The impact of localization and observation averaging for convective-scale data assimilation in a simple stochastic model

George C. Craig, Michael Würsch*
Meteorological Institute, University of Munich, Germany
*Correspondence to: M. Würsch, Meteorological Institute, University of Munich, Theresienstr. 37, 80333 München.
E-Mail: michael.wuersch@lmu.de

A simple stochastic model, based on a Poisson birth-death process, is proposed as a test-bed for convective-scale data assimilation methods. The simple model mimics the extreme nonlinearity and non-Gaussianity associated with rapidly developing and intermittent convective storms. In this framework, we evaluate the ETKF (Ensemble Transform Kalman Filter) and SIR (Sequential Importance Resampling) filters, and assess the impact of two strategies to improve their performance and efficiency: localization and observation averaging. In their basic implementations, both filters perform poorly. The SIR filter rapidly collapses, then very gradually converges to the observations as random perturbations introduced by resampling occasionally improve the analysis. The ETKF rapidly assimilates the correct locations of convective storms, but has large errors due to creation of spurious clouds by nonlinear amplification of small data assimilation increments. Localization, i.e. assimilating only local observations to produce the analysis at a given grid point, dramatically improves the performance of the SIR filter, but does not reduce errors in the ETKF. Observation averaging, i.e. spatially smoothing the observations before assimilation and thus making the distribution more Gaussian, is also effective for the SIR filter, and improves convergence of the ETKF. **Copyright © 2011 Royal Meteorological Society**

Key Words: Data Assimilation, Ensemble Transform Kalman Filter, Particle Filter

Received . . .

Citation: . . .

1. Introduction

Convective-scale NWP (Numerical Weather Prediction), based on models with a horizontal resolution of order 1 km, is motivated to a large extent by the desire to predict precipitation and winds associated with cumulus convection. As conventional observations are sparse at the convective-scale, radar is an important source of information. The first operational NWP systems kilometre-scale resolution use simple methods to assimilate radar data, such as Latent Heat Nudging (LHN) (Jones and Macpherson 1997; Macpherson 2001; Leuenberger and Rossa 2007; Stephan *et al.* 2008), but research is ongoing using more sophisticated techniques.

In a perfect model context, simulated observations of convective storms have been assimilated using four-dimensional variational assimilation (4DVAR) (Sun and Crook 1997, 1998) and Ensemble Kalman Filter (EnKF) methods (Caya *et al.* 2005). Real observations of an isolated storm have been assimilated using EnKF by Dowell *et al.* (2004, 2010) and Aksoy *et al.* (2009, 2010). These studies represent significant progress towards operational systems, but also indicate the difficulty of inserting observed storm cells into the models and suppressing spurious simulated cells.

In general one would expect problems, since the dynamics of convective clouds at small scales strongly violate key assumptions that the methods depend on. In 4DVAR (Talagrand and Courtier 1983; Bouttier and Rabier 1998; Bouttier and Kelly 2001) and Kalman Filtering (Kalman 1960; Kalman and Bucy 1961), it is assumed

that the background error has a Gaussian distribution and can thus be described by an error covariance matrix. The size of this matrix can then be reduced to a manageable number of degrees of freedom by balance assumptions and observations of correlations, or through representation with a small ensemble in EnKF (Evensen 1994; Houtekamer and Mitchell 1998, 2001). Furthermore it is assumed that the temporal evolution of the error distribution can be represented by tangent linear dynamics or the evolution of a small ensemble of forecasts. Alternatively, more general methods such as the particle filter (Van Leeuwen 2009) do not require these assumptions, although they gain generality at the cost of computational efficiency, potentially requiring prohibitively large ensemble sizes (Snyder *et al.* 2008; Bickel *et al.* 2008; Bengtsson *et al.* 2008). These issues are reviewed by Bocquet *et al.* (2010).

In exploring new data assimilation methods, it is often convenient to complement tests using full atmospheric models with test problems using idealised models. Popular choices include the models of Lorenz (1963, 1995, 2005), which include coupling of fast and slow variables in a low-dimensional dynamical system, or the quasigeostrophic equations (Ehrendorfer and Errico 2008). However both of these systems were designed to represent key processes of synoptic-scale dynamics, rather than to capture the particular characteristics of the convective-scale that make data assimilation difficult. To make progress, it is necessary to consider the nature of the non-Gaussianity and nonlinearity found in the convecting atmosphere.

The essence of this problem can be seen by considering assimilation of radar reflectivity for convective storms. Radar data has a high spatial resolution comparable to the model resolution, but the field of precipitation particles it observes is highly intermittent. Large areas contain no precipitation at all, and strong gradients over distances comparable to a couple of grid points are common. This results in a highly non-Gaussian forecast error distribution, with long tails associated with displacement errors, where a position error of a few grid points produces an order one error in reflectivity. A consequence of this spatial intermittency of precipitation fields is a lack of spatial correlations, that would reduce the effective number of degrees of freedom. This can be contrasted with the situation on synoptic scales, where dynamical balance introduces correlations in space and between model variables. The problem is a version of the "curse of dimensionality" – the number of possible states of the system increases exponentially as the dimensionality of the system grows (Bellman 1957, 1961).

