

Morphing ensemble Kalman filters

By JONATHAN D. BEEZLEY^{1,2} and JAN MANDEL^{1,2*}, ¹*Center for Computational Mathematics, University of Colorado at Denver and Health Sciences Center, Denver, CO, USA;* ²*Mesoscale and Microscale Meteorology Division, National Center for Atmospheric Research, Boulder, CO, USA*

(Manuscript received 6 February 2007; in final form 20 August 2007)

ABSTRACT

A new type of ensemble filter is proposed, which combines an ensemble Kalman filter (EnKF) with the ideas of morphing and registration from image processing. This results in filters suitable for non-linear problems whose solutions exhibit moving coherent features, such as thin interfaces in wildfire modelling. The ensemble members are represented as the composition of one common state with a spatial transformation, called registration mapping, plus a residual. A fully automatic registration method is used that requires only gridded data, so the features in the model state do not need to be identified by the user. The morphing EnKF operates on a transformed state consisting of the registration mapping and the residual. Essentially, the morphing EnKF uses intermediate states obtained by morphing instead of linear combinations of the states.

1. Introduction

The research reported here has been motivated by data assimilation into wildfire models (Mandel et al., 2006). Wildfire modelling presents a challenge to data assimilation because of non-Gaussian probability distributions centred around the burning and not burning states, and because of movements of thin reaction fronts with sharp interfaces. This work is a part of an effort to build a Dynamic Data Driven Application System (DDDAS, Darema, 2004) for wildfires (Mandel et al., 2007). Data assimilation is one of the building blocks of the DDDAS concept, which involves a symbiotic network of computer simulations and sensors.

The standard ensemble Kalman filter (EnKF) approach (Evensen, 2003) fails for highly non-linear problems that have solutions with coherent features, such as firelines, rain fronts, or vortices, because it is limited to a linear update of state. This can be ameliorated to some degree by penalization of non-physical states (Johns and Mandel, 2005), localization of EnKF (Anderson, 2003; Ott et al., 2004), and employing the location of the feature as an observation function (Chen and Snyder, 2007), but the basic problem remains: EnKF works only when the increment in the location of the feature is small (Chen and Snyder, 2007).

EnKF analysis formulas are based on the assumption that all probability distribution involved are Gaussian, so even if an en-

semble can approximate non-Gaussian distribution, the closer the system state distribution is to Gaussian the better. One mechanism how non-Gaussian distributions arise in strongly non-linear systems is by evolution of coherent features. While the location and the size of the feature may have an error distribution that is approximately Gaussian, this is not necessarily the case for the value of the state at a given point. For example, the typical probability density of the temperature at a point near the reaction region in a wildfire model is concentrated around the burning and the ambient temperature (Fig. 4a). There is clearly a need to adjust the simulation state by distorting the simulation state in space rather than employing an additive correction to the state. Therefore, alternative error models that include the position of features were considered in the literature (Hoffman et al., 1995; Davis et al., 2006) and a number of works emerged that achieve more efficient movement of features by using a spatial transformation as the field to which additive corrections are made, such as a transformation of the space by a global low order polynomial mapping to achieve alignment (Alexander et al., 1998), and two-step models to use alignment as pre-processing to an additive correction (Lawson and Hansen, 2005; Ravela et al., 2007).

Moving and stretching one given image to become another given image is known in image processing as registration (Brown, 1992). Once the two images are registered, one can easily create intermediate images, which is known as morphing.

The essence of the new method described here is to replace the linear combinations of states in an ensemble filter by intermediate states obtained by morphing. This method provides additive and position correction in a single step. For the analysis step (the Bayesian update), the state is transformed into an extended

*Corresponding author.

e-mail: Jan.Mandel@cudenver.edu

DOI: 10.1111/j.1600-0870.2007.00275.x

state consisting of additive and position components. After the analysis step, the state is converted back and advanced in time. The purpose of this article is to demonstrate the potential of this approach.

The paper is organized as follows. In Section 2, we briefly recall data assimilation by EnKF but we do not present the formulation of the EnKF in detail and refer to the literature instead. In Section 3, we describe image morphing and the automatic registration algorithm used. Section 4 contains the formulation of the morphing EnKF. Numerical results on a wildfire model problem are reported in Section 5. Section 6 contains the conclusion and a discussion of future extensions.

2. Data assimilation and ensemble filters

The purpose of data assimilation is to estimate the system state using all data available up to the current time. A discrete filter, considered here, works by advancing in time a probability distribution of the model state until a given *analysis time*. At the analysis time, the probability distribution, now called the *prior* or the *forecast*, is modified by accounting for the data, which are considered to be observed at that time. The new probability distribution, called the *posterior* or the *analysis*, is given by the Bayes theorem,

$$p_a(u) \propto p(d|u)p_f(u), \quad (1)$$

where $\propto$ means proportional, u is the model state, $p_f(u)$ is the forecast probability density, $p_a(u)$ is the analysis probability density, d is the data, and $p(d|u)$ is the data likelihood. The data likelihood is the density of the probability that the data value is d assuming that the state is u . Assuming an additive data error model, the data likelihood is found from the probability density p_ε of data error, which is assumed to be known (every measurement must be accompanied by an error estimate), and from an *observation function* h , by

$$p(d|u) = p_\varepsilon[d - h(u)]. \quad (2)$$

The value $h(u)$ of the observation function is what the correct value of the data would be if the model state u were exact. For a tutorial on data assimilation, see Kalnay (2003).

The well-known Kalman filter (Kalman, 1960) reduces the Bayesian update (1) to linear algebra in the case when all probability distributions are Gaussian. The Kalman filter must advance the covariance of state, which is possible only when the model is linear, and it is computationally very expensive. The EnKF (Evensen, 1994; Houtekamer and Mitchell, 1998) approximates the probability distribution by the empirical measure $\frac{1}{N} \sum_{i=1}^N \delta_{u_i}$, where u_i are members of an *ensemble* of simulations states, and δ_u denotes the Dirac delta measure concentrated at u . Each ensemble member is advanced in time between the Bayesian updates independently. The EnKF approximates the mean and the covariance of the forecast by the mean and the covariance of the ensemble, while still making the assumption that all probability

distributions are Gaussian. The EnKF works by forming the analysis ensemble as *linear combinations of the forecast ensemble*, and the Bayesian update is implemented by linear algebra operating on the matrix created from ensemble members as columns. This allows an efficient implementation using high-performance matrix operations. We have used the EnKF version from Burgers et al. (1998), but any other variant could be used as well. For surveys of EnKF techniques, see Evensen (2007) and Tippett et al. (2003).

