Non-Smooth Observation Operators in Data Assimilation of a Shallow Water Equations Model

J. L. Steward, I. M. Navon, M. Zupanski & N. Karmitsa

Abstract At any one time, approximately 50% of the earth is covered with clouds. However, currently operational data assimilation schemes – i.e. algorithms seeking to maximize utilization of meteorological observations in a numerical weather prediction system – throw away satellite observations that are suspected to be cloud-contaminated. Because much of the severe weather of human interest – such as hurricanes – occurs in the presence of clouds, this is a crucial area of research.

There are several difficulties associated with assimilating cloudy radiances. One such issue is a jump in the derivative of the operator modeling the radiation transfer that a satellite experiences. This creates issues in the data assimilation as current algorithms are based on the assumption of smoothness in this observation operator.

This poster presents preliminary work on assimilating observations that have a non-smooth observation operator. This work is done using a two-dimensional limited-area shallow-water equation model and its adjoint. We test the performance of the "Four-Dimensional" Variational Approach (4D-Var, here: two dimensions plus time) compared to that of the Maximum Likelihood Ensemble Filter (MLEF), a hybrid ensemble/variational method.

We also investigate minimization of the data assimilation cost functional using the Limited Memory BFGS (L-BFGS) quasi-Newton algorithm originally intended for smooth optimization, the non-linear conjugate gradient method also originally intended for smooth optimization, and the Limited-Memory Bundle Method algorithm (LMBM) specifically designed to address large-scale non-smooth minimization problems. Numerical results obtained show that both the CG, L-BFGS and LMBM algorithms give excellent results when the non-smoothness is not extreme. However, CG and L-BFGS both fail for non-smooth functions with large Lipschitz constants. The LMBM method is found to be suitable choice for large-scale non-smooth optimization.

Department of SCIENTIFIC COMPUTING THE FLORIDA STATE UNIVERSITY

FLORIDA STATE UNIVERSITY 1851 VIRES ARTES MORES

Shallow water equations

For our tests, we use the limited area shallow water equation model:

$$\begin{aligned} u_t &= -uu_x - vu_y + fv - \phi_x \\ v_t &= -uv_x - vv_y + fv - \phi_y \\ \phi_t &= -u\phi_x - v\phi_y \end{aligned} \qquad (1)$$

where u and v are the two components of the horizontal velocity in m/s, ϕ is the geopotential field in $\mathrm{m}^2/\mathrm{s}^2$, and f is the Coriolis factor in s^{-1}.

The initial conditions used were a β plane of length L and depth D, with the height of the free surface, in meters, given by

$$\begin{aligned} h(x,y) &= h_0 + h_1 \tanh\left(\frac{9(y-y_0)}{2D}\right) \\ &+ h_2 \operatorname{sech}^2\left(\frac{9(y-y_0)}{2D}\right)\sin\left(\frac{2\pi x}{L}\right) \end{aligned} \qquad (2)$$

where $h_0 = 2000$ m, $h_1 = -220$ m, $h_2 = 133$ m, $L = 6000$ km, $D = 4400$ km, and $y_0 = D/2$.

The initial conditions are derived through geostrophic balance. The model is discretized using a second-order quadratic conservation advective scheme. The space and time increments are $\Delta x = 300$ km, $\Delta y = 220$ km, and $\Delta t = 600$ s, respectively, resulting in a mesh comprising 21×21 spatial grid points. The boundary conditions are a rigid wall in the north-south direction and periodic flow in the east-west directions. The model is integrated for 80 time steps, i.e. a window of assimilation of 13 hours 20 minutes

