On the use of an idealised convective column model for ensemble data assimilation

S. Vetra-Carvalho, S. Migliorini, N. K. Nichols, S. P. Ballard

DA Workshop 2019, Leeds

May 16, 2019

Funding: NERC & Met Office

Motivation

In my PhD I looked at convective scale data assimilation from two angles:

- Hydrostatic balance properties in high resolution ensemble forecast errors in the presence of convection;
- *How well does EnSRF capture sudden model regime changes (linear, non-linear, discontinuous) using an idealised column model.*

Model variables

The 1+1D model in time and space is based on the 1D Unified Model from Met Office and was first developed by A. Rudd [Rudd et al., 2012]

Symbol	Description	Units
T	temperature	K
Ψ	change of temp. with height	K/km
p	pressure	hPa
q_t	total water	kg/kg
RH_t	total relative humidity	%
f	cloud fraction	%

Figure: Model variables

Prescribed model parameters:

- RHc: critical relative humidity in %;
- $w(z)$: vertical velocity in m/s (varies in z but not in time).

Model flow

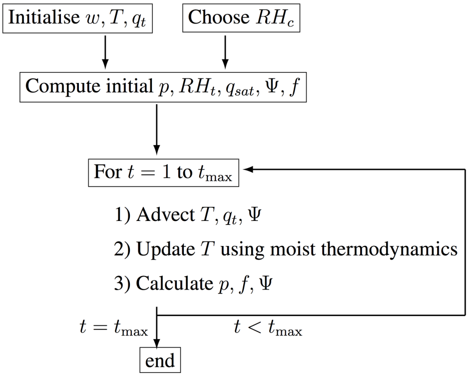


Figure: Flow of the non-linear model with rain not permitted.

Initial profiles

Vertical wind is a constant function in time, given by

$$w_j = 0.1 \sin\left(\frac{z_j}{z_t}\pi\right).$$

The temperature profile is given by

$$T_j = T_1 - (j-1)80.0/Z,$$

where $T_1 = 300$ K is ground temperature, and $j = 1, ..., 51$.

Total water is given by

$$q_t = 0.01 q_{sat}.$$

Model profiles

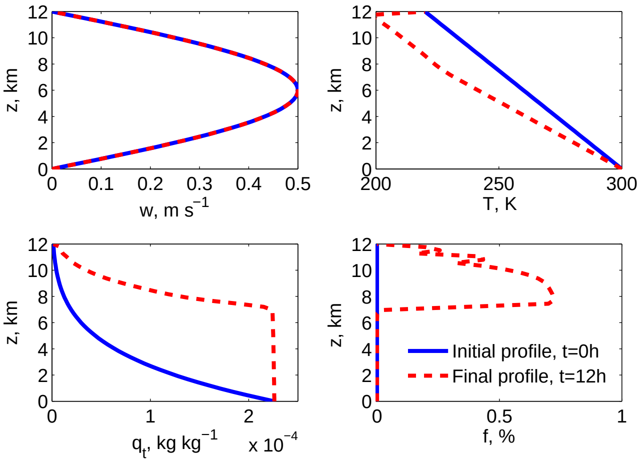


Figure: Model $t = 0$ and $t = 12$h profiles for wind, temperature, total

We use 1st order SL advection scheme to advect q_t, T, and Φ. Their discretised advection equations are as follows:

$$
\begin{aligned}
q_j(t) &= \left(1 - \frac{w_j \Delta t}{\Delta z}\right) q_j(t-1) + \frac{w_j \Delta t}{\Delta z} q_{j-1}(t-1) \\
T_j^a(t) &= \left(1 - \frac{w_j \Delta t}{\Delta z}\right) T_j(t-1) + \frac{w_j \Delta t}{\Delta z} T_{j-1}(t-1) \\
\Psi_j(t) &= \left(1 - \frac{w_j \Delta t}{\Delta z}\right) \Psi_j(t-1) + \frac{w_j \Delta t}{\Delta z} \Psi_{j-1}(t-1)
\end{aligned}
$$

where $T_j^a(t)$ is the advected temperature quantity and forms part of temperature profile at time t:

$$
\begin{aligned}
T_j(t) &= T_j^a(t) + \Delta T_j(t) \\
T_j(t) &= T_j^a(t) - \Psi_j(t) w \Delta t.
\end{aligned}
$$

We use hydrostatic balance to calculate the **pressure** variable p at each gridpoint.

