An Improved Data Assimilation Localization Method Based on Fuzzy Logic Control

Yu-long BAI*, Irtaza SHAHID, Xiao-yan MA and Yue WANG
College of Physics and Electrical Engineering, Northwest Normal University, Lanzhou 730070, China
*Corresponding author

Keywords: Ensemble transform Kalman filter, Fuzzy Logic control, Covariance fuzzy, Fuzzy analysis.

Abstract. To determine the best estimates of the ocean or atmospheric state, Data Assimilation (DA) is a methodology to combine all available information from numerical models and observation information. With small ensemble numbers in the Ensemble Kalman filter, the more effective utilization the observation data are, the higher the assimilation performance become. In this study, we proposed a new fuzzy control-based data assimilation system, named FA (fuzzy analysis) and CF (covariance fuzzy), which coupled with fuzzy logic control algorithms to improve assimilation performance. A number of numerical experiments are designed using a classical nonlinear model (the Lorenz-96 model) to explore the effects of the new algorithms on the ensemble transform matrices. The experiments show that the FA algorithm can be selected when the system is in weak assimilation, whereas both algorithms can be implemented in medium assimilation situations. If the system is strongly assimilated, the CF algorithm has demonstrated more robust performance.

Introduction

Data assimilation refers to continue integration the new observation information to obtain the optimal estimation of the current state, which can improve the model forecast accuracy and reduce uncertainty and describe as accurately as possible the true state of the atmosphere, sea and land surface [1]. It has a wide range of applications in atmospheric, oceanic, geophysical and nonlinear dynamical systems. Ensemble Kalman filter (EnKF) is an advanced method that uses several ensembles of model states to propagate and update the estimates of the state covariance[2]. Realistic data assimilation problems are usually implemented with small ensembles to reduce the computational cost. However, using small ensembles can cause a lot of problems, such as rank-deficient with forecast/background ensemble and important Monte Carlo sampling errors [3,4], as well as the ubiquitous nonlinearity of the dynamics system. Several auxiliary techniques have been proposed to alleviate these problems, among others include covariance inflation and local analysis. A number of localisation methods for further improvements have been explored [4]. Although these algorithms to improve the assimilation effect, to avoid the false correlation between observed variables and status updates, in practice, due to the different choices, these algorithms need to be improved. Thus, a data assimilation method of coupled fuzzy algorithm is firstly proposed in [5], which proved the effectiveness of the FETKF (Fuzzy Ensemble Transform Kalman Filter) method with the high-dimensional chaotic Lorenz-96 model. Compared to the former algorithms, it can significantly improve the filtering accuracy in the assimilation system.

Based on fuzzy logic control algorithms, two new algorithms are proposed in this paper, namely, CF (covariance fuzzy) and FA (fuzzy analysis). In the simulation experiment with typical Lorenz chaotic systems, we explored the effect of the CF and FA algorithms on the ensemble transform matrix. It is worth noting that the new algorithm can eliminated the spurious correlation and filter divergence between the state variables in the final assimilation process.
Theoretical Background

**Ensemble Transform Kalman Filter.** The Ensemble transform Kalman (ETKF) method was originally put forward by Bishop in [6]. In the formulation, it is expressed that multiply the forecast perturbation by a transform matrix to obtain the analysis perturbation. The ETKF update equations are commonly written as follows:

\[
x_i^a = x_i^f + K_i(y - H x_i^f).
\]

(1)

\[
A_i^a = A_i^f T
\]

(2)

where all superscript \(i\) represented local, \(x_i^a\) is the state analysis at \(ith\) time; \(x_i^f\) is the state forecast at \(ith\) time; \(y\) is the local vector of observations; \(K_i\) is the local Kalman gain at \(ith\) row; \(H\) is the local observation operator; \(A_i^a\) is the \(ith\) row of the analysis perturbed matrix; \(A_i^f\) is the \(ith\) row of the forecast perturbed matrix; \(T\) is the local ensemble transform matrix. The Kalman gain and the Ensemble transform matrix can be expressed as follows:

\[
S = R^{-1/2} H A^f / \sqrt{m-1}
\]

(3)

\[
K_i = A_i^f S^T (I + S S^T)^{-1} R^{-1/2} / \sqrt{m-1}.
\]

(4)

\[
T = (I + S^T S)^{-1/2}.
\]

(5)

Where \(m\) is the ensemble number, \(S\) are the local ensemble anomalies, and the superscript \(i\) indicates localisation, which will be used to update the \(ith\) state variable; \(R\) is diagonal, the local observation error covariance Matrix; \(I\) is the identity matrix; superscript “T” denotes matrix transposition.