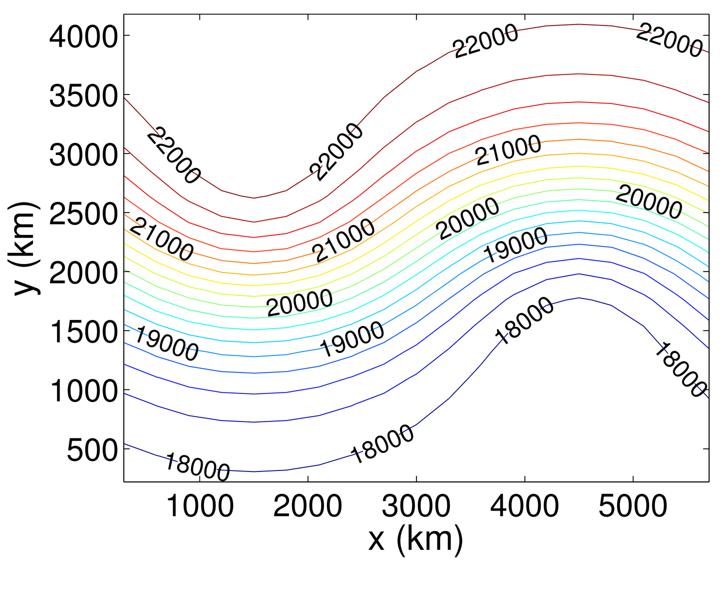

Figure 1: ϕ contours

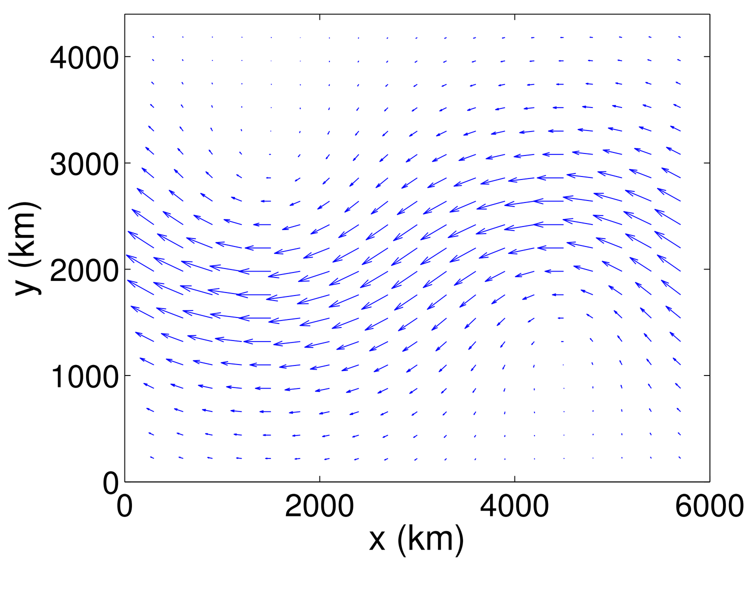

Figure 2: u and v wind-field

Observation operator

We create non-smooth observation operators that highlight the performance of the non-smooth data assimilation techniques.

