IOP Conference Series: Earth and Environmental Science

PAPER • OPEN ACCESS

A Comparative Study of Localization Methods in EnkF Data Assimilation

To cite this article: Yulong Bai *et al* 2019 *IOP Conf. Ser.: Earth Environ. Sci.* **310** 022035

A Comparative Study of Localization Methods in EnkF Data Assimilation

Yulong Bai[1,*]**, Xiaoyan Ma**[1]**, and Lihong Tang**[1]

[1] College of Physics and Electrical Engineering, Northwest Normal University, Lanzhou 730070, China

[*] Corresponding author: yulongbai@gmail.com

Abstract. This study investigates the relation between some common localization methods in ensemble Kalman filter (EnKF) systems including covariance localization (CL) and local analysis (LA), which are popular used in large-scale Geo-science applications. Two fuzzy-based localization methods, named covariance fuzzy (CF) and fuzzy analysis (FA), are formulated in terms of tapering of ensemble covariance in a fuzzy logic way. To explore the effects of all algorithms on the error covariance matrix, numerical experiments are designed using a classical nonlinear model (i.e., the Lorenz-96 model) by determining the Power Spectrum Density (PSD) of the corresponding systems. The experiments show that the new algorithms can eliminate spurious correlation of the background error covariance matrix. The results of PSD demonstrate that the new fuzzy based methods have more robust performance than the CL or LA algorithm.

1. Introduction

In a numerical weather prediction system, Data Assimilation (DA) provides an improved initial state for the next forecast and obtains an optimal analysis state based on the statistical treatments of available observations and current forecasts (Martin et al. 2015). Ensemble Kalman filter (EnKF) was developed by Evensen (1994) for a quasigeostrophic model. Due to spurious correlations caused by a small ensemble, a technique named “localization” was developed to taper the long-range correlations in the EnKFs’ ensemble covariance matrices (Hamill et al., 2001; Houtekamer and Mitchell, 1998). To calculate the distances between the physical locations of model variables and/or observations, most localization schemes were proposed (Miyoshi, 2007; Sakov, 2011; Anderson, 2012). The assimilation methods coupled with the fuzzy control method are proposed to verify the effectiveness of algorithms (Lu et al. 2017). However, there is no detailed comparison between the new algorithms and the traditional methods. Further on, little is known about if the those localization methods, such as covariance localisation (CL), local analysis (LA), covariance fuzzy(CF) and fuzzy analysis (FA), equivalent or not in practice. This study is rather specific in its goals; it aims at giving an answer to that question.

2. Theoretical Background

2.1 Ensemble Kalman filter

Ensemble Kalman filter (EnKF) methods are based on the Kalman filter analysis equations as follows:

$$x^a = x^f + K(d - Hx^f) \quad (1)$$

$$K = P^f H^T \left(HP^f H^T + R\right)^{-1} \qquad (2)$$

$$P = \frac{1}{m-1} AA^T \qquad (3)$$

where the superscript *a* and *f* represent analysis and forecast, respectively; the superscript *T* represents a matrix transpose and *m* is the ensemble size. x^a and x^f represent the state analysis and forecast value, respectively; *d* is the vector of observations; *K* is the Kalman gain; *H* is the observation operator; *R* is the observation error covariance matrix. *P* is the covariance matrix. *A* is the ensemble anomalies.

2.2 The Localization techniques

The localization techniques are applied in the ensemble assimilation, which means the observation data only works on the nearby model points. The general implementation method is to observe covariance of the variables at a certain pattern point multiplied by a continuous function which decreases with the increase of distance. The traditional local methods (Houtekamer and Mitchell 2001) are applied to the EnKF as follows:

$$K_B = (\rho \circ P^f)H^T (H(\rho \circ P^f)H^T + R)^{-1} \qquad (4)$$

$$K_R = (ZY^T)(YY^T + R/\rho)^{-1} \qquad (5)$$

Where K_B and K_R are the Kalman gain of background and observation localization, respectively. ρ represents correlation coefficient matrix based on distance.

2.3 The implementation process with coupling fuzzy logic control

In general, the fuzzy logic control controller consists of four parts: database, fuzzy quantification, inference mechanism and defuzzification. Fig. 1 shows the structure of the data assimilation system coupled with Fuzzy logic control. Firstly, the fuzzy logic database is established, then input distance and output weight variables are fuzzy quantification. According to fuzzy rule, the equivalent weight is obtained by fuzzy inference and finally defuzzification. Subsequently, the weights obtained above are used for status updates of data assimilation.