3. Registration, warping, and morphing

In this section, we build the tools from image processing that we are going to use for the morphing EnKF later in Section 4. The registration problem in image processing is to find a spatial mapping that turns two given images into each other (Brown, 1992). Classical registration requires the user to pick manually points in the two images that should be mapped onto each other, and the registration mapping is interpolated from that. Here we are interested in a *fully automatic registration procedure that does not require any user input*. The specific feature or objects, such as fire fronts or hurricane vortices, do not need to be specified. The method takes as input only the pixel values in an image (i.e. gridded arrays). Of course, for the method to work well, the images being registered should be sufficiently similar.

3.1. Notation

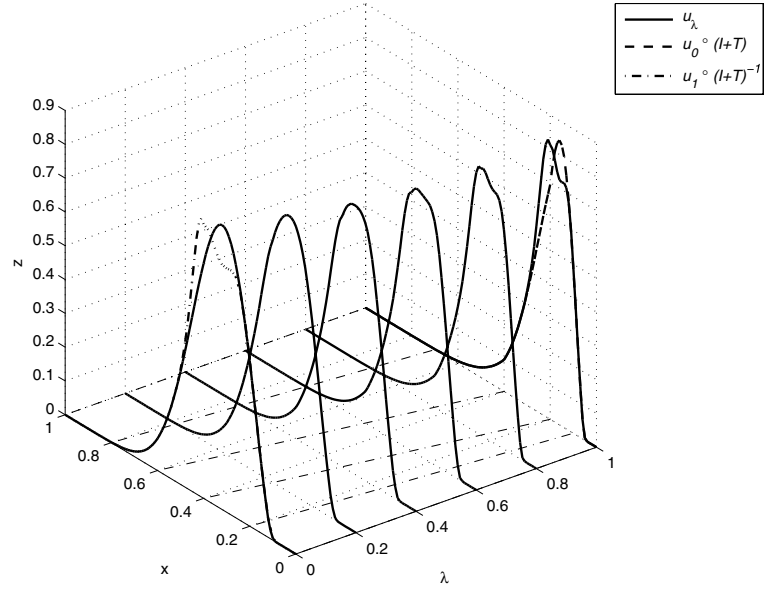
We will find it useful to use mappings as a convenient notation, so we review few basics in an informal manner. The symbol ' $\mapsto$ ' is read 'maps to', and we sometimes find it convenient to write $A : z \mapsto w$ instead of $A(z) = w$. The domain of a mapping A is the set of its arguments z such that $A(z)$ is defined. For two mappings A and B , their composition $A \circ B$ is defined by $A \circ B : z \mapsto A[B(z)]$; I is the identity mapping defined by $I : z \mapsto z$ (for all z from its domain); and the inverse A^{-1} of A is a mapping such that $A^{-1} \circ A = I$ and $A \circ A^{-1} = I$. A real function on a domain D is also a mapping, namely one that maps D into the reals, $\mathbb{R}$. Given two mappings A and B from the same domain into the same linear space (say, $\mathbb{R}$ or $\mathbb{R}^k$) and two scalars a and b , the linear combination of the mappings is defined by taking the linear combination of their values,

$$aA + bB : z \mapsto aA(z) + bB(z).$$

3.2. Image registration

Consider, for example, two grey scale images with intensities u and v , given as functions on some domain D (such as a rectangle in the plane, $\mathbb{R}^2$). For simplicity, assume that both u and v are equal to some constant at and near the boundary of the domain D . In our application, u and v will be temperature fields from two states of our wildfire model, the fire will be inside the domain, and the temperature near the boundary of the domain D will be

Fig. 1. Example of morphing procedure (8) in one dimension. In this example, all variables are non-dimensional. The functions $u_0(x)$ and $u_1(x)$ are given on the interval $[0, 1]$. The intermediate functions u_λ are computed by the morphing formula (8), which is seen to interpolate the difference in position as well as the difference in magnitude. The registration mapping T is piecewise linear on the morphing grid in the interval $[0, 1]$ with spacing equal to 0.2. The thin dash-dotted lines in the horizontal plane connect the nodes x of the morphing grid with the values of the registration mapping $(I + T)(x)$.



equal to the ambient temperature (assumed to be the same everywhere). In image processing, u and v can be the darkness levels of two photographs of objects with the same solid background. The functions u and v will be also referred to as images in the following description.

The registration then becomes the problem to find two functions $T_x(x, y)$ and $T_y(x, y)$ such that the transformation of the argument of u by

$$(x, y) \mapsto [x + T_x(x, y), y + T_y(x, y)] \quad (3)$$

transforms u into a function approximately equal to v on the domain D ,

$$v(x, y) \approx u[x + T_x(x, y), y + T_y(x, y)], \quad (4)$$

for all $(x, y) \in D$.

Define the mapping T from D to $\mathbb{R}^2$ by

$$T : (x, y) \mapsto [T_x(x, y), T_y(x, y)]. \quad (5)$$

Then (4) can be written compactly as

$$v \approx u \circ (I + T) \text{ on } D,$$

or

$$v - u \circ (I + T) \approx 0 \text{ on } D. \quad (6)$$

The mapping $I + T$ will be called the *registration mapping*, and the mapping T will be called *warping*. The reason for writing the registration mapping as $I + T$ is that the zero warping $T = 0$ is the neutral element of the operation of addition, and so linear combinations of warpings have a meaningful interpretation as blends of the warpings. This will be important in the development of the morphing EnKF.

To avoid unnecessarily large and complicated warping, the warping T should be as close to zero and as smooth as

possible,

$$T \approx 0, \quad \nabla T \approx 0, \quad (7)$$

where ∇T denotes the matrix of the first derivatives (the Jacobian matrix) of T ,

$$\nabla T = \begin{pmatrix} \frac{\partial T_x}{\partial x} & \frac{\partial T_x}{\partial y} \\ \frac{\partial T_y}{\partial x} & \frac{\partial T_y}{\partial y} \end{pmatrix}.$$

In addition, we require that the registration mapping $I + T$ is one-to-one, so the inverse $(I + T)^{-1}$ exists. However, we do not require that the values of $I + T$ are always in D or the inverse $(I + T)^{-1}$ is defined on all of D , so $u \circ (I + T)$ and $u \circ (I + T)^{-1}$ may not be defined on all of D . Therefore, we consider all functions $u, v, u \circ (I + T), u \circ (I + T)^{-1}$, etc., extended on the whole $\mathbb{R}^2$ by the constant value of u and v on the boundary of D .