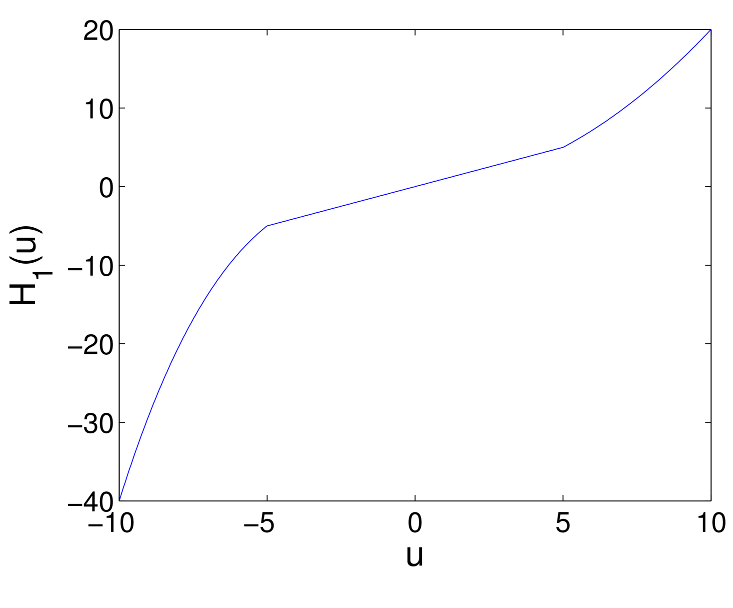

Figure 3: $\mathcal{H}_1(u) = u^3/25$ if $u < -5$, $u^2/5$ if $u \geq 5$, u else

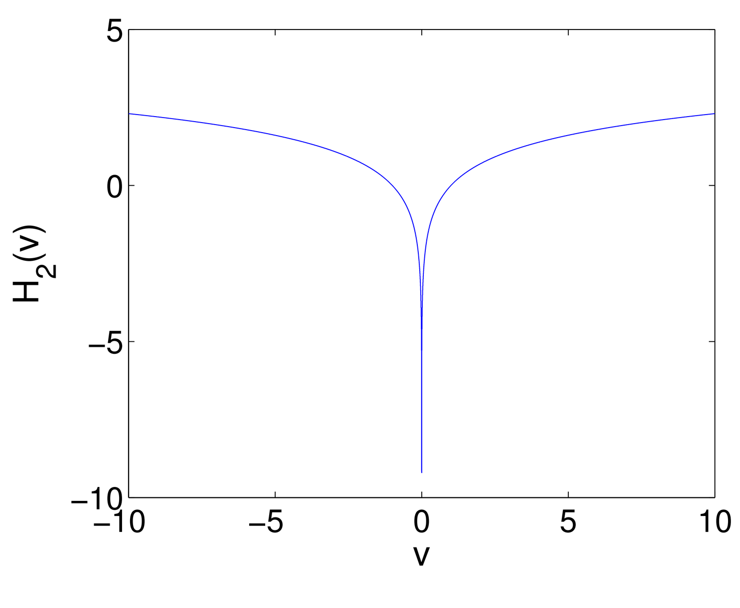

Figure 4: $\mathcal{H}_2(v) = \log(v+\delta)$ if $v \geq 0$, $\log(-v_i + \delta)$ if $v_i < 0$

4DVar data assimilation

Data assimilation is used to integrate observations with the model prediction to come up with an optimal estimate of the state. This is accomplished by minimizing a the constrained cost function:

$$\begin{aligned} J(\mathbf{x}) &= \tfrac{1}{2}\,\delta_b(\mathbf{x})^{\mathrm{T}}\mathbf{B}^{-1}\delta_b(\mathbf{x}) \\ &+ \tfrac{1}{2}\textstyle\sum_{k=0}^{NT}\delta_{y_k}(\mathbf{x})^{\mathrm{T}}\mathbf{R}^{-1}\delta_{y_k}(\mathbf{x}) \end{aligned} \qquad (3)$$

where:

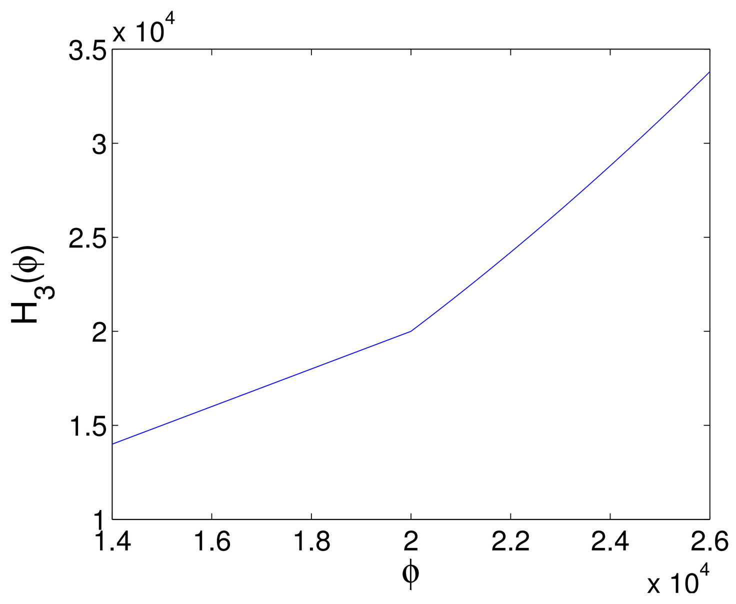

Figure 5: $\mathcal{H}_3(\phi) = \phi$ if $\phi < 20000$, $\phi_i^2/20000$ else

- $\mathbf{x}$ is the control variable, the initial model state
- $\mathbf{x}_k$ is the model state at time k with the strong constraint $\mathbf{x}_k = \mathcal{M}(\mathbf{x}_{k-1})$, $\mathbf{x}_0 = \mathbf{x}$
- $\mathcal{M}$ is the (generally non-linear) model operator
- $\delta_b(\mathbf{x}) = \mathbf{x} - \mathbf{x}_b$ is difference between the background prediction $\mathbf{x}_b$ (from a previous analysis cycle)
- $\delta_{y_k}(\mathbf{x}) = \mathbf{y}_k - \mathcal{H}(\mathbf{x}_k)$ is difference between the observation $\mathbf{y}_k$ at time k and $\mathcal{H}(\mathbf{x}_k)$
- $\mathcal{H}$ is the (generally non-smooth and non-linear) observation operator
- $\mathbf{B}$ is the background covariance matrix, $\mathbf{R}$ is the observation covariance matrix
- NT is the number of observation batches (time steps that have observations)

Maximum Likelihood Ensemble Filter

MLEF is a hybrid ensemble/variational filter that directly minimizes the likelihood of the posterior pdf directly in a manner reminiscent of 3DVar (4DVar without the time dependence). It takes several best practices from other ensemble filter methods, including:

- Using reduced-rank square-root forecast ($\hat{\mathbf{P}}_i^{1/2}$) and analysis ($\mathbf{P}_i^{1/2}$) error covariances
- Minimizing the maximum likelihood problem

$$J(\mathbf{x}) = \frac{1}{2}\,\delta_b(\mathbf{x})^{\mathrm{T}}\mathbf{B}^{-1}\delta_b(\mathbf{x}) + \frac{1}{2}\,\delta_{y_k}(\mathbf{x})^{\mathrm{T}}\mathbf{R}^{-1}\delta_{y_k}(\mathbf{x}) \qquad (4)$$

  - here the definitions are the same as in 4DVar, but the observations are only for the current timestep
  - uses conjugate gradient to solve the optimization problem
- Sophisticated Hessian preconditioning
- Calculates the square-root analysis error covariance similar to Ensemble Transform Kalman Filter (ETKF)
- **Does not required the Jacobian or adjoint of $\mathcal{M}$ or $\mathcal{H}$**

Non-smooth Optimization

A non-smooth optimization (NSO) problem is one where the function or its derivatives have discontinuities. Two main classes of optimization algorithms, bundle and sub-gradient methods, show significant promise. In this work, we investigate using CG, L-BFGS and LMBM with sub-gradients.

