Amplitude-Position Formulation of Data Assimilation⋆

Sai Ravela

Earth, Atmospheric and Planetary Sciences &
Computer Science and Artificial Intelligence Laboratory
Massachusetts Institute of Technology
`ravela@mit.edu`

Abstract. Classical formulations of data-assimilation perform poorly when forecast locations of weather systems are displaced from their observations. They compensate position errors by adjusting amplitudes, which can produce unacceptably "distorted" states, adversely affecting analysis, verification and subsequent forecasts. It is non-trivial to identify sources of position error, but correcting misplaced forecasts is essential for operationally predicting strong, localized weather events such as tropical cyclones. In this paper, we propose a method that accounts for both position and amplitude errors. The proposed method assimilates observations in two steps. The first step is *field alignment*, where the current model state is aligned with observations by adjusting a continuous field of local displacements, subject to certain constraints. The second step is amplitude adjustment, where contemporary assimilation approaches are used. Our method shows improvements in analyses, with sparse and uncertain observations.

1 Introduction

Environmental data assimilation is the methodology for combining imperfect model predictions with uncertain data in a way that acknowledges their respective uncertainties. Data assimilation can be used to continually correct imperfect forecasts as well as guide a dynamic data acquisition procedure by indicating where and when new observations will be most helpful. However, data assimilation can only work when the estimation process properly represents all sources of error.

The difficulties created by improperly represented error are particularly apparent in mesoscale meteorological phenomena such as thunderstorms, squalllines, hurricanes, precipitation, and fronts. Errors in mesoscale models can arise in many ways but they often manifest themselves as errors in the position of these features. Such a problem can be seen in Figure 1, which shows predicted and observed tracks for hurricane KATE. The various operational forecasts shown of the left side of this figure all contain significant position errors, giving trajectories that move in completely different directions. We would like to use data

---

assimilation to dynamically correct inaccurate predictions such as the hurricane KATE trajectories. But we typically don't know where these errors arise or exactly how they affect position. If we could attribute position error to a single (or even a small number) of incorrectly specified model inputs we could adjust these inputs using a classical parameter estimation procedure. It is, however, more likely that the position errors we observe are the aggregate result of errors in parameter values, initial conditions, and boundary conditions.

Since it is generally not possible to isolate the sources of position error in this way operational forecasters resort to ad hoc position adjustment procedures. One such procedure is commonly called *bogussing* [4], which removes the incorrectly positioned cyclone forecast and inserts a new "bogus" cyclone at the correct location, estimated from satellite imagery. Bogussing seems simple but it creates a host of problems. The bogus cyclone can introduce significant shocks in the dynamical system and if the original incorrectly positioned cyclone is not properly removed the resulting "ghost" vortex can severely degrade the track forecast.

A more sophisticated alternative to bogussing is to use data assimilation methods. Unfortunately, sequential [6], ensemble-based [5] and variational [9, 2] state estimation methods used in data assimilation applications adjust amplitudes to deal with position error. Adjusting amplitudes doesn't really fix position error, and instead, can produce unacceptably distorted estimates. These can, in turn, adversely affect the quality of subsequent forecasts [8, 10].

To see why this is, consider a Bayesian formulation of the data assimilation problem [7]. The state vector X_n at a discrete time t_n can be estimated using all measurements $Y_{0:n}$ from an associated conditional probability density $P(X_n|Y_{0:n})$. If we suppose that the dynamics is Markov in time, that is $P(X_n|X_{0:n-1}) = P(X_n|X_{n-1})$ and the observation errors are uncorrelated in time, and therefore $P(Y_{0:n}) = \prod_{i=0}^{n} P(Y_i)$, then $P(X_n|Y_{0:n})$ can be evaluated as:
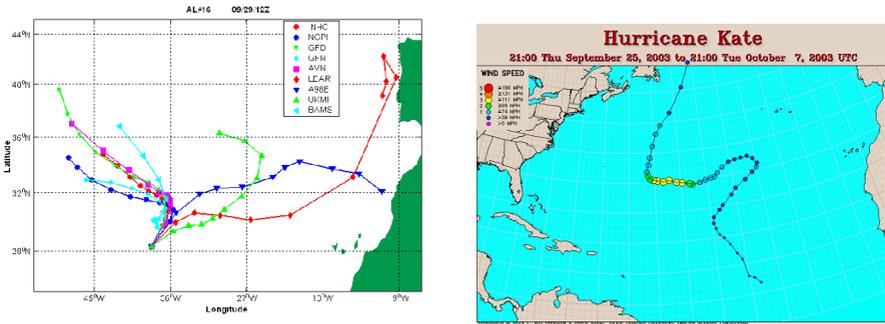$$P(X_n|Y_{0:n}) \propto P(Y_n|X_n)P(X_n|Y_{0:n-1}).$$

