Ensemble data assimilation in the

Ensemble data assimilation in the presence of cloud

S. Vetra-Carvalho*, S. Migliorini, N. K. Nichols

Sc f Ma e a c, Me e g, a d P c

phenomena often impact on very localised regions and models are required to have high spatial resolution to maximise the chances of skillful forecasts. At high resolutions orography can be resolved much better, which allows the

[9]. An ensemble could be used to investigate the relationships between model variables to estimate the initial covariance matrix for convective flow in variational systems. It is also important to study the performance of the ensemble method itself at convective scales, for example, the EnSRF performance in cases where the parameterized cloud growth is a strongly non-linear function of the state variables. Of particular interest, is to investigate the ensemble response to a regime switch from linear to highly-nonlinear (i.e. a switch from linear advection of state variables with no cloud to a sudden cloud growth in the system) and the EnSRF ability to trace the true solution in the presence of parametrized variables such as cloud fraction and rain.

Here we apply the EnSRF to an idealised 1+1D convective-scale model first developed by A. Rudd [5] with parametrized cloud and rain to investigate the performance of the EnSRF in the case of a sudden regime change from linear to highly non-linear. We examine the frequency of observations and the number of ensemble members needed for the EnSRF to capture the solution in the linear phase, as well as the ability of the ensemble to detect the switch and capture the cloud growth after the change of regime. We are especially interested in whether the cloud growth is indicated in the forecast error correlations before it actually happens. To test the ensemble further we consider the case where part of the growing cloud is allowed to rain out. This case is more complex as it correlates the two control variables, temperature and total water, which were previously independent of each other. It is also of interest to see if there is an optimal way to initialise the ensemble, which would give better results without increasing its size.

2. The EnSRF

An Ensemble Square Root Filter (EnSRF) has been built and implemented, as given by [2], making sure that the filter is unbiased and does not collapse [4], [1]. The EnSRF is based on the Kalman Filter (KF) equations,

$$\mathrm{x}^a = \mathrm{x}^f + \mathrm{K}\left(\mathrm{y} \quad \mathrm{H}\mathrm{x}^f\right) \tag{1}$$

$$\mathrm{P}^a = (\mathrm{I} \quad \mathrm{K}\mathrm{H})\,\mathrm{P}^f \tag{2}$$

$$\mathrm{K} = \mathrm{P}^f\mathrm{H}^T\left(\mathrm{H}\mathrm{P}^f\mathrm{H}^T + \mathrm{R}\right)^{-} \tag{3}$$

where x is the state vector, K is the Kalman gain, y is the observation vector, H is the observation operator, P denotes an error correlation matrix and superscripts , stand for analysis and forecast, respectively. The ensemble matrix is defined as

$$\mathbf{X} = [\mathbf{x}\ , \mathbf{x}\ , ..., \mathbf{x}_N] \quad \mathcal{R}^{n \times N} \tag{4}$$

where $\mathbf{x}_i$ are state vectors (ensemble members), n is the total number of variables in the control vector, and is the number of ensemble members. The model state is assumed to be represented by the ensemble mean, which we define as

$$\overline{\mathbf{x}} = \frac{1}{\ } \sum_{i}^{N} \mathbf{x}_{(i)} \qquad \text{or} \qquad \overline{\mathbf{X}} = (\overline{\mathbf{x}}, \overline{\mathbf{x}}, ..., \overline{\mathbf{x}})$$

where the terms V , , S, Z, come from the decompositions (see [2] for more detail). Finally the analysis of the ensemble is

$$\mathbf{X}^a = \overline{\mathbf{X}}^a + \mathbf{X}'^a \tag{7}$$

and the ensemble analysis forecast error covariance matrix is

$$\mathbf{P}_e^a = \frac{1}{ 1}\mathbf{X}'^a\mathbf{X}'^{aT}. \tag{8}$$

The main di erence between the traditional EnKF and the square root version EnSRF (and the reason for using EnSRF) is that in the square root algorithm the perturbation of measurements is avoided as this can lead to more errors. Also, we do not need to perform inversion of the observation error correlation matrix, R, nor do we need the assumption of uncorrelated measurement error covariance matrix [2].

