

# Application of Data Assimilation Algorithms Based on Kalman Ensemble Filters for the Lorenz Attractor

Yulia S. Timoshenkova, Nikolay T. Safiullin, and Sergey V. Porshnev

Yeltsin Ural Federal University, Yekaterinburg, Russian Federation  
julia.timoshenkova@urfu.ru n.t.safiullin@urfu.ru s.v.porshnev@urfu.ru

**Abstract.** The paper describes application of the Data Assimilation methods based on the Kalman ensemble filtering to forecast the Lorentz system. The quality of the forecast and the influence of the method for the parameter choice on the result of forecasting the system was assessed. The paper shows the advantages and disadvantages of the chosen methods of Data Assimilation. The recommendations for increasing accuracy of those methods are presented.

**Keywords:** EnKF, LEnKF, Data Assimilation, Lorenz attractor, Kalman filter, Forecast.

## 1 Introduction

Correction of the forecast based on mathematical models with using new observations is an actual problem. This approach avoids growth of the forecast confidence interval as its duration increases.

One of the instruments for solving this task is the Data Assimilation (DA) technique.

There are areas of science, in which the forecast models are built not on a one-dimensional time series generated by the multidimensional one, but on a multidimensional underlying process itself.

The use of methods of assimilation and correction of data ensures correct prediction and modeling of the complex systems.

Since the DA methods are based on the mathematical model of the system, the choice of different methods of data assimilation can significantly affect the final results of the forecast correction.

There are no recommendations for the choice of a particular family of DA methods because of their novelty [1]. Therefore, the problem of comparative analysis of data assimilation methods is actual. In order to compare those methods and to test their accuracy one needs a standard suitable system to forecast.

The Lorentz system was designed to construct a simplified model of atmospheric convection for the long-term weather forecasting. Nowadays, the Lorenz attractor is a standard model for testing the DA methods; thus, it is used as a model to forecast in this work.

The mathematical model of the Lorenz attractor is described by a system consisting of three nonlinear ordinary differential equations that represent a finite amplitude of convection [2]

$$\begin{aligned}\frac{dx}{dt} &= \sigma(y - x), \\ \frac{dy}{dt} &= x(r - y) - y, \\ \frac{dz}{dt} &= xy - bz,\end{aligned}\tag{1}$$

where  $\sigma = \nu/k$  is the Prandtl number,  $r = R_a/R_{ac}$  is the normalized Rayleigh number (normalized),  $b = 4/(1 + a_2)$  is the geometric factor  $x$  is the convection intensity,  $y$  is the temperature difference between ascending and descending wind flows,  $z$  is the deviation of the vertical temperature profile from the linear one.

The peculiarity of this system is that when the parameters are  $\sigma = 10$ ,  $b = 8/3$ , and  $r \geq 24.06$ , the behavior of the system becomes random. In the phase space, the attractor has the topology of the tangle of trajectories, in which two regions can be distinguished [2]. At any time point, the solution is in one of these regions, and the transition of the system state to another region is unpredictable.

The purpose of the work is to carry out a comparative analysis of applying two DA methods, such as the Ensemble Kalman Filter and Local Ensemble Transform Kalman Filter to the Lorenz attractor.

## 2 Mterials and Methods

In the Data Assimilation method, the original model is formulated in accordance with the following system of equations

$$\begin{aligned}x_{k+1} &= M(x_k) + w_k, \\ y_k &= H(x_k) + v_k,\end{aligned}\tag{2}$$

where  $M$  is an operator or transition function that determines the evolution of the system in time,  $x_k$  is the known model state at the time point  $t_k$ ,  $x_{k+1}$  is the model state at the next time point  $t_{k+1}$ ,  $y_k$  is the observation vector,  $H$  is the observation operator that describes the multi-dimensional state of the system based on the observation vector,  $w_k$  is the model error,  $v_k$  is the observation error.

In the DA method, the analysis step is based on the system current state  $x_k^f$ . The corrected system forecasting state  $x_k^a$  for the next time point is based on  $x_k^f$ . For calculation of  $x_k^a$ , only observation vector for the current time point is used [3].

$$x_k^a = x_k^f - K(y_k - H(x_k^f)),\tag{3}$$

where  $K$  is the transmission coefficient,  $y_k$  is the observation,  $e = y_k - H(x_k^f)$  is the error between the known observations and values obtained using the model.

Various DA methods minimize error  $e$  using different techniques described in this paper. The transmission coefficient is also called the Kalman coefficient. There is a family of DA methods based on the Kalman filter (KF). One of those methods is the Ensemble Kalman Filter (EnKF) [4].

The EnKF is a variant of the generalized Kalman filter, in which the covariance of the prediction errors is estimated using the ensemble of forecasts [5]. The Kalman ensemble filtration methods are widely used to assimilate observations in dynamic models.