Fig. 1. The left picture shows forecasts of hurricane KATE launched on 09/29/2003 at 12Z using several operational models. The right picture shows its true path. The errors in positions are significant.

To simplify the ensuing discussion, we adopt a linear observation model, $Y_n = HX_n + \eta$, where η is additive measurement noise and uncorrelated in space and time, and H is a linear observation operator. Further, we drop the explicit dependence on time by writing $X = X_n$ and observation $Y = Y_n$, and the forecast $X^f = X_n|Y_{0:n-1}$. Then, if the distributions in question are assumed Gaussian, with $P(X^f) \sim N(X, B)$ and $\eta \sim N(0, R)$, the mean of $P(X_n|Y_{0:n})$ is equal to its mode, which is the value of X obtained from the solution: $X = X^f + BH^T(HBH^T + R)^{-1}(Y - HX^f)$.

This equation can be interpreted deterministically, where Y and X^f are fixed vectors called the observation and first guess, and X is the estimated state. The matrices B and R are the respective uncertainties in state and observations. As shown, these equations are a simplified version (due to the linear observation operator) of what is commonly known as "3DVAR" in the meteorological community [2, 9].

It can also be interpreted probabilistically, which forms the basis for the Ensemble Kalman Filter [5]. In this case, an ensemble of estimates at time t_{n-1} are forecast to time t_n using the model. Let us call the forecast ensemble $A^f = [X_1^f \ldots X_S^f]$, where the columns of A^f are the S replicates of the ensemble. To implement the Ensemble Kalman Filter, we will assume that observation equation is linear (as before) with spatially and temporally uncorrelated additive noise. We let Z represent a matrix of perturbed observations, $\tilde{A}^f$ be the deviation from mean $\bar{A}^f$ of A^f, and write: $A = A^f + \tilde{A}^f(H\tilde{A}^f)^T \left[(H\tilde{A}^f)(H\tilde{A}^f)^T + R\right]^{-1} (Z - HA^f)$.

Both these methods perform equally poorly in the presence of position errors, because position errors introduce bias and/or degrade the forecast amplitude error-covariance. To see this, suppose that we have forecasts X_i, and $i = 1 \ldots S$, which we deem a perfect Gaussian ensemble. Without loss of generality, we will represent this ensemble by a random variable X. We think of X as a spatially one-dimensional vector quantity, also referenced by a position variable p as $X(p)$. Let us suppose that $X \sim N(\bar{X}, B)$. That is, the covariance $B = E_X[\tilde{X}\tilde{X}^T]$, where E_X is the expectation under the distribution of X and $\tilde{X} = X - \bar{X}$. Let us consider a scalar position perturbation λ to this perfect forecast distribution, that is $X(p + \lambda)$. We now expand this via Taylor series (assuming the expansion is possible) as $X(p + \lambda) = X(p) + \lambda\frac{\partial X}{\partial p}$. The amplitude mean under the position perturbation can be written as $\bar{X}_\lambda = E_X[X(p + \lambda)] = \bar{X} + \lambda E_X[\frac{\partial X}{\partial p}]$. Further, we define the deviation of the gradients as $\Delta = \frac{\partial X}{\partial p} - E[\frac{\partial X}{\partial p}]$. Then the covariance as a result of the position perturbation is $C = B + \lambda E_X[\tilde{X}\Delta^T] + \lambda E_X[\Delta\tilde{X}^T] + \lambda^2 E_X[\Delta\Delta^T]$. If λ is assumed to be a Gaussian random variable with mean 0 and standard deviation σ_λ, then we can say that the expected amplitude covariance under the position perturbation is $E_\lambda[C] = B + \sigma_\lambda^2 E_X[\Delta\Delta^T]$. We define $C_\lambda = E_\lambda[C]$ and define $C_{\Delta\Delta} = E_X[\Delta\Delta^T]$ and therefore, we have $C_\lambda = B + \sigma_\lambda^2 C_{\Delta\Delta}$.