A second issue is that the typical temporal resolution of radar observations of 5-15 minutes is coarse in comparison to the model time step, which is determined by the numerical requirement that the model state does not change too much over the interval, and is typically less than a minute. Furthermore, clouds do not appear on radar until large precipitation particles have had time to form, by which time the dynamical circulation of the cloud is well developed. The result is that there is significant error growth between observation times, and indeed well developed cumulus clouds can appear from one observation time to the next. This lack of temporal correlation between observation times results in an essentially stochastic evolution of the field between observation times, as previously unobserved features suddenly appear as precipitating clouds.

Two potential strategies have been discussed to cope with the lack of spatial and temporal correlations. The first is localization, where the analysis at a given location is only influenced by observations that are close by in space and time (Ott *et al.* 2004). Patil *et al.* (2001) showed that the atmosphere often has a local low dimensionality and therefore localization reduces a high dimensional problem to a set of problems of lower dimension. The second strategy is observation averaging. By averaging the observations over a region in space to create a so-called super-observation (Alpert and Kumar 2007; Zhang *et al.* 2009), the intermittency is reduced. Upscaling the observations not only reduces the effective dimensionality of the system by introducing spatial correlations, it also produces more smoothly varying fields leading to better (more Gaussian) error statistics. The cost of this improvement is that the observations lose detail and may no longer resolve individual convective cells, so that even an analysis that "perfectly" matches the averaged observations is not a perfect analysis when considered at full resolution.

The purpose of this paper is to introduce a minimal model that represents these key features of spatial intermittency and stochastic time evolution. This will provide a simple testbed to examine the performance of various data assimilation methods. The model can be regarded as a minimal version of the stochastic cumulus parametrisation scheme of Plant and Craig (2008), which is based on a statistical mechanics theory of convective fluctuations (Craig and Cohen 2006). The convecting atmosphere is represented by a stochastic birth-death process in space, where cumulus clouds appear at random locations with a certain triggering frequency, and existing clouds disappear with a certain frequency. The result is a field of randomly located clouds (a spatial Poisson process) that have a random lifetime, but with the average density of clouds in space and the average cloud lifetime determined by the birth and death rates. This highly simplified model does not attempt to provide a detailed description of the processes responsible for triggering convective clouds, nor to represent the coupling of convection to the larger-scale flow.

Two data assimilation methods will be applied to this simple model, in basic and localized forms, and with averaged observations. These are the Ensemble Transform Kalman Filter (ETKF) of Bishop and Toth (1999) and its local version as described by Hunt *et al.* (2007), and the Sequential Importance Resampling (SIR) particle filter (Van Leeuwen 2009) and its local version. These two methods were chosen because they have a very different theoretical basis and are likely to show different behaviours. The behaviour of the SIR filter should be easy to anticipate since it is expected respond directly to the effective dimensionality of the system, while the ETKF is being applied well outside of its regime of validity and may not work at all.

The organisation of the paper is as follows. First the simple model is introduced, along with the implementation details of the two data assimilation algorithms. The ability of the basic schemes to converge to the correct state for stationary and time-varying cloud fields is then examined in detail for a representative ensemble size. The dependence on ensemble size is then considered, followed by the impact of localization and averaging.

2. Methods

2.1. The stochastic convection model

A simple stochastic model is used to produce a changing number of clouds at a set of n grid points. At each grid point we define an integer number of clouds present. The convective dynamics is specified as a birth-death process, with, at every time step, a given probability, λ, of a cloud being initiated at each grid point, and a given probability, μ, of each existing cloud disappearing. The probabilities λ and μ are chosen to give a desired mean cloud half life (*hl*) and an average density of clouds per grid point (ρ). In this initial study we assume that the grid points are arranged on a one-dimensional line, but since cloud positions are uncorrelated we could equally have arranged the grid points in a two-dimensional array.

One realization of this model is integrated as a truth simulation, and an ensemble of k simulations is used for data assimilation. The initial distribution of the truth and the ensemble members is given by drawing at every grid point from a Poisson distribution with the average cloud density ρ. This is a perfect model scenario in the statistical sense that all ensemble members are governed by the same rules as the truth run, but since the random numbers are different in each member, model unpredictability plays a key role. At each time step, the stochastic birth-death process is updated, and observations are taken.

For the experiments in this paper, the number of grid points is fixed at $n = 100$, and the birth and death probabilities are chosen to give a mean cloud density of $\rho = 0.1$ clouds per grid point, thus providing a realistic degree of intermittency. The birth-death parameters are also chosen to give an average cloud lifetime of *hl* = 30 time steps, or in some experiments *hl* = 3000 time steps, as discussed below.

Each time step of the model corresponds to an observation time. An observation is the complete state of the truth run at that instant, i.e. the number of clouds at each grid point, with no added error. This is motivated by the characteristics of network radar observations that provide a spatially complete view of precipitation, but relatively little information about the dynamical variables associated with its evolution. Indeed, for the stochastic dynamics used here, the observations contain no information about which clouds will die out in future or where new clouds will later be initiated. If radar observations were available every 5 minutes, the mean cloud lifetime of *hl* = 30 steps would correspond to 2.5 hours, making it possible, in principle, for the data assimilation to lock on to an observed cloud before it randomly disappears. For comparison, experiments were also done with a mean cloud lifetime of *hl* = 3000 observation times, which will be referred to as a stationary cloud field since the mean lifetime is much longer than the duration of the experiments. For the runs with observation averaging, the observations are computed as the total number of clouds in non-overlapping subregions of size 10 grid points. Synthetic observations are computed from the ensemble members in the same way.