The EnKF algorithm is implemented by the following sequence of steps.

1. Forming an ensemble of the initial data at the initial time point with the help of a mathematical model.
2. Calculation of  $N$  predictions for ensemble of the initial data in order to obtain the values of the observed parameters at the time point for the state vector obtained on the previous step. Next, the method of successive corrections for the model is used.
3. Calculation of the covariance error matrices for the forecast correction. The step consists of the following items: formation of a covariance error matrix of the forecast vector; calculation of the Kalman weight operator  $K$ ; update of the ensemble model values in accordance with the weight operator and the observation vector; update of the covariance matrix of the forecast errors in accordance with the specified ensemble; calculation of value of the analyzed components and the covariance matrix of the analysis errors.
4. Steps 2 and 3 are executed sequentially for each time point with each new observation obtained.

The Local Ensemble Filters use localization of the observational error covariance, which takes into account only those observations that are located in a certain region around the desired point. The main idea of the Local Ensemble Transform Kalman Filter (LETKF or LEnKF) is to perform calculations not in the physical space of the model, but in the space of the ensemble. Typically, the ensemble space is less than physical one of the model [6].

The LEnKF algorithm is implemented by the following sequence of steps.

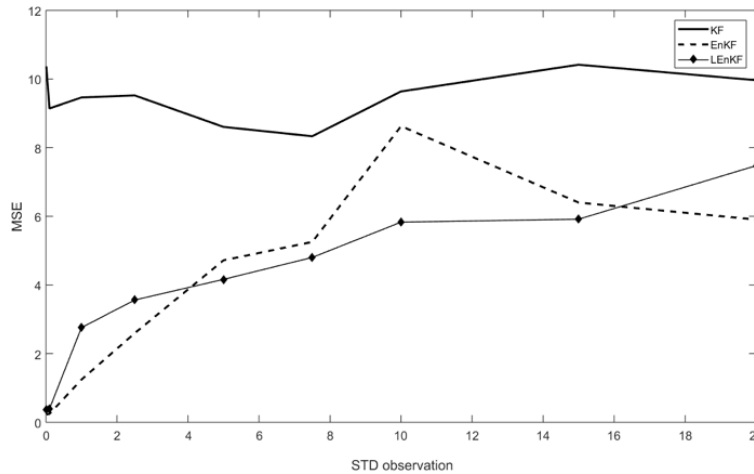
1. Formation of an ensemble of model data of size  $N$  at the initial time point with the help of a mathematical model.
2. Calculation of the ensemble forecast of the system state at the next instant with the help of the model equation and the model state vector at the previous time point.
3. Creating a local plane in the state space, and constructing the projection of each point of the ensemble of the system state onto this plane.
4. For each state vector in the ensemble obtained on step 2, calculate the distance from the ensemble average to the current state vector and project this distance into the low-dimensional state subspace that represents the ensemble in that region in the best way.

5. Perform the assimilation of data in each of the local low-dimensional subspaces, obtaining the average analysis value and covariance of the state error in each local region according to the EnKF.
6. Obtaining the necessary local analysis of the state ensembles using the mean value of the local analysis and covariance.
7. Formation of a new global analysis ensemble using the local analyses obtained on step 6.
8. Repeat steps 1–6 until the observations are complete.

### 3 Results

During the study, the standard deviations of the Lorenz attractor were taken equal to 0.5, 1, and 5, which corresponds to 4.5%, 8.8% and 44% of the coordinate amplitude of the Lorenz system. These values may seem small, but the standard deviations greater than 50% significantly affect the system and generate very inaccurate observations for all three components of the system. The deviation values of the observed states are taken to be from 0.01 to 20, which correspond to 0.08% to 176% of the Lorenz system coordinate amplitude, or approximately from 60 dB to  $-5$  dB in the ratio of the real coordinate to the error of its observation.

Evaluation of the noise influence on the result of the algorithms operation was carried out for several simulation parameters. To analyse the quality of the algorithm, the mean square error (MSE) parameter is calculated. Results of the MSE parameter estimations are shown in Fig. 1.



