The Extended and Ensemble Kalman Filters on Lorenz Equations

Deepayan Banik*
UofT Department of Physics
(Dated: December 17, 2022)

We present some background on data assimilation, its steps, and applications in the context of nonlinear dynamical systems. To this end, we discuss and apply the Extended and Ensemble Kalman Filters on the set of Lorenz equations that describe a preliminary chaotic atmosphere and compare their efficacies in assimilation. We find the Ensemble filter to be more accurate compared with the Extended filter albeit at the cost of increased computational effort.

1. INTRODUCTION

Imagine you have, by means of fundamental mathematics and physics, figured out the set of differential equations that govern a particular system. If it has just one dependent and one independent variable, then it may be easy to determine the state of the system at any future time. For example, a simple pendulum is relatively simple to track. However, if your system, albeit deterministic, is more complex, involves multiple variables, and is defined by a set of nonlinear partial differential equations, which may exhibit sensitive dependence to initial conditions aka chaos, determining the future state is difficult. This is because it is impossible to know the initial conditions to infinite precision. The mathematical model is bound to eventually diverge from reality. This is why weather prediction does not work past a specific timescale.

1.1. What is data assimilation and retrieval theory (DART)?

To solve the above problem we combine real data with model output under the broad heading of *data assimilation and retrieval theory*. For that, one must be sure that real data is accurate and reliable and to what extent it is so. *Retrieval theory* combines data from multiple sources to produce a more accurate observational representation of a real system, while, *data assimilation* merges these retrievals optimally with model output to predict the future accurately for an extended period of time than is possible by the model alone. It is most commonly used in meteorology and oceanography to improve the performance of weather and climate models, but it also has many other applications, including remote sensing, radar, medical imaging, and finance.

*Electronic address: deepayan.banik@mail.utoronto.ca; URL: https://www.deepayanbanik.info/

1.2. The covariance matrix

While compiling data from various sources, one must be aware of the error associated with them. The algorithm of assimilation accounts for the imperfectness of the retrievals by constructing what is called an *error covariance matrix*.

What is a covariance matrix (CovM)? A covariance matrix is a square matrix that describes the variability and relationship between *two* same-length sets of random variables, say X and Y. Additionally, if X equals Y, the CovM is symmetric, with the diagonal representing the covariance of an element with itself. If this matrix is an identity matrix, it implies that there is zero cross-correlation between the elements of the random variable Y ie. they are not related to one another.

For DART, the random variable could be the *state* of the system at any point in time, the *observations*, the *truth*, or the *error* associated with the state. For now, let us talk about observations ie. data acquired from instruments. The covariance matrix represents how much the different observations are related to one another. Different observations could mean, observations from different sources or at different locations or of different variables. Consider the following scenario: we obtain observations from an instrument but cannot trust the data completely because the instrument is prone to error. We must have a prior understanding of how the acquired data at a particular location in space is related to its values in neighborhood locations. Historically, let us say there exists a strong correlation between observations at some location with those from its neighborhood and a weak correlation with those away from it. In case, that is occasionally not true according to instrument readings, we know how to apply a correction based on the previously constructed covariance matrix. Essentially for evolving systems, the nature of the correlation between different members of the state/observation/error might change over time and that would be captured by the covariance matrix.

1.3. Lorenz equations

The Lorenz equations are a set of three nonlinear ordinary differential equations that form a mathematical model of the behavior of a fluid in the atmosphere

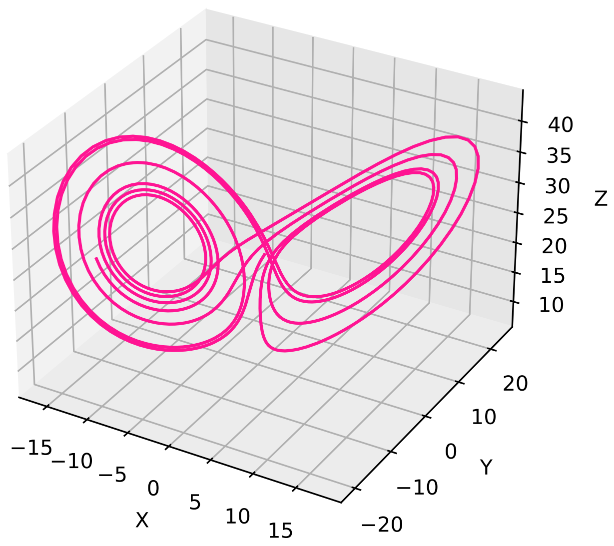

FIG. 1: Three-dimensional phase space representation of the evolution of the Lorenz system. The famous butterfly.