This equation says that the expected amplitude-covariance as a result of random position perturbations is less certain or "inflated" than without. Thus it is immaterial whether position errors show up in ensemble forecasts as a result of

model error, or they are introduced to account for model error in the forecast uncertainty. In either case, it weakens the forecast uncertainty, causing poor estimates. Therefore, rather than try to deal with position error indirectly, only through adjustments of amplitudes, we must also adjust positions! In particular, we can determine how the values predicted at model grid points need to be adjusted in amplitude and moved in space in order to achieve a better fit to observations. This perspective is the key our approach. The solution is general and applies equally well to meteorological, oceanographic, and hydrological applications.

2   Data Assimilation by Field Alignment

We reformulate the classical assimilation objective to allow position and amplitude adjustments. To make this framework more explicit it is useful to introduce some notation. Let $X = X(\mathbf{r}) = \{X[\underline{r}_1^T] \ldots X[\underline{r}_m^T]\}$ be the model-state vector defined over a spatially discretized computational grid Ω, and $\mathbf{r^T} = \{\underline{r}_i = (x_i, y_i)^T, i \in \Omega\}$ be the position indices. Similarly, let $\mathbf{q}$ be a *vector* of displacements. That is, $\mathbf{q^T} = \{\underline{q}_i = (\Delta x_i, \Delta y_i)^T, i \in \Omega\}$. Then the notation $X(\mathbf{r} - \mathbf{q})$ represents *displacement* of X by $\mathbf{q}$. The displacement field $\mathbf{q}$ is real-valued, so $X(\mathbf{r} - \mathbf{q})$ must be evaluated by interpolation if necessary.

In a probabilistic sense, we may suppose that finding $(X, \mathbf{q})$ that has the maximum a posteriori probability in the distribution $P(X, \mathbf{q}|Y)$ is appropriate. Using Bayes rule we obtain $P(X, \mathbf{q}|Y) \propto P(Y|X, \mathbf{q})P(X^f|\mathbf{q})P(\mathbf{q})$. As before, we assume a linear observation model with uncorrelated noise in space and time, and Markov dynamics. If we make a Gaussian assumption of the component densities, we can write:

$$P(Y|X, \mathbf{q}) = \frac{1}{(2\pi)^{\frac{n}{2}} |R|^{\frac{1}{2}}} e^{-\frac{1}{2}(Y - H\,X(\mathbf{r}-\mathbf{q}))^T R^{-1}(Y - H\,X(\mathbf{r}-\mathbf{q}))} \qquad (1)$$

This equation is the data-likelihood term. It implies that the observations can be related using a Gaussian model to the displaced state X(**r**- **q**), where X(**r**) is defined on the original grid, and **q** is a displacement field. We use the linear observation model here, and therefore, $Y = HX(\mathbf{r} - \mathbf{q}) + \eta, \eta \sim N(0, R)$.. We should emphasize here that the observation vector is fixed. It's elements are always defined from the original grid.

$$P(X^f|\mathbf{q}) = \frac{1}{(2\pi)^{\frac{n}{2}} |B(\mathbf{q})|^{\frac{1}{2}}} e^{-\frac{1}{2}\left(X(\mathbf{r}-\mathbf{q}) - X^f(\mathbf{r}-\mathbf{q})\right)^T B(\mathbf{q})^{-1}\left(X(\mathbf{r}-\mathbf{q}) - X^f(\mathbf{r}-\mathbf{q})\right)} \quad (2)$$