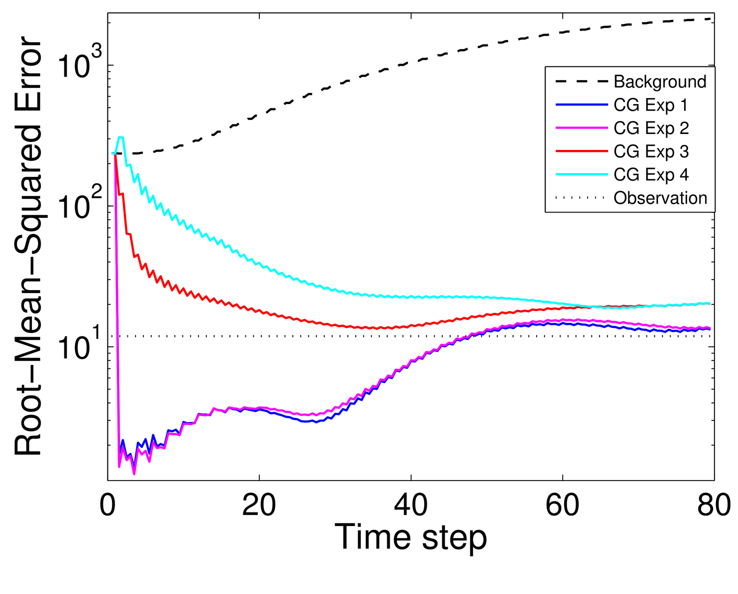

Figure 6: MLEF CG, RMSE for ϕ

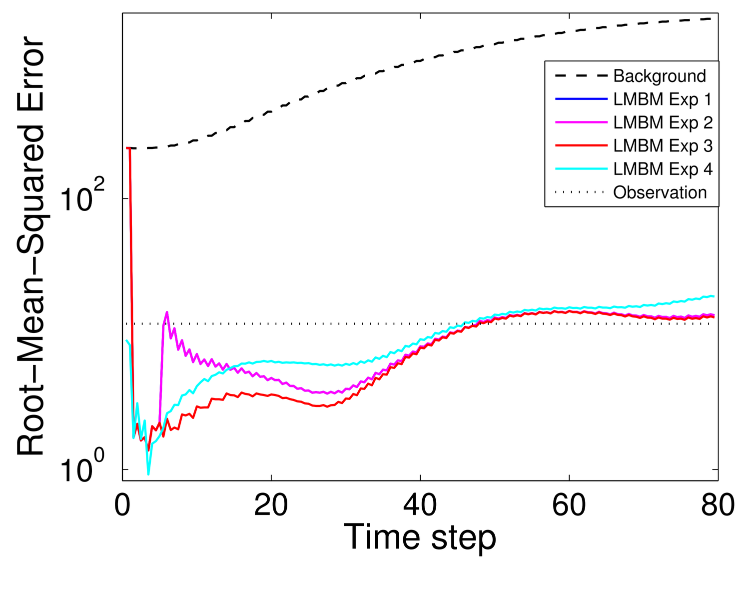

Figure 7: MLEF LMBM, RMSE for ϕ

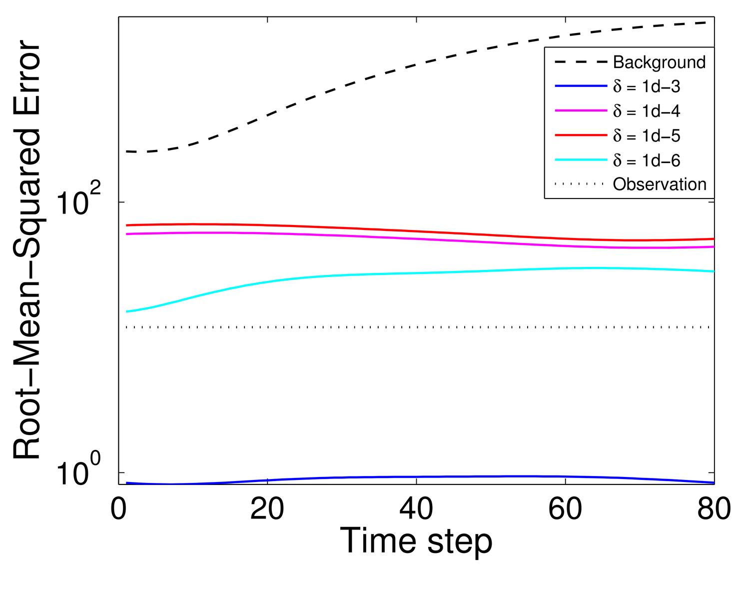

Figure 8: 4D-Var L-BFGS, RMSE for ϕ

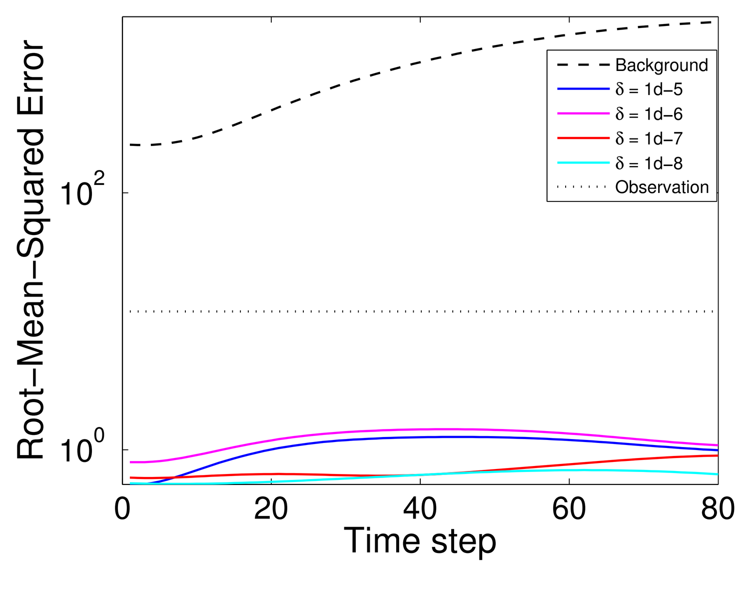

Figure 9: 4D-Var LMBM, RMSE for ϕ

Conclusions and Future Work

We tested the impact of non-differentiable observation operators on the data assimilation of a limited-area shallow water equations model. By simply replacing the gradient of the cost function with the sub-gradient (see figure 10), both 4D-Var and MLEF are able to assimilate the non-smooth observations to varying degrees of success with a smooth optimization algorithm, especially when the non-smoothness is not severe. However, both methodologies encounter difficulties with the more sharply non-smooth experiments. This difficulty can be remedied in both MLEF and 4D-Var with the