Is the Local Ensemble Transform Kalman Filter suitable for operational data assimilation?

Mikhail Tsyrulnikov

Hydrometcentre of Russia, Moscow, Russia

Abstract

Principal approximations inherent in the Local Ensemble Transform Kalman Filter (LETKF) are examined, and conclusions about the related approximation errors are drawn. First, the major difficulty in implementing the LETKF technique in the operational context is its lack of efficiency in assimilation of non-local, primarily satellite, observations. The cause of this difficulty is the restriction of the analysis increment to low-dimensional ensemble space, on the one hand, and large spatial supports of satellite data, on the other hand. Second, making analysis in small local boxes gives rise to small-scale noise due to local data selection—as in Optimum Interpolation. Third, an attempt to account for realistically correlated observational errors makes LETKF and other ensemble-space and model-space Ensemble Kalman Filters (EnKF) computationally inefficient. Published experimental results are discussed in the light of the theoretical and experimental findings made in this study.

1 Introduction

In the recent years, a kind of Ensemble Kalman Filter (EnKF), the Local Ensemble (Transform) Kalman Filter (LEKF/LETKF) has become popular. Numerous papers, starting from (Ott *et al.* 2004), have explored capabilities of LETKF to deal with different data assimilation issues, including real-data assimilation (Hunt *et al.* 2007, Szunyogh *et al.* 2007b, Miyoshi and Yamane 2007, Bonavita *et al.* 2008), satellite data assimilation (Szunyogh 2007a, Fertig *et al.* 2007), four-dimensional assimilation (4D-LETKF, Hunt *et al.* 2004, Kalnay *et al.* 2007, Harlim and Hunt 2007b), non-Gaussian background-error distributions (Harlim and Hunt 2007a), forecast bias (Baek *et al.* 2006) and others. The LETKF technique is simple, computationally efficient, and allows for flow-dependent background-error covariances. Different extensions to the basic LETKF formulation have been proposed. Some of the above papers report on promising results in near-operational setting. But still there are no operational implementations.

The questions arise: Is the LETKF technique a real alternative to variational methods (3D-Var and 4D-Var) as a means of operational (global and limited-area) data assimilation? What are the conditions under which LETKF can be considered to be really competitive?

To answer these questions, we consider below the details of the LETKF algorithm and draw some conclusions about its capabilities, especially in the areas of satellite and meso-scale data assimilation. The focus is on the *analysis* step, where, as we will see below, the major limitations of the technique lie.

2 LETKF. Analysis step: brief description

In this section, we derive the LETKF analysis equations, emphasizing approximations (and thus imperfections) inherent in this technique.

At each analysis step, we have: the deterministic forecast $\mathbf{x}^f$, an ensemble of perturbed forecasts $\{\mathbf{x}_i^f\}$ ($i = 1, \ldots, n_e$, n_e is the ensemble size), and observations $\mathbf{x}^o$. The ensemble mean is usually used as a background field $\mathbf{x}^b$ in the analysis. The deviation fields $\mathbf{x}_i^f - \mathbf{x}^b$ are deemed to be independent samples from the unknown true probability distribution of the (minus) *background error* $\mathbf{x}^b - \mathbf{x}$, where $\mathbf{x}$ is the true field to be estimated. Here, we assume that $\mathbf{x}_i^f$ are already undergone a procedure like 'variance inflation', which attempts to take into account errors due to forecast model imperfections (model errors).

The distinctive features of the LETKF analysis are:

1. At each analysis grid point, the analysis is performed *locally*: using only nearby observations from a box (cylinder, ellipsoid, ...) surrounding the grid point—as in Optimum Interpolation (OI) (Gandin 1963, Lorenc 1981).
2. The analysis is performed in *ensemble space*: within each local box, the analysis increment belongs to the subspace spanned by ensemble deviations, $\mathbf{x}_i^f - \mathbf{x}^b$.

The details of the technique follow.

2.1 Ensemble space

We start by computing the (global) *ensemble vectors*,

$$\mathbf{e}_i := \frac{1}{\sqrt{n_e - 1}}(\mathbf{x}_i^f - \mathbf{x}^b), \tag{1}$$

where the sign := means 'equal by definition'. The normalization by $\sqrt{n_e - 1}$ is introduced into Eq.(1) only with the intention to simplify the analysis-algorithm formulae.

Next, we proceed by forming (theoretically) the ensemble matrix

$$\mathbf{E} = (\mathbf{e}_1 \cdots \mathbf{e}_{n_e}). \tag{2}$$

Note that, as it follows from Eqs.(1) and (2), the ensemble sample covariance matrix $\mathbf{B}^e$ is

$$\mathbf{B}^e = \mathbf{E} \cdot \mathbf{E}^T. \tag{3}$$

Now, we introduce *ensemble space* $\mathcal{E} := Span\{\mathbf{e}_1, \cdots, \mathbf{e}_{n_e}\}$, so that any $\mathbf{z} \in \mathcal{E}$ can be expanded in the ensemble vectors:

$$\mathbf{z} = \sum_{i=1}^{n_e} \tilde{z}_i \mathbf{e}_i \equiv \mathbf{E}\tilde{\mathbf{z}} \tag{4}$$

Here are below, by tilde, we denote coordinates of a vector in the ensemble 'basis', $\{\mathbf{e}_i\}$. Note that the set of ensemble vectors $\{\mathbf{e}_i\}_{i=1}^{n_e}$ does not constitute a true basis in $\mathcal{E}$ if $\mathbf{x}^b$ is the ensemble mean (as $\{\mathbf{e}_i\}$ sum up to zero and are thus not linearly independent). This implies that the expansion Eq.(4) does exist but can be not unique.

Thus, the control variable we wish to estimate in the analysis is $\tilde{\mathbf{x}}$ such that $\mathbf{x} = \mathbf{E}\tilde{\mathbf{x}}$. The dimensionality of $\tilde{\mathbf{x}}$ is as low as n_e (several tens, in practice).