2.2. Sequential Importance Resampling Filter

The first data assimilation method to be used is a Sequential Importance Resampling (SIR) particle filter, as representative of methods that are, in principle, appropriate for very nonlinear, non-Gaussian problems.

The implementation of the SIR filter follows Van Leeuwen (2009), and in particular the five steps listed in section 3a of that paper, with probabilistic resampling. In the analysis step at time t, each ensemble member k is assigned a weight, $w(t, k)$, according to

$$w(t, k) = w(t - 1, k) \exp[-e_{rms}(k)/\sigma],$$

where $e_{rms}(k)$ is the root mean square difference between the ensemble member and the truth run, and $\sigma = 0.05$ is a memory timescale for the weights. The weights are then normalized so that the sum over all ensemble members is equal to one.

The new analysis ensemble is then formed by resampling. New ensemble members are chosen by randomly drawing from the old ensemble, with each member given a probability according to its weight. If one member has a sufficiently high weight, it is possible that all other members will be replaced by its copies. To maintain diversity in the ensemble all members are then perturbed with an additive noise of the form $a \cdot \epsilon \cdot e_{rms}(k)$, where a is an amplitude factor set to 0.1 for the global particle filter and 0.25 for the local version, ϵ is a random number drawn from a uniform distribution between -0.5 and 0.5. The local version of the filter is obtained by dividing the domain into equally sized subregions, and performing the analysis and resampling steps described above on each subregion individually.

2.3. Ensemble Transform Kalman Filter

The second data assimilation method is an Ensemble Transform Kalman Filter (ETKF), as an example of methods that depend on linear and Gaussian assumptions. The ETKF and its local version (LETKF) are implemented as described by Hunt *et al.* (2007), but with one additional step required by the stochastic dynamics. The LETKF decomposes the initial ensemble into a mean and deviations. To construct a new analysis ensemble, an updated ensemble mean is produced, and the updated ensemble members are constructed by adding linear combinations of the deviations. The cloud number acquired in this way will contain non-integer values which must be converted to integers to produce a valid model state. This is done by treating the non-integer part of the cloud number as a probability for a cloud being present, and randomly choosing to assign a cloud or not according to this probability. No covariance inflation factor is used, except for the simulations with observation averaging, where it was found useful to use a deflation of 0.7 to avoid an excess of spread due to the probabilistic conversion to integer cloud numbers.

In the local version of the filter, the size of the local region used is one grid point, as in Hunt *et al.* (2007), but only one observation is used for each region. This is contrary to the recommendation of Hunt *et al.* (2007), who suggest using observations from a region centred on the grid point being updated in order to ensure that the increments at neighbouring grid points vary smoothly. However, concerns about smoothness do not arise for the stochastic birth-death process used here, since it produces fields that are uncorrelated between grid points.

Since we seek to investigate generic behaviours of the two data assimilation methods, rather than make

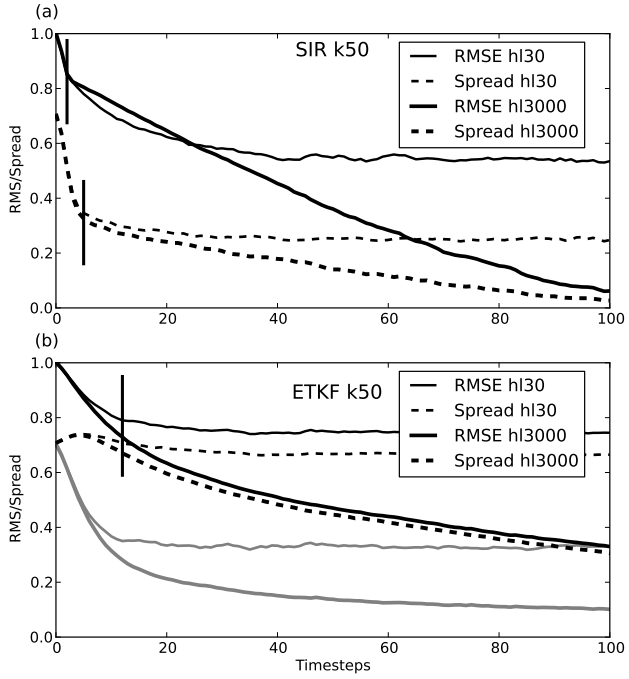

Figure 1. Root mean squared error (solid lines) and ensemble spread (dashed lines) as function of time, for an ensemble size of 50, for (a) SIR and (b) ETKF with half-lives of 3000 (thick) and 30 (thin) time steps. Vertical lines indicate regime boundaries (see text). Grey lines in (b) show RMS error of the ensemble mean.