Cloud fraction is calculated using a simplified version of the Sharpe cloud scheme [Rudd 2012, Sharepe 2004]:

$$f\left(T, q_t\right) = 0.5\left(1 + \tanh\left(\frac{2(RH_t - 1)}{1 - RHc}\right)\right) \tag{1}$$

Using the new cloud fraction values, the **change of temperature** is then calculated:

$$\Psi(f, T, q_t) = \left(1 - f(T, q_t)\right)\Gamma_d + f(T, q_t)\Gamma_s, \tag{2}$$

where Γ_d is the dry adiabatic lapse rate (*a constant*) and Γ_s is the saturated adiabatic lapse rate (*nonlinear function*).

Using the definition on total relative humidity, $RH_t = q_t/q_{sat}$, we can rewrite cloud scheme as follows,

$$f_j = 0.5\left(1 + \tanh\left(\frac{2q_{cl,j}}{q_{sat,j}(1 - RHc)}\right)\right),$$

- RHc is the critical relative humidity for cloud formation
- q_{sat} is saturation specific humidity
- $q_{cl} = q_t - q_{sat}$ represents cloud liquid water.

When $f_j > 0.2$, cloud fraction is reduced (rain) by fraction a

$$\begin{aligned} f_j &:= af_j \\ q_{cl,j} &:= 0.5q_{sat,j}\left(1 - RHc\right)\operatorname{arctanh}\left(2f_j - 1\right) \\ q_t &:= q_{cl} + q_{sat}. \end{aligned}$$

We choose initial profiles and parameters (w and $RH_c = 85\%$) such that the model exhibits a clear linear period with no cloud followed by a sudden cloud growth.

Model also retains part of the lower domain in a linear regime for the entire 12h run, i.e. this setup cloud never forms below $z = 6.5$km.

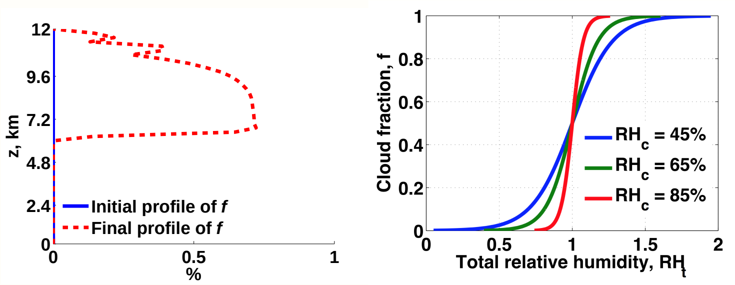


Figure: **Left:** Linear and non-linear regimes seen at $t = 0, 12$h cloud fraction profiles with rain scheme. **Right:** Change in cloud growth speed with varying RHc.

Data assimilation setup

- State variable: $\mathbf{x} = (\mathbf{T}(z), \mathbf{q}_t(z)) \in \mathcal{R}^{102\times 1}$.
- $z = 1, ..., 51$ vertical levels
- 12h assimilation window
- Parametrized cloud and rain
- $\mathbf{R}$ diagonal with $\sigma_o \in [5\%, 25\%]$
- Observing both T and q_t

The ensemble is defined as $\mathbf{X} = (\mathbf{x}_1 \;\; \mathbf{x}_2 \;\; \ldots \;\; \mathbf{x}_N) \in \mathcal{R}^{102\times N_e}$

and perturbations are given by $\mathbf{X}'_i = \mathbf{X}_i - \bar{\mathbf{x}}$ for $i = 1, ..., N_e$.

EnSRF

Ensemble forecast error covariance matrix is given by

$$\mathbf{P}_e^f = \frac{\mathbf{X}'\mathbf{X}'^T}{N-1}$$

The ensemble analysis error covariance can be written as

$$(\mathbf{P}^a \approx)\,\mathbf{X}'^a\mathbf{X}'^{aT} = (\mathbf{I} - \mathbf{KH})\,\mathbf{P_e}^f.$$

Decompositing **KH** we can express ensemble analysis perturbations explicitly

$$\mathbf{X}'^a = \mathbf{X}'^f\mathbf{V}_2\sqrt{\mathbf{I} - \mathbf{\Sigma}_2^T\mathbf{\Sigma}_2}\mathbf{V}_2^T.$$