3.3. Morphing

Once the registration mapping $I + T$ is found, one can construct intermediate functions u_λ between u_0 and u_1 by *morphing* (Figs 1 and 2),

$$u_\lambda = (u + \lambda r) \circ (I + \lambda T), \quad 0 \leq \lambda \leq 1, \quad (8)$$

where

$$r = v \circ (I + T)^{-1} - u \quad (9)$$

will be called the *registration residual*; it is easy to see that r is linked to the approximation in (6) by

$$r = [v - u \circ (I + T)] \circ (I + T)^{-1},$$

thus (6) also implies that $r \approx 0$.

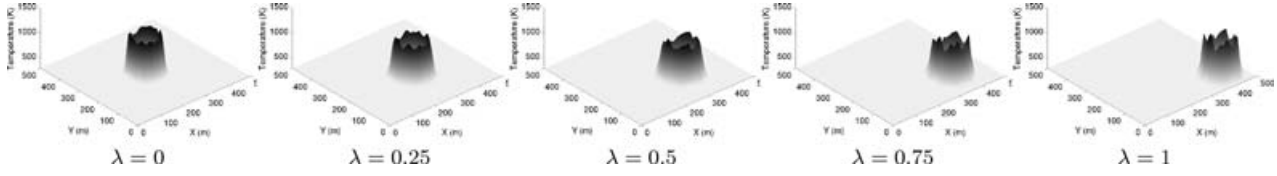


Fig. 2. Morphing of two solutions of a reaction-diffusion equation system used in a wildfire simulation. The states with $\lambda = 0$ and $\lambda = 1$ are given. The intermediate states are created automatically. The horizontal plane is the earth surface. The vertical axis and the colour map are the temperature. The morphing algorithm combines the values as well as the positions.

The original functions u and v are recovered by choosing in (8) $\lambda = 0$ and $\lambda = 1$, respectively,

$$u_0 = u \circ I = u, \quad (10)$$

$$\begin{aligned} u_1 &= (u + r) \circ (I + T) \\ &= (u + v \circ (I + T)^{-1} - u) \circ (I + T) \\ &= v \circ (I + T)^{-1} \circ (I + T) = v. \end{aligned} \quad (11)$$

Remark 1. In the registration and morphing literature, the residual is often neglected. Then the morphed function is given simply by the transformation of the argument, $u_\lambda = u \circ (I + \lambda T)$. The simplest way how to account for the residual is to add a correction term to u_λ . This gives the morphing formula

$$u_\lambda = u \circ (I + \lambda T) + \lambda[v - u \circ (I + T)], \quad 0 \leq \lambda \leq 1, \quad (12)$$

which is much easier to use because does not require the inverse $(I + T)^{-1}$ like (8). The formula (12) also recovers $u = u_0$ and $v = u_1$, but, in our computations, we have found it unsatisfactory for tracking features and therefore we do not use it. The reason is that when the residual is not negligibly small, the intermediate functions u_λ will have a spurious feature in the fixed location where the residual is large. On the other hand, the more expensive improved morphing formula (8) moves the contribution to the change in amplitude along with the change of the position.

3.4. Grids

An array of values associated with a rectangular grid is called a *gridded array*. The functions u_λ are represented by a gridded array on a *pixel grid*, while the mapping T is represented by two gridded arrays on a coarser *morphing grid*. In our application, the domain D is a rectangle, discretized by a uniform $n_x \times n_y$ pixel grid with $n = n_x n_y$ nodes, and, for $i = 1, \dots, M$, by a $(2^i + 1) \times (2^i + 1)$ uniform grid D_i with nodes denoted by (x_j, y_k) and total $m_i = (2^i + 1)^2$ nodes. The morphing grid is the finest grid D_M , with $m = m_M$ nodes. Denote by T_i the gridded array T restricted to the grid D_i , and by I_{i-1}^i the bilinear interpolation operator from grid D_{i-1} to grid D_i . All grids contain nodes on the boundary of D . It is assumed that $n \gg m$.

The values of the functions and of the mappings away from their respective grid points are evaluated by bilinear interpolation without mentioning the interpolation explicitly. So, composed

functions like $u \circ (I + T)$ are calculated in a straightforward manner: for an arbitrary $(x, y) \in D$, first $(I + T)(x, y) = (x', y')$ is computed by bilinear interpolation on the morphing grid and then $u(x', y')$ is evaluated by bilinear interpolation on the pixel grid. The calculation of $(I + T)^{-1}(x', y')$ is done by inverse interpolation as follows. For uniformly spaced nodes (x_j, y_k) on the morphing grid, the images $(x'_j, y'_k) = (I + T)(x_j, y_k)$ form a non-uniform grid, and the value of $(I + T)^{-1}(x', y')$ is approximated by interpolating the original coordinates x and y as functions on the non-uniform grid. This can be accomplished efficiently, for example, by using the MATLAB function `griddata`, which works exactly like interpolation, but allows for non-uniformly spaced data. We then have $(I + T)^{-1} \circ (I + T) \approx I$ and $u_1 \approx v$ as in (11) only up to the error of interpolation from the morphing grid.

3.5. An automatic registration procedure

The formulation of registration as (6)–(7) naturally leads to a construction of the mapping T by optimization. So, suppose we are given u and v on the pixel grid and wish to find a warping T that is an approximate solution of

$$J(T, u, v) = \|v - u \circ (I + T)\| + C_1 \|T\| + C_2 \|\nabla T\| \rightarrow \min_T, \quad (13)$$

where the norms are chosen as

$$\|v - u \circ (I + T)\| = \int_D |v - u \circ (I + T)| \, dx \, dy, \quad (14)$$

$$\|T\| = \int_D |T_x| + |T_y| \, dx \, dy, \quad (15)$$

$$\|\nabla T\| = \int_D \left| \frac{\partial T_x}{\partial x} \right| + \left| \frac{\partial T_x}{\partial y} \right| + \left| \frac{\partial T_y}{\partial x} \right| + \left| \frac{\partial T_y}{\partial y} \right| \, dx \, dy. \quad (16)$$

The optimization formulation tries to balance the conflicting objectives of good approximation by the registered image, and as small and smooth warping as possible. The objective function $J(T, u, v)$ is in general not a convex function of T , and so there are many local minima. For example, a local minimum of may occur when some small part of u and v matches, while the overall match is still not very good.

