The purpose of (53) is to derive the σ^2 for the other feature:

$$c^2 = \frac{1}{\Delta k \sum_{l,p} e^{-2(k_l^2 + r_p^2)/\sigma^2}} \qquad (54)$$

ii) After c and σ^2 are obtained, we can construct a 2-dimension pseudorandom field:

$$W(x_n, y_m) = \sum_{l,p} \frac{c}{\Delta k} e^{-(k_l^2 + r_p^2)/\sigma^2} e^{2\pi i \varphi(l,p)} e^{i(k_l x_n + r_p y_m)} \Delta k \quad (55)$$

While x_n, y_m cover the whole domain, (55) produce a $n * m$ 2-dimensional random filed with spatial coherence structure and the variance of 1 and mean of zero. If the realistic uncertainty (error) has an amplitude β, the perturbation should be βW. Similarly, (55) is often used for the error perturbation γ_i used in (52).

Sometimes, we need to consider the vertical coherence of pseudorandom fields between adjacent levels in oceanic models. A simple method for this purpose is to construct the pseudorandom field at the *kth* level $\boldsymbol{\varepsilon}_k$ (k=1,2,3,…,L) by following equation:

$$\varepsilon_k = \alpha \varepsilon_{k-1} + \sqrt{1-\alpha^2} W_k \qquad (56)$$

where $\mathbf{W}_k$ (k=1,…,L) is the pseudorandom field at the *kth* level without considering vertical coherence, constructed using the above method. Initially, for the surface perturbation (k=1), the vertical coherence is not considered, i.e., $\alpha = 0$, since $\boldsymbol{\varepsilon}_{k-1}$ does not exist. Eq (4) indicates that a practical pseudorandom at the *kth* level ($\boldsymbol{\varepsilon}_k$) is composed of $\mathbf{W}_k$ and $\boldsymbol{\varepsilon}_{k-1}$. As such the $\boldsymbol{\varepsilon}_k$ is correlated with $\boldsymbol{\varepsilon}_{k-1}$, i.e., the practical pseudorandom fields of two adjacent levels ($\boldsymbol{\varepsilon}_{k-1}$ and $\boldsymbol{\varepsilon}_k$) are coherent with each other.

Their correlation or coherent structure is determined by the coefficient $\alpha \in [0,1]$. Eq (56) generates a sequence that is white in the vertical direction if $\alpha = 0$ (i.e., $\boldsymbol{\varepsilon}_k = \mathbf{W}_k$), but a sequence that is perfect correlated in vertical if $\alpha = 1$ (i.e., $\boldsymbol{\varepsilon}_k = \boldsymbol{\varepsilon}_{k-1}$). Eq (56) is also often used to construct random fields that is temporally coherent, for example, a continuous random noise that has coherence in time , as used for γ_i in (52) (Evensen 2003). The random noise γ_i in (52) can also be replaced by the random noise imposed in model forcing. For example, the random noise is continuously added to wind forcing for oceanic models. Even for some atmospheric models with transition processes, there are inherent random noises making γ_i not necessary. One important criteria for γ_i and the amplitude β is to examine ensemble spread by some sensitivity experiments.

(2) The computational cost of Kalman Gain.

The Kalman gain, as expressed by (48), has dimension of *m*m*, where *m* is the number of model variables of observation. In many realistic problems, *m* is a very large number (m >> M), making the inversion very expensive.

A simple procedure is to rewrite the Kalman gain K, as below:

$$K = \mathbf{x}'^{*}\mathbf{x}'^{T} h^{T} (h\mathbf{x}'^{*}\mathbf{x}'^{T} h^{T} + \boldsymbol{\varepsilon} * \boldsymbol{\varepsilon}^{T})^{-1} \qquad (57)$$

Where $\mathbf{X}'$ indicates the model ensemble predictions removed the ensemble mean ($\mathbf{X}'$ = $[x_i^b - \bar{x}^b]$, i=1,2…M, M is the ensemble size); $R = \frac{1}{M}\boldsymbol{\varepsilon} * \boldsymbol{\varepsilon}^T$ was invoked here. If we assume the ensemble prediction error ($\mathbf{x}^b - \bar{\mathbf{x}}^{\mathbf{tr}} \approx \mathbf{x}^{\mathbf{b}} - \bar{\mathbf{x}}^{\mathbf{b}} = \mathbf{x}'$) is not correlated to observation error, i.e., $h\mathbf{x}'^{*}\boldsymbol{\varepsilon}^T = 0$, the following is valid (Evensen 2003),

$$(h\mathbf{x}'^{*}\mathbf{x}'^{T} h^{T} + \boldsymbol{\varepsilon} * \boldsymbol{\varepsilon}^{T}) = (h\mathbf{x}'+\boldsymbol{\varepsilon})(h\mathbf{x}'+\boldsymbol{\varepsilon})^{\mathbf{T}} \qquad (58)$$

$(h\mathbf{x}'+\boldsymbol{\varepsilon})$ has dimension m*M. Usually ensemble size M is much less than *m*. Using SVD technique, we have

$$(h\mathbf{x}'+\boldsymbol{\varepsilon}) = U\sum V^T$$

The Eq. (58) then becomes

$$(h\mathbf{x}'^{*}\mathbf{x}'^{T} h^{T} + \boldsymbol{\varepsilon} * \boldsymbol{\varepsilon}^{T}) = (h\mathbf{x}'+\boldsymbol{\varepsilon})(h\mathbf{x}'+\boldsymbol{\varepsilon})^{\mathbf{T}} = \mathrm{U}\sum \mathrm{V}^{\mathrm{T}} V \sum {}^{T} U^{T} = U\sum\sum {}^{T} U^{T} = \mathrm{U}\Lambda\mathrm{U}^{\mathrm{T}}$$

So,

$$(h\mathbf{x}'*\mathbf{x}'^{T}\,h^{T}+\varepsilon*\varepsilon^{T})^{-1}=U\Lambda^{-1}U^{T} \qquad (59)$$

where U and Λ are the eigenvector and the square of eigenvalues of $(h\mathbf{x}'+\varepsilon)$. There are *M* non-zero eigenvalues for $(h\mathbf{x}'+\varepsilon)$, therefore the dimension is not large, allowing us to efficiently compute the inversion for a global analysis in most practical situation.