which is approximated as a two-dimensional fluid layer uniformly warmed from below and cooled from above. The equations are named after the meteorologist Edward Lorenz, who first derived them in 1963 while investigating the predictability of weather systems. The equations are given by:

$$\begin{pmatrix} \dot{x} \\ \dot{y} \\ \dot{z} \end{pmatrix} = \begin{pmatrix} \sigma(y-x) \\ x(\rho - z) - y \\ xy - \beta z \end{pmatrix} \tag{1}$$

where, x, y, and z are the variables that represent the state of the system: x is proportional to the rate of convection, y to the horizontal temperature variation, and z to the vertical temperature variation. Therefore the state is a three-component vector, and so are the observations, truth, initial guess, and analysis. σ is the Prandtl number, ρ is the normalized Rayleigh number, and β is a dimensionless wavenumber. These are parameters that determine the behavior of the system. The equations describe how these variables evolve over time.

The system has three equilibrium points: the origin represents no motion or conduction (unstable node), and the two others are $(\pm\sqrt{\beta(\rho-1)}, \pm\sqrt{\beta(\rho-1)}, \rho-1)$ representing two counter-rotating convection cells (unstable spirals) for the following: $\sigma = 10, \rho = 28$ and $\beta = 8/3$. One of the key features of the Lorenz equations is that they exhibit chaotic behavior meaning that small differences in initial conditions can lead to vastly different outcomes over time. This phenomenon is a consequence of the nonlinearity of the equations and sensitivity to initial conditions.

1.4. The tangent linear model

In a tangent linear model, the evolution of the system's state variables is represented as linear equations perturbed around some solution of the state at a specific time. The equations are derived by taking the derivative (or tangent) of the nonlinear equations that describe the full system. This allows the model to accurately capture the system's behavior in the immediate vicinity of the reference point, but it may not be as accurate for predicting behavior at other points in time or for other initial conditions.

A tangent linear model is that it is computationally efficient. Because the model uses linear equations, it can be solved quickly and accurately using well-known mathematical techniques such as matrix inversion. If the model, let us call ϕ is only weakly nonlinear, then we can approximate them using the first two terms of a Taylor expansion:

$$\phi(\overline{x} + \delta x) \approx \phi(\overline{x}) + \frac{\partial \phi(\overline{x})}{\partial x}\delta x \tag{2}$$

where, δx is a small perturbation around state $\overline{x}$. Here, the tangent linear model $\Phi = \frac{\partial \phi(\overline{x})}{\partial x}$. Typically, a tangent linear model breaks down after correctly predicting the state for a while. Figure 2 shows the divergence of the tangent linear form from the nonlinear model at about 5 units of non-dimensional time for the set of Lorenz equations for a small perturbation of magnitude 0.02. Different parameters of the system, like the initial conditions or the constants σ, ρ, and β affect the point of divergence.

2. METHODS

2.1. Components of an assimilation process

- State - the set of variables in the system at every spatial location; e.g. horizontal velocities u and v and height h at 10 grid points would amount to 30 variables in all for the shallow water equations; for the Lorenz system, it is simply the vector $[x \;\; y \;\; z]$

- Initial guess for the state - the initial state where we start the analysis; this acts as the analysis for the first iteration

- Model - the vector representing the RHS of the system of equations where the LHS is just the time derivative of the state expressed as functions of time and other state variables ie. $[\dot{x} \;\; \dot{y} \;\; \dot{z}]$

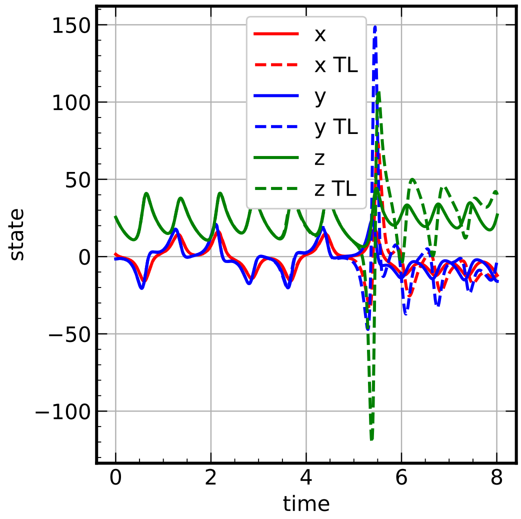

FIG. 2: Evolution of states x, y, and z of the Lorenz eq
by the nonlinear (solid) and tangent linear (dashed) li