The great advantage of making the analysis in ensemble space is that the dimensionality of $\mathcal{E}$ can be chosen much less than both the number of local influencing observations and the number of grid points in the local box. As a result of this, the analysis equations can be solved substantially faster than both in observation space and model space. But this acceleration comes at a price, which can be high: (low-dimensional) ensemble space can be too poor to reproduce the background-error spatial variability, resulting in poor analysis. In more detail, this issue is considered in section 4.2.

2.2 The ensemble-space observation model

We start with the conventional global-space observation model:

$$\mathbf{x}^o = \mathcal{H}(\mathbf{x}) + \eta, \tag{5}$$

where $\mathcal{H}$ is the observation operator, $\mathbf{x}$ the global-grid state variable (vector), and η the observation error (which consists of the measurement error and the observation-operator error). We linearize $\mathcal{H}(\mathbf{x})$ around the background $\mathbf{x}^b$:

$$\mathbf{x}^o = \mathcal{H}(\mathbf{x}^b) + \mathbf{H}(\mathbf{x} - \mathbf{x}^b) + \eta + \eta_{lin}, \tag{6}$$

where $\mathbf{H}$ is the tangent linear observation operator and η_{lin} is the error due to truncation of the Taylor series in Eq.(6). To simplify the notation, we turn to *increments*—all denoted by $\mathbf{y}$:

$$\mathbf{y} := \mathbf{x} - \mathbf{x}^b, \tag{7}$$

$$\mathbf{y}^o := \mathbf{x}^o - \mathcal{H}(\mathbf{x}^b). \tag{8}$$

With these increment variables, Eq.(6) writes

$$\mathbf{y}^o = \mathbf{H}\mathbf{y} + \eta + \eta_{lin}. \tag{9}$$

This is the linearized global-space observation-increment model. Now, we have to transform it to ensemble space.

2.2.1 The ensemble truncation (representativeness) error

An observation model in ensemble space relates the state vector in ensemble space, $\tilde{\mathbf{y}}$, to the observation increment $\mathbf{y}^o$. To build this model, we project (with an appropriately selected scalar product) $\mathbf{y}$ onto $\mathcal{E}$, getting

$$\mathbf{y} = \mathbf{y}_e + \mathbf{y}_{res}, \tag{10}$$

where $\mathbf{y}_e \in \mathcal{E}$ and $\mathbf{y}_{res} \perp \mathcal{E}$. Being in $\mathcal{E}$, $\mathbf{y}_e$ can be expanded in the ensemble vectors (see Eq.(4)), $\mathbf{y}_e = \mathbf{E}\tilde{\mathbf{y}}$. The residual in Eq.(10), $\mathbf{y}_{res}$, cannot be represented within ensemble space and, as a result of this, manifests itself as a source of *observation representativeness* error. Indeed, from Eqs.(9) and (10), we obtain

$$\mathbf{y}^o = \tilde{\mathbf{H}}\tilde{\mathbf{y}} + \eta + \eta_{lin} + \eta_{et}, \tag{11}$$

where

$$\tilde{\mathbf{H}} = \mathbf{H} \cdot \mathbf{E} \tag{12}$$

and

$$\eta_{et} = \mathbf{H}\mathbf{y}_{res} \tag{13}$$

is the error due to *ensemble truncation*.

To complete the LETKF ensemble-space observation model, we approximate the tangent-linear observation operator in Eq.(12) as follows.

2.2.2 A finite-difference approximation to H

As it is proposed in the LETKF literature, working in ensemble space enables another simplification of the analysis technique: the tangent linear observation operator (and, in the 4-D case, the tangent linear forecast model) can be usefully approximated by finite differencing. To do so, we write the first term on the r.h.s. of Eq.(11), using Eq.(12), as

$$\tilde{\mathbf{H}}\tilde{\tilde{y}} \equiv \mathbf{H}\mathbf{E}\tilde{\tilde{y}} \equiv \sum \tilde{y}_i \cdot \mathbf{H}\mathbf{e}_i \tag{14}$$

and approximate $\mathbf{H}\mathbf{e}_i$ by $\frac{1}{s}(\mathcal{H}(\mathbf{x}^b + s\mathbf{e}_i) - \mathcal{H}(\mathbf{x}^b))$, where $s := \sqrt{n_e - 1}$ and $\mathbf{x}^b + s\mathbf{e}_i \equiv \mathbf{x}_i^f$, getting

$$\mathcal{H}(\mathbf{x}) - \mathcal{H}(\mathbf{x}^b) = \frac{1}{s}\sum \tilde{y}_i \cdot (\mathcal{H}(\mathbf{x}_i^f) - \mathcal{H}(\mathbf{x}^b)) + \eta_{lin} + \eta'_{fd}. \tag{15}$$

Here, η'_{fd} represents the finite-differencing error.

It can be shown that if $\tilde{\mathbf{y}}$ is multivariate Gaussian with zero mean and $\mathcal{H}$ is only second-order non-linear, then the covariance matrix of the sum $\eta_{lin} + \eta'_{fd}$ is larger than that of η_{lin} (in the sense that difference of the two covariance matrices is a positive-definite matrix). In order to find out how important this loss of accuracy is, we need to know intricate properties of the tensor $\partial^2\mathcal{H}_k/\partial x_i \partial x_j$, which is beyond the scope of this article. For realistic weakly non-linear observation operators the difference is expected to be small, e.g. for linear $\mathcal{H}$, both η_{lin} and η'_{fd} are equal to zero. In addition, Lorenc (2003b) noted that the linearized observation operator is encountered in the analysis equations only combined with error covariance matrices. As the latter are known only approximately, there is little sense in making the linearized observation operator 'very accurate'. So, the finite differencing error is only a minor problem, at least for not too non-linear observation operators.

In Eq.(15), it can be preferential to replace $\mathcal{H}(\mathbf{x}_i^f) - \mathcal{H}(\mathbf{x}^b)$ by $\mathcal{H}(\mathbf{x}_i^f) - \bar{\mathcal{H}}$ with $\bar{\mathcal{H}} = \frac{1}{n_e}\sum_{j=1}^{n_e}\mathcal{H}(\mathbf{x}_j^f)$, as it is done in the LETKF literature. Denoting the error of this approximation by η_{fd}, we finally obtain

$$\mathbf{y}^o = \mathbf{Z}_{glob} \cdot \tilde{\mathbf{y}} + \eta^+ \equiv \sum \tilde{y}_i \cdot \mathbf{z}_i + \eta^+, \tag{16}$$

where $\mathbf{Z}_{glob}$ is the matrix of the same size as $\tilde{\mathbf{H}}$ whose columns are $\mathbf{z}_i := (\mathcal{H}(\mathbf{x}_i^f) - \bar{\mathcal{H}})/s$, and

$$\eta^+ = \eta + \eta_{lin} + \eta_{et} + \eta_{fd} \tag{17}$$

is the total effective observation error.