**The Principle of Fuzzy Logic Control.** The concept of Fuzzy Logic Control was proposed by Zadeh in [7]. To obtain better assimilation effect, we use a new localisation method of coupling the fuzzy logic control. The data assimilation system coupled with fuzzy logic control is comprised of principal blocks as follow:

Normally, fuzzy logic control includes fuzzy variable, fuzzy sets and fuzzy membership functions. In this paper, a set of fuzzy conditions (if...then...) is used to describe the mapping relationship between the Euclidean distance and the weight of the observation point and the state update grid point to obtain the weight of each observation point. The following is the principle of constructing equivalent observation weights using fuzzy theory.

As shown in Figure 1, the state update location and the observation position are indicated by “blue seven angle star” and “red pentacle”, respectively. The purple rectangle represents the traditional localized assimilation scheme which is based on sequential filtering. It passes through the state variable and all observation data in turn, then, to all analysis grids, where each analysis grid is independent of the others. The dashed circles of different colors represent different observation weights with fuzzy control. Meanwhile, we define \(dist\) as the distance between state update location and observation position:
\[ dist_i = \sqrt{(O_i - V_i)^2 + (O_j - V_j)^2} \quad (i, j = 1, \ldots, 20). \] (6)

Where \( O_i \) and \( O_j \) are the abscissa and ordinate of observation points respectively, \( V_i \) and \( V_j \) are the abscissa and ordinate of state update points respectively. It can be seen from the graph that \( dist_i \) represents the Euclidean distance between the \( ith \) observation point and the \( ith \) state update point, the observation weights can decreases with the distance increases.

In this paper, the fuzzy subset of the input variable "dist" is divided into \{the nearest distance, second closest distance..., second farthest distance, farthest distance\}, which abbreviated as \( \{I_1, I_2, \ldots, I_{20}\} \). After input quantification, the domain of "dist" is \( D = [0, 20] \). The fuzzy subset of the output "coeffs" is \{highest weight, higher weight... the second lowest weight, lowest weight\}, which is abbreviated as \( \{O_1, O_2, \ldots, O_{20}\} \), and its domain is \( C = [0, 1] \). The connectivity between input fuzzy variable is input fuzzy set and output fuzzy variable is output fuzzy set is represented by a fuzzy implication relation,

\[
M(dist, coeffs) = \begin{array}{ccccccc}
coeffs & O_1 & O_2 & \cdots & O_j & \cdots & O_m \\
I_1 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
I_2 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
I_i & \cdots & \cdots & \cdots & M(I_i, O_j) & \cdots & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
I_n & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots 
\end{array}
\]

where \( dist \in \{I_1, I_2, \ldots, I_n\} = D, coeffs \in \{O_1, O_2, \ldots, O_m\} = C \) and \( M(dist, coeffs) \) denotes the strength of fuzzy relation for \( dist = I_i \) and \( coeffs = O_j \), satisfying the implication between input fuzzy variable is input fuzzy set and output fuzzy variable is output fuzzy set.

Different implication functions are used in the direct use to describe the fuzzy if-then connectivity. In this paper, the well-known Mamdani implication relations is used to illustrate the principle of fuzzy reasoning. Thus, the fuzzy relationship \( M \) can be determined based on the fuzzy rules as:

\[
M = (I_1 \times O_1) + (I_2 \times O_1) + (I_3 \times O_1) \cdots + (I_n \times O_n).
\] (7)
where “×” refers to the Cartesian product of the fuzzy vector. It is a matrix multiplication operator (not a fuzzy operator). According to fuzzy logic control theory, a fuzzy set of equivalent observation weights "coeffs” can be obtained when the fuzzy subsets of input variables "dist” and fuzzy relationship matrix $M$ complete the fuzzy inference. Thus, the final control variable (output fuzzy variable) can be written as:

$$coeffs = \text{dist} \circ M$$  \hspace{1cm} (8)

where $\circ$ is a max-min compositional operation, which is similar to matrix multiplication operation with summation replaced by maximum and product replaced by minimum operators. We employ the maximum membership degree principle in the defuzzification process, the maximum membership degree principle can be written as:

$$coeffs_i = \inf_{j \in \{1, \ldots, n\}} \max\left(\text{coeffs}_j \circ (\text{dist}_i \circ M), \text{coeffs}_j \circ (\text{dist}_j \circ M)\right).$$  \hspace{1cm} (9)

**An Improved Localisation Method Based on Fuzzy Logic Control.** Compared with the original LA algorithm, Fuzzy Analysis (FA) gives an approximation of the background error covariance that has completed fuzzy control for each updated state vector element. The fuzzy updating process of the $l$-th state variable is the same with the above LA algorithm process the covariance with fuzzy control algorithms as follows:

$$C \cdot P = \begin{pmatrix}
coeffs(x_i - x_i) p_{11} & \cdots & \coeffs(x_i - x_i) p_{1m} \\
\vdots & \ddots & \vdots \\
\coeffs(x_m - x_i) p_{m1} & \cdots & \coeffs(x_m - x_m) p_{mm}
\end{pmatrix}$$  \hspace{1cm} (10)

The modified matrix $C \cdot P$ will involve in assimilation operations as the background error covariance matrix; Using fuzzy control methods, these improved approaches seek to eliminate the distanced spurious correlations between the state variables in the background error covariance matrix and to improve the quality of the background error covariance matrix.