To solve the minimization problem (13), we have used the algorithm from Gao and Sederberg (1998) with several modifications. We have also filled in some details that were not provided. The description of the resulting algorithm is the subject of this section. It is presented for completeness only; any other automatic registration procedure from the literature could be used as well. The algorithm guarantees by construction that $I + T$ is invertible.

For speed and to decrease the chance that the minimization gets stuck in a local minimum, the method proceeds by building T on a nested hierarchy of meshes D_i , starting with T_1 on the coarsest mesh D_1 and ending with the mapping $T = T_M$ built on the morphing grid D_M . In the computation, $\|v - u \circ (I + T_i)\|$ in (14) is integrated numerically on the pixel grid, while $\|T\|$ and $\|\nabla T\|$ in (15) and (16) are integrated on one of the grids D_i , with the derivatives approximated by finite differences. The integrals are approximated by the scaled sums of the values of the integrands at grid nodes of the respective grids. Denote this version of the objective function on the grid D_i by $J_i(T_i, u, v)$. Assume that an initial guess $\tilde{T}$ of T on the morphing grid is known. If none is given, use $\tilde{T} = 0$. The notation $\tilde{T}_i$ means the gridded array $\tilde{T}$ restricted to the grid D_i , just like T_i .

On each mesh D_i , the method proceeds as follows. In order not to overload the notation with many iteration indices, the values of T , and thus also T_i , change during the computation just like a variable in a computer program.

- (i) Smooth u and v by convolution to get $\tilde{u}_i$ and $\tilde{v}_i$ on the pixel grid.
- (ii) If $i = 1$, initialize $T_1 = \tilde{T}_1$. Otherwise interpolate T_i from the coarse grid T_{i-1} as a correction to $\tilde{T}$.
- (iii) Minimize the objective function $J_i(T_i, \tilde{u}_i, \tilde{v}_i)$ by adjusting the value of T at one node of D_i at a time. Stop when a maximum number of sweeps through all nodes has been reached, or a stopping test based on the decrease of the objective function or residual size has been met.

We now describe each step in more detail.

- (i) *Smoothing.* The purpose of registration on a coarse mesh first is to capture coarse similarities between the images u and v . In order to force coarse grids to capture coarse features only and to disregard fine features, on each grid, we first smooth the images by convolution with a Gaussian kernel. This allows to track large-scale perturbations on coarse grids even for a thin feature such as a fireline, while maintaining small scale accuracy on fine grids. For the use with the grid D_i , we create the smoothed image $\tilde{u}_i$ with resolution on the scale of D_i , by

$$\tilde{u}_i(x_j, y_k) = \sum_{j'=-n_x+1}^{2n_x} \sum_{k'=-n_y+1}^{2n_y} \hat{\phi}_j(x_{j'}) u(x_{j'}, y_{k'}) \hat{\psi}_k(y_{k'}), \quad (17)$$

where

$$\hat{\phi}_j(x) = c_j \exp \left[-\frac{(x_j - x)^2}{\alpha_i} \right],$$

$$\hat{\psi}_k(y) = d_k \exp \left[-\frac{(y_j - y)^2}{\alpha_i} \right],$$

the constant $\alpha_i = 0.25/(2^i + 1)$ is tuned so that there is more smoothing on the coarser grids, the values $u(y_{j'}, y_{k'})$ outside of D are replaced by the constant boundary value, and the normalization constants c_j and d_k are determined so that

$$\sum_{j'=-n_x+1}^{2n_x} \hat{\phi}_j^2(x_{j'}) = \sum_{k'=-n_y+1}^{2n_y} \hat{\psi}_k^2(y_{k'}) = 1.$$

We also compute $\tilde{v}_i$ from v in the same way.

- (ii) *Initialization.* Consider the grid D_i , $i > 1$, with the nodes (x_j, y_k) . The values of $I + T_i$ are already known at the nodes on the coarse grid D_{i-1} . The correction that the optimization on grid D_{i-1} applied to the initial guess is $T_{i-1} - \tilde{T}_{i-1}$. To apply this correction to $\tilde{T}$ everywhere on D_i , we interpolate the correction from the grid D_{i-1} to D_i by the bilinear interpolation I_{i-1} to get the initial guess on the grid D_i , that is,

$$T_i = \tilde{T}_i + I_{i-1}^i(T_{i-1} - \tilde{T}_{i-1}).$$

- (iii) *Optimization.* The value of $A' = (I + T)(x_j, y_k)$ is optimized by first evaluating the local objective function $f(A') = J_i(T_i, \tilde{u}_i, \tilde{v}_i)$ at the nodes of a local grid inside the mapped local rectangle $(I + T_i)[(x_{j-1}, x_{j+1}) \times (y_{k-1}, y_{k+1})]$. The location of A' is then refined by several iterations of non-linear optimization, starting from the local grid node with the least value of $f(A')$ found. We have used coordinate descent alternating in the x and y direction by calling MATLAB function `minfnb` for 1-D constrained optimization (cf. Fig. 3). For nodes (x_j, y_k) on the boundary of the domain D , the location of A' is constrained within D , but it is allowed to move inside the domain D .

The differences between the method described here and the method by Gao and Sederberg (1998) are: the refinement of node positions by non-linear optimization; the gradient term in the objective function; smoothing of the images before registration on coarse levels; the use of an initial guess $\tilde{T}$; and allowing the nodes on the boundary to move inside of the domain.

3.6. Computational complexity

The operation (17) is the multiplication of three dense matrices. Assuming bounded aspect ratio of the image, $n_x \approx \text{const}$, $n_y \approx \text{const}$, the computation (17) takes $O(n^{1.5})$ operations. Using FFT to replace the convolution of functions by multiplication of their Fourier coefficients, it can be implemented in $O(n \log n)$ operations.