Thus, the observation model relating the local ensemble-space increment vector $\tilde{\mathbf{y}}$ to the observation increments $\mathbf{y}^o$ is derived—Eq.(16). Note that the corresponding observation error, η^+, has one major additional LETKF specific component, η_{et} (due to finite ensemble size), and one minor additional component, η_{fd} (due to the finite-differencing error). The larger the observation error, the less information it bears, the less accurate the analysis. How important these additional errors are for the performance of LETKF is further discussed in section .

2.3 The analysis equations

Note that all the above equations are written for the *global* state variable and the *global* ensemble vectors. Now, it is time to *localize* the analysis. We achieve this by performing the analysis for each analysis grid point independently and limiting the number of observations

used in each local analysis (only observations in a local box surrounding the analysis grid point are selected).

Specifically, from the global matrix $\mathbf{Z}_{glob}$, we build its local version, $\mathbf{Z}$, by leaving only rows that correspond to the selected observations. We also build the local version of $\mathbf{E}$, $\mathbf{E}_{loc}$, by dropping rows that correspond to grid points not within the support of any of the selected observations. Thus, the number of rows in $\mathbf{E}_{loc}$ equals the number of observed degrees of freedom in the current local analysis. It is important to notice that if non-local satellite observations are used, the $\mathbf{E}_{loc}$ matrix is enlarged due to the grid points that can be outside the selected box but within the supports of these observations. This enlargement is further discussed below in section .

For any analysis grid point, having selected 'influencing' observations, and with the observation model, Eq.(16), in hand, we easily write down the (linear) analysis equation

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{E}_{loc} \cdot \tilde{\mathbf{K}} \cdot \mathbf{y}^o, \tag{18}$$

where $\tilde{\mathbf{K}}$ is the gain matrix in ensemble space,

$$\tilde{\mathbf{K}} = (\mathbf{I} + \mathbf{Z}^T\mathbf{R}^{-1}\mathbf{Z})^{-1}\mathbf{Z}^T\mathbf{R}^{-1}, \tag{19}$$

$\mathbf{R} = \mathbb{E}\eta^+(\eta^+)^T$ is the (local) observation-error covariance matrix, $\mathbb{E}$ denotes the mathematical expectation, and the equality $\tilde{\mathbf{B}} \equiv \mathbb{E}\tilde{\mathbf{y}} \cdot \tilde{\mathbf{y}}^T = \mathbf{I}$ is used following, e.g. (Hunt *et al.* 2007). Note that if $\mathbf{R}$ is diagonal, then the only matrix that requires non-trivial inversion in Eq.(19) is $\mathbf{I}+\mathbf{Z}^T\mathbf{R}^{-1}\mathbf{Z}$, its order being as small as n_e. This results in the very fast numerical algorithm and, essentially, makes it possible to perform multiple analyses at all grid point separately.

An additional 'observation localization' is sometimes applied: the $\mathbf{R}^{-1}$ entries are multiplied by a monotonically decaying function of the distance between the center of the box and the particular observation (within the box). This *ad hoc* device acts to reduce the influence of distant observations, which, first, diminishes the detrimental impact of sampling noise in the background error covariances at large distances, and second, makes the transition from one analysis grid point to the adjacent more smooth.

3. Advantages of the LETKF analysis technique

- Simplicity

  The LETKF analysis algorithm does not require any background-error covariance *model*, which is the major simplification as compared to 3D-Var or 4D-Var.

- Computational efficiency

  First, the LETKF analysis algorithm is fast because of working in low-dimensional ensemble space. Second, as in local OI, LETKF analysis computations at all grid points are completely mutually independent, which enables very efficient parallelization in implementations on present and future massively parallel computers.

- Flow-dependent covariances

  Ensemble statistics allows to easily introduce flow dependence in the background constraint, which is a very desirable feature.

- 4D-LETKF

  The technique is easily and naturally extended to the 4-D case.

4. Weaknesses of LETKF

The above mentioned strengths of the LETKF analysis algorithm are detailed in the LETKF papers, so let us concentrate here on LETKF problems.

4.1 With non-local satellite observations, the effective box size becomes *large*

As it follows from the above description of the LETKF analysis and as it was proposed by Fertig *et al.* (2007), if the support of an observation does not lie within a local box, *global* ensemble ($\mathbf{z}_i$) vectors are used to fit the observation in the local analysis. This implies that the effective box includes supports of all non-local (non-point-support) observations. In other words, if we wish to fit a non-local observation (with the support larger than the set of grid points surrounding the observation point), there in no choice other than to control the fields within its support. As a result, the actual domain in physical space where the local analysis is done (which we call the effective box) is enlarged:

$$a_i \geq |supp\mathcal{H}|_i, \tag{20}$$

where a_i is the effective box extent (diameter), $i = x, y, z$, $|A|_i$ denotes the diameter of the set A along the coordinate axis i, and $supp\mathcal{H}$ denotes support of $\mathcal{H}$.

Supports of real observation types can be quite large:

(i) Nadir radiances are often influenced by a part of the atmospheric vertical column comparable in depth to the whole atmosphere, e.g. AMSU-A channels 6-10 have supports as large as 20-30 km in the vertical (e.g. Goldberg *et al.* 2001). So, including such observations in a *local* analysis makes it, effectively, *global* in the vertical.

(ii) Limb observations (which look through the Earth atmosphere from space to space) imply large effective *horizontal* extents of the local boxes. Indeed, a limb radiance measurement or a radio-occultation observation both depend on a *horizontal* path in the Earth atmosphere. For a thin atmosphere, the length of this tangent path is $l \approx 2\sqrt{2R_e d}$, where R_e is the Earth radius and d the effective depth of the atmosphere. If $d = 30$ km, $l \approx 1200$ km. If $d = 15$ km, $l \approx 900$ km.