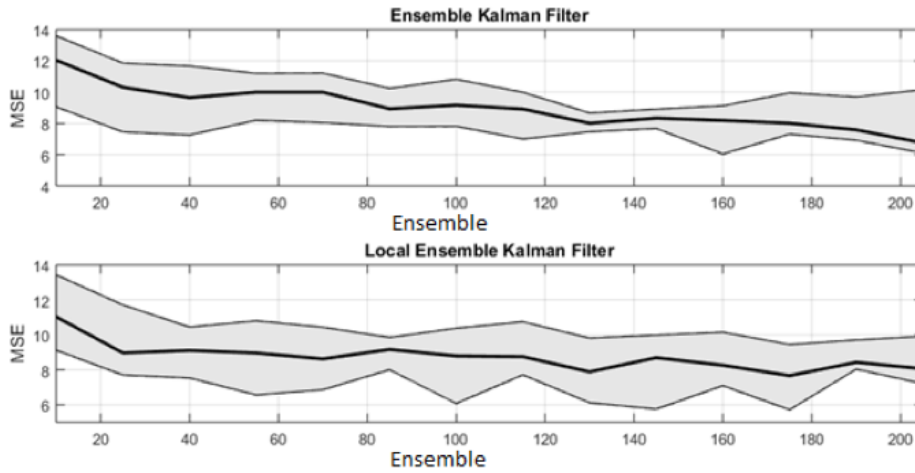
**Fig. 1.** Dependence of the MSE prognosis on the variance of the observation error dispersion

Fig. 1 shows that the choice of the DA method mainly influences on the MSE prognosis. The use of the EnKF gives the best result. For the experiment, in which the highest values of the observation noise dispersion are taken, the worst results of the simulation are obtained. This confirms the theory that the observation error greatly influences the obtained forecast.

It should be noted that the results of the experiment are also influenced by the fact that all components of the attractor are related to each other; that is, presence of errors in one of the coordinates affects the result of the forecast of each other component.

Fig. 2 shows the MSE dependence on the number of ensembles for the EnKF and LEnKF filters. As the number of filter ensembles grows, the resulting value of the MSE decreases.

Under the minimum error in the covariance of observations, the effect of the observation noise variance on the MSE is minimal, as seen in Fig. 1. However, with an increase of the observation error variance and model error, the MSE value becomes approximately equal for all DA methods. This dependence suggests that with increasing the noise variance to achieve an acceptable MSE, it is necessary to increase the number of ensembles in the KF.



**Fig. 2.** Dependence of the MSE prognosis on the number of filter ensembles for the component  $x$  for the EnKF and LEnKF

Based on the obtained results, it can be concluded that when the MSE of the observation error is less than 11, the EnKF should be used to obtain a smaller MSE. When the standard deviation of the observation error is greater than 11, or more than 100% of the system coordinate amplitude, the best smaller MSE are given by the ensemble filters. In the case of using the ensemble filters, the value of the MSE ceases to change significantly with the number of ensembles

larger than 100. Thus, when using the EnKF, the optimum number of ensembles is 100 for predicting the behavior of the Lorenz attractor.

## 4 Conclusion

During the research of the capabilities of under applying the ensemble Kalman filter methods to perform prediction of the coordinates of the Lorenz attractor, the following practical results were obtained.

The increase in variance of the observation error (greater than 70%) significantly affects the MSE prediction. Accuracy of the prediction falls also with the increase in the observation error, as the obtained results shown.

The result of forecast is also significantly influenced by the choice of the filter type. To obtain the minimum MSE, it is recommended to use the LEnKF. A small MSE (lower than 10) of the EnKF can be explained by the fact that the system is modeled several times that averages the error of the modeled chaotic system.

A significant increase in MSE (up to 2–3 times) occurs when the value of the standard deviation of system errors increases by larger than 50%. As a consequence, for high deviation values (higher than 100% of the basic amplitude), the largest possible number of ensembles in the EnKF and LEnKF should be used. But after the value of 100, the value of the MSE ceases to change significantly. Thus it is recommended to use the value of 100 ensembles in the EnKF and LEnKF methods, as a balance between accuracy and speed of the Data Assimilation.

Compared to the simple Kalman Filtering, for higher observational errors, the methods EnKF and LEnKF provide a better accuracy (up to 3 times for small deviations). For observational errors below 40%, the LEnKF performs better in accuracy than the EnKF. But with higher observational errors their accuracy is on the same level. Thus, the choice between those two methods must be made based on their performance and memory consumption.

## References

1. P. Leeuwen, S. Vetra-Carvalho, SANGOMA: Stochastic Assimilation for the Next Generation Ocean Model Applications, Leeuwen Vetra-Carvalho, (2012)
2. W. Van den Bossche, Data assimilation toolbox for Matlab., (2012)
3. G. Evensen, Data assimilation: the ensemble Kalman filter. Springer Science & Business Media, 2009.
4. Y. Timoshenkova, N. Safiullin, On the possibility of the forecast correction for inaccurate observations based on data assimilation., Dynamics of Systems, Mechanisms and Machines (Dynamics), 1–4, (2017). DOI: 10.1109/Dynamics.2017.8239519
5. M. Bocquet and . des P. P. CEREIA, Introduction to the principles and methods of data assimilation in geosciences, Notes Cours c. Ponts ParisTech, (2014)
6. P. L. Houtekamer, H. L. Mitchell, Data assimilation using an ensemble Kalman filter technique, Mon. Weather Rev., Vol. 126, No. 3, 796–811, (1998)