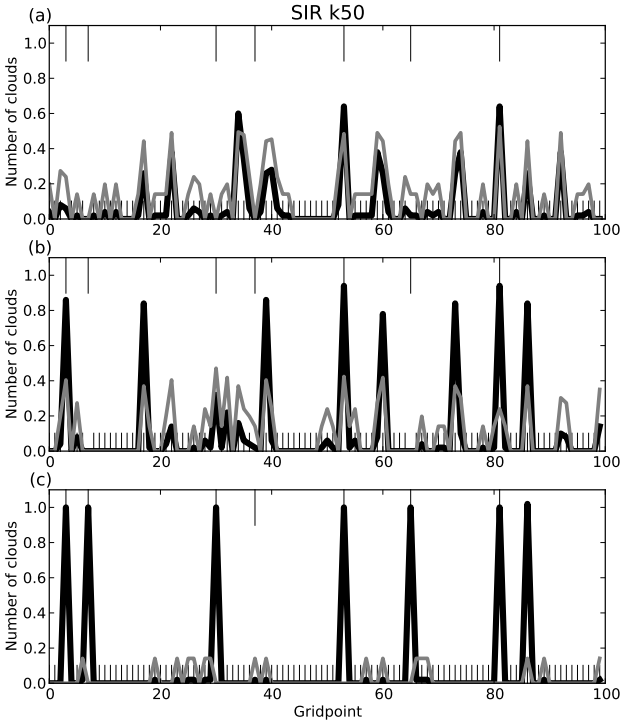

Figure 2. Ensemble mean cloud number (black line) and spread (grey line) for a sample run of the SIR filter with stationary cloud field and 50 ensemble members at time step (a) 3, (b) 7, and (c) 45. Thin vertical lines indicate locations of clouds in the observations.

judgements that one is better than the other, no attempt has been made to tune parameters or otherwise optimize the two schemes.

3. Results

3.1. Convergence for stationary and time-varying cloud fields

We first consider the ability of the two assimilation schemes, in their basic forms, to converge to the observed state. A representative ensemble size of 50 members is used. The RMS error of the individual ensemble members is computed at every timestep. To reduce the noise level in the figures, the errors are averaged over 100 repetitions of the experiment with different realisations of the stochastic process (400 repetitions in the case of the SIR hl30 experiment). The average error is then scaled so that a value of one corresponds to the RMS average difference between two randomly chosen realisations of the model state.

The thick solid line in Fig. 1a shows the evolution of the mean error of the SIR filter for the stationary cloud field. The SIR filter converges, although rather slowly, and the decrease in error continues beyond 100 time units, eventually saturating at a value close to zero. The ensemble spread (thick dashed line) shows an initial rapid decrease to a fixed fraction of the error, which is maintained through the rest of the experiment.

One can identify three phases in this process, separated by the vertical lines in Fig. 1a. In the first stage, resampling removes members with no correct clouds, replacing them with perturbed copies of members with a correct cloud (with the parameters used here it is rare to obtain a member with more than one correct cloud in the initial ensemble). By the end of the first phase, the ensemble consists of descendents of only a few members of the initial ensemble. An example realisation at the end of the first phase is shown in Fig. 2a, where a set of 3 clouds (two correct) is seen to be present in about 60% of the members, showing their common parentage. The average error and spread decrease rapidly during this phase as the members with no correct cloud are eliminated.

The second stage in Fig. 1a is characterised by a slower rate of decrease in error, but a continuing rapid decrease in spread. Since the correct clouds are part of the subset common to most members, differences in error between the members are determined by the number of incorrect clouds. Resampling removes the members with the largest error, which are those with the largest number of clouds. The number of gridpoints without a cloud in any member increases, corresponding to a rapid drop in spread. By the end of this phase, the filter has essentially collapsed, with spread only being maintained by the stochastic perturbations introduced at each resampling stage. The example state shown in Fig. 2b shows that nearly all members have a common subset of eight clouds (three in correct locations), and at many locations there are no clouds in any ensemble member. The mean error decreases slowly during this phase, since occasionally a perturbation during the resampling will change the subset of clouds common to all members, producing a member that has an additional correct cloud, or one that does not have an incorrect cloud that all other members have. Descendents of this better member take over the ensemble within a few resampling steps, but since such events are rare, the average error decreases slowly.

In the third phase, the mean error and spread both decrease slowly (Fig. 1a). The rate of creation of new clouds by resampling perturbations approximately balances

the tendency to reduce the number of clouds by selectively removing the members with the most clouds, and further changes in error and spread are associated only with occasional changes to the common subset of clouds. Each time a new correct cloud is produced by the resampling perturbations it rapidly spreads to the rest of the ensemble, but since the number of possible locations is large, many timesteps are required until all the correct cloud locations are found. In the example in Fig. 2c, most members have the correct solution at most locations, with one incorrect cloud and occasional additional clouds at random locations.

The rapid collapse of the SIR filter to include a common subset of clouds in all members is as expected for small ensemble sizes. Subsequently, the filter behaves as a Markov Chain Monte Carlo simulation, randomly exploring the state space, but retaining correct information from previous timesteps. The effectiveness of this behaviour depends crucially on the strategy for perturbing duplicate members introduced during resampling. In this simple model, the resampling perturbations have been chosen consistently with the stochastic model dynamics, and the filter continues to converge, albeit slowly.

Figure 1a shows that the time to converge for a stationary cloud field is long compared to the physically motivated mean cloud lifetime of up to 30 time steps. This suggests that, for this ensemble size, the SIR filter will not be able to track changes in a time-varying cloud field. As seen in the thin line in Fig. 1a, the error initially decays at a rate similar to that for stationary clouds, but reaches a minimum value of about 55% of the error of a random field. The initial behaviour of the filter is similar, with the initial random noise disappearing within a few time steps and a common subset of clouds appearing in the majority of ensemble members. However, the evolution of the common subset is too slow to track changes in the evolving observed state and the error remains large.