The cost of evaluating the entire objective function is linear in the number of pixels in the image, which is not practical. Fortunately, computing the entire objective function is unnecessary:

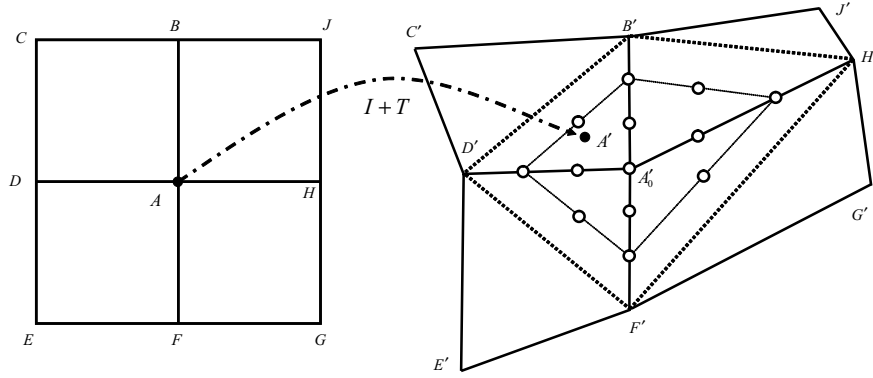


Fig. 3. Update of the value of $I + T$ at one point by local optimization. The mapping $I + T$ transforms the rectangle $BCDEFGJ$ into the deformed shape $B'C'D'E'F'G'H'$. The value $A' = (I + T)(A)$ is to be determined by minimizing the objective function $J_D(T)$ defined by (13), varying only the location of the point A' , starting from a given initial value A'_0 . First, the objective function is evaluated at several search points (empty circles) and then, starting from the best location found, it is locally optimized by coordinate descent alternating in the x and y directions, using a standard library function; we have used `fminbnd` from MATLAB. The mapping $I + T$ is defined by bilinear interpolation in each quadrant (e.g. $A'B'C'D'$), and the inverse $(I + T)^{-1}$ on the quadrant exists if and only if the quadrant is convex (Frey et al., 1978). Thus, A' is constrained within the quadrilateral $B'D'F'H'$ (thick dot lines) formed by the midpoints of the sides the deformed rectangle. The search points are constructed by putting several local grid points (here, 2) at equal distance on the segment between the initial position A'_0 and the midpoint (e.g. $A'_0 B'$), and then adding equidistant grid of search points on each of the lines that connect corresponding points, to form a triangle as shown.

changing $T(x_i, y_k)$ on the grid D_i can only influence the terms in the objective function associated with the region $D_{jk} = [x_{j-1}, y_{k-1}] \times [x_{j+1}, y_{k+1}]$, which requires only $O[n/(2^i + 1)^2]$ operations. Since there are $(2^i + 1)^2$ nodes to optimize, the cost of one optimization sweep is $O(n)$.

Recall that $m = (2^M + 1)^2$ is the number of the morphing grid points. Since the optimization on each grid cost $O(n)$ operations, smoothing on each grid costs $O(n \log n)$ operations, and there is $M = O(\log m)$ grids, the total complexity of the registration algorithm is $O(n \log m \log n)$. Thus, the method is suitable for a large number of pixels as well as a large number of nodes on the morphing grid.

4. Morphing ensemble filter

The state of the model in general consists of several gridded arrays, $U = (w, z, \dots)$. For simplicity, suppose that all arrays are defined over the same grid and that the registration is applied only to the first array, w ; this will be the case in the model application in Section 5. The general case will be discussed in Section 6.

Let $\{U_k\} = \{U_1, \dots, U_N\}$ be an ensemble of states, with the ensemble member U_k consisting of the gridded arrays

$$U_k = (w_k, z_k, \dots).$$

The subscript k in this section means the number of the ensemble member, and it is not associated with the hierarchy of grids as in Section 3. The concept of the hierarchy of grids is relevant only to the internal working of the particular automatic registration algorithm described in Section 3; here we just use the result of the registration, which is a mapping defined by gridded arrays on the morphing grid D_M .

Given one fixed state $U_0 = (w_0, z_0, \dots)$, the automatic registration (13) of the first array w defines the *registration representations* $[R_k, T_k]$ of the ensemble members as morphs of U_0 , with the registration residual

$$\begin{aligned} R_k &= (r_{w_k}, r_{z_k}, \dots), \\ r_{w_k} &= w_k \circ (I + T_k)^{-1} - w_0, \\ r_{z_k} &= z_k \circ (I + T_k)^{-1} - z_0, \\ &\vdots \end{aligned}$$

and warpings T_k determined as approximate solutions of independent optimization problems based on the state array w ,

$$\|w_k - w_0 \circ (I + T_k)\| + \|T_k\| + \|\nabla T_k\| \rightarrow \min_{T_k}.$$

The mapping T_k from the previous analysis cycle is used as the initial $\tilde{T}_k$ in the automatic registration. In our tests, this all but guarantees good registration and a speedup of one or more orders of magnitude compared to starting from zero.

Instead of EnKF operating on the ensemble $\{U_k\}$ and making linear combinations of its members, the morphing EnKF applies the EnKF algorithm to the ensemble of registration representations $\{[R_k, T_k]\}$, resulting in the analysis ensemble in registration representation, $\{[R_k^a, T_k^a]\}$, with $R_k^a = (r_{w_k}^a, r_{z_k}^a, \dots)$. The analysis ensemble is then transformed back by (11), which here becomes

$$w_k^a = (w_k + r_{w_k}^a) \circ (I + T_k^a), \quad (18)$$

$$z_k^a = (z_k + r_{z_k}^a) \circ (I + T_k^a), \quad (19)$$

$\vdots$

Remark 2. Note that the registration representations $\{[R_k^a, T_k^a]\}$ of the analysis ensemble are linear combinations of the registration representations $\{[R_k, T_k]\}$ of the forecast ensemble. Denote by λ_k the coefficients of one such linear combination; then a member of the analysis ensemble has the form

$$\left(w_0 + \sum_{k=1}^N \lambda_k r_{w_k}\right) \circ \left(I + \sum_{k=1}^N \lambda_k T_k\right), \quad (20)$$

and similarly for the other state arrays. This imposes certain constraints, for example, in general it may not be possible to write the zero state as (20), and thus the potential for amplitude corrections might be limited. This limitation does not seem to be important in the application of interest here (wildfire), and its effect will be studied elsewhere.

Given an observation function h , cf., (2), the transformed observation function for EnKF on the registration representations can be obtained directly by substituting from (18) into the observation function,

$$\tilde{h}([R, T]) = h[(w + r_w) \circ (I + T), (z + r_z) \circ (I + T), \dots].$$

However, constructing the observation function this way may not be the best. Consider the case of one point observation, such as the temperature at some location. Then the difference between the observed temperature and the value of the observation function gives little indication which way should the transformed state be adjusted. Suppose the temperature reading is high and the ensemble members have high temperature only in some small location (fireline). Then it is quite possible that the observation function (temperature at the given location) evaluated on ensemble members will miss the fire in the ensemble members completely. This is, however, a reflection of the inherent difficulty of localizing small features from point observations.