This equation defines the *amplitude prior*. Given a fixed displacement field **q** and a forecast $X^f(\mathbf{r})$ defined on the original grid, it states that the forecast distribution is assumed to be Gaussian in the position-corrected space, even if it isn't in the uncorrected space. Once we assume Gaussian statistics in a position corrected space, it is immediately clear that the forecast statistics are conditioned

on the displacement field. In particular, its second moment, the covariance B, is dependent on $\mathbf{q}$. By simple analogy, if we associate an error covariance on the original forecast grid, this error covariance will have to be remapped when the forecast grid is deformed. Thus, we write the forecast covariance as $B(\mathbf{q})$.

$$P(\mathbf{q}) = \frac{1}{C} e^{-L(\mathbf{q})} \tag{3}$$

This equation specifies a *displacement prior*. This prior is constructed from an energy function $L(\mathbf{q})$ which expresses constraints on the displacement field. The proposed method for constructing L is drawn from the nature of the expected displacement field. Displacements can be represented as smooth flow fields in many fluid flows and often arise from systematic and large-scale1 *background flow* errors, fore example see [1]. Smoothness naturally leads to a Tikhonov type formulation [11] and, in particular, $L(\mathbf{q})$ is designed as a gradient and a divergence penalty term. These constraints, expressed in quadratic form are:

$$L(\mathbf{q}) = \frac{w_1}{2} \sum_{j \in \Omega} \mathbf{tr}\{[\nabla \underline{q}_j][\nabla \underline{q}_j]^T\} + \frac{w_2}{2} \sum_{j \in \Omega} [\nabla \cdot \underline{q}_j]^2 \tag{4}$$

In Equation 4, $\mathbf{q}_j$ refers to the jth grid index and $\mathbf{tr}$ is the trace. Equation 4 is a *weak constraint*, weighted by the corresponding weights w_1 and w_2. Note that the constant C can be defined to make Equation 3 a proper probability density. In particular, define $Z(\mathbf{q}) = e^{-L(\mathbf{q})}$ and define $C = \int_{\mathbf{q}} Z(\mathbf{q})d\mathbf{q}$. This integral exists and converges.

With these definitions of probabilities, we are in a position to construct an objective by evaluating the log probability. To develop a solution easily, we make use of a statistical representation of uncertainty, using an *ensemble of forecast states* (this assumption is relaxed later). So suppose that S samples $\mathbf{X} = X_1 \ldots X_S$ are to be estimated along with associated displacements $Q = \mathbf{q_1}, \ldots \mathbf{q_S}$, from S forecasts $X_s^f, s = 1 \ldots S$. Let $\mathbf{p_s} = \mathbf{r} - \mathbf{q_s}$ and $\bar{X}^f = \frac{1}{S} \sum_{s=1}^{S} X_s^f(\mathbf{p_s})$. The background error covariance is:

$$B_Q = B(\mathbf{X^f}; Q) = \frac{1}{S-1} \sum_{s=1}^{S} (X_s^f(\mathbf{p_s}) - \bar{X}^f)(X^f(\mathbf{p_s}) - \bar{X}^f)^T \tag{5}$$

The ensemble framework leads to a quadratic objective from the log-likelihood, which we can write as $J(\mathbf{X}, Q) = \frac{1}{S} \sum_{s=1}^{S} J_s(\mathbf{X}, Q)$, where J_s is defined as, *for s =* $1 \ldots S$

$$J_s(\mathbf{X}, Q) = \frac{1}{2} \left(X_s(\mathbf{p_s}) - X_s^f(\mathbf{p_s}) \right)^T B_Q^{-1} \left(X_s(\mathbf{p_s}) - X_s^f(\mathbf{p_s}) \right)$$
$$+ \frac{1}{2} (Y_s - H X_s(\mathbf{p_s}))^T R^{-1} (Y_s - H X_s(\mathbf{p_s}))$$
$$+ L(\mathbf{q_s}) - ln(|B_Q|) \tag{6}$$