3. 1+1D column model

In this work we use a 1+1D model (1D in space and time) describing the atmospheric flow in a vertical column. The model state variables are vertical velocity, ( ), temperature, ( ), total water, $_t$( ), pressure, ( ), temperature change with height, Ψ( ), liquid cloud water, $_{cl}$( ), saturated vapour, $_{sat}$( ), and cloud fraction, ( ), where [0, 12] km is vertical height. From all of the model variables, only ( ) and $_t$( ) are used in the data assimilation process to update the system, and these are known as the control variables. In vector form we define the control vector x, as

$$\mathbf{x} = \begin{pmatrix} \mathbf{T} \\ \mathbf{q}_t \end{pmatrix} \quad \mathcal{R}^{\,0\,\times}, \tag{9}$$

with height being discretised with 51 equal levels and T, $\mathbf{q}_t$ $\mathcal{R}^{\times}$.

Control variables, and $_t$, at a given height are linearly advected by a known vertical velocity ( ) = $0.5\sin((/ _{top})\pi)$, constant in time with maximum speed in the middle atmosphere of 0.5 m/s. The model uses a cloud scheme [7], to compute a strongly non-linear cloud fraction, , given by

$$( ) = 0.5\left(1 + \tanh\left(\frac{2\ _{cl}( )}{_{sat}(z)(1 \quad H)}\right)\right), \tag{10}$$

[1]Here we do use uncorrelated **R**.

where H is critical relative humidity, ${}_{cl} = {}_{t}$ ${}_{sat}$ and ${}_{sat} = \epsilon e_s/$ with ϵ = 0.622 being the ratio of molecular weights of water and dry air and $e_s = e_s(\ )$ being the saturation vapour pressure. Thus, cloud fraction, , depends on both control variables and ${}_{t}$, and the range of is [0,1] with = 0 meaning no cloud and = 1 meaning full cloud.

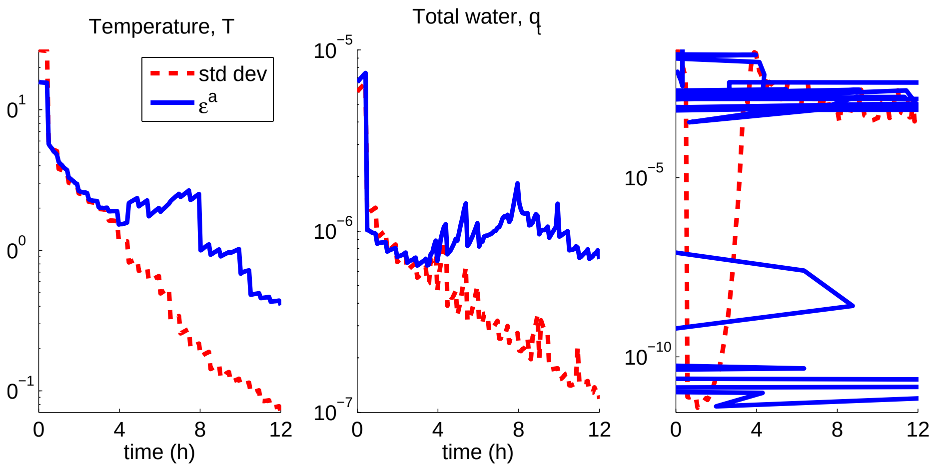
The model exhibits linear and non-linear regimes. If the model exhibits no cloud, the system is in a linear state and we refer to this as a 'no cloud regime' or 'linear regime'. However, if the model has cloud growth, the non-linear. The model parameters are chosen so that initially for the first 3 hours there is no cloud in the solution, but then a very sudden and fast cloud growth occurs in the upper half of the atmosphere. To examine the EnSRF performance in these conditions we use a twin experiment. We create a 12 hour long reference solution ('truth') from initial chosen profiles of (, 0) and $_t$(, 0). We then sample observations with some time frequency from the reference solution and add noise with a σ_o = 1% variance. We also perturb the initial profiles with σ_e = 10% variance to give an initial ensemble mean around which we create the initial ensemble. Using a twin experiment allows us to deduce the accuracy of the ensemble estimate as a di erence between the ensemble analysis and the truth.