**Numerical Tests**

**The Lorenz-96 Model.** In numerical tests section, we describe experiments with the 40-variable Lorenz-96 model. The coupled set of ordinary differential equations as follows:

$$\frac{dx_i}{dt} = (x_{i+1} - x_{i-2})x_{i-1} - x_i + F$$  \hspace{1cm} (11)

This equation applied for $t > 0$, and $i = 1:m$, with $m = 40$ and periodic boundary conditions. The model parameters are chosen to be identical to those of [4]. $F$ is a forcing term that can be tuned to produce chaotic behavior for a given model dimension. In this study, we set $F = 8$.

To verify the validity of fuzzy logic algorithm, the experiments were carried out based on Lorenz-96 model. As seen in Fig. 2, compared with the traditional LETKF, the FLETKF with fuzzy logic control can effectively reduce Root Mean Squared Errors (RMSE) and improve assimilation effect [4].
Influence of CF and FA on the Ensemble Transform Matrices. In this section, the proposed new algorithms CF and FA are demonstrated the effectiveness on the ensemble transform matrix. In terms of the EnSRF assimilation process for the three cases with observation errors $R$ of 10, 1, and 0.001, respectively, the ensemble transform matrices from these two algorithms are extracted for analysis. The results corresponding to the three cases are shown in Figs. 3a, 3b and 3c respectively. The ensemble transform matrices of CF is abbreviated as TCF; The ensemble transform matrices of FA is abbreviated as TFA; and the ensemble transform matrix of the 40-th state variables using the three types of algorithms (i.e., no localisation (none), CF and FA) is abbreviated as $T_{ii}$.

Figure 3. The contrast of the CF and FA algorithms on ensemble transform matrix ($T$). (a) the observation errors $R$ selected sets 10, the intensity of assimilation is called “weak assimilation”. (b) the observation errors $R$ selected sets 1, the intensity of assimilation is called “medium assimilation”. (c) the observation errors $R$ selected sets 0.001, the intensity of assimilation is called “strong assimilation”.

When $R = 10$, the intensity of assimilation is called “weak assimilation”. As seen in Fig. 3(a): 1) The spurious correlation between the state variables is maintained in the ensemble transform matrix with no-fuzzy control. However, the CF and FA algorithms eliminate the distanced spurious correlation in the ensemble transform matrix and is both symmetric and diagonal; 2) Compared
with the CF algorithm, the FA algorithm updates the \( T \) only for the center point of the defined domain, and other \( T \) of non-central elements are not updated. Thus, the FA algorithm performs an asynchronous updating process on \( T \), whereas the CF algorithm performs a synchronous updating process on \( T \); 3) According to Fig. 3(a), the ensemble transform matrices \( T_{ii} \) following updates using the CF and FA algorithms on the same state variable are approximately equal, and the difference between the two algorithms is negligible.

When \( R = 1 \), the intensity of assimilation is called “medium assimilation”. As seen in Fig. 3(b): 1) The transform matrix \( T_{CF} \) from the CF algorithm retains its symmetry and the FA algorithm has little influence; 2) Under medium assimilation, differences gradually emerge in the ensemble transform matrix following updates by the CF and FA algorithms on the same state variable.

When \( R = 0.001 \), the intensity of assimilation is called “strong assimilation”. As seen in Fig. 3(c): 1) The ensemble transform matrix \( T_{CF} \) of the CF algorithm retains its symmetry. Compared with the case observation error: \( R = 1 \), the transform matrix from the FA algorithm is sparse; 2) With an increase of assimilation intensity, which will cause divergence of the two algorithms (i.e., CF and FA), differences in the ensemble transform matrix appear more obvious; 3) According to Fig. 3(c), the difference of the ensemble transform matrix \( T_{ii} \) gradually increases following updates by the two algorithms on the same state variable.

Summary
This paper introduces two common processing methods for the covariance matrix. Specifically, the fuzzy control algorithm is coupled to form the new algorithms CF and FA. In the simulation experiments, the paper explores the effects of the CF and FA algorithms on the ensemble transform matrix. Overall, the FA algorithm can be selected when the system is in weak assimilation, whereas both algorithms can be implemented in medium assimilation situations. If the system is strongly assimilated, the CF algorithm has demonstrated more robust performance.

Acknowledgement
This work is supported by the NSFC (National Natural Science Foundation of China) project (grant number: 41861047 and 41461078) and Northwest Normal University young teachers' scientific research capability upgrading program (NWNU-LKQN-17-6).

References