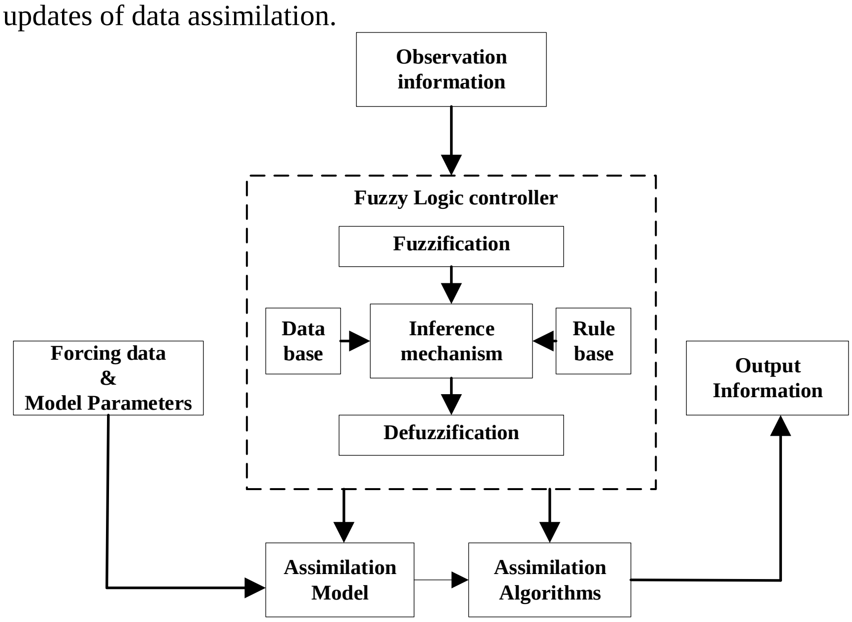


Fig. 1 The structure of the data assimilation system coupled with Fuzzy logic control.

2.4 The Localization method in the EnKF based on Fuzzy Logic Control

The EnKF is an effective data assimilation method. To reduce the error caused by the inefficient use of observation data, we introduce a new localization method with coupling fuzzy logic control to improve the situation. The schematic diagram of the EnKF embedded fuzzy logic control is shown in Table 1.

Before calculating the local coefficients, the fuzzy logic control algorithm is performed to find the more accurate observation weights. Then, the obtained observation weights are used for updating the ensemble mean and ensemble anomalies. Finally, the assimilated analysis values and covariance matrix are obtained for calculating the analysis state Root Mean Square Error(RMSE).

Table 1. The schematic diagram of the EnKF embedded fuzzy logic control

```
Algorithm1: Localization method coupled with the Fuzzy logic control
            algorithm
function main(prm, x, x_true, E, stats)
  initialize x_true0, E0,t (by ens.mat), step0, step1. Create y, H, pos.
  for step = step0: step1 do
    Update E (by a fourth-order Runge-Kutta method)
    generate synthetic observation of the Lorenz96 system.
    function [dx, A] = assimilation(prm, A, HA, pos, dy, stats)
    Begin local process:
    Algorithm2:Calculate the coeffs by the fuzzy control algorithm,
               obtained more accurate observation weight.
      function[coeffs]=calc_loccoeffs(radius,tag,dist)
          switch tag
          | Case 'Gfuzzy'
          |   Design a new fuzzy controller fz6;
          |   Set fuzzy rules rulelist;
          |   Set defuzzy method centroid;
          |   Perform fuzzy reasoning calculation to get coeffs.
          end switch
      end function
    Update the ensemble mean
    Update ensemble anomalies
    end function
    Calculate analysis stats RMSE.
  end for
end function
```

3. Numerical experiments

3.1 Performance comparison of the traditional and new localization methods

In this section, we compare the performance of the traditional localization methods (CL, LA) with the new algorithms (CF, FA) on the 40-variable Lorenz96 model. The assimilation results of the time series of the daily averaged RMSE for the four local assimilation techniques are shown in Fig.2.