Thus, we see that assimilation of (very informative in practice) non-local nadir and limb satellite observations makes the effective size of the local boxes *large*. We cannot avoid this enlargement without cutting distant parts of the observational support (i.e. without nullifying their influence on the observation operator), which would have a serious detrimental impact on their assimilation. With non-local satellite data, the effective boxes can cover (almost) the whole model atmosphere in the vertical and have horizontal extents greater than 1–2 thousand km.

4.2 Within *large* effective boxes, affordable ensemble size implies poor analysis resolution and hence accuracy

As discussed, defining the control variable in ensemble space $\mathcal{E}$ is essential for LETKF because it yields the very fast computational algorithm, but, at the same time, it is the major limitation of the technique. Indeed, as we have seen in section 2.2.1, working in $\mathcal{E}$ entails the additional observation representativeness error, η_{et}, due to the inability of a small number of ensemble vectors to span the physical space within the (effective) local box. This error reduces the analysis accuracy, especially in cases with large difference between the (low)

dimensionality of $\mathcal{E}$ and the (large) number of observed degrees of freedom in the local analysis.

The implications of the ensemble being too small can be also explained from a more conventional perspective. Namely, as it is stressed many times in the LETKF and other literature and as it follows from Eq.(18) above, the local analysis increment is a linear combination of the local ensemble vectors. So, as noticed by Lorenc (2003b), in order for an EnKF analysis to be capable of fitting observation, the ensemble size should be comparable with the number of observations within the localization domain. If observations are plentiful whilst the ensemble size is small, the analysis will inevitably *smooth* the observational information, which can lead to loss of analysis accuracy. Note that this problem is also encountered in particle filters (Tsyrulnikov 2007). So, we need to make the ensemble size commensurable with the number of *observed* degrees of freedom within an effective box:

$$n_e \sim n_{odof}. \tag{21}$$

If the local spatial variability is high (the most practically important case) *and* this high variability is captured by the existing high-resolution observing systems, then we need high-dimensional analysis space (not available in LETKF) in order to represent the observed variability in the analysis increments. This is especially important on the meso scale, where small-scale phenomena are abundant both in the horizontal and in the vertical: fronts, jets, inversion layers, convective systems, polar lows, etc.

In other words, small dimensionality of analysis space implies poor resolution in a *local* analysis. In a local box with the effective extents a_i, not more than n_e features can be resolved in the local analysis. The resulting local analysis resolution measured by the effective mesh sizes h_i^{eff} is

$$h_i^{eff} \sim a_i / \sqrt[3]{n_e}. \tag{22}$$

Realizing that the typical ensemble size is $n_e = 30 - 100$, we conclude that the effective local resolution is only about $\sqrt[3]{n_e} \div 4$ pieces of information in each of the three spatial dimensions. Note that in the 4-D version of LETKF, the local resolution can be even less ($\sqrt[3]{n_e}$ should be replaced by $\sqrt[4]{n_e}$ in Eq.(22)).

On the other hand, we have seen in the previous subsection that assimilating non-local satellite data implies that local boxes become, effectively, as large as $(1-2)\cdot 10^3$ km in the horizontal and (almost) the whole model atmosphere in the vertical. Therefore, if these observations are assimilated, the analysis resolution within a box appears to be very poor: 3–6 km in the vertical and 200–500 km in the horizontal, which is far less than we would require even from a *global* analysis, let alone limited-area *meso-scale* assimilation.

It is worth noting that poor resolution in a local analysis means that even if there are (useful) high-accuracy observations near the centre of the box, they simply can be not resolved. As a result, the accuracy of the resulting analysis at the centre of the box is reduced. It is important to stress that the very attempt to assimilate non-local satellite data using an ensemble-space analysis technique can make assimilation of *all* observations inefficient.

The above $\sqrt[3]{n_e}$ dependency implies that high resolution in a local box cannot be even reached with the ensemble-space analysis technique in practical NWP applications. Indeed, the effective 20-km mesh size in the horizontal (100 pieces of information for a 2000-km box) and 50 levels in the vertical would require unimaginable $n_e \approx 100\cdot 100\cdot 50 = 500,000$ ensemble members, which is, certainly, not feasible. Bishop and Hodyss (2009) proposed to use a kind of resampling technique in order to 'statistically' increase the ensemble size up to thousands. In its simplest form, their idea can be used to generate 'synthetic' ensemble perturbations by $\mathbf{y}_s^* = \mathbf{E}\mathbf{a}$, where $\mathbf{a}$ is a pseudo-random vector with the unit covariance matrix and $s = 1, \ldots, S$

denotes the number of the generated perturbation. Thus defined, any $\mathbf{y}_s^*$ has the covariance matrix exactly equal to $\mathbf{B}^e$ (conditional on the forecast ensemble). Appending the 'synthetic' perturbations to the forecast ensemble perturbations increases ensemble size, improving the local resolution and reducing the ensemble representativeness error. But not dramatically, because the required very large 'synthetic' ensembles are expensive: with the dimensionality of a practical problem as large as $n = 10^8$ and $n_e = 10^2$, just generating $S = 10^4$ 'synthetic' perturbations would require huge $n \cdot n_e \cdot S = 10^{14}$ flops. So, we conclude that an ensemble-space EnKF analysis technique cannot yield high spatial resolution in practice.

Summarizing, we claim that with LETKF, it is not possible to efficiently assimilate non-local satellite data in the presence of high local spatial variability when this variability is observed.

4.3 *Small* local boxes can led to small-scale noise in the analysis increments

The LETKF analysis algorithm can be viewed as an attempt to reintroduce OI but with ensemble covariances instead of analytic ones. As a result, many OI drawbacks are inherited by LETKF. In particular, changes in the sets of influencing observations from one analysis grid point to another can lead to small-scale noise: horizontal and vertical analysis gradients become contaminated. As the gradients directly enter the prognostic equations, their accuracy is as important as accuracy of the fields themselves. In addition, inaccurate gradients can destroy *balances* (hydrostatic, geostrophic, ...).