4.1. Results for regime switch

The ensemble exhibits a high ability to capture the solution in the linear phase (i.e. no cloud) even with a small ensemble size (= 10) with respect to the size of the state space (n = 102), provided enough good observations of both control variables are given at a suitable time frequency, e.g every 30 min. For larger ensemble size the observation quality and/or frequency can vary to obtain the same accuracy of the solution. Interestingly, the EnSRF is able to capture the regime switch in most cases, with the accuracy and the rate of ensemble convergence to the true solution depending mainly on the ensemble size and secondly on the observation frequency. However, it is important that both control variables are observed at least once before the cloud growth. Note that, in cases where the ensemble size is small, we find that the EnSRF over-predicts the cloud growth, i.e. in the ensemble estimate, cloud growth is initiated sooner than in the reference solution. This clearly is an issue, especially in an operational setting, where the ensemble size is much smaller than the size of the state space. However, in this idealised model, unless liquid is removed from the system, the model would just saturate in the levels where cloud is present, thus becoming linear again. If this is allowed, the ensemble, even with = 10, can capture the solution very well after a few observation cycles.

4.2. Results with rain parametrization

To keep nonlinearity in the system we introduce parametrized rain as explained in section 3. This reduces the cloud fraction after the cloud has grown to a given threshold in the column. In this case, a large ensemble (say $>$ 102) will be able to capture the solution with parametrized rain well if good observations (at least every 30min) are used. However, it is not realistic, for practical applications, to have an ensemble size larger than the state space. With a small ensemble size and many good observations at the initial phase (no cloud regime), the ensemble is not capable of increasing its



Temperature, T
std dev
ε^a
0^1
0^0
0^{-1}
0
4
8
12
time (h)
Total water, q_t
10^{-5}
10^{-6}
10^{-7}
0
4
8
12
time (h)
10^{-5}
10^{-10}
0
4
8
12

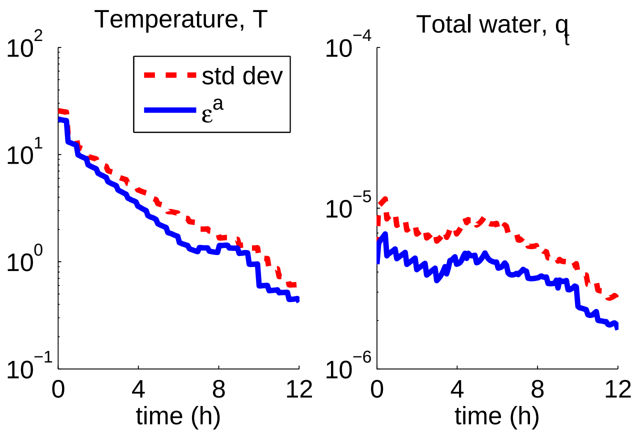


Figure 2: Ensemble analysis error (blue) compared to ensemble standard deviation (red). Observing every tenth level of T and q_t every 30min, $N = 30$. Rain at 3h.

system was observed at least once before the cloud developed, even with a small ensemble size, the EnSRF was able to detect the regime switch. However, the accuracy with which the switch was captured depended mainly on the ensemble size and secondly on the observation frequency. An important note is that for small ensemble sizes the EnSRF developed cloud in its solution sooner than in the reference solution. The reason for this and the impact it has on the ensemble solution needs to be investigated further.

In the cases where parametrized rain was permitted after the initial cloud development, we found that for ensembles that have more ensemble members than variables with good frequent observations, the performance of the Ensembles and, in particular, when rain was parametrized, the correlations and cross-correlations at the time of cloud development were shifted in time or incorrect as the ensemble either developed cloud too soon and hence developed cross-correlations too soon, or its spread was too small, thus resulting in incorrect correlations.

References

[1] L. Baker, *Properties of the Ensemble Kalman Filter*, 2007, University of Reading, Department of Mathematics, MSc Thesis

[2] G. Evensen, *Sampling strategies and square root analysis schemes for the EnKF*, Ocean Dynamics, 2004, Vol 54, pp 539 - 560

[3] R.E. Kalman, *A new approach to linear filtering and prediction problems*,