(3) Local analysis: Localization of EnKF

To avoid the problems associate with a large *m*, a so-called local EnKF (LEKF) is proposed to use. In LEKF, only measurements located within a certain distance from a grid point will impact the analysis in this grid point. This allows for an algorithm where the analysis is computed grid point by grid point. Mitchell et al. (2002) found that severe localization could lead to imbalance, but with large enough radius of influence (de-correlation length) for the measurements, this was not a problem.

(4) Inflation approach

The forecast error covariance is defined by (41):

$$B=\frac{1}{M-1}\sum_{i=1}^{M}(\mathbf{x}_i^b-\overline{\mathbf{x}^b})(\mathbf{x}_i^b-\overline{\mathbf{x}^b})^T=\frac{1}{M-1}\mathbf{X}'*\mathbf{X}'^{\mathbf{T}} \qquad (60)$$

Eq. (60) is an approximation to B using forecast ensemble. Due to limited computational source, the ensemble size M is often restricted to a small value for many realistic issues. A small ensemble size may cause a very small ensemble spread, causing the approximation of B by (60) is seriously underestimated. To solve this problem, the B is multiplied by a inflator factor, λ (slightly greater than 1). λ is an empirically determined, such as some sensitivity experiments, with the typical value of 1.01. The λB is used to replace B in EnKF formula (46) – (52). This approach is equivalent to the below approach

$$\mathbf{x}_i^b=\lambda(\mathbf{x}_i^b-\overline{\mathbf{x}^b})+\overline{\mathbf{x}^{\mathbf{b}}}$$

(5) Ensemble square root filter (EnSRF)

In EnKF introduced in section 7.2.4, the observation assimilated into dynamical model should be treated to be stochastic variable, as expressed by $y_0+\varepsilon_i$ in (46). It is a must if the classic EnKF algorithm, as expressed by (46) – (52), is used. It has been proven that if the EnKF assimilates deterministic observations (i.e., observation y_0 not changed at each ensemble member), the analysis error covariance will be systematically underestimated, typically leading to filter divergence, as indicated by below (Houtekamer and Mitchell 1998; Burgers et al. 1998).

$$P^{*}_{a} = (I - Kh)B(1 - Kh)^{T} \quad (61)$$

Eq (61) gives the analysis error covariance if the observed is not perturbed. Comparing (61) with (51), a theoretically unbiased estimate, P^{*}_{a} is always less than P_a.

However, the perturbed observation approach (i.e., $y_0 + \varepsilon_i$) introduces an additional source of sampling error that reduces analysis error covariance accuracy and increases the probability of understanding analysis error covariance (Whitaker and Hamill 2002; Tippett et al. 2003). Thus an approach that only uses a single observation realization but avoids systematical underestimation of analysis error covariance was pursued. There are several approaches to implement this goal, as summarized by Tippett (2003). Below, we will introduce an approach developed by Whitaker and Hamill (2002), called EnSRF.

Denote the deviation of analysis from the analysis mean by $x'_a = x_a - \bar{x}_a$. It is easy to write

$$x'_a = x'_b + \tilde{K}[y'_0 - hx'_b] \quad (62)$$

Where $y'_0 = y_0 - \bar{y}_0$. If a single observation realization is assimilated in all ensemble members, $y'_0 = 0$ and

$$x'_a = x'_b - \tilde{K}hx'_b = (I - \tilde{K}h)x'_b,$$

$$P^{*}_{a} = (I - \tilde{K}h)B(1 - \tilde{K}h)^{T}$$

We seek a definition for $\tilde{K}$ that will result in an ensemble whose analysis error covariance satisfies (51), i.e.,

$$(I - \tilde{K}h)B(1 - \tilde{K}h)^{T} = (I - \tilde{K}h)B \quad (63)$$

The solution of (63) is

$$\tilde{K} = (1 + \sqrt{\frac{\mathrm{R}}{\mathrm{hBh}^{\mathrm{T}} + \mathbf{R}}})^{-1} K \quad (64)$$

Therefore, EnSRF is summarized as below:

$$\bar{x}_a = \bar{x}_b + K[y_0 - h\bar{x}_b] \quad (65)$$

$$x_a^{'} = x_b^{'} - \tilde{K}hx^{'}_b \quad (66)$$

$$x_a = \bar{x}_a + x_a^{'} \quad (67)$$

$$B = \frac{1}{M-1}\sum_{i=1}^{M}(\mathbf{x}_i^b - \overline{\mathbf{x}^b})(\mathbf{x}_i^b - \overline{\mathbf{x}^b})^T \quad (68)$$

$$K = Bh^T(hBh^T + R)^{-1} \quad (69)$$

$$\tilde{K} = (1+\sqrt{\frac{\mathrm{R}}{\mathrm{hBh^T} + \mathbf{R}}})^{-1} K \quad (70)$$

$$Bh^T = \frac{1}{M}\sum_{i=1}^{M}(\mathbf{x}_i^b - \overline{\mathbf{x}^b})(H(\mathbf{x}_i^b) - \overline{H(\mathbf{x}^b)})^T \quad (71)$$

$$h\mathbf{P}^b h^T = \frac{1}{M}\sum_{i=1}^{M}(H(\mathbf{x}_i^b) - \overline{H(x^b)})(H(\mathbf{x}_i^b) - \overline{H(\mathbf{x}^b)})^T \quad (72)$$

$$P_a = (I - \tilde{K}h)B \quad (73)$$

$$x^i{}_{t+1} = F(x_a^i + \gamma_i) \quad (74)$$

It should be noted that there are two Kalman Gains used in EnSRF, the original *K* for ensemble analysis mean, and a new $\tilde{K}$. Eq (65) indicates that one single observation realization of classic EnKF has the same ensemble analysis mean as stochastic observations or EnSRF.