In order to make the resulting small-scale noise reasonably low, one should ensure that in each box, observations close to its boundaries have low weights in the analysis. With observation errors similar in magnitude to background errors (a typical situation in modern data assimilation), this is the case if background-error correlations between the center of the box and the boundaries are sufficiently low. This can be achieved either by increasing the box size or by implementing the above 'observation localization'. It is worth noting, however, that severe localization distorts the effective background-error correlations: their length scales L_i may become unrealistically small, which can substantially reduce the accuracy of the analysis. So, we cannot use 'observation localization' to *radically* reduce the background-error length scales, which, thus, appear to be the approximate lower bounds for the respective box radia:

$$\frac{a_i}{2} \geq L_i. \tag{23}$$

Otherwise, we say the box is small. Now, we show that with small boxes, the noise can really be generated. We carried out a simple 1-D experiment. We supposed that the analyzed field had analytic correlation function $C(r) = (1 + r/L)\exp(-r/L)$, where r is the distance on the circle and L the length scale. We selected $L = 300$ km as a typical value for background-error correlations. The field was assumed to have unit variance. The ensemble size was chosen to be $n_e = 50$. The analysis domain was a (latitude) circle. The grid spacing was $h \approx 55$ km (1° at latitude 60°). Realizations of both the pseudo-random 'truth' and forecast perturbations were generated by forming the covariance matrix of the grid-point values, whose entries are $\mathbf{B}_{ij} = C(|r_i - r_j|)$, performing its eigen-decomposition, and perturbing the 'principal components' according to their variances given by the respective eigen-values of $\mathbf{B}$. Observations were placed at grid points with the spacing $h_{obs} = m \cdot h$. The pseudo-random Gaussian observation noise with unit variance was added. Observation errors were assumed to be uncorrelated. 'Observation localization' was implemented with the Gaussian localization function, $\exp(-0.5(r/l)^2)$, where l is the localization length scale.

In Fig.1. we show results of the experiment with $m = 4$ (each fourth grid point was observed), box diameter as small as $12h$, and $l = 3h$. The effective box diameter with the observation

localization can be roughly assessed as $a_{eff} = 2l = 6h \approx 330$ km, which is close to $L = 300$ km, so the boxes are nearly small in the above sense. The solid thick curve represents one arbitrarily chosen realization of a reference-analysis increment produced using large local boxes (with diameter $50h$) and exact background-error statistics (OI, essentially). The solid thin curve represents the LETKF analysis increment. The dotted curve corresponds to the analysis produced on a coarser grid (here, 8 times coarser than the analysis grid) and post-interpolated using the weight interpolation proposed by Yang *et al.* (2009).

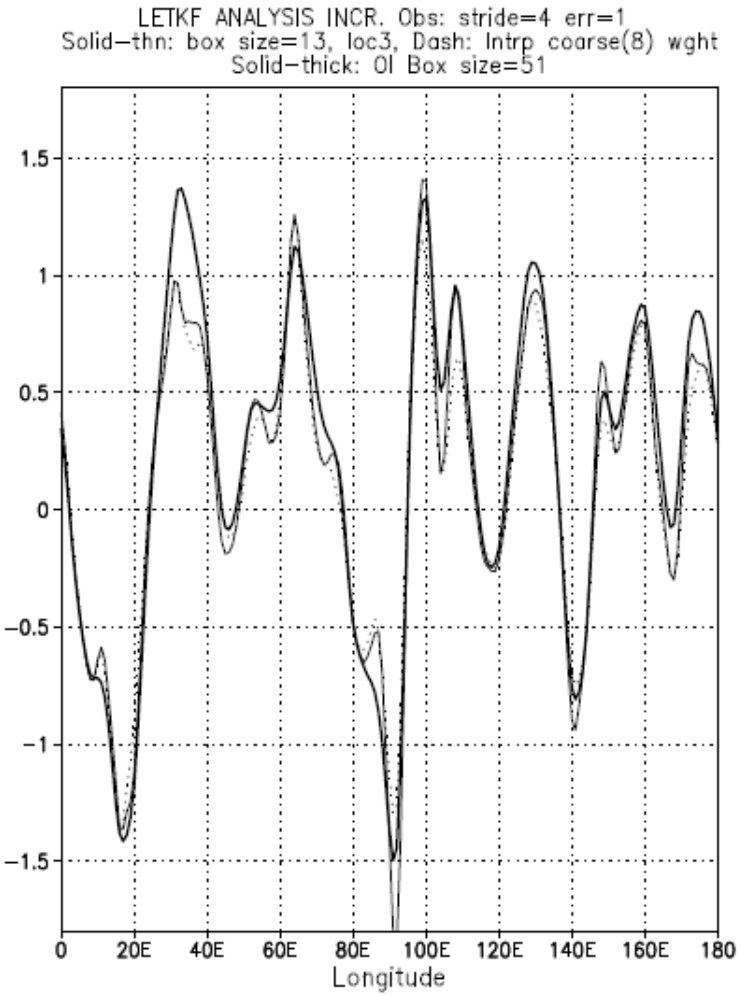

Figure 1: Analysis increment, one realization: OI box diameter 50 (solid thick curve), LETKF box diameter 12, localization length 3 (solid thin), and LETKF on coarse grid with post-interpolated ensemble weights (dotted).

From Fig.1, one can see that indeed small boxes can cause excessive and spurious extrema in LETKF analysis increments. Weight interpolation is seen to smooth somewhat the noisy analysis increments but only slightly. With the coarse-grid mesh size equal to $3h$, the resulting interpolated increment (not shown) appeared to be almost indistinguishable from the full-grid (solid thin) LETKF curve. So, the noise is really generated in spite of 'observation localization' and weight interpolation. With larger boxes, denser observation network and smaller observation errors, the noise in the increment field reduces. We remark that the these experimental results depend only on the ratio L/h, so they are equally valid for, say, $h = 1.8$ km and $L = 10$ km.