The analysis error covariance matrix is given by

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

Geir Evensen at http://enkf.nersc.no/

Initial ensemble

The initial true error covariance matrix is

$$\mathbf{B} = \mathbf{\Sigma}\mathbf{C}\mathbf{\Sigma},$$

where

- $\mathbf{\Sigma}^2$ is a diagonal matrix with initial $\sigma_e \in [10\%, 15\%]$ of ref. values;
- $\mathbf{C}$ is a correlation matrix given by SOAR function

$$\mathbf{C}_{i,j} = \left(1 + \frac{|i-j|dz}{L}\right) e^{-|i-j|dz/L}.$$

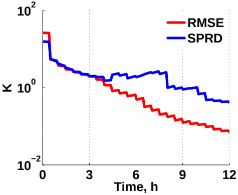
- Initially there are no cross-correlations between $\mathbf{T}$ and $\mathbf{q}_t$.

The initial ensemble perturbations are randomly sampled from this $\mathbf{B}$ matrix with Gaussian distribution.

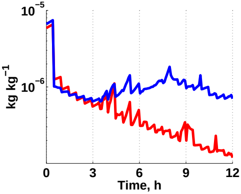
Results - no rain

- Initial profiles have important effect on cloud top and base heights, as well as time and speed of cloud growth;
- Observing the entire state vector produces an ensemble that is too confident unable to increase its spread in the cloud growth regime sufficiently even with $N_e = 100$;
- Reducing number of spatial observations allows the ensemble spread to respond to regime change and capture the cloud growth;
- Correlation between T and q_t is one-sided, with T affected by q_t through cloud fraction.

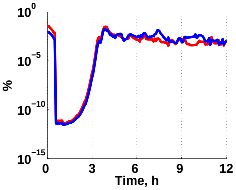
RMSE vs. Ens. Std., $N_e = 30$, observing entire state vector



((a)) T

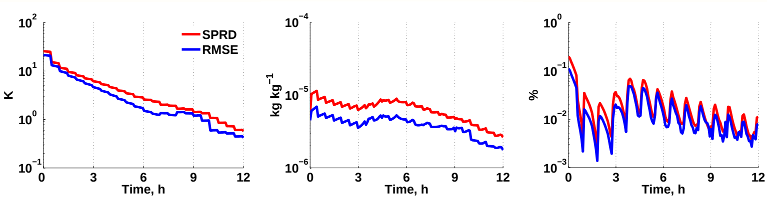


((b)) q_t



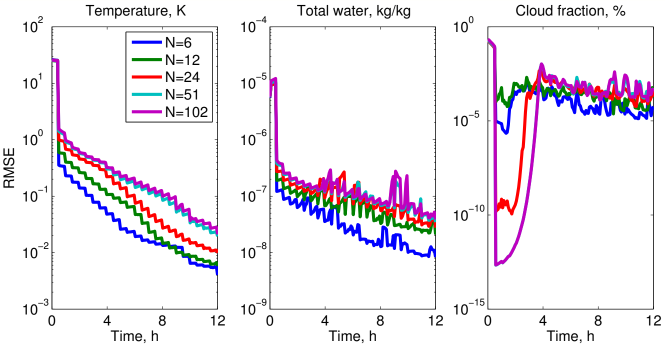
((c)) f

RMSE vs. Ens. Std., $N_e = 30$, observing every 5th element in state vector

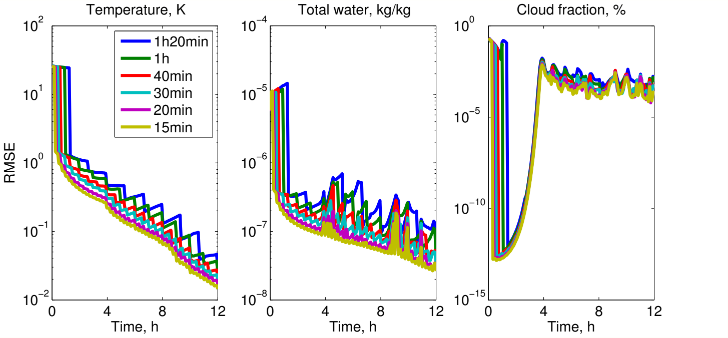


((a)) T ((b)) q_t ((c)) f

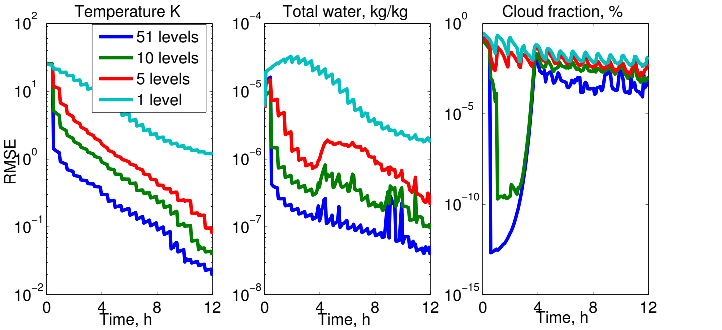
RMSE varying ensemble size when observing entire state space every 30min