7.2.6 General form of ensemble-based filters for Gaussian models

In proceeding sections, we introduced Kalman based filters. Originally Kalman filter applies linear model and linear measurement function. Further, EKF and EnKF were developed to address nonlinear models. However the measurement functions are still assumed to be linear. Eqs (49) and (50) can directly evaluate nonlinear measurement functions but they were proposed intuitionally and not proven yet. In this section, we will present a general form for nonlinear measurement function, and further prove Eqs (49) and (50) mathematically using the general form.

For generality, we assume the below model:

$$\begin{aligned} x_{t+1} &= M(x_t, \eta_t) \\ y_t &= h(x_t, \varsigma_t) \end{aligned} \quad (75)$$

where *M* and *h* are nonlinear operator of model and measurement. The *x* is model state and y is the observation. The η_t and ς_t are the model errors and observed errors, respectively, which have variance: var(η_t)=<η_t, η_t^T > = Q, var(ς_t)=<ς_t, ς_t^T > =R.

Assuming the estimate of model state x_a at a time step is a linear combination of model forecast x_b and observation y_0, i.e., the filter itself is linear, so

$$x_a = x_b + K[y_0 - h(x_b)] \tag{76}$$

Denote $x'_a = x_t - x_a;\ x'_b = x_t - x_b;\ y' = y_0 - h(x_b)$

We have

$$x'_a = x'_b - Ky' \tag{77}$$

$$P_a = E[x'_a x'^T_a] = E(x'_b - Ky')(x'_b - Ky')^T$$
$$= E(x'_b x'^T_b - x'_b y'^T K^T - Ky' x'^T_b + Ky' y'^T K^T)$$

$$P_a = P_b - P_{x'y'}K^T - KP_{y'x'} + KP_{y'y'^T}K^T$$

The optimal estimate asks the trace of P_a minimum, namely,

$$\frac{\partial[trace(P_a)]}{\partial K} = 0$$

$$-P_{x'y'} - P_{y'x'} + 2KP_{y'y'^T} = 0$$

$$K = P_{x'y'}P_{y'y'^T}^{-1}$$
$$P_a = P_b - P_{x'y'}$$

where we invoked the below properties:

$$\frac{\partial(trace[xAx^T])}{\partial x} = 2xA;$$
$$\frac{\partial(trace[xA^T])}{\partial x} = \frac{\partial(trace[Ax^T])}{\partial x} = A$$

Thus we have the optimal estimate filter:

$$x_a = x_b + K[y_0 - h(x_b)] \tag{78}$$

$$K = P_{x'y'}P_{y'y'^T}^{-1} \tag{79}$$

$$P_a = P_b - KP_{x'y'} \tag{80}$$

$$\begin{aligned} x^b{}_{t+1} &= E[M(x_t, \eta_t)] = E[M(x_a)] \\ y_t &= E[h(x_t, \varsigma_t)] = E[h(x_a)] \end{aligned} \tag{81}$$

Eqs. (78)-(81) give a general algorithm for Gaussian nonlinear model and nonlinear measurement function. Eq. (81) assumes the noise is additive. Eq. (79) can be interpreted as the cross-covariance $P_{x'y'}$ between the state and observation errors, and the remaining expression can be interpreted as the error covariance $P_{y'y'^T}$ of the difference between model observation and observation itself. Here, *y*' is defined as the error between the noisy observation y_0 and its prediction $h(x_b)$.

In the standard KF the state error covariance is calculated during the time update process and is updated during the measurement update process. Updating the error covariance matrix is important because it represents the change in forecast error covariance when a measurement is performed. The EnKF implementation does not require the covariance update equation because it can directly calculate the updated error covariance matrix from a set of ensembles. Evensen (2003) has derived the analysis covariance equation, which is consistent with the standard KF error covariance update Eq. (37). But the true representation of the updated error covariance requires a large ensemble size, which is often computationally infeasible. The above algorithm makes use of the reformulated error covariance to update Eq. (80) and chooses the ensembles deterministically in such a way that they can capture the statistical moments of the nonlinear model accurately; in other words, the forecast error covariance Eq. (78) is computed using deterministically chosen samples, called sigma points. In a broad sense, the above algorithm implicitly uses the prior covariance update equation (or the analysis error covariance matrix) to calculate the forecast error covariance. Thus, the above algorithm is fully consistent with the time update and measurement update formulation of the Kalman filter algorithm. In the next subsection we will discuss a new type of Kalman filter, called Sigma-Point Kalman filter, based on the above algorithm.

If the model is linear, obviously

$$\begin{aligned} x^b{}_{t+1} &= Ax_t^a + \eta_t \\ B_{t+1} &= AP_t^a A^T + Q \end{aligned}$$

If the measurement function is linear, i.e.,

$$y' = y_0 - h(x_b, \varsigma) = y_0 - hx_b + \varsigma = hx_t - hx_b + \varsigma = hx_b^{'} + \varsigma$$

$$P_{x'y'} = P_{x'x'^T} h^T$$

$$P_{y'y'^T} = hP_{x'x'^T} h^T + R$$

So, Kalman filter

$$K = P_{x'x'^T} h^T (hP_{x'x'^T} h^T + R)^{-1} \qquad (82)$$

Eq. (82) is identical to Eq. (48). Therefore, Eq. (48), or KF, EKF and EnKF, is a special case of Eq. (82) under the assumption of linear measurement function.

7.2.7 Sigma-Point Kalman filters

EnKF was developed in order to overcome the linearization of nonlinear models. As introduced earlier, the idea behind EnKF is to 'integrate' Fokker-Plank equation using ensemble technique to estimate the forecast error covariance. Theoretically if the ensemble size is infinite, the estimate approaches the true value. However in reality, we can only use finite ensemble size, even very small size for many problems, leading to transaction errors. Thus, some concerns exist such as how to wisely generate finite samples for the optimal estimate of prediction error covariance?; how much the least ensemble size is for an efficient estimate of error covariance?; and how much the true error covariance can be taken into account in the EnKF, given an ensemble size? In this section, we will introduce a new ensemble technique for EnKF, which is called Sigma-point Kalman filter (SPKF).