Finally, we note that, as it was proposed by Lorenc (1981) for OI and by Bishop and Hodyss (2009) for LETKF, one can suppress the small-scale noise by simultaneously updating a number of grid points near the centre of the box. But this can be only achieved at the expense of increasing the box size, which can exacerbate the ensemble-space-analysis resolution problem discussed in the previous section.

4.4 Observation-error correlations can destroy the LETKF computational efficiency

At first glance, the number of floating-point operations involved in the LETKF analysis algorithm Eqs.(18)–(19) does not critically depend on the number of observations, n_{obs}, because the (small) ensemble size (and thus the dimensionality of the system of algebraic analysis equations to be solved) remains untouched. With the increasing n_{obs}, the number of entries in the $\mathbf{Z}$ matrix and thus the number of floating-point operations (flops) in the analysis grows as $O(n_{obs})$. But this slow growth of the floating-point operation count with the box size becomes badly fast if we try to account for spatial dependencies (correlations) between the *observation errors*.

Indeed, if the local $\mathbf{R}$ matrix is dense [2], its factorization involved in the application of $\mathbf{R}^{-1}$ in Eq.(19)) requires as large as $O(n_{obs}^3)$ flops (e.g. Golub and van Loan 1989). In this case, we state that *the computational advantage of the LETKF algorithm disappears*. This issue is most important for satellite data assimilation and on the meso scale (where radar data do have spatially correlated errors (Xu *et al.* 2007)).

Hunt *et al.* (2007) discussed using 'batches' of observations with zero correlations of observational errors between different 'batches'. But this assumption is, generally, not met in case of data with spatially and temporally almost *continuous* coverage (satellites and radars). So, the 'batch' approach cannot help with these data (the 'batches' will be too large, say, all AMSU-A data will be one 'batch').

Another device capable, in principle, of removing the problem is the diagonalization of the $\mathbf{R}$ matrix before the analysis (e.g. Fertig *et al.* 2007). However, this approach again requires $O(n_{obs}^3)$ flops to compute the matrix $\mathbf{R}^{-1/2}$, which transforms the original observations $\mathbf{x}^o$ to new 'pseudo-observations' with uncorrelated errors:

$$\check{\mathbf{x}}^0 := \mathbf{R}^{-1/2} \cdot \mathbf{x}^o. \tag{24}$$

In addition, these 'pseudo-observations' become substantially *non-local* because their observation operator involves left-multiplication by $\mathbf{R}^{-1/2}$ (as it follows from Eq.(24)), which complicates their efficient assimilation in an *ensemble-space* analysis (as discussed above in section ()). So, diagonalizing the $\mathbf{R}$ matrix is not likely to give rise to a computationally efficient LETKF algorithm with correlated observation errors.

Thus, with the existing methodologies, allowing for correlated satellite and radar observation errors makes the LETKF analysis algorithm computationally inefficient (we have to solve large systems of linear algebraic equations at every analysis grid point).

Finally, we note that $\mathbf{R}^{-1}$ enters the equations of the (L)ETKF analysis ensemble generation scheme (e.g. Hunt *et al.* 2007), which can compromise the applicability of the Ensemble Transform analysis technique in the presence of correlated observational errors.

[2] Bormann *et al.* (2003) report that for AMV (atmospheric motion vectors) observations, the horizontal observation-error correlation falls to 0.5 at 300 km and to zero at about 1000 km.

4.5 Meso-scale issues

On the meso scale, a kind of *scale separation* technique can be a remedy for the above problems with the LETKF assimilation of non-local satellite data. Namely, satellite observations can be assimilated in a global variational analysis, which is used as a first guess for the 'meso' analysis, which, in turn, updates meso scales by using only locally supported 'meso' observations. In this case, the above limitations are greatly relaxed.

Indeed, first, large supports of satellite observations no longer limit the effective box size from below (as there are no 'meso' data with comparably large supports). The boxes can be made small so that the effective resolution within a box, see Eq.(22), can be reasonably high. Second, analyzing only meso scales implies small background-error length scales, so it can be easy to satisfy the requirement Eq.(23). Third, balances are known to be weak on the meso scale (e.g. geostrophy is not valid because the Rossby number is large), so there is little danger to destroy them by localization.

Thus, boxes can be made sufficiently small, which would result in enhanced spatial resolution within a box without excessive smoothing of observations, without generating small-scale noise, and without destroying multivariate balance constraints. However, at scales less than about 1 km, the radar error correlations appear to be significant (Xu *et al.* (2007) report on the 2-km decorrelation radius). In this case, the LETKF algorithm can lose its computational efficiency and its real-data applicability becomes questionable.

We also note that on the meso scale, strong non-linearity of forecast equations can lead to highly non-Gaussian background-error distributions. But Lawson and Hansen (2004) showed that the ensemble transform technique is incapable of properly blending the non-Gaussian prior distribution and the largely Gaussian observation-error distribution. So, on the meso scale, a perturbed-observations technique seems to be more suitable than an ensemble-transform based technique.

5. Published experimental evidence

(1) Concerning satellite data, we state that the only attempt to assimilate real radiances with the LETKF methodology was reported in (Szunyogh *et al.* 2007a) but without detailed presentation of the results. Simulated radiances were assimilated by Fertig *et al.* (2007). All peer-reviewed papers reporting on real-data assimilation (listed in the Introduction) do *not* involve satellite radiances, in spite of the fact that these latter are known to be the crucial part of the global atmospheric observing system (e.g. Kelly and Thépaut 2007).

(2) The necessity of the above relationship, Eq.(21), between the ensemble size and the number of observed degrees of freedom within a box is largely confirmed by the published experimental results: practical schemes appear to work reasonably well if this condition is met. Specifically, Szunyogh *et al.* (2005) found that for $n_e = 40$, when every 9-th horizontal grid point was observed, the best results were obtained with about $n_{dof} = 500$ model variables within the box (see their Fig.8). We note that in this optimal configuration, the number of *observed* degrees of freedom within a box is about $n_{odof} = 500/9 \approx 56$, which is very close to $n_e = 40$.