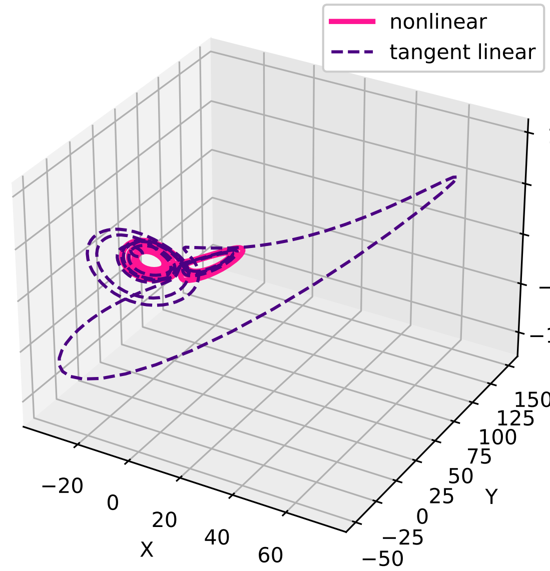

FIG. 3: Phase space representation of the evolution of the Lorenz system (solid pink, same as in Fig 1) and the tangent linear model for the same (dashed violet, shows significant divergence from the nonlinear model).

- Forecast/background/prior - the analysis fro previous timestep propagated forward by the by one timestep
- Observations/measurement - vector of real data coming into the assimilation
- Observation operator matrix - the observation may not be collocated with the model grid points; this operator accounts for such differences in discretization in assimilation and observation spaces
- Gain matrix - describes the relative weightage of observations and the background/prior to generate the current analysis
- Analysis/posterior/estimate - the updated assimilated state comprising contributions from the model forecast/prior/background and observations
- Error covariance matrices of the forecast, observation, analysis - denotes contribution and correlation of errors between state variables in the different stages of assimilation
- Bias - implies whether the mean of errors is zero or not; unbiased is zero mean
- Truth - real model output starting from specific initial conditions (different from the initial guess mentioned above) just for the sake of comparison

2.2. The Kalman Filter

The Kalman filter is a mathematical tool used to estimate the state of a system, given noisy measurements of that system over time. It combines model state and data via a linear superposition to improve the accuracy of the state estimate. This is done using a series of equations that propagate error covariances through the system to account for uncertainty in the state and the measurements. At each time step, the Kalman filter uses a combination of the model state, the previous state estimate, and the current measurement to update the state estimate.

2.3. Nomenclature and equations

The following are the equations for the Kalman filter (k represents timesteps):

System model: $x_{k+1} = \phi_k x_k + w_k$
Measurement model: $z_k = H_k x_k + v_k$
Initial guess/conditions: x_0^f, P_0^f
Kalman gain: $K_k = P_k^f H_k^T (H_k P_k^f H_k^T + R_k^{-1})$
Analysis step 1: $x_k^a = x_k^f + K_k(z_k - H_k x_k^f)$
Analysis step 2: $P_k^a = (I - K_k H_k) P_k^f$
Forecast step 1: $x_{k+1}^f = \phi_k x_k^a$
Forecast step 2: $P_{k+1}^f = \phi_k P_k^a \phi_k^T + Q_k$

Assumptions: The model and the observation errors are unbiased, uncorrelated, and white in time. This implies: $(< w_k >, < v_k >) = (0, 0)$, $< w_k w_l^T >= QI$,

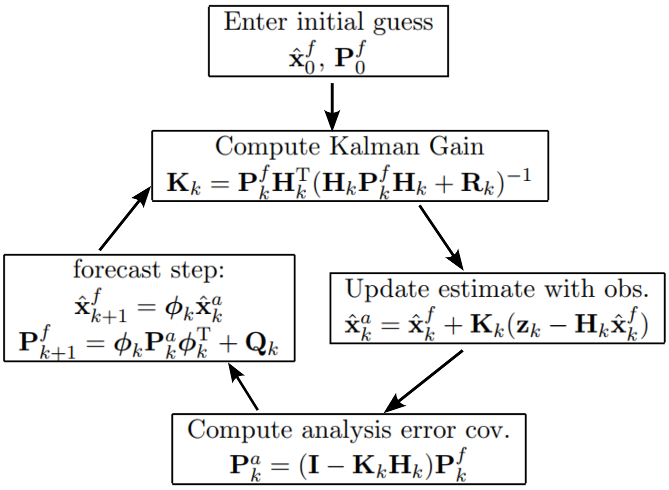

FIG. 4: Flow chart for the Kalman Filter.