The so-called sigma-point approach is based on deterministic sampling of state distribution to calculate the approximate covariance matrices for the standard Kalman filter equations. The family of SPKF algorithms includes the unscented Kalman filter (UKF; Julier et al. 1995; Wan and van der Merwe 2000), the central difference Kalman filter (CDKF; Nørgaard et al. 2000b; Ito and Xiong 2000), and their square root versions (Haykin 2001; van der Merwe and Wan 2001a,b). Another interpretation of the sigma-point approach is that it implicitly performs a statistical linearization (Gelb 1974; Lefebvre et al. 2002) of the nonlinear model through a weighted statistical linear regression (WSLR) to calculate the covariance matrices (van der Merwe and Wan 2001a,b; van der Merwe et al. 2004). In SPKF, the model linearization is done through a linear regression between n number of points (called sigma points) drawn from a prior distribution of a random variable rather than through a truncated Taylor series expansion at a single point (van der Merwe et al. 2004). It has been found that this linearization is much more accurate than a truncated Taylor series linearization (Schei 1997; Lefebvre et al. 2002; van der Merwe and Wan 2001a).

Eqs. (78) – (81) construct a core of SPKF. A central issue here is how to generate the optimal ensemble members for applying these equations. There are two basic approaches aforementioned, UKF and CDKF. For an L-dimensional dynamical system represented by a set of discretized state space equations of (75), it has been proven that 2*L+1 ensemble members, constructed by UKF or CDKF, can precisely estimate the mean and covariance. We ignore the theoretical proof and only outline the UKF scheme as below.

Denote t2L+1 sigma -points at time *k* for producing ensemble members by $\chi_k = [\chi_{k,0}, \chi^+_{k,i}, \chi^-_{k,i}]$, i=1,2,…,L ; the model state by x_k as in (75). For simplicity, we ignore the time *k* in below expressions:

$$\chi_0^k = \bar{X}_k^a; \tag{83}$$

$$\chi^+_{i,k} = \bar{X}_k^a + [c\sqrt{P^a_{X,k}}]_i \tag{84}$$

$$\chi^-_{i,k} = \bar{X}_k^a - [c\sqrt{P^a_{X,k}}]_i \tag{85}$$

where i=1,2,…L, $L = N_x + N_\eta + N_\varsigma$: the dimension of model states, model noise and measurement noise. P_x is the covariance of model state x (analysis) at the previous step. The augmented state vector $X = [x, \eta, \varsigma]$, a L-dimensional vector. $[\sqrt{P^a_{X,k}}]_i$ is the ith row (column) of the weighted matrix square root of the covariance matrix (L dimension). c is a scale parameter which will be specified later. The key point here is to produce (2L+1) ensembles by integrating model (80) with 2L+1 initial conditions of (83)-(85); by the ensembles the filter (78)-(80) will be performed.

The procedure is summarized as below:

1) Initially, perturb a small amount, denoted by $\tilde{x}_0$ on initial condition x_0, using Evensen (2003) method; also randomly generate perturbation for q and r, drawn from normal distributions of N(0,Q) and N(0,R). Thus, we can construct the augmented state vector and corresponding covariance (k=0);

$$\bar{X}_0^a = [x_0, 0, 0]$$

$$P_x^0 = \tilde{x}_0 \tilde{x}_0^T$$

$$P_{X,0} = \begin{bmatrix} P_x^0 & 0 & 0 \\ 0 & Q & 0 \\ 0 & 0 & R \end{bmatrix}$$

From the above formula, we can calculate sigma-points using (83) and (85). Note

that each set of sigma-points, denoted by χ_k, has dimension L, e. g, the *ith* sigma-point can be expressed by $(\chi_k)_i = [x_{k,i}^{sig}, \eta_{k,i}^{sig}, \varsigma_{k,i}^{sig}]$

2) Using the 2L+1 sigma-points to integrates state-space model. For the *i*th sigma-point, we have $(x_{k+1}^f)_i = M(x_{k,i}^{sig}, \eta_{k,i}^{sig})$. When *i* varies from 1 to 2L+1, we produce 2L+1 ensemble members, from which analysis mean and covariance will be obtained, which are in turn used to produce sigma-points for next step (k+1), to form a recursive algorithm.

Supposed we have 2L+1 ensembles, the analysis mean and the covariance are calculated as follows

$$\bar{x}_{k+1}^f = \sum_i^{2L} w_i^{(m)} (x_{k+1}^f)_i \tag{86}$$

$$(P_{x,x}^f)_{k+1} = \sum_i^{2L} w_i^{(c)} [(x_{k+1}^f)_i - \bar{x}_{k+1}^f][(x_{k+1}^f)_i - \bar{x}_{k+1}^f]^T \tag{87}$$

$$(y_{k+1}^f)_i = h(x_{k,i}^{sig}, \varsigma_{k,i}^{sig}) \tag{88}$$

$$\bar{y}_{k+1}^f = \sum_i^{2L} w_i^{(m)} (y_{k+1}^f)_i \tag{89}$$

$$(P_{y,y})_{k+1} = \sum_i^{2L} w_i^{(c)} [(y_{k+1}^f)_i - \bar{y}_{k+1}^f][(y_{k+1}^f)_i - \bar{y}_{k+1}^f]^T \tag{90}$$

$$(P_{x,y})_{k+1} = \sum_i^{2L} w_i^{(c)} [(x_{k+1}^f)_i - \bar{x}_{k+1}^f][(y_{k+1}^f)_i - \bar{y}_{k+1}^f]^T \tag{91}$$

$$K_{k+1} = P_{x,y} P_{y,y}^{-1} \tag{92}$$

$$\bar{x}_a^{k+1} = \bar{x}_{k+1}^f + K_{k+1}[y_{k+1} - \bar{y}_{k+1}^f] \tag{93}$$

$$P_a^{k+1} = (P_{x,x}^f)_{k+1} - K_{k+1} P_{y,y} K_{k+1}^T \tag{94}$$

$$c = \sqrt{L+\lambda}$$

$$w_0^{(m)} = \frac{\lambda}{L+\lambda}$$

$$w_0^{(c)} = \frac{\lambda}{(L+\lambda)} + 1 - \alpha^2 + \beta$$

Where (95)

$$w_i^{(m)} = w_i^{(c)} = \frac{1}{2(L+\lambda)} \quad , \qquad \mathrm{i} = 1,2,....,2\mathrm{L}$$

$$\lambda = \alpha^2 (L+\kappa) - L;$$

α and κ are tuning paremeter. $0 \le \alpha \le 1$ and $\kappa \ge 0$. Often κ is chosen 0 as a default value; $\beta = 2$.