We propose a solution using Euler-Lagrange equations. These can be written as:

$$\frac{\partial J}{\partial \mathbf{q_s}} = (\nabla X_s|_{\mathbf{p_s}} - \nabla X_s^f|_{\mathbf{p_s}})^T B_Q^{-1} \left(X_s\left(\mathbf{p_s}\right) - X_s^f\left(\mathbf{p_s}\right)\right)$$

$$+ \nabla X_s|_{\mathbf{p_s}} H^T R^{-1} \left(H\ X_s\left(\mathbf{p_s}\right) - Y_s\right) + \frac{\partial L}{\partial \mathbf{q_s}} = 0 \qquad (7)$$

$$\frac{\partial J}{\partial X_s} = B_Q^{-1} \left(X_s(\mathbf{p_s}) - X_s^f(\mathbf{p_s})\right) + H^T R^{-1} \left(H\ X_s\left(\mathbf{p_s}\right) - Y_s\right) = 0 \qquad (8)$$

These equations are highly non-linear functions in X_s and $\mathbf{q_s}$. We solve them sequentially, in exactly two steps. In the first step we fix X_s to X_s^f in Equation 7 and solve for $\hat{Q}$. We then use this solution to solve Equation 8. The sequential approach is a *two-step* approximation. The first-step is the the displacement or alignment equation, written as:

$$\frac{\partial L}{\partial \mathbf{q_s}} + \nabla X_s^{fT}|_{\mathbf{p_s}} H^T R^{-1} \left(H\ X_s^f\left(\mathbf{p_s}\right) - Y_s\right) = 0 \qquad (9)$$

Using the regularization constraints the alignment equation at a node i now becomes:

$$w_1 \nabla^2 \underline{q}_{s,i} + w_2 \nabla(\nabla \cdot \underline{q}_{s,i}) + \left[\nabla X_s^{fT}|_{\mathbf{p_s}} H^T R^{-1} \left(H\left[X_s^f\left(\mathbf{p_s}\right)\right] - Y_s\right)\right]_i = 0 \quad (10)$$

Equation 10 is the field alignment formulation. It introduces a forcing based on the residual between the model- and observation-fields. The constraints on the displacement field allow the forcing to propagate to a consistent solution. Equation 10 is also non-linear, and is solved iteratively, as a Poisson equation. During each iteration $\mathbf{q_s}$ is computed by holding the forcing term constant. The estimate of displacement at each iteration is then used to deform a copy of the original forecast model-field using bi-cubic interpolation for the next iteration. The process is repeated until a small displacement residual is obtained, the misfit with observations does not improve, or an iteration limit is reached. Upon convergence, we have an aligned forecast ensemble $X_s^f(\hat{\mathbf{p}}_\mathbf{s})$, $s = 1 \ldots S$, from which $B_{\hat{Q}}$ is computed. With these quantities, the amplitude recovery is written as:

$$X_s(\hat{\mathbf{p}}_\mathbf{s}) = X_s^f(\hat{\mathbf{p}}_\mathbf{s}) + B_{\hat{Q}} H^T (H B_{\hat{Q}} H^T + R)^{-1} \left(Y_s - H\ X_s^f(\hat{\mathbf{p}}_\mathbf{s})\right) \qquad (11)$$

Equation 11 can be implemented using several familiar schemes. Here we outline three:

Ensemble Scheme: Equation 11 can be an ensemble Kalman filter, because $\hat{Q}$ is fixed. Of course, several EnKF schemes can be implemented using Equation 11.

Deterministic Ensemble Scheme: The background can be computed from individual, aligned forecast replicates. Aligned replicates can be used to compute a statistical background error covariance. Once this is done, deterministic assimilation proceeds from the background.