RMSE varying observation frequency in *time* for $N = 51$ when observing entire state space



RMSE varying observation frequency in *space* for $N = 51$ when observing every 30min

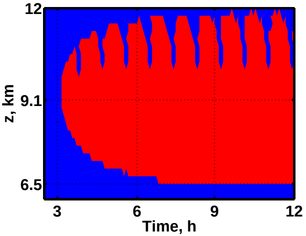


Results - rain

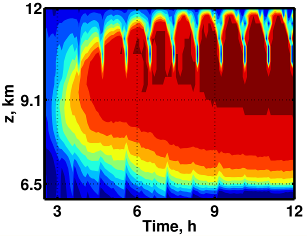
- Initial profiles have important effect on cloud top and base heights, as well as time and speed of cloud growth;
- Observing the entire state vector produces an ensemble that is too confident unable to increase its spread in the cloud growth regime sufficiently even with $N_e = 100$;
- Reducing number of spatial observations allows the ensemble spread to respond to regime change and capture the cloud growth and discontinuous rain regime;
- Correlation between T and q_t is two-sided now.

Presence of rain in the ensemble

Here, $N_e = 30$, observing every 5th vertical level of T and q_t.
Ensemble results average of 20 runs.



((a)) Reference



((b)) Ens.

Figure: In ensemble solutions colours represent a fraction of ensemble members that are raining at a point in space and time: with dark blue meaning none of the ensemble members are raining and dark red all of the ensemble members are raining.

Conclusions

- Small ensemble can capture no cloud solution however, if many good observations given at each assimilation time, ensemble spread does not increase sufficiently to capture the solution in the regime switch. *Ensemble too confident.*
- Increasing ensemble size has an important impact on the ensemble performance, however benefit from running ensemble larger than the size of the state space is small.
- Given some observations ensemble can capture the 'no cloud' case.
- Given frequent in time (not fully observed state) set of observations it is possible to reasonably capture the discontinuous rain regime.
- Ensemble in most of cases over-predicts cloud growth.

Other research using the model

Alison Rudd developed the model in Fortran for use with variational data assimilation for satellite observations [Rudd 2009, 2012].

The aim of her work was to investigate if it is possible to recover a simple profile of vertical motion from observations of cloud-affected brightness temperatures, via a standard VAR scheme.

Showed that:

- Cost function sensitivity to prognostic variables depended on the wetness of the initial profile;
- Vertical wind profile could not be recovered by the variational scheme through observing only TB where $RH_t > 60\%$;
- In more moist and non-linear conditions the accuracy of TLM became questionable as more higher order terms were being neglected.

What could we do more?

- Investigate inflation and localisation methods to allow ensemble to make use of more observations with small ensemble size.
- Compare various DA methods, e.g. variational and ensemble methods with this model.
- Augment the state with cloud top and bottom height observations.
- Further, the model can be made more complex by making the vertical velocity time-dependent and including it in the sate vector.

- G. Evensen, *Sampling strategies and square root analysis schemes for the EnKF*, 2004, Vol. 54, Nr 6, pp. 539-560(22)
- A.C. Rudd, I. Roulstone, J.R. Eyre, *A simple column model to explore anticipated problems in variational assimilation of satellite observations*, Environmental Modelling Software, Vol. 27-28, pp. 23-39
- P. Sakov, P. R. Oke, *Implications of the form of the ensemble transformation in the ensemble square root filters*, 2007, Mon. Wea. Rev., Vol. 136, pp. 1042-1053
- M. K. Tippett, J. L. Anderson, C. H. Bishop, T. M. Hamill, J. S. Whitaker, *Ensemble Square Root Filters*, 2003, AMS, Mon. Weath. rev., Vol. 131, No 17, pp 1485 - 1490