.

3) From P_a^{k+1}, as well choosing random perturbation for model noise η and observation noise ς, drawn from Gaussian distribution of N(0,Q) and N(0,R), we calculate sigma-points using (83)-(85), and repeat 2), …and so on until the assimilation is completed for the entire period.

Remakes of SPKF

1) SPKF was introduced into the earth sciences very recently (e.g., Ambadan and Tang, 2009; Luo and Moroz 2009). Until now, there have not a realistic application of SPKF on atmos./ocean. sciences, although some basic experiments were performed using a highly simplified Lorenz model.

2) In SPKF, sigma-points are 2L+1, here L is the dimension of the augmented state vector $X = [x, \eta, \varsigma]$, i.e., $L = N_x + N_\eta + N_\varsigma$: model state, model noise and observation noise. Usually L is the odder10^{3-4}, so the computational expense is a huge challenge in SPKF for realistic problems. A solution is to use the truncated singular value decomposition (TSVD) to reduce the sigma-points. As seen the eqs (83) –(85), the $P_{X,k}^a$ is a L*L matrix, where the first subscript indicates the ensemble member and the second subscript means a L dimensional vector for model inputs used in $(x_{k+1}^f)_i = M(\chi_{k,i}^{sig})$. Thus the dimension of $P_{X,k}^a$ determines the ensemble size. Supposed that $P_{X,k}^a$ can be expressed as

$$P_{X,k}^a = E_{X,k}^a \Sigma_k (E_{X,k}^a)^T \qquad (96)$$

where $\Lambda_k = diag(\sigma_k^1, \sigma_k^2 ... \sigma_k^L)$ is a diagonal matrix of eigenvalues which are sorted in descending order, i.e., $\sigma_k^1 \ge \sigma_k^2 \ge ... \ge \sigma_k^L$, and

$$E^a_{X,k} = (e^a_{X,k,1}, e^a_{X,k,2}, ..., e^a_{X,k,L})$$

Truncating the first *m* modes, so we can write the sigma-points (83) –(85) as below:

$$\chi^k_0 = \overline{X}^a_k; \tag{97}$$

$$\chi^+_{i,k} = \overline{X}^a_k + c\sqrt{\sigma^i_k}\, e^a_{X,k,i} \tag{98}$$

$$\chi^-_{i,k} = \overline{X}^a_k - c\sqrt{\sigma^i_k}\, e^a_{X,k,i} \tag{99}$$

i=1,2…m. Thus the ensemble size becomes *2*m+1*. Some fast SVD algorithms can be used here such as Lanczos and block Lanczos (Chapter 9 of Golub and Loan). The application of the truncated SVD was also found in Hansen 1987; Ehrendorfer and Tribbia 1997).

3) Further simplification of $P^a_{X,k}$ based on its definition (or Cholesky decomposition), i.e., $P^a_{X,k} = A^a_{X,k} * (A^a_{X,k})^T$. Thus, (83)-(84) can write

$$\chi^k_0 = \overline{X}^a_k; \tag{100}$$

$$\chi^+_{i,k} = \overline{X}^a_k + (cA^a_{X,k})_i \tag{101}$$

$$\chi^-_{i,k} = \overline{X}^a_k - (cA^a_{X,k})_i \tag{102}$$

where $(A^a_{X,k})_i = [\, x^k_a, \eta^k, \varsigma^k\,]_i$, i=1,2,…2L.

$$(x^k_a)_i = (x^k_k)_i + K_k[y_k - y^f_k]$$

(100)-(102) transfer the covariance matrix $P^a_{X,k}$ to data matrix $A^a_{X,k}$ in constructing sigma-points. The largest advantage is to avoid explicit expression of $P^a_{X,k}$, which could be a very large matrix beyond memory of current computers. However (100)-(102) can't reduce the ensemble size (2L+1). A solution is to decompose $A^a_{X,k}$ such as Principal Component Analysis, as used in Ambadan and Tang, 2009). Further discussions on optimal construction of sigma-point should be conducted for a realistic application of SPKF.

4) Again, we look at Sigma-points generations, i.e., eqs (97)-(99) or (100)-(102). As we defined, an augmented matrix is applied here: $X = [x, \eta, \varsigma]$. Without losing

the generality, we take eqs (100) –(101) and rewrite them as below:

$$\begin{bmatrix} x_0^k \\ \eta_0^k \\ \varsigma_0^k \end{bmatrix} = \begin{bmatrix} \bar{x}_0^k \\ 0 \\ 0 \end{bmatrix} \tag{103}$$

$$\begin{bmatrix} x_i^k \\ \eta_i^k \\ \varsigma_i^k \end{bmatrix} = \begin{bmatrix} \bar{x}_0^k \\ 0 \\ 0 \end{bmatrix} + \mathrm{c}\begin{bmatrix} (x_a^k)_i \\ \eta_i^k \\ \varsigma_i^k \end{bmatrix} \tag{104}$$