$< v_k v_l^T >= RI$ and $< w_k v_l^T >= 0$, where $Q = StdQ^2$, $R = StdO^2$ and I is the identity matrix.

The flow of the Kalman filter is shown in Fig. 4. It is as follows:

1. Choose an initial guess and initial error covariance matrix for the state
2. Set up the observation operator and observation error covariance matrix
3. Compute the Kalman gain using the observation operator and the error covariances
4. Generate analysis/estimate by combining the background/forecast (initial guess for the first timestep) and observations
5. Compute the analysis error covariance matrix by applying a combination of the Kalman gain and observation operator on the forecast error covariance
6. Propagate the analysis and the analysis error covariance by the model to generate the next forecast and the forecast error covariance while accounting for model errors
7. Repeat steps 3 to 6

The Kalman filter described so far works well for linear systems. A lot of real systems are nonlinear and chaotic, which makes the linear discrete KF ineffective.

2.4. The EXTENDED Kalman Filter (EKF)

When the model ϕ is nonlinear, as is true for the Lorenz equations, one uses the extended version of the Kalman Filter. The state is propagated using the full nonlinear model ϕ but the EKF uses the tangent linear model Φ to propagate the analysis error covariance matrix P_k^a. This allows the generation of the dynamics matrix ψ which can be simply applied to P_k^a. The EKF is supposed to work decently well for weakly nonlinear systems because the tangent linear model is just a first-order approximation to the full nonlinear model.

Modified forecast step 2: $P_{k+1}^f = \Phi_k P_k^a \Phi_k^T + Q_k$

2.5. Calculation of dynamics matrix Φ for nonlinear systems

For linear systems, Φ is simply the model, ϕ, operating on an identity matrix. The use of the identity matrix is justified by the fact that the operation converts the linear model ϕ from a 1D vector to a 2D matrix which can be directly applied on the true state to evolve it through one timestep. However, this is not possible for a nonlinear model. The true state has to be evolved separately through the model. While evolving the analysis error covariance, applying a nonlinear model is expensive and hence we construct a dynamics matrix based on the tangent linear model Φ for ϕ. This is done by adding the identity matrix propagated through the nonlinear model to a perturbation about the identity matrix propagated through the tangent linear model. The perturbation is approximately of the same magnitude as the standard deviation in the initial analysis error. See code.

2.6. The ENSEMBLE Kalman Filter (EnKF)

The EnKF was prescribed to do away with the difficult process of propagating the analysis error covariance by a tangent linear model when the state size is huge. The propagation equation of the EKF is removed and the forecast error covariance matrix P_f using an ensemble of forecasts. Since a nonlinear model is used for the forecasts, the covariance thus calculated does not involve any linearization and hence is better than the EKF.

The mean forecast of the ensemble is thus

$$< x_k^f >= \frac{1}{s}\sum_{i=1}^{s} x_k^{f,i} \tag{3}$$

and the forecast error covariance is

$$P_k^f \approx \frac{1}{s-1}(x_k^{f,i} - < x_k^f >)(x_k^{f,i} - < x_k^f >)^T \tag{4}$$

There is no assumption of weakly nonlinear dynamics for the EnKF. It can be applied to highly nonlinear models and was introduced to the meteorological literature by [1] to make up for the divergence of the EKF because of dynamic instabilities.

