

Combining Observations with Models: Penalized Likelihood and Related Methods in Numerical Weather Prediction

Grace Wahba

<http://www.stat.wisc.edu/~wahba>

→ *TRLIST*

Interface 2000, the Millenium Interface

April 6, 2000

Based on joint work with
Jianjian Gong, Joseph Tribbia and Donald R.
Johnson

Abstract

We will look at variational data assimilation as practiced by atmospheric scientists, with the eyes of a statistician. Recent operational numerical weather prediction models operate on what might be considered a very grand penalized likelihood point of view: A variational problem is set up and solved to obtain the evolving state of the atmosphere, given heterogeneous observations in time and space, a numerical model embodying the nonlinear equations of motion of the atmosphere