Similarly, we can write (102) using individual variables. From (103) and (104), we can draw,

i) Noise and model state analysis in constructing sigma-points at k step are independent; It should be noted that x_a^k is from eq. (93) and noise are draw from a Gaussian distribution. If we assume that x_a^k is independent with model noise and observation noise, $P_{X,k}$ is a diagonal block matrix, i.e.,

$$P_{X,k} = \begin{bmatrix} P_x^k & 0 & 0 \\ 0 & \mathrm{Q} & 0 \\ 0 & 0 & \mathrm{R} \end{bmatrix} \tag{105}$$

ii) There are not update eqs. for noise so they are randomly taken from Gaussian distribution, i.e., the index *i* in η_i and ς_i actually does not have meaning. Thus, it should be a reasonable assumption that the η_i and ς_i , Used for constructing sigma-point at time step k+1, is not related to $P_{X,k}$ (time step of k). Thus, (105) always hold unless the noise is designed with the consideration of temporal coherence such as red noise in time.

iii) Based on i) and ii), the actual ensemble size is 2* N_x +1, not 2*L+1. This is because neither model noise nor observation noise can produce ensemble alone. Model error η_i and $(x_k^f)_i$ must jointly together to produce ensemble of N_x

.

Let's see this in details:

Initial time: initial perturbation on model states x_0, plus drawn noise for model errors and measurement errors

$$\overline{X}_0^a = [x_0, 0, 0] \qquad P_{X,0} = \begin{bmatrix} P_x^0 & 0 & 0 \\ 0 & Q & 0 \\ 0 & 0 & R \end{bmatrix}$$

Theoretically there are $2(N_x + N_\eta + N\varsigma)+1$ ensembles, denoted by the *i*th column of $P_{X,0}$ (i=1,…, N_x; N_x+1,...N_η; $N_x + N_\eta + 1$,...., $N_x + N_\eta + N\varsigma$) and formula (83) – (85). However, at the *ith* column, the elements of the row, indicating the model inputs (x, η, ς), have the non-zero values of N_x. Obviously, the sigma-point of zero-values makes $(x_{k+1}^f)_i = M(\chi_{k,i}^{sig})$ not sense, thus, the actual ensemble size is 2* N_x +1.

5) When truncation technique is applied to reduce the ensemble size, the ensemble spread might be shrunk due to relatively small ensemble size. Like EnKF, an inflation approach of SPKF might be helpful. It is interested in developing such a scheme for SPKF.

6) Similar to EnKF, it is interested to develop a square root SPKF which can avoid the perturbation of observation. There is already the square root SPKF but it is not clear such a square root SPKF has the same sense as EnKF. A further configuration is expected from readers.

7) Again, we can localize SPKF, like localized EnKF, to solve memory and computation issues.

8) All of the remarks of SPKF are from my personal thinking and understanding. It is interesting to further test and validate these ideas and properties using simple models.

References:

1. Aksoy, A., F. Zhang, and J. Neilsen-Gammon, 2005a: Ensemble based simultaneous state and parameter estimation with mm5. Geophys. Res. Lett., 33, L12 801, doi:10.1029/2006GL026 186.
2. Aksoy, A., F. Zhang, and J. Neilsen-Gammon, 2006: Ensemble based simultaneous state and parameter estimation in a two-dimensional sea-breeze model. Mon. Wea. Rev., 134, 2951–2970.
3. Aksoy, A., F. Zhang, J. Neilsen-Gammon, and C. Epifanio, 2005b: Ensemble based data assimilation for thermally forced circulations. J. Geophys. Res., 110, D16 105, doi:10.1029/2004JD005 718.

4. Ambadan, J. and Y. Tang, 2009: Sigma-point kalman filter data assimilation methods for strongly nonlinear systems. J. Atmos. Sci., 66, 261–285.
5. Anderson, J. L., 2001: An ensemble adjustment kalman filter for data assimilation. Mon. Wea. Rev., 129, 2884–2903.
6. Anderson, J. L., 2002: A local least square framework for ensemble filtering. Mon. Wea. Rev., 131, 634–642.
7. Anderson, J. B. Moore, Optimal Filtering, Information and System Science Series, Prentice Hall, Englewood Cliffs, NJ, USA., 1979.
8. Annan, J., 2005: Paramater estimation using chaotic time series. Tellus, 57A, 709–714.
9. Annan, J. and J. Hargreaves, 2004: Efficient paramater estimation for a highly chaotic system.Tellus, 56A, 520–526.
10. Annan, J., J. Hargreaves, N. Edwards, and R. Marsh, 2005a: Paramater estimation in an intermediate complexity earth system model using an ensemble kalman filter. Ocean Modell, 8, 135–154. 25
11. Annan, J., D. Lunt, J. Hargreaves, and P. Valdes, 2005b: Paramater estimation in an atmospheric gcm using the ensemble kalman filter. Nonlinear Proc. Geoph., 12, 363–371.
12. Arulampalam, S. Maskell, N. Gordon, T. Clapp, A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking, IEEE Transactions on Signal Processing 50(2) (2002) 174–188.
13. Bengtsson, T., P. Bickel, B. Li, Curse-of-dimensionality revisited: Collapse of the particle filter in very large scale systems, Probability and Statistics: Essays in Honor of David A. Freedman, IMS Collections 2 (2008) 316–334.
14. G. Burgers, P. J. van Leeuwen, G. Evensen, Analysis scheme in the Ensemble Kalman Filter, Monthly Weather Review 126 (1998) 1719–1724.
15. Casella, G., C. P. Robert, Rao-Blackwellisation of sampling schemes,Biometrika 83(1) (1996) 81–94.
16. Cohn, S. E., An Introduction to Estimation Theory, Journal of the Meteorological Society of Japan 75 (1997) 147–178.
17. F. X. L. Dimet, O. Talagrand, Variational Algorithms for Analysis and Assimilation of Meteorological Observations: Theoretical Aspects, Tellus 38A (1986) 97–100.
18. Courtier, P. E. Andersson, W. Heckley, J. Pailleux, D. Vasiljevic, M. Hamrud, A. Hollingsworth, F. Rabier, M. Fisher, The ECMWF implementation of three-dimensional variational assimilation (3D-Var). I: Formulation, Quarterly Journal of the Royal Meteorological Society 124(550) (1998) 1783–1808.
19. Dee, D. P., 1995: On-line estimation of error covariance parameters for atmospheric data assimilation. Mon. Wea. Rev., 123, 11281145.
20. Doucet, N. de Freitas, N. Gordon, Sequential Monte Carlo Methods in Practice, Springer-Verlag, New York, USA., 2001.
21. Doucet, S. J. Godsill, C. Andrieu, On sequential Monte Carlo sampling methods for Bayesian filtering, Statistics and Computing 10(3) (2000) 197–208.
22. Evensen, G., 1992: Using the extended kalman filter with a multilayer quasi-geostrophic ocean model. J. Geophys. Res., 97 (C11), 17 905–17 924.