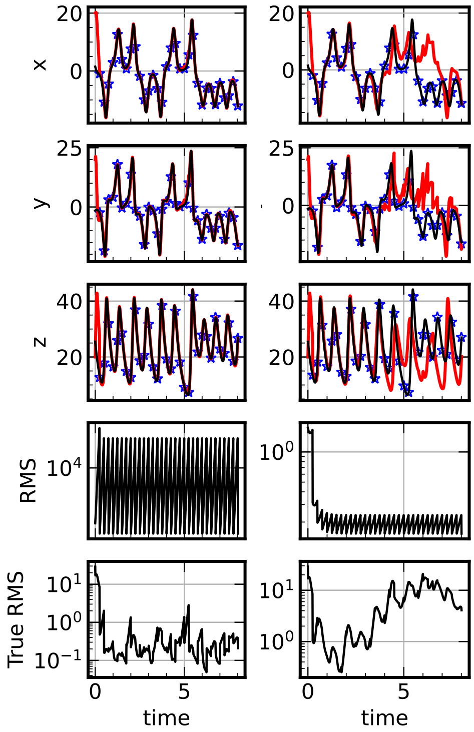

FIG. 5: EKF: StdO = 0.2, tobs = 25, StdP = 1 (column 1), 0.1 (column 2).

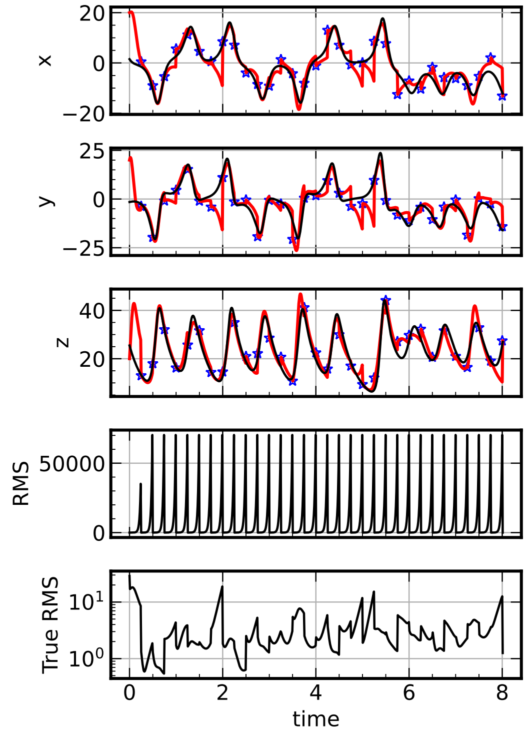

FIG. 6: EKF: StdO = 2.0, tobs = 25, StdP = 1.

2.7. Setup of the assimilation schemes

We will use both the Extended and the Ensemble Kalman filters on the set of Lorenz equations and compare their efficacies in terms of RMS errors and overall performance. The same numerical settings as [2] are applied. The model timestep is 0.01 and is integrated in time using a 4th-order Runge-Kutta scheme. For ease of understanding, the model parameters are kept at standard values of $\sigma = 10, \rho = 28$, and $\beta = 8/3$, and the model error standard deviation *StdQ* is set to 0. The inital conditions are $(x, y, z) = (1.508870, -1.531271, 25.46091)$ and the initial guess for the assimilation scheme is $(x, y, z) = (20, 20, 20)$. Since all the equations are non-dimensionalized, running the model for 800 timesteps gives us a total non-dimensional run time of 8 units. These parameters remain constant across all runs. Observations are generated by perturbing the truth with elements from random Gaussian distribution of mean 0 and standard deviation of 2.

The parameters that can be varied are the observation frequency, the observation error, and the ensemble size.

3. RESULTS

Format of plots: we plot the truth and the analysis of states x, y, and z in black and red, respectively. We also show the expected RMS — the squared root of the trace of the analysis error covariance matrix and the true RMS error — the RMS of the difference between the analysis and the truth.

3.1. Results: EKF

The value of the perturbation p_{pert} to be used for the tangent linear model was not known. What is the right value to be assumed for p_{pert} in the code? We have considered it as a parameter to understand its role in assimilation.

- The smaller the magnitude of the dynamics matrix, ie. the smaller the perturbation p_{pert}, the greater the chance that the analysis diverges from the truth.

- For smaller perturbations, divergence first occurs near the time when the tangent linear model is seen to diverge from the nonlinear model in Fig. 2 ie. approximately 5 units of non-dimensional time.

- Increasing the observation error *StdO* reduces the accuracy of the analysis at all times.