Fig. 1b shows the corresponding results for the ETKF. For the stationary cloud field, the mean error eventually approaches a small value (well beyond the 100 timesteps shown in the figure), with spread very similar to the error. Also shown is the error in the ensemble mean, which should represent the ”best estimate” of the observed state. Two stages are visible in the figure, distinguished mainly by the behaviour of the ensemble spread. While the error decreases continuously, the spread remains roughly constant for approximately 15 timesteps, then decreases together with the error.

As shown for an example realisation in Fig. 3b, by the end of the first phase clouds appear at the correct locations in the majority of the ensemble members. However, the variability at other locations, associated with incorrect clouds in random members, remains largely unchanged. Since the incorrect clouds are at different locations in the different ensemble members (in contrast to the common subset in the SIR filter), the error of the ensemble mean (also shown in Fig. 1b) is smaller than the mean of the errors of the individual ensemble members. This reflects the well-known reduction of RMS error for smoother fields, and does not imply that the ensemble mean is a satisfactory estimate of the observed state since the presence of low rainrates everywhere in the domain is not consistent with the dynamics of the physical model.

The gradual decrease of spread and error in the second phase is associated with the disappearance of incorrect

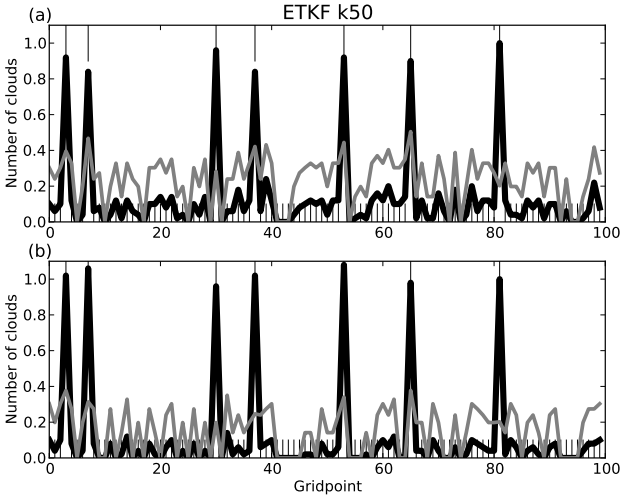

Figure 3. As Fig. 2, but for the ETKF at timestep (a) 15, and (b) 45.

clouds from more and more members. This process is much slower than the introduction of clouds at the correct locations. The ETKF analysis step takes each ensemble member and adds positive and negative perturbations from other members. Since clouds occupy a small fraction of gridpoints in each member, there is more chance of introducing a new incorrect cloud than to remove an existing one. The slow convergence is thus a direct result of the non-Gaussianity of the errors once the locations of the observed clouds have been captured by the ensemble.

For time-varying cloud fields (thin line in Fig. 1b), a significant diversity is retained in the ensemble, but the error saturates at a high level since the filter is not able to remove the noise within the half-life of the clouds.

Although both the SIR filter and ETKF have large errors with a time-varying cloud field, the nature of the errors is quite different. The error in the SIR filter comes primarily from wrongly positioned clouds that are present in almost all ensemble members, while the error in the ETKF is associated with a high level of background noise. This suggests that the optimal method for producing probabilistic predictions from the two ensembles will be different.

3.2. *Ensemble size*

With an ensemble size of 50, neither data assimilation method is able to converge to the time-varying cloud field with any degree of accuracy. To illustrate how the results change with ensemble size, a *final error* was computed for each experiment. This was estimated as the error after 500 time steps for experiments with a stationary cloud field and 100 time steps for the 30 time step lifetime cases, since changes after this time were found to be negligible.

In Fig. 4a it can be seen that any ensemble size greater than about 10 is sufficient for the SIR filter to converge for stationary clouds, while for the time-varying cloud field the error decreases rapidly with increasing ensemble size up to about 20 members, after which the decrease in error is slow. Except for very small ensemble sizes, the spread is about half the magnitude of the error.

The ETKF collapses (vanishing spread) for small ensembles with a stationary cloud field (Fig 4b). The results improve significantly with ensemble size, up to about 40 members, with final error reaching about 20% of the random

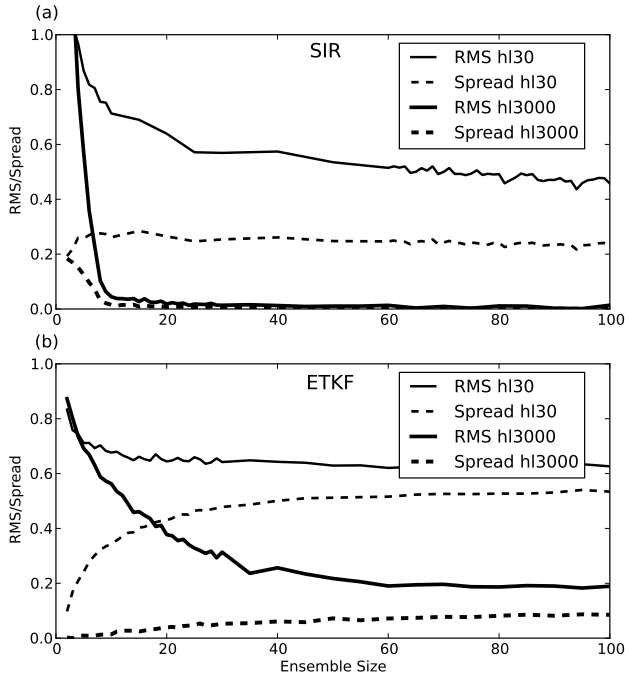

Figure 4. Final error (solid, as defined in text) and ensemble spread (dashed) as function of ensemble size for (a) SIR and (b) ETKF with half-lives of 3000 (thick) and 30 (thin) time steps.