For data that is given as gridded arrays (e.g. images, or a dense array of measurements), there is a better way. Suppose the data d is a measurement of the first array in the state, w . Then, transforming the data d into its registration representation $[r_d, T_d]$ just like the registration of the state array w , the observation eq. (2) becomes the comparison between the registration representations of the data d and the state array w ,

$$\tilde{h}([R, T]) = [r_d, T_d] \approx [r_w, T]. \quad (21)$$

Data given on a part of the domain can be registered and used in the same way. Note that no manual identification of the location of the feature either in the data or in the ensemble members is needed.

5. Numerical results

We have applied the morphing EnKF to an ensemble from the wildland fire model in Mandel et al. (2006). The simulation state consists of the temperature and the fuel fraction remaining on a 500×500 m domain with a 2 m uniform grid. The model has

two state arrays, the temperature w and fuel supply z . An initial fire solution U_0 was created by raising a small square in the centre of the domain above the ignition temperature and applying a small amount of ambient wind until a fire front developed. The simulated data consisted of the whole temperature field of another solution, started from U_0 and morphed so that the fire was moved to a different location of the domain compared with the average location of the ensemble members. The observation eq. (21) was used, with Gaussian error in the registration residual of the temperature and in the registration mapping. The standard deviation of the data error was 50 K and 5 m, respectively. This large discrepancy is used to show that the morphing EnKF can be effective even when data is very different from the ensemble mean. The image registration algorithm was applied with a 17×17 morphing grid, that is, $M = 4$ refinement levels. We have performed up to five optimization sweeps, stopping if the relative improvement of the objective function for the last sweep was less than 0.001 or if the infinity norm of the residual r_w fell below 1 K. The optimization parameters used for scaling the norms in the objective function (13) were $C_1 = 10\,000 \text{ mK}^{-1}$ and $C_2 = 1000 \text{ m}^2 \text{ K}^{-1}$. For simplicity in the computation, the fuel supply variables were not included in the data assimilation. Although the fuel supply was warped spatially as in (19), the registration residual of the fuel supply, r_{z_k} , was taken to be zero.

The 50 member ensemble shown in Fig. 4 was generated by morphing the initial state U_0 using smooth random fields r_{w_k} and T_k of the form

$$\text{const} \sum_{j=1}^d \sum_{\ell=1}^d \lambda_{j,\ell} d_{j,\ell} \sin j\pi x \sin \ell\pi y \quad (22)$$

with $d_{j,\ell} \sim N(0, 1)$ and $\lambda_{j,\ell} = (1 + \sqrt{j^2 + \ell^2})^{-2}$, for each ensemble member. The constant in (22) was 50 K for the residual r_{w_k} and 5 m for the warping T_k . Since it is not guaranteed that $(I + T)^{-1}$ exists for a smooth random T , we have tested if $I + T$ is one to one and generated another random T if not. The resulting 17×17 and 250×250 matrices are appended to form $17^2 + 250^2$ element vectors representing an ensemble state $[R_k, T_k]$ for the EnKF. The same state U_0 was advanced in time along with the ensemble. (Of course, other choices of U_0 are possible.)

The ensemble was advanced in 3 min analysis cycles. The new registration representation $[r_k, T_k]$ was then calculated using the previous analysis values as an initial guess and incremented by EnKF. The ensemble after the first analysis cycle is shown in Fig. 4. The results after five analysis cycles were similar, indicating no filter divergence (Fig. 6). Numerical results indicate that the error distribution of the registration representation is much closer to Gaussian than the error distribution of the temperature itself. This is demonstrated in Fig. 5, where the estimated probability density functions for the temperature, the registration residual of the temperature, and the registration mapping for Fig. 6 are computed at a single point in the domain using a Gaussian kernel with bandwidth 0.3 times the sample standard

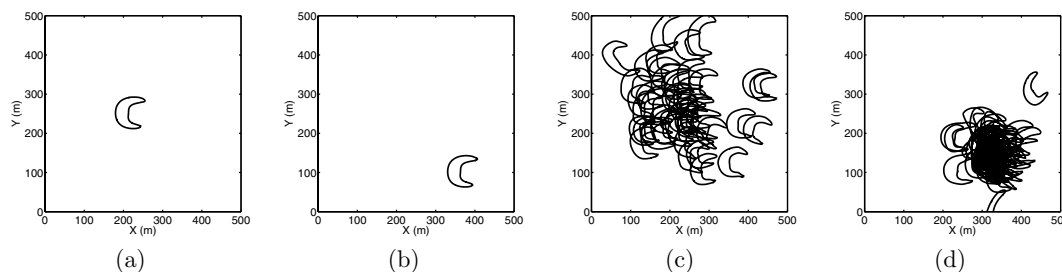


Fig. 4. Data assimilation by the morphing ensemble filter. The forecast ensemble (b) was created by smooth random morphing of the initial temperature profile (a). The analysis ensemble (d) was obtained by the EnKF applied to the transformed state. The data for the EnKF was the morphing transformation of the simulated data (c), and the observation function was the identity mapping. Contours are at 800 K, indicating the location of the fireline. The reaction zone is approximately between the two curves.

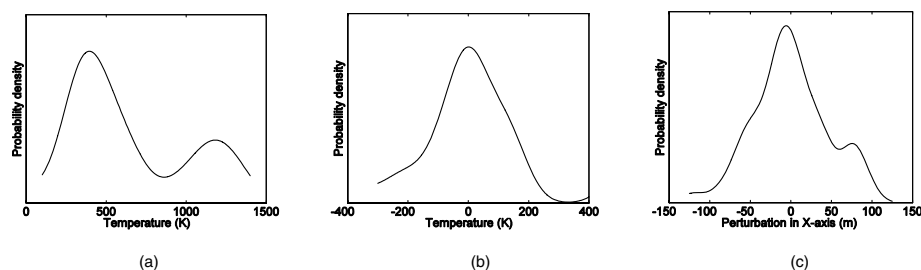


Fig. 5. Probability densities estimated by a Gaussian kernel with bandwidths 37 K, 19 K and 30 m. Data was collected from the ensemble shown in Fig. 5c. Displayed are typical marginal probability densities at a point near the reaction area of the original temperature (a), the registration residual of temperature (b) and the registration mapping component in the x direction (c). The transformation has vastly improved Gaussian nature of the densities.