use of an algorithm specifically designed for non-smooth optimization, which in this research was the limited memory bundle algorithm (LMBM).

The next steps are to apply these results, which appear encouraging, to the problem of all-sky satellite radiance observation assimilation. Modeling and simulating satellite radiative transfer with clouds is a challenge. However, if successful in this application, it is anticipated that non-smooth optimization methods may eventually take hold in an operational setting.

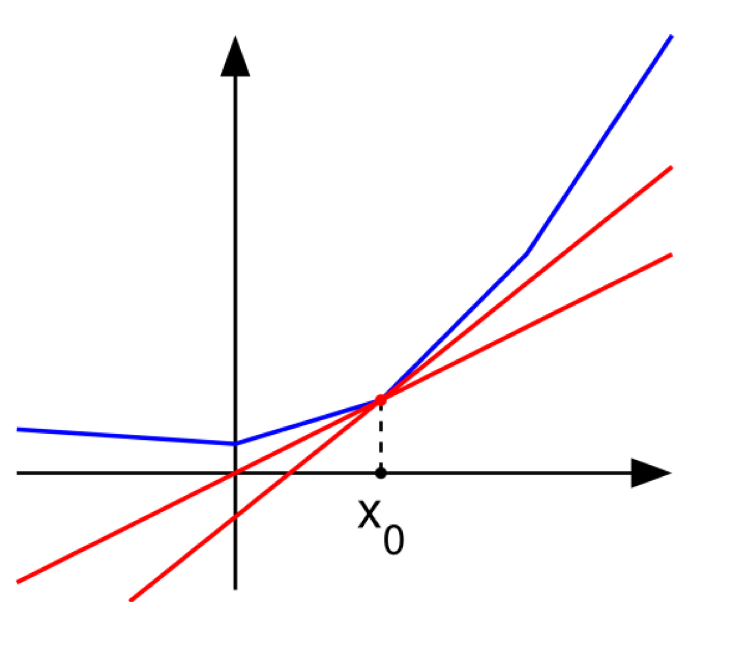

Figure 10: Sub-derivative example. Any line that remains below the function in the neighborhood of a non-smooth point $\mathbf{x}_0$ can be considered a sub-derivative. A sub-gradient is the vector of sub-derivatives with respect to each independent variable. At smooth points, there is only one sub-gradient, and it corresponds to the traditional gradient.