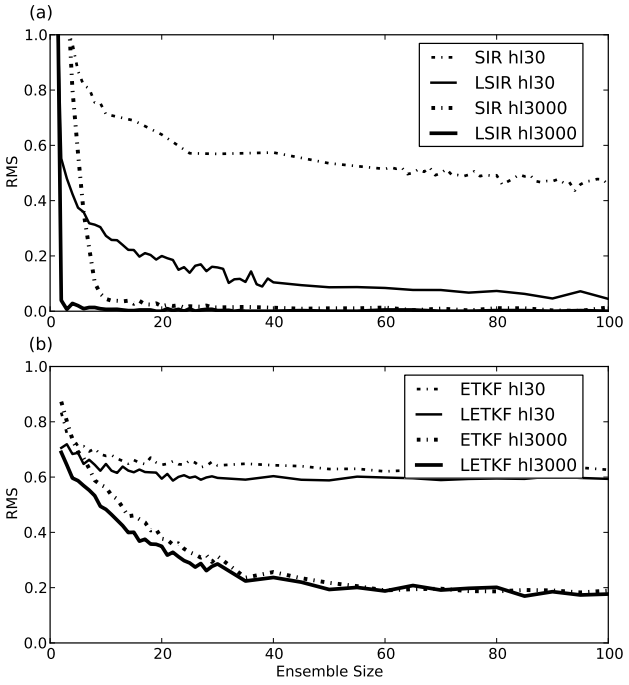

Figure 5. Final error as function of ensemble size for (a) SIR and (b) ETKF, with half-life 30 (thin dash-dotted) and localisation (thin) and half-life 3000 (thick dash-dotted) and localisation (thick)

value and ensemble spread about half the final error. For the time-varying cloud field there is almost no improvement with ensemble size, and even for an ensemble size of 100, the final error is only about 5% better than for an ensemble size of 15.

The behaviour of ensemble spread relative to error shown shown in Fig. 4 is also found in the subsequent experiments, therefore spread will not be plotted on the remaining figures.

3.3. *Localization*

Assimilating data in local regions independently can drastically reduce the number of degrees of freedom in the system, potentially improving the performance of small ensembles. As can be seen in Fig. 5a, it has a major effect on the performance of the SIR filter, leading to convergence with even smaller ensemble sizes than achieved by the non-local filter for the stationary cloud field. A major reduction in final error is found for the time-varying cloud field, with errors less than 20% for ensemble sizes larger than 20. This result is not surprising: on average, the observations have 10 clouds scattered over 100 possible locations, for a total of 100^{10} possible states. This is vastly larger than the ensemble sizes used here. localization decomposes the domain into 100 subdomains, each with 2 likely states (cloud or no-cloud), for a total of about $100 \cdot 2$ possibilities. An ensemble of a few tens of members can sample this space within the cloud half-life of 30 timesteps.

On the other hand, the effects for the LETKF are modest, with small improvements in final error, particularly for smaller ensemble sizes. The primary source of error for the ETKF is the continual creation of clouds at incorrect locations by the assimilation increments, which is unaffected by localization.

3.4. *Averaging*

In contrast to localization, observation averaging should have the effect of making the distribution of the errors more Gaussian, potentially leading to better results for the ETKF. Figure 6 shows the final errors for experiments with observation averaging over 10 grid point regions. For this figure RMS errors have been computed from averaged observations and model states, and normalized by the difference between two random states in this measure. An error of zero would thus imply that the number of clouds within each 10 grid point region was correct, but not necessarily their locations. The errors thus reflect the ability of the methods to solve the simpler problem of producing the correct density of convective clouds over subregions, and are not directly comparable to the errors shown on the previous figures.

With averaged observations the SIR filter converges even with small ensemble sizes (Fig. 6a). As with localization, the problem is decomposed into a set of smaller problems that are solved independently. Unlike localization, averaging changes the statistical character of the smaller problems, but the reduction in dimensionality is similar.

For a stationary cloud field, the AETKF (Fig. 6b) shows a large reduction in the ensemble size required to obtain small errors in the region of small ensembles (10-20 members). The increase in error for larger ensemble sizes is surprising, but preliminary experiments (not shown) suggest that this could be corrected by increasing the covariance deflation factor with increasing ensemble size to produce errors at least as small as those for smaller ensembles. Even for a time-varying cloud field the performance of the ETKF is significantly improved by averaging, although for the relatively small averaging region used here the errors remain large.