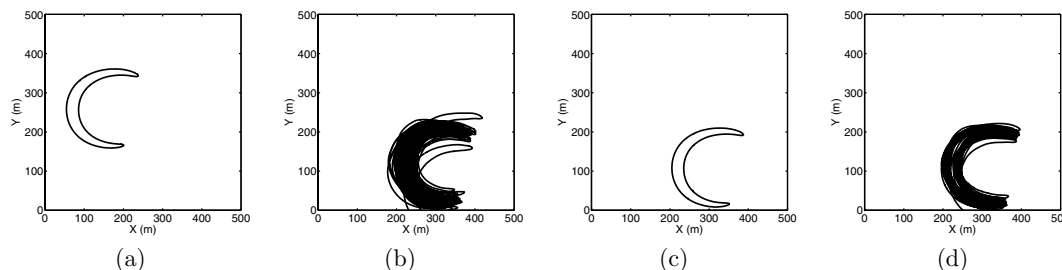


Fig. 6. After five analysis cycles, the ensemble shows less spread and follows the data reliably. Ensemble members were registered using the initial state, advanced in time without data assimilation (a). The forecast ensemble (b) is closer to the simulated data (c) because of preceding analysis steps that have attracted the ensemble to the truth. The analysis ensemble (d) has a little less spread than the forecast, and the change between the forecast and the analysis is well within the ensemble spread. Contours are at 800 K, indicating the location of the fireline. The reaction zone is approximately between the two curves.

deviation. The point was chosen to be on the boundary of the fire shape, where the non-Gaussianity may be expected to be the strongest. In Fig. 7, the Anderson–Darling test for normality was applied to each point on the domain for the analysis step from Fig. 6. The resulting p -values were plotted on their corresponding locations in the domain with darkness determined on a log scale with black as 10^{-8} (highly non-Gaussian) and white 1 (highly Gaussian). While the Anderson–Darling test is intended to be a hypothesis test for normality, it is used here to visualize on a continuous scale the closeness to normality of the marginal probability distribution any point in the domain. Again, strongest

departure from normality of the distribution is seen around the fire.

The implementation was a prototype done in Matlab. Therefore, we do not report timings.

6. Conclusion

The numerical results show that the morphing EnKF is useful for a highly non-linear problem (a model problem for wildfire simulation) with a coherent spatial feature of the solution (propagating fireline). In previous work (Johns and Mandel, 2005;

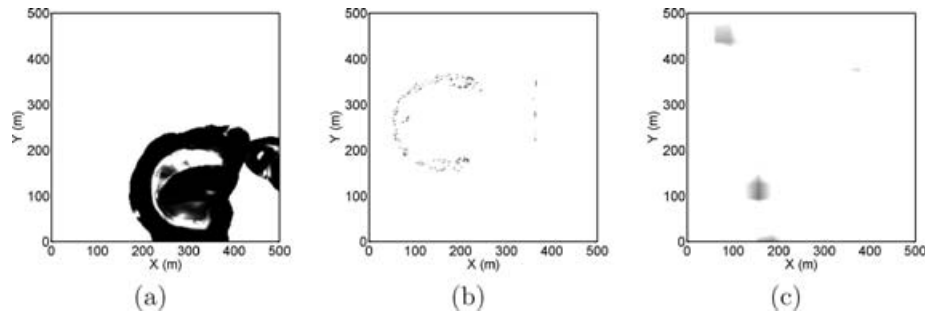


Fig. 7. The p -value from the Anderson–Darling test of the data from the ensemble after five morphing EnKF analysis cycles shows the ensemble transformed into its registration representation, the registration residual of the temperature (b) and the registration mapping (c), has distribution much closer to Gaussian than the original ensemble (a). The shading at a point indicates where the marginal distribution of the ensemble at that point is highly Gaussian (white) or highly non-Gaussian (black).

Mandel et al., 2006), we have used penalization of non-physical solutions, but the location of the fire in the data could not be too far from the location in the ensemble, artificial perturbation had to be added to retain the spread of the ensemble, and the penalty constant, the amount of additional spread, and the data variance had to be finely tuned for acceptable results. This new method does not appear to have the same limitations. The registration works automatically from gridded data and no objects need to be specified. The difference between the feature location in the data and in the ensemble can be large and the data variance can be as small as necessary, without causing filter divergence. One essential limitation is that the registered images need to be sufficiently similar, and the registration mapping should be sufficiently similar to the initial guess. This will eventually impose a limitation on how long can the ensemble go without an analysis step. However, compared to previous results for the same problem (Johns and Mandel, 2005; Mandel et al., 2006), the convergence of the present filter is much better.

It was shown that the number of operations grows almost linearly with the number of degrees of freedom, so the method is suitable for very large problems. The method may be useful in wildfire modelling as well as in data assimilation for other problems with strongly non-Gaussian distributions and moving coherent features, such as rain fronts or hurricane vortices. This paper only presents the basic method and reports on results for a highly non-linear, but still a fairly simple model problem. Further enhancements, needed for practical applications, such as general atmospheric science problems and coupled wildfire and atmosphere models, will be studied elsewhere. These enhancements should include the following.

In registration (Section 3), small residual can be forced on a smaller domain than the whole domain to allow a global shift and rotation of the image. This will be important in weather applications, where there is no solid background; the ambient temperature served as the background in the wildfire model tested here. The registration method guarantees that the inverse $(I + T)^{-1}$ exists, but the derivatives of the inverse could be very large, resulting in a loss of stability. Therefore, the inverse should be

involved in the objective function. For example, penalty functions can be used to force in the local optimization step the value of $T(x_j, y_k)$ to stay well inside the region where $(I + T)^{-1}$ exists, or the reciprocal of the Jacobian of $I + T$ or the norm of the inverse $(I + T)^{-1}$, multiplied by a constant, can be added as a penalty term to the objective function directly. Also, the registration does not treat the input images u and v in the same way; an algorithm that is symmetric with respect to swapping the input images would take care of the issue automatically. Finally, the registration mapping is piecewise bilinear, and therefore the morphed state will have kinks—not very good for differential equations—unless the morphing grid is as fine as the pixel grid. In order to be able to use a relatively coarse morphing grid (and thus cheap registration), one could use smooth interpolation, such as B -splines, used in the image registration context, for example, by Arganda-Carreras et al. (2006).