23. M. Ehrendorfer and J.J Tribbia 1997: Optimal prediction of forecast error covariance through singular vectors, J. Atmos. Sci., 286-312.
24. Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics. J. Geophys. Res., 99 (C5), 10 143–10 162.
25. Evensen, G., 1997: Advanced data assimilation for strongly nonlinear dynamics. Mon. Wea. Rev., 125, 1342–1354.
26. Evensen, G., 2003: The ensemble kalman filter: theoretical formulation and practical implementation. Ocean Dynam., 53, 343–367.
27. Evensen, G., D. Dee, J. Schrter, and J. S. Oter, 1998: Parameter estimation in dynamical models.
28. Gauthier, P., 1992: Chaos and quadric-dimensional data assimilation:a study based on the Lorenz model. Tellus, 44A, 2–17.
29. N. J. Gordon, D. J. Salmond, A. F. M. Smith, Novel approach to nonlinear/non-Gaussian Bayesian state estimation, IEE proceedings. F, Radar and signal processing 140 (1993) 107–113.
30. G. H. Golub and C.F. Van Loan: Matrix Computation, JHU Press, 1996
31. Hacker, J. and C. Snyder, 2005: Ensemble kalman filter assimilation of fixed screen-height observations in a parameterized pbl. Mon. Wea. Rev., 133, 3260–3275.
32. Hamill, T. M., 2006: Ensemble-based atmospheric data assimilation. Predictability of Weather and Climate, T. Palmer and R. Hagedorn, Eds., Cambridge University Press, 124–156.
33. Haykin, E., S., 2001: Kalman Filtering and Neural Networks. Wiley., 128 pp.
34. P. C. Hansen, 1987: The truncated SVD as a method for regularization, BIT 27, 534-553.
35. Houtekamer, P. and H. Mitchell, 1998: Data assimilation using an ensemble kalman filter technique. Mon. Wea. Rev., 126, 796–811.
36. Houtekamer, P. and H. Mitchell, 2001: A sequential ensemble kalman filter for atmospheric data assimilation. Mon. Wea. Rev., 129, 123–137.
37. P. L. Houtekamer, H. L. Mitchell, G. Pellerin, M. Buehner, M. Charron, L. Spacek, B. Hansen, Atmospheric data assimilation with an Ensemble Kalman Filter: Results with real observations, MonthlyWeather Review 133 (2005) 604–620.
38. P. L. Houtekamer, H. L. Mitchell, Ensemble Kalman filtering, Quarterly Journal of the Royal Meteorological Society 131 (2005) 3269–3289.
39. Hunt, B., E. Kostelich, and I. Szunyogh, 2007: Efficient data assimilation for spatiotemporal chaos: A local ensemble transform kalman filter. Physica D, 230, 112–126.
40. Ito, K. and K. Xiong, 2000: Gaussian filters for nonlinear filtering problems. IEEE T. Automat. Contr., 45(5), 910–927.
41. M. Isard, A. Blake, A smoothing filter for condensation, Proceedings of 5th European Conference on Computer Vision 1 (1998) 767–781.

42. Jazwinsky, A., 1970: Stochastic Processes and Filtering Theory. Academic Press, New York. Julier, S., J. Uhlmann, and H. Durrant-Whyte, 1995: A new approach for filtering nonlinear systems.
43. A. H. Jazwinski, Stochastic Processes and Filtering Theory, Academic Press, New York, USA., 1970.
44. Proceedings of the American Control Conference, Seattle, WA, USA, IEEE, 1628–1632.
45. Kalman, R., 1960: A new approach to linear filtering and prediction problems. Transaction of the ASME, Series D. J. Basic Eng., 82, 35–45.
46. Keppenne, C. L., 2000: Data assimilation into a primitive-equation model with a parallel ensemble kalman filter. Mon. Wea. Rev., 128, 1971–1981.
47. Kivman, G. A., 2003: Sequential parameter estimation for stochastic systems. Nonlinear Proc.Geoph., 10, 253–259.
48. E. Klinker, F. Rabier, G. Kelly, J.-F. Mahfouf, The ECMWF operational implementation of four-dimensional variational assimilation – Part III: Experimental results and diagnostics with operational configuration, Quarterly Journal of the Royal Meteorological Society 126 (2000) 1191–1215.
49. G. Kitagawa, Monte Carlo filter and smoother for non-Gaussian nonlinear state space models, Journal of Computational and Graphical Statistics 5(1) (1996) 1–25.
50. Lefebvre, T., H. Bruyninckx, and J. De Schutter, 2002: Comment on - new method for the nonlinear transformation of means and covariances in filters and estimators deterministic non-periodic flow.¨ . IEEE T. Automat. Contr., 47(8), 1406– 1409.
51. Lermusiaux, P. F. J., 1997: Error subspace data assimilation methods for ocean field estimation: Theory, validation and applications. Ph.D. thesis, Harvard University, Cambridge, MA., 402 pp.,[Available from Harvard Archives Microfilm, Harvard Depository HU 90.14492.9].
52. Lermusiaux, P. F. J. and A. R. Robinson, 1999: Data assimilation via error subspace statistical estimation. part i: Theory and schemes. Mon. Wea. Rev., 127, 1385–1407.
53. H. Lilliefors, On the Kolmogorov-Smirnov test for normality with mean and variance unknown, Journal of the American Statistical Association 62 (1967) 399–402.
54. J. S. Liu, E. Chen, Sequential Monte Carlo methods for dynamics systems, Journal of the American Statistical Association 93 (1998) 1032–1044.
55. Lorenz, E., 1963: Deterministic nonperiodic flow. J. Atmos. Sci., 20, 13–141.
56. Lorenz, E., 2005: Designing chaotic models. J. Atmos. Sci., 62, 1574–1587.
57. Lorenz, E. and K. A. Emmanuel, 1998: Optimal sites for supplementary weather observations: Simulations with a small model. J. Atmos. Sci., 55, 399–414.
58. Lorenz, E. N., 2006: Predictability - a problem partly solved. Predictability of Weather and Climate, T. Palmer and R. Hagedorn, Eds., Cambridge University Press, 40–58.
59. E. N. Lorenz, Predictability: A problem partially solved, Proceedings of the Seminar on Predictability, ECMWF 1 (1996) 1–18.