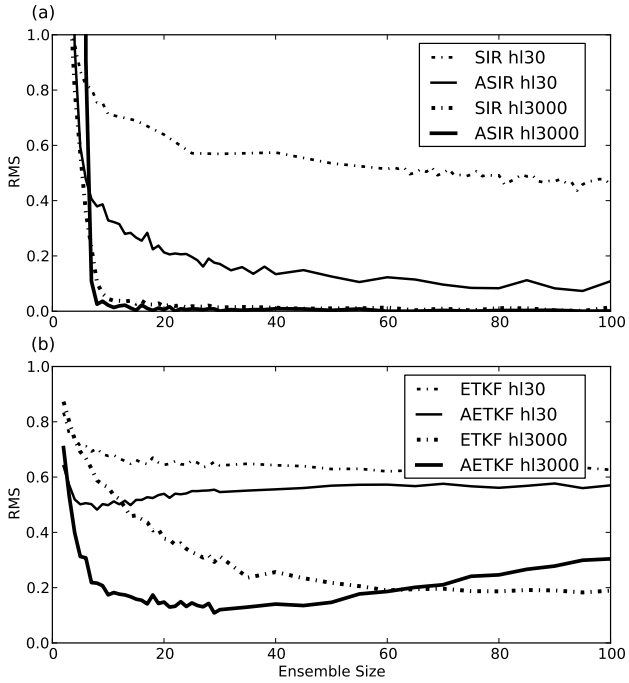

Figure 6. As Fig. 5 but with dash-dotted lines indicating experiments with observation averaging.

4. Conclusions

The aim of this work was to test convective-scale data assimilation algorithms using a simple model based on a stochastic birth-death process. The stochastic dynamics models the extreme nonlinearity of the convecting atmosphere, where a storm can appear from one observation time to the next. The spatial Poisson distribution in the simple model represents the intermittency and lack of correlation characteristic of remote sensing observations such as Radar. Since the model does not include any dynamical balances that might or might not be present at the convective-scale, it constitutes an extreme test for data assimilation algorithms, and both the SIR filter and ETKF fail to produce good results for realistic parameter values of cloud density and lifetime. The performance could undoubtedly be improved somewhat by optimising parameters such as the covariance inflation or the resampling probability in the particle filter, but the results would be unlikely to change qualitatively.

The SIR filter rapidly collapses to a state in which variance is only maintained by the perturbations associated with resampling after members are eliminated. Interestingly, the filter can eventually converge to the observed state, since any correct cloud locations found by the random perturbations are retained in the ensemble. Since the rate of convergence is controlled by the resampling perturbations, rather than the importance weighting, strategies like an improved proposal distribution (Van Leeuwen 2009) may have the greatest potential to improve filter performance. In any case the true statistical properties of the collapsed ensemble need to be taken into account in generating probabilitstic forecast products, since the weights of the ensemble members produced by the filter provide little information.

Localization and observation averaging both produced dramatic improvements in the performance of the SIR filter since they both drastically reduce the dimensionality of the space that must be explored by the resampling perturbations. In more realistic models, however, these methods might cause problems by violating dynamical balances that couple different spatial regions.

The ETKF collapses for small ensemble sizes, but otherwise captures the correct cloud locations quickly. However, the ensemble is plagued by large numbers of incorrect clouds that contaminate the ensemble mean. This occurs because negative clouds are not possible, so that the nonlinear dynamics rectify the analysis increments, producing a non-Gaussian distribution of variability in the ensemble. Averaging of observations has the potential to correct this problem, since the Poisson distribution will converge to Gaussian as the averaging region becomes large enough to contain many clouds. Indeed, significant improvement with averaging was found, although for the relatively small averaging region used here, the errors remain large. localization, on the other hand, seems to have no benefit for the ETKF in this environment.

While the stochastic test problem proposed here is highly idealised, the results point to specific problems that are likely to limit the performance of particular data assimilation methods for cumulus convection, and provide indications regarding how the methods can be improved. This could be a starting point for hierarchy of models, where a stochastic process representing convection is introduced into simple dynamical models (e.g. shallow water, or quasi-geostrophic), and the convection is coupled to the large-scale dynamics by a simple closure assumption, where the rate parameter depends on the large-scale fields.

References

Aksoy A, Dowell DC, Snyder C. 2009. A multicase comparative assessment of the ensemble kalman filter for assimilation of radar observations. part I: Storm-scale analyses. *Mon. Weather Rev.* **137**: 1805–1824.

Aksoy A, Dowell DC, Snyder C. 2010. A multicase comparative assessment of the ensemble kalman filter for assimilation of radar observations. part II: Short-range ensemble forecasts. *Mon. Weather Rev.* **138**: 1273–1292.

Alpert JC, Kumar VK. 2007. Radial Wind Super-Obs from the WSR-88D Radars in the NCEP Operational Assimilation System. *Mon. Weather Rev.* **135**: 1090–1109.

Bellman R. 1957. *Dynamic programming*. Rand Corporation research study, Princeton University Press, ISBN 9780691079516.

Bellman R. 1961. *Adaptive control processes: a guided tour*. A Rand Corporation Research Study Series, Princeton University Press.

Bengtsson T, Bickel P, Li B. 2008. Curse-of-dimensionality revisited: Collapse of the particle lter in very large scale systems. *IMS Collections* **2**: 316–334.

Bickel P, Li B, Bengtsson T. 2008. Sharp failure rates for the bootstrap particle filter in high dimensions. pushing the limits of contemporary statistics. *Contributions in Honor of Jayanta K. Ghosh* **3**: 318–329.

Bishop C, Toth Z. 1999. Adaptive Sampling with the Ensemble Transform Kalman Filter. Part I: Theoretical Aspects. *Mon. Weather Rev.* **56**: 1748–1765.

Bocquet M, Pires CA, Wu L. 2010. Beyond Gaussian Statistical Modeling in Geophysical Data Assimilation. *Mon. Weather Rev.* **138**(8): 2997–3023.

Bouttier F, Kelly G. 2001. Observing-system experiments in the ECMWF 4D-Var data assimilation system. *Q. J. R. Meteorol. Soc.* **127**: 1469–1488.