Fertig *et al.* (2007) found that simulated radiosonde observations were most effectively assimilated with $n_e = 20$ if $n_x = n_y = 7$ and the boxes were only one level deep. The analysis variable comprised 4 three-dimensional fields, so a box contained $n_{dof} = 7 \cdot 7 \cdot 4 \approx 200$ degrees of freedom. But not all of them were observed. Concretely, they used about 600 simulated radiosonde profiles, so that with their analysis grid having $48 \cdot 96 = 4600$ grid points, only a

portion of 600/4600 $\approx 1/8$ was observed. Hence $n_{odof} = n_{dof}/8 \approx 25$, which is in excellent agreement with $n_e = 20$.

Miyoshi and Yamane (2007) came up with the following optimized parameters of their LETKF scheme. In their OSSE1 experiments, the box size for $n_e = 40$ was: $n_x = 11$ with the Gaussian observation localization length equal to 3 mesh sizes and $n_z = 7$ with the vertical localization length 2 mesh sizes. As the localization lengths are smaller than the respective box extents, we easily assess the effective box sizes in each of the two horizontal dimensions as $2 \cdot 3 + 1 = 7$ grid points and, in the vertical, $2 \cdot 2 + 1 = 5$ grid points, so that there were, roughly, $7 \cdot 7 \cdot 5 = 245$ grid points within a box. With 5 three-dimensional analysis fields, we obtain $n_{dof} = 5 \cdot 245 = 1225$ effective analysis degrees of freedom. But only each hundredth degree of freedom was observed, which yields $n_{odof} = 1225/100 \approx 12$. This small number explains why Miyoshi and Yamane (2007) found that even n_e as small as 10 could be used.

(3) As regards the possible generation of small-scale noise due to the local nature of the LETKF analysis, Liu *et al.* 2008 reported on a very successful behaviour of LETKF. They used low resolution in the horizontal (about 450-500 km), so that the box size was large ($a_x \approx$ 3000 km). With the box size this large, it is likely that our condition Eq.(23) was satisfied. Indeed, this is the case if the horizontal background-error length scale is as large as $L \leq 1500$ km, which is significantly higher than the typical length scales (300–500 km). Similarly, local boxes in (Szunyogh *et al.* 2005) were, in their base experiment, about $1000 \div 1400$ km broad.

On the other hand, tiny $n_x = n_y = 3$ boxes in (Harlim and Hunt 2007b) *did* lead to small-scale noise, as we would expect (see their Figs. 3 and 5).

(4) No attempt to account for realistic observation-error correlation was reported on yet, so we cannot check the conclusions of section 4.4.

Summarizing this section, we note that our theoretical inference on the applicability of the LETKF analysis does not contradict to the existing experimental evidence and seems to provide explanation for some experimental results.

6. Discussion

We have shown that the major deficiency of the present formulation of the LETKF analysis is its inability to efficiently assimilate non-local satellite observations. This is because in each local LETKF analysis, the analysis increment is confined to be a linear combination of (a small number of) the forecast ensemble perturbations. With other EnKF formulations, this limitation can be relaxed. Solving the analysis equations in observation space allows us to apply covariance localization (e.g. Houtekamer and Mitchell 2006), so that the analysis increment no longer belongs to the low-dimensional ensemble space. Another suitable approach is to use spatially averaged covariances (Raynaud *et al.* 2008). Hybrid EnKF-3DVar schemes also allow us to avoid the 'curse of (low) dimensionality' of ensemble space (Hamill and Snyder 2000, Lorenc 2003b, Wang *et al.* 2007).

The experimental fact that LETKF can quite successfully (despite locality and ensemble-space restrictions) assimilate conventional observations, suggests that the *ensemble* data assimilation principle is indeed promising for operational and other purposes. We notice, however, that good results reported in the LETKF and some other ensemble data assimilation papers (e.g. Whitaker *et al.* 2008) were obtained without satellite data, i.e. for poorly observed flows, where the errors have time to develop complicated anisotropic structures. With the addition of frequent and ubiquitous satellite observations, it is likely that the effect of flow-dependent covariances may appear to be less dramatic.

7. Conclusions

The main findings of this study are:

- Non-local satellite observations (both nadir and limb) are shown to make the effective size of local LETKF boxes large: extended by supports of all used observations.
- Small ensemble size implies small number of resolvable features in local LETKF analyses. Large effective boxes and small ensemble size imply, thus, low spatial resolution within a local box, which can make the assimilation of all observations inefficient— if non-local satellite data (radiances, in particular) are assimilated.
- Without non-local satellite data, the local-box extents are limited from below by the respective background-error length scales. Smaller boxes can give rise to significant small-scale noise in the analysis increments.
- Allowing for realistic correlations in observation errors (e.g. for satellite and radar data) removes the advantage of LETKF in computational efficiency as compared to observation-space formulations.

All in all, we state that the LETKF approach in its present formulation does not seem to be a good general method of choice for operational data assimilation on the global scale. LETKF can be utilized as a means of meso-scale data assimilation in a scale-separation scheme without non-local satellite observations. The LETKF technique, being simple and cheap, can be successfully used to solve some other particular data assimilation problems and in other applications (research, education, etc.).

8. Acknowledgments

I appreciate a discussion with B. Hunt on the issues raised in the article. This review is written at request of COSMO working group WG1.

References

[1] Baek S-J, Hunt BR, Kalnay E, Ott E, and Szunyogh I. 2006. Local ensemble Kalman filtering in the presence of model bias. *Tellus*, **58A**: 293-306.

[2] Bishop CH., Hodyss D. 2009. Ensemble covariances adaptively localized with ECO-RAP. Part 2: a strategy for the atmosphere. *Tellus*, **61A**: 97-111.

[3] Bonavita M, Torrisi L, Marcucci F. 2008. The ensemble Kalman Filter in an operational regional NWP system: Preliminary results with real observations. *Q.J.Roy. Meteorol. Soc.*, **134**: 1733-1744.

[4] Bormann N, Saarinen S, Kelly G and Thepaut J-N. 2003. The spatial structure of observation errors in atmospheric motion vectors from geostationary satellite data. *Mon. Wea. Rev.*, **131**: 706-718.

[5] Fertig EJ, Hunt BR, Ott E, and Szunyogh I. 2007. Assimilating non-local observations with a local ensemble Kalman Filter. *Tellus*, **59A**: 719-730.