- Increasing the observation frequency *tobs* does not help for low values of p_{pert}, but better constrains

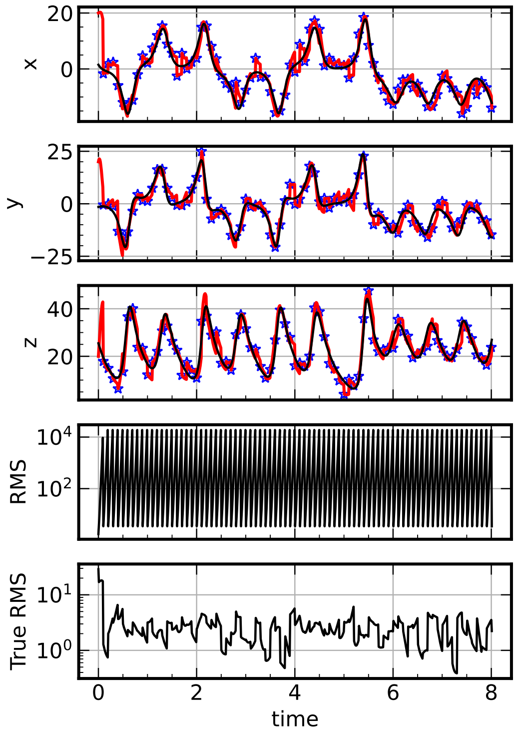

FIG. 7: EKF: StdO = 2.0, tobs = 10, StdP = 1.

the evolution of the analysis even for large observation errors when p_{pert} is of the order of initial background error or larger.

- The expected RMS error is large when the true RMS error is low and vice versa (see Fig. 5; this points towards filter divergence. The expected RMS keeps increasing until an observation comes in when it plummets to a small value, after which it starts increasing again; this is justified in the following way: if the dynamics matrix is, on the whole, greater than 1 (at least the diagonal elements are > 1: corresponds to greater p_{pert}), P_f keeps increasing so long there are no observations, consequently increasing the expected RMS; a larger P_f just before the assimilation step implies a larger Kalman gain, thereby weighing in the observations when available, following which there is a sharp reduction in P_f, meaning that the observations need not be weighted as much for some time and the model can just propagate on its own, till the next observation comes in.

3.2. Results: EnKF

In the Ensemble Kalman Filter, an ensemble of forecasts is propagated forward in association with an ensemble of observations, to obtain an ensemble of analyses,

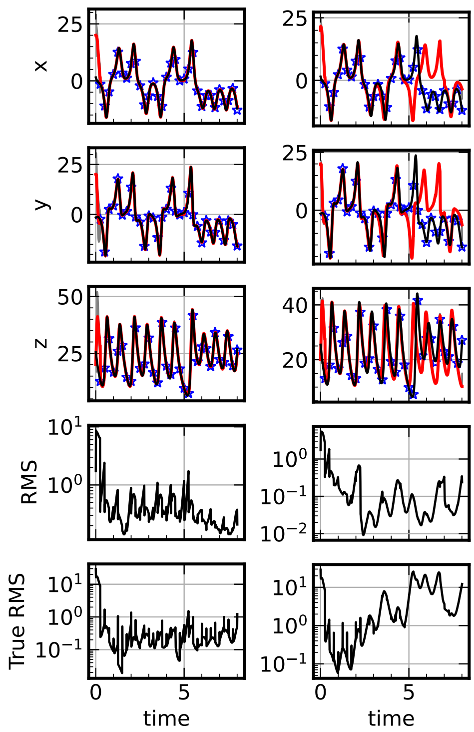

FIG. 8: EnKF: StdO = 0.2, tobs = 25, ensize = 10 (column 1), 2 (column 2).

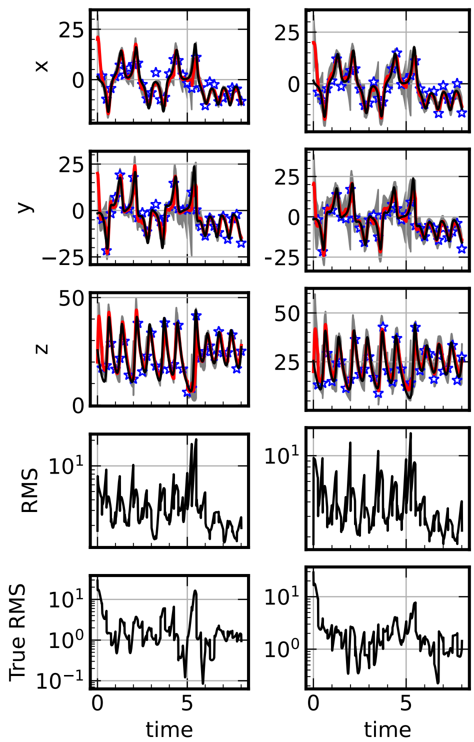

FIG. 9: EnKF: StdO = 2.0, tobs = 25, ensize = 10 (column 1), 100 (column 2).