The registration might be further improved by updating the registration mapping more often, whether there are new data or not, and thus assuring that the initial guess for the registration algorithm is always sufficiently close. Also, the common state U_0 to register the ensemble members against has been evolved from one initial condition regardless of the data. Over time, such U_0 could diverge significantly from the ensemble (which tracks the data), resulting in more strenuous registration. If this becomes an issue, a better U_0 might be constructed from the analysis directly (perhaps as the mean of registration representations of the analysis ensemble members) and then evolved in time until the next analysis step.

When some ensemble members totally miss the feature (e.g. the fire), the registration mapping does not matter much and all error will be in the registration residual. This is not a problem, because those ensemble members have low data likelihood, so they do not influence the posterior pdf much. They do affect the ensemble mean and covariance, so the EnKF analysis might change significantly if too many members miss the feature. Currently, the method was tested for the case when there is a single significant feature (the fire), which is essentially characterized by its position and strength. Further research will be

needed to deal with more general cases. The method will need to be generalized to use more state arrays for the registration at the same time, work with nested grids, and perhaps use different registration mappings applied to different fields, for example, in a coupled atmosphere-fire model, the state of the atmosphere and the state of the fire might require different position adjustments. 3-D registration might be needed for atmospheric problems, especially those with strong buoyancy as over a wildfire.

7. Acknowledgments

This research was supported by the National Science Foundation under grants CNS-0325314 and CNS-0719641, and by NCAR Faculty Fellowship.

References

- Alexander, G. D., Weinman, J. A. and Schols, J. L. 1998. The use of digital warping of microwave integrated water vapor imagery to improve forecasts of marine extratropical cyclones. *Mon. Wea. Rev.* **126**, 1469–1496.
- Anderson, J. L. 2003. A local least squares framework for ensemble filtering. *Mon. Wea. Rev.* **131**, 634–642.
- Arganda-Carreras, I., Ó. Sánchez Sorzano, C., Marabini, R., Carazo, J. M., Ortiz-de Solorzano, C. and co-authors. 2006. Consistent and elastic registration of histological sections using vector-spline regularization. *Computer Vision Approaches to Medical Image Analysis* Springer Berlin/Heidelberg, volume 4241 of *Lecture Notes in Computer Science* 85–95.
- Brown, L. G. 1992. A survey of image registration techniques. *ACM Comput. Surveys* **24**, 325–376.
- Burgers, G., van Leeuwen, P. J. and Evensen, G. 1998. Analysis scheme in the ensemble Kalman filter. *Mon. Wea. Rev.* **126**, 1719–1724.
- Chen, Y. and Snyder, C. 2007. Assimilating vortex position with an ensemble Kalman filter. *Mon. Wea. Rev.* **135**, 1828–1845.
- Darema, F., 2004. Dynamic data driven applications systems: A new paradigm for application simulations and measurements. In: *Computational Science-ICCS 2004: 4th International Conference* (eds M. Bubak, G. D. van Albada, P. M. A. Soot, and J. J. Dongarra), Springer, volume 3038 of *Lecture Notes in Computer Science* 662–669.
- Davis, C., Brown, B. and Bullock, R. 2006. Object-based verification of precipitation forecasts. Part I: methodology and application to mesoscale rain areas. *Mon. Wea. Rev.* **134**, 1772–1784.
- Evensen, G., 1994. Sequential data assimilation with nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.* **99**(C5), 143–162.
- Evensen, G. 2003. The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dyn.* **53**, 343–367.
- Evensen, G. 2007. *Data Assimilation: The Ensemble Kalman Filter*. Springer, Berlin.
- Frey, A. E., Hall, C. A. and Porsching, T. A. 1978. Some results on the global inversion of bilinear and quadratic isoparametric finite element transformations. *Math. Comput.* **32**, 725–749.
- Gao, P. and Sederberg, T. W. 1998. A work minimization approach to image morphing. *Visual Comput.* **14**, 390–400.
- Hoffman, R. N., Liu, Z., Louis, J.-F. and Grassotti, C. 1995. Distortion representation of forecast errors. *Mon. Wea. Rev.* **123**, 2758–2770.
- Houtekamer, P. and Mitchell, H. L. 1998. Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.* **126**, 796–811.
- Johns, C. J. and Mandel, J. 2005. A two-stage ensemble Kalman filter for smooth data assimilation. Environmental and Ecological Statistics, in print. CCM Report 221, <http://www.math.cudenver.edu/ccm/reports/rep221.pdf>, conference on New Developments of Statistical Analysis in Wildlife, Fisheries, and Ecological Research, Oct 13–16, 2004, Columbia, MI.
- Kalman, R. E. 1960. A new approach to linear filtering and prediction problems. *Trans. ASME-J. Basic Eng., Ser. D* **82**, 35–45.
- Kalnay, E. 2003. *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press, Cambridge, UK.
- Lawson, W. G. and Hansen, J. A. 2005. Alignment error models and ensemble-based data assimilation. *Mon. Wea. Rev.* **133**, 1687–1709.
- Mandel, J., Beezley, J. D., Bennethum, L. S., Soham Chakraborty, J. L. C., Douglas, C. C. and co-authors. 2007. A dynamic data driven wildland fire model. In: *Computational Science-ICCS 2007: 7th International Conference* (eds Y. Shi, G. D. van Albada, P. M. A. Soot and J. J. Dongarra), Springer, volume 4487 of *Lecture Notes in Computer Science* 1042–1049.
- Mandel, J., Bennethum, L. S., Beezley, J. D., Coen, J. L., Douglas, C. C. and co-authors. 2006. A wildfire model with data assimilation. CCM Report 233, <http://www.math.cudenver.edu/ccm/reports>.
- Ott, E., Hunt, B. R., Szunyogh, I., Zimin, A. V., Kostelich, E. J. and co-authors. 2004. A local ensemble Kalman filter for atmospheric data assimilation. *Tellus* **56A**, 415–428.
- Ravela, S., Emanuel, K. A. and McLaughlin, D. 2007. Data assimilation by field alignment. *Physica D* **230**, 127–145.
- Tippett, M. K., Anderson, J. L., Bishop, C. H., Hamill, T. M. and Whitaker, J. S. 2003. Ensemble square root filters. *Mon. Wea. Rev.* **131**, 1485–1490.