[6] Gandin LS. 1965. *Objective analysis of meteorological fields*. Gidrometeorologicheskoe izdatelstvo, Leningrad, 1963. Translated from Russian, Israeli Program for Scientific Translations: Jerusalem.

[7] Golub GH and van Loan CF. 1989. *Matrix computations*. The John Hopkins University Press.

[8] Goldberg MD, Crosby DS, and Zhou L. 2001. The limb adjustment of AMSU-A observations: methodology and validation J. Appl. Meteorol., **40**: 70-83.

[9] Hamill TM and Snyder C. 2000. A hybrid ensemble Kalman Filter–3D variational analysis scheme. *Mon. Wea. Rev.*, **128**: 2905-2919.

[10] Harlim J and Hunt BR. 2007a. A non-Gaussian ensemble filter for assimilating infrequent noisy observations. *Tellus*, 59A, **59A**: 225-237.

[11] Harlim J and Hunt BR. 2007b. Four-dimensional local transform Kalman Filter: numerical experiments with a global circulation model. *Tellus*, **59A**: 731-748.

[12] Houtekamer PL and Mitchell HL. 2006. Ensemble Kalman filtering. *Q.J.Roy. Meteorol. Soc.*, **131**: 3269-3289.

[13] Hunt BR, Kalnay E, Kostelich EJ, Ott E, Patil DJ, Sauer T, Szunyogh I, Yorke JA, and Zimin AV. 2004. Four-dimensional ensemble Kalman filtering. *Tellus*, **56A**: 273-277.

[14] Hunt BR, Kostelich EJ, and Szunyogh I. 2007. Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman Filter. *Physica D*, **230**: 112-126.

[15] Kalnay E, Li H, Miyoshi T, Yang S-C, and Ballabrera-Poy J. 2007. 4-D-Var of ensemble Kalman Filter? *Tellus*, **59A**: 758-773.

[16] Kelly G and Thépaut J-N. 2007. Evaluation of the impact of the space component of the Global Observing System through Observing System Experiments. *EUMETSAT / ECMWF Report Series Final report SOW EUM.MET.SOW.04.0290.*

[17] Lawson WG and Hansen JA. 2004. Implications of stochastic and deterministic filters as ensemble-based data assimilation methods in varying regimes of error growth. *Mon. Wea. Rev.*, **132**: 1966-1981.

[18] Liu J, Fertig EJ, Li H, Kalnay E, Hunt BR, Kostelich EJ, Szunyogh I, Todling R. 2008. Comparison between local ensemble transform Kalman Filter and PSAS in the NASA finite volume GCM – perfect model experiments. *Nonlin. Processes Geophys.*, **15**: 645-659.

[19] Lorenc A. 1981. A global three-dimensional multivariate statistical interpolation scheme. *Mon. Wea. Rev.*, **109**: 701-721.

[20] Lorenc AC. 2003a. Modelling of error covariances by 4D-Var data assimilation. *Q.J.R. Meteorol. Soc.*, **129**: 3176-3182.

[21] Lorenc AC. 2003b. The potential of the ensemble Kalman Filter for NWP – a comparison with 4D-Var. *Q.J.R. Meteorol. Soc.*, **129**: 3183-3203.

[22] Miyoshi T and Yamane S. 2007. Local ensemble transform Kalman filtering with an AGCM at a T159/L48 resolution. *Mon. Wea. Rev.*, **135**: 3841-3861.

[23] Ott E, Hunt BR, Szunyogh I, Zimin AV, Kostelich EJ, Corazza M, Kalnay E, Patil DJ, Yorke JA. 2004. A local ensemble Kalman Filter for atmospheric data assimilation. *Tellus*, **56A**: 415-428.

[24] Raynaud L, Berre L, and Desrosiers G. 2008. Spatial averaging of ensemble-based background-error variances. *Q.J.Roy. Meteorol. Soc.*, **134**: 1003-1014.

[25] Szunyogh I, Kostelich EJ, Gyarmati G, Patil DJ, Hunt BR, Kalnay E, Ott E, Yorke JA. 2005. Assessing a local ensemble Kalman Filter: perfect model experiments with the National Centers for Environmental Prediction global model. *Tellus*, **57A**: 528-545.

[26] Szunyogh I, Satterfield EA, Aravequia JA, Fertig EJ, Guarmati G, Kalnay E, Hunt BR, Kostelich EJ, Kuhl DD, Ott E, Yorke JA. 2007a. The local ensemble transform Kalman Filter and its implementation on the NCEP global model at the University of Maryland. *Proceedings of the ECMWF Workshop on Flow dependent aspects of data assimilation*, ECMWF, Reading, UK, 47-64.

[27] Szunyogh I, Kostelich EJ, Gyarmati G, Kalnay E, Hunt BR, Ott E, Satterfield E, Yorke JA. 2007b. A local ensemble transform Kalman Filter data assimilation system for the NCEP global model. *Tellus*, **59A**: 1-18.

[28] Tsyrulnikov MD. 2007. Is the particle filtering approach appropriate for meso-scale data assimilation? *COSMO Technical Report N 10*, 1-16 (http://www.cosmo-model.org/content/model/documentation/techReports/docs/techReport10.pdf).

[29] Wang X, Hamill TM, Whitaker JS, Bishop CH. 2007. A comparison of hybrid ensemble transform Kalman Filter–Optimum Interpolation and ensemble square root filter analysis schemes. *Mon. Wea. Rev.*, **135**: 1055-1076.

[30] Whitaker JS, Hamill TM, Wei X, Song Y and Toth Z. 2008. Ensemble data assimilation with the NCEP global forecast system. *Mon. Wea. Rev.*, **136**: 463-482.

[31] Xu Q, Nai K, Wei L. 2007. An innovation method for estimating radar radial-velocity observation error and background wind error covariances. *Q.J.Roy. Meteorol. Soc.*, **133**: 407-415.

[32] Yang S-C, Kalnay E, Hunt BR, Bowler N. 2009. Weight interpolation for efficient data assimilation with the Local Ensemble Transform Kalman Filter. *Q. J. Roy. Meteorol. Soc.* 135, 251-262.