Purely Deterministic Scheme: The proposed two-step method can also be used when a forecast ensemble is not available. Since the alignment does not depend on the background error covariance, one can align the first-guess field and use the aligned first-guess to craft a state-dependent background error covariance [3]. In essence, the same procedure that would be used with the unaligned first-guess can be also used with the aligned first-guess.

3   Example

Figure 3a(A) depicts the vorticity contours of a true vortex field of size 32×32. Figure 3a(B) depicts observed locations (+), and the observation field is gen-
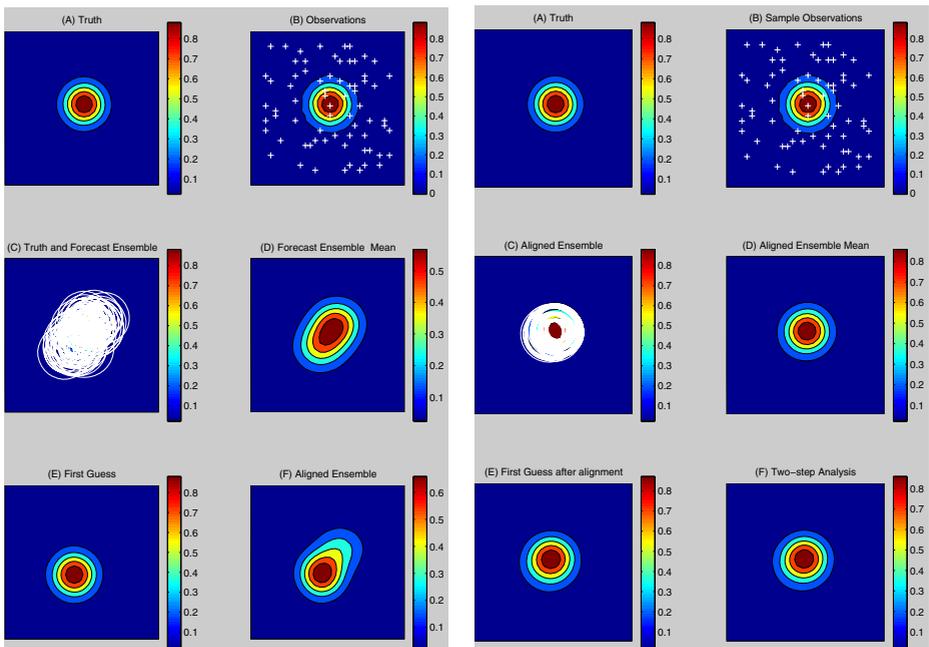
(a) 3DVAR: Panel (A) is a true vortex field, observed at sparse locations (+ sign)shown in panel (B). Panel (C) shows an overlay of the forecast ensemble (white rings). Panel (D) depicts the forecast ensemble mean. The first guess is selected to be a coherent vortex of similar structure as the observations from the ensemble, as shown in Panel (E). Panel (F) depicts the result of a deterministic assimilation scheme.

(b) Two-step method: Panel (A) and (B) are the same as Figure 3a. Panel (C) is the the result of field alignment. This produces a better forecast ensemble mean shown in panel (D). The first guess is selected from the ensemble as shown in panel (E). Panel (F) depicts the result of a deterministic assimilation scheme after alignment.

Fig. 2. Comparison of 3DVAR and two-step approaches

erated by adding to truth a 1% (of peak amplitude) uncorrelated noise field in space. Only 6.8% of the state is observed (70 observations), at randomly selected points. In particular, about 12 observations lie on significant contour levels of the vortex.

Figure 3a(C) depicts the contours (white) of a forecast ensemble (25 ensemble members) overlaid on truth. They completely "cover" truth, suggesting that truth is well-captured by the ensemble members. However, all the ensemble members have position errors from truth as well as amplitude errors. Their peak amplitude can vary by as much as 10% and their position by 10 pixels. Figure 3a(D) depicts the forecast ensemble mean, which is broader than the true vortex. Figure 3a(E) depicts the first-guess used for a 3DVAR procedure. The background error covariance (B) is computed from the forecast ensemble, and the observation uncertainty is constructed from a 0.01 standard deviation, iid. Figure 3a(F) depicts the analysis produced by 3DVAR. The smearing of the vortex can be explained by the fact that position errors produce an artificially broad (or more uncertain) background error covariance.