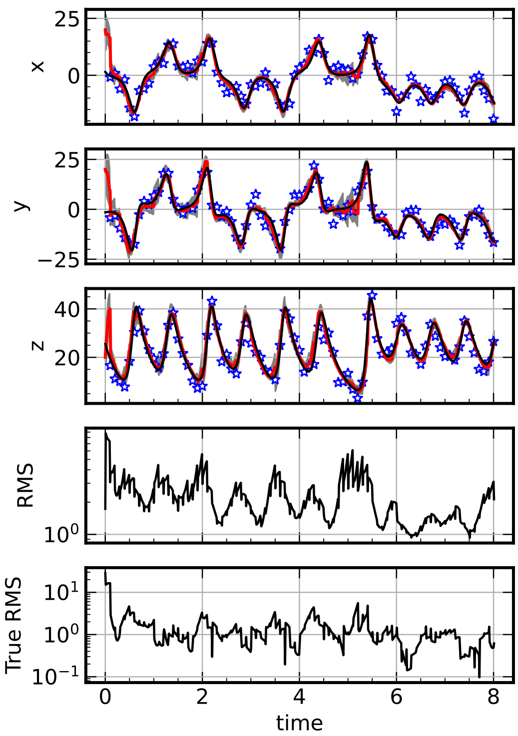

FIG. 10: EnKF: StdO = 2.0, tobs = 10, ensize = 10.

and a mean is calculated that represents the final analysis. The observation ensemble is created by first perturbing the truth with elements of a random normal distribution with a factor of $stdO^2$, as for the EKF, and then perturbed several times by the same magnitude around the original observation to produce many realizations for the observation. The plots for the states x, y, and z also show the spread of the state around the mean evolving with time in grey, in addition to the analysis in red and the truth in black.

- As is expected, the EnKF performs better than the EKF for similar parameter sets, especially when *StdO* is large (the true RMS in Fig. 6 is larger than Fig. 9 column 1).

- Ensemble size: increasing ensemble size (*ensize*) improves the performance of the EnKF. See Fig. 8 comparing 10 ensemble members versus 2. For the latter case the analysis diverges beyond a certain point in time. However, increasing *ensize* beyond a certain limit will only make the scheme computationally expensive without improving the accuracy any further (see Fig. 9).

- Expected RMS error shows better behavior and is commensurate with the true RMS error than is the case for the EKF, where the expected RMS is unusually large even though the assimilation performs relatively fine. The EnKF is therefore a nondivergent filter.
- Increased observation frequency *tobs* results in better analysis. In Fig. 10 we see both the expected and true RMS decreasing with time for $tobs = 10$.

4. CONCLUSIONS

In this report, we have applied data assimilation on the Lorenz system of equations, which are nonlinear in nature and exhibit sensitive dependence to initial conditions or chaos. The assimilation schemes used here are the Extended and Ensemble Kalman Filters. The basic concept of the Kalman Filter in data assimilation is that it combines observations with output from a mathematical model based on a gain calculated by weighing in several factors such as the error in observations, the error in the previous state (variables being solved for) and the pattern of taking observations, all of which are known apriori. If the gain is large the scheme weighs observations in more than if the gain is small. The model itself updates the state when there are no observations. The Kalman Filter is advantageous because it also propagates errors in the analysis using the model. However, this works only for linear dynamical systems.

The Extended KF is applicable to weakly nonlinear systems and uses a tangent linear model to propagate error covariances. This becomes difficult for large systems with several variables. The Ensemble KF is more useful because it gets rid of the tangent linear model and error covariance propagation altogether. It directly evaluates the forecast error covariance matrix using an ensemble of forecasts.

We have compared the two filters for the Lorenz equations and found that the Ensemble Filter performs better especially when the observations are not very reliable which is the case for most practical applications.

[1] G. Evensen, Ocean dynamics **53**, 343 (2003).
[2] R. N. Miller, M. Ghil, and F. Gauthiez, Journal of Atmospheric Sciences **51**, 1037 (1994).