60. A. C. Lorenc, The potential of the Ensemble Kalman Filter for NWP: A comparison with 4D-Var, Quarterly Journal of the Royal Meteorological Society 129 (2003) 3183–3203.
61. A. J. Majda, J. Harlim, Filtering Nonlinear Dynamical Systems with Linear Stochastic Models, Nonlinearity 21 (2008) 1281–1306.
62. J.-F. Mahfouf, F. Rabier, The ECMWF operational implementation of four-dimensional variational assimilation - Part II: Experimental results with improved physics, Quarterly Journal of the Royal Meteorological Society 126 (2000) 1171–1190.
63. Miller, R., M. Ghil, and F. Gauthiez, 1994: Advanced data assimilation in strongly nonlinear dynamical systems. J. Atmos. Sci., 51, 1037–1056.
64. Navon, I., 1998: Practical and theoretical aspects of adjoint parameter estimation and identifiability in meteorology and oceanography. Dyn. Atmos. Oceans, 27, 55–79.
65. Nørgåd Magnus, P., N. K. Poulsen, and O. Ravn, 2000: New developments in state estimation of nonlinear systems. Automatica., 36, 1627–1638.
66. Palmer, T., 1993: Extended-range atmospheric prediction and the lorenz model. Bull. Amer. Meteor Soc., 74, 49–65.
67. E. Parzen, On estimation of a probability density function and mode, Annals of Mathematical Statistics 33 (1962) 1065–1076.
68. F. Rabier, H. Jarvinen, E. Klinker, J.-F. Mahfouf, A. Simmons, The ECMWF operational implementation of four-dimensional variational assimilation Part I: Experimental results with simplified physics, Quarterly Journal of the Royal Meteorological Society 126 (2000) 1143–1170.
69. Recipes in C: The Art of Scientific Computing. Cambridge University Press, 2 Edition. Simon, D., 2006: Optimal State Estimation, Kalman, H_1, and Nonlinear Approaches. 1st ed.,Wiley-Interscience, 318-320 pp.
70. Talagrand, O., 2007: Data Assimilation in Meteorology And Oceanography. Academic Press.
71. C. Snyder, T. Bengtsson, P. Bickel, J. Anderson, Obstacles to High-Dimensional Particle Filtering, Monthly Weather Review 136 (2008) 4629–4640.
72. T. B. Sch¨on, Estimation of Nonlinear Dynamic Systems - Theory and Applications, Link¨oping Studies in Science and Technology, Ph.D. Dissertation, 2006.
73. Talagrand, Assimilation of observations, an introduction, Journal of the Meteorological Society of Japan 75 (1997) 191–209.
74. Tang, Y. and J. Ambadan, 2009: Reply to comment on “sigma-point kalman filter data assimilation methods for strongly nonlinear systems”: in press. J. Atmos. Sci.
75. Tang, Y. and W. Hsieh, 2001: Coupling neural networks to incomplete dynamical systems via variational data assimilation. Mon. Wea. Rev., 129, 818–834.
76. Teman, R., 1991: Approximation of attractors, large eddy simulations and multiscale methods. Proc. Roy. Soc. London., 434A, 23–29.
77. Tippett, M. K., J. Anderson, C. Bishop, T. Hamill, and J. Whitaker, 2003: Ensemble square root filters. Mon. Wea. Rev., 131, 1485–1490.
78. Tong, M. and M. Xue, 2008a: Simultaneous estimation of microphysical parameters and atmospheric state with simulated radar data and ensemble square

root kalman filter. part i: Sensitivity analysis and parameter identifiability. Mon. Wea. Rev., 136, 1630–1648.
79. B. W. Silverman, Density Estimation for Statistics and Data Analysis, Monographs on Statistics and Applied Probability, Chapman and Hall, London, 1986.
80. Tong, M. and M. Xue, 2008b: Simultaneous estimation of microphysical parameters and atmospheric state with simulated radar data and ensemble square root kalman filter. part ii: Parameter estimation experiments. Mon. Wea. Rev., 136, 1630–1648.
81. Van der Merwe, R. and E. A. Wan, 2001a: Efficient derivative-free kalman filters for online learning. Proceedings of ESANN, Bruges.
82. Van der Merwe, R. and E. A. Wan, 2001b: The square-root unscented kalman filter for state and parameter estimation. Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP)., Salt Lake City, Utah, IEEE.
83. Van der Merwe, R., E. A. Wan, and S. I. Julier, 2004: Sigma-point kalman filters for nonlinear estimation and sensor fusion: Applications to integrated navigation. AIAA Guidance, Navigation and Control Conference, Providence, RI, American Institute of Aeronautics and Astronautics, 5120–5122.
84. Wan, E. and R. Van Der Merve, 2000: The unscented kalman filter for nonlinear estimation. Adaptive Systems for Signal Processing, Communications, and Control Symposium ASSPCC 2000,
85. Lake Louise, Alberta, Canada, IEEE Communications Society, 153–158.
86. Zhang, F. and C. Snyder, 2007: Ensemble-based data assimilation. Bull. Amer. Meteor. Soc., 88, 565–568.