Figure 3b depicts the same example using the proposed two-step approach. Figure 3b(A) and Figure 3b(B) are identical to corresponding plots in Figure 3a. Figure 3b(C) depicts the *aligned* ensemble, using the alignment formulation developed in the previous section. In particular, equation 10 is used to align each ensemble member with a perturbed observation (of which Figure 3b(B) is one sample). The ensemble mean after alignment is shown in Figure 3b(D). An ensemble member is is selected as the first guess for 3DVAR, as shown in Figure 3b(E). The analysis is shown in Figure 3b(F). The reason why the two-step approach works is because the aligned ensemble may be considered to be a far better depiction of background uncertainty than the one containing position errors. Experiments using the EnKF to produce the analysis leaves us with the same conclusions, as the one-dimensional examples also show. These are not repeated here.

4   Conclusions

Our method is a Bayesian perspective of the assimilation problem that simultaneously considers position and amplitude errors. Our approach has several distinct advantages: (a) It is useful for a wide range of environmental assimilation problems because a more general description of error than current methods. (b) Our approach does not require features to be identified for correcting position errors. This is a significant advantage because features cannot always be clearly delineated from sparse observations and they are not explicitly defined in the Eulerian models most commonly used for environmental forecasting. (c) our approach can be integrated easily with current operational implementations, thereby making this effort more likely to have a real impact. (d) We have examined extensions of the two-step algorithm to multivariate fields and 3D fields. In the former case, we see that alignment does not perturb a preexisting dynamical balance. Finally, the regularization constraint in field alignment is a weak

constraint and the weights determine how strongly the constraints influence the flow field. The constraint in L is modeled as such because we expect the fluid flow to be smooth. From a regularization point of view, there can be other choices [12] as well.

References

1. G. D. Alexander, J. A. Weinman, and J. L. Schols. The use of digital warping of microwave integrated water vapor imagery to improve forecasts of marine extratropical cyclones. *Monthly Weather Review*, 126:1469–1495, June 1998.
2. P. Courtier. Variational methods. *J. Meteor. Soc. Japan*, 75, 1997.
3. R. Daley. *Atmospheric Data Analysis*. Cambridge University Press, 1994.
4. C. Davis and S. Low-Nam. The ncar-afwa tropical cyclone bogussing scheme. *Technical Memorandum, Air Force Weather Agency (AFWA), Omaha, NE, [http://www.mmm.ucar.edu/mm5/mm5v3/tc-report.pdf]*, 2001.
5. G. Evensen. The ensemble kalman filter: Theoretical formulation and practical implementation. *Ocean Dynamics*, 53:342–367, 2003.
6. A. Gelb. *Applied Optimal Estimation*. MIT Press, 1974.
7. A. H. Jazwinski. *Stochastic Processes and Filtering Theory*. Academic Press, 1970.
8. C. D. Jones and B. MacPherson. A latent heat nudging scheme for assimilation of precipitation data into an operation mesoscale model. *Meteorol. Appl.*, pages 269–277, 1997.
9. A. C. Lorenc. Analysis method for numerical weather predictin. *Q. J. R. Meteorol. Soc.*, 112:1177–1194, 1986.
10. H. J. Thiebaux, P. R. Julian, and G. J. DiMego. Areal versus collocation data quality control. In *Intl. Symp. on Assimilation of Observations in Meteorolgy and Oceanography*, pages 255–260, Clermont-Ferrand, France, WMO, 1990.
11. A.N. Tikhonov and V. Y. Arsenin. *Solutions of Ill-Posed Problems*. Wiley, New York, 1977.
12. G. Wabha and J. Wendelberger. Some new mathematical methods for variational objective analysis using splines and cross-validation. *Monthly Weather Review*, 108, 1980.