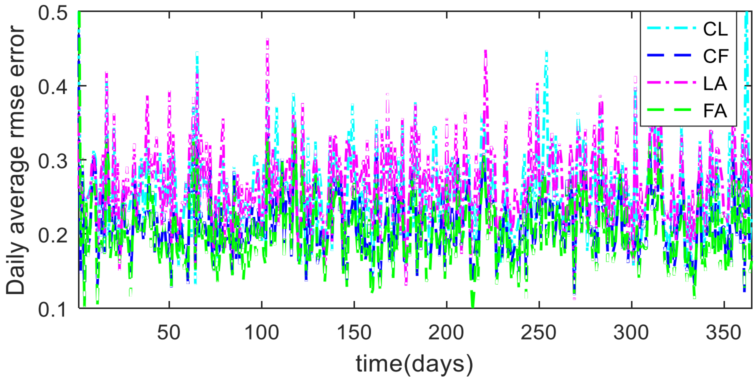


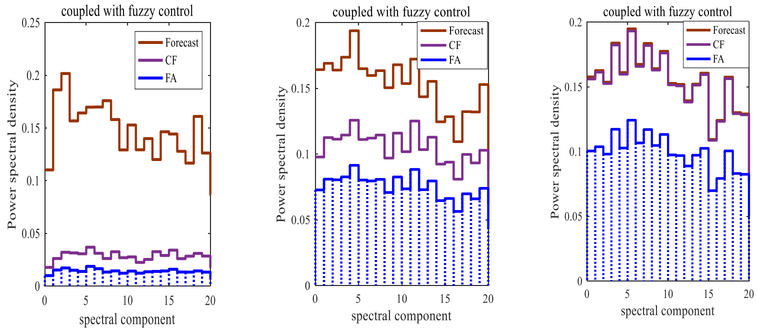
Fig.2. Time series of the daily averaged root mean square error (*RMSE*) for the four local assimilation techniques (CL, LA, CF, FA) using default parameter setups with the Lorenz-96 model.

It shows that all methods behave considerably satisfactory in terms of overall low *RMSE* errors.

However, the FA and CF algorithms perform slightly better than the traditional methods.

3.2 Influence of new localization methods on the Power spectral density (PSD)

The Power Spectrum Density (PSD) method could be used to test the performance of the above algorithms. As the observation error (R) changes the new algorithms of PSD changed correspondingly. As shown in the following Fig. 3, under the different condition of observation error, we compared performance of the CF and FA algorithm.



(a)R=0.0001 (b)R=0.01 (c)R=1

Fig.3. PSD evaluate of the CF and FA algorithms by the effect of Fuzzy control.

It can be seen from the results: (1) When the observation error is relatively small, the PSD of the algorithm with fuzzy control is lower. (2) The PSD value of the algorithm with fuzzy control increased as the value of the error increases. However, the amplitude of increase of the PSD value decreases when the error increases to a certain value. (3) With the increase of the observation error R, the area under the curve for the FA algorithm is always slightly smaller than that for the CF algorithm. Overall, the FA algorithm is more robust following the changes of the error, and the assimilation efficiency can be improved by this algorithm when the error is small.

4. Summary

This paper introduces several kinds of the traditional and new localization methods. In particular, the theoretically concrete implementation process of the new localization method coupling fuzzy logic control is elaborated. Subsequently, several localization methods are compared with the Lorenz-96 model from the view of model error and observation error. In addition, using PSD as the performance evaluation index, the robustness with fuzzy logic control algorithm is proved.

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

[1] Martin M J , Balmaseda M , Bertino L , et al. Status and future of data assimilation in operational oceanography[J]. Journal of Operational Oceanography, 8(sup1): s28-s48 (2015)

[2] Evensen G. Sequential data assimilation with a nonlinear quasigeostrophic model using Monte-Carlo methods to forecast error statistics[J]. Journal of Geophysical Research, 99(55) 10143-10162 (1994)

[3] Hamill T M, Whitaker J S, Snyder C. Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter[J]. Monthly Weather Review, 129（11）: 2776-2790(2001)

[4] Houtekamer P L, Mitchell H L8. Data assimilation using an Ensemble Kalman Filter technique[J]. Mon. Wea. Rev,26: 796-811(1998)

[5] Sakov P, Bertino L. Relation between two common localisation methods for the EnKF[J]. Computational Geosciences, 15 (2): 225-237(2011)

IOP Conf. Series: Earth and Environmental Science **310** (2019) 022035 doi:10.1088/1755-1315/310/2/022035

[6] Lu Yongnan, Bai Yulong, Xu Baoxiong. Observation Error Handling Methods of Data Assimilation Coupled with Fuzzy Control Algorithms[J]. Remote Sensing Technology and Application, 32(3): 459-465(2017)

[7] EPSTEIN E S. Stochastic Dynamic Prediction1 ［J］. Tellus, 21 (6): 739-759(1969)

[8] Anderson, J. L. Localization and Sampling Error Correction in Ensemble Kalman Filter Data Assimilation, J. Monthly Weather Review. 140 (7):2359-2371(2012)

[9] Bishop CH, Etherton BJ, Majumdar SJ. Adaptive Sampling with the Ensemble Transform Kalman Filter. Part I: Theoretical Aspects. Monthly Weather Review .129:420-436(2001)

[10] Miyoshi T, Yamane S, Enomoto T. Localizing the Error Covariance by Physical Distances within a Local Ensemble Transform Kalman Filter (LETKF) [J]. Sola. 3(623): 89-92(2007)