Bouttier F, Rabier F. 1998. The operational implementation of 4D-Var. *ECMWF Newsletter* **78 (Summer 1997/1998)**: 2–5. Available online at http://ecmwf.int/publications/newsletters/pdf/78.pdf.

Caya A, Sun J, Snyder C. 2005. A comparison between the 4dvar and the ensemble kalman filter techniques for radar data assimilation. *Mon. Weather Rev.* **133**: 3081–3094.

Craig GC, Cohen BG. 2006. Fluctuations in an equilibrium convective ensemble. Part I: Theoretical formulation. *J. Atmos. Sci.* **63**: 1996–2004.

Dowell D, Wicker L, Snyder C. 2010. Ensemble Kalman Filter Assimilation of Radar Observations of the 8 May 2003 Oklahoma City Supercell: Influences of Reflectivity Observations on Storm-Scale Analyses. *Mon. Wea. Rev., accepted* .

Dowell DC, Zhang F, Wicker LJ, Snyder C, Crook NA. 2004. Wind and Temperature Retrievals in the 17 May 1981 Arcadia, Oklahoma, Supercell: Ensemble Kalman Filter Experiments. *Mon. Weather Rev.* **132**: 1982–2005.

Ehrendorfer M, Errico R. 2008. An atmospheric model of intermediate complexity for data assimilation studies. *Q. J. R. Meteorol. Soc.* **134**: 1717–1732.

Evensen G. 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.* **99**: 10 143–10 162.

Houtekamer PL, Mitchell HL. 1998. Data assimilation using an ensemble kalman filter technique. *Mon. Weather Rev.* **126**: 796–811.

Houtekamer PL, Mitchell HL. 2001. A sequential ensemble kalman filter for atmospheric data assimilation. *Mon. Weather Rev.* **129**: 123–137.

Hunt BR, Kostelich EJ, Szunyogh I. 2007. Efficient data assimilation for spatiotemporal chaos: A local ensemble transform kalman filter. *Physica D* **230**: 112–126.

Jones C, Macpherson B. 1997. A latent heat nudging scheme for the assimilation of precipitation data into an operational mesoscale model. *Met. Apps* **4**: 269–277.

Kalman R. 1960. A new approach to linear filtering and prediction problems. *Transactions of the ASME–Journal of Basic Engineering* **82**: 35–45.

Kalman R, Bucy R. 1961. New results in linear filtering and prediction theory. *Transactions of the ASME–Journal of Basic Engineering* **83**: 95–107.

Leuenberger D, Rossa A. 2007. Revisiting the latent heat nudging scheme for the rainfall assimilation of a simulated convective storm. *Meteorol. Atmos. Phys.* **98**: 195–215.

Lorenz E. 1963. Deterministic nonperiodic flow. *J. Atmos. Sci.* **20**: 130–141.

Lorenz E. 1995. Predictability: A problem partly solved. *In Proc. Sem. Predictability* **1**: 1–18.

Lorenz E. 2005. Designing chaotic models. *J. Atmos. Sci.* **62**: 1574–1587.

Macpherson B. 2001. Operational experience with assimilation of rainfall data in the Met Office Mesoscale model. *Meteorol. Atmos. Phys.* **76**: 3–8.

Ott E, Hunt B, Szunyogh I, Zimin A, Kostelich E, Corazza M, Kalnay E, Patil D, Yorke J. 2004. A local ensemble kalman filter for atmospheric data assimilation. *Tellus A* **56**: 415–428.

Patil DJ, Hunt BR, Kalnay E, Yorke JA, Ott E. 2001. Local low dimensionality of atmospheric dynamics. *Phys. Rev. Lett.* **86**: 5878–5881.

Plant RS, Craig GC. 2008. A stochastic parameterization for deep convection based on equilibrium statistics. *J. Atmos. Sci.* **65**: 87–105.

Snyder C, Bengtsson T, Bickel P, Anderson J. 2008. Obstacles to high-dimensional particle filtering. *Mon. Weather Rev.* **136**: 4629–4640.

Stephan K, Klink S, Schraff C. 2008. Assimilation of radar-derived rain rates into the convective-scale model COSMO-DE at DWD. *Q. J. R. Meteorol. Soc.* **134**: 1315–1326.

Sun J, Crook N. 1997. Dynamical and microphysical retrieval from doppler radar observations using a cloud model and its adjoint. Part I: Model development and simulated data experiments. *J. Atmos. Sci.* **54**: 1642–1661.

Sun J, Crook N. 1998. Dynamical and microphysical retrieval from doppler radar observations using a cloud model and its adjoint. Part II: Retrieval experiments of an observed Florida convective storm. *J. Atmos. Sci.* **55**: 835–852.

Talagrand O, Courtier P. 1983. Variational assimilation of meteorological observations with the adjoint vorticity equation. I: Theory. *Q. J. R. Meteorol. Soc.* **113**: 1311–1328.

Van Leeuwen PJ. 2009. Particle filtering in geophysical systems. *Mon. Weather Rev.* **137**: 4089–4114.

Zhang F, Weng Y, Sippel JA, Meng Z, Bishop CH. 2009. Cloud-resolving hurricane initialization and prediction through assimilation of doppler radar observations with an ensemble kalman filter. *Mon. Weather Rev.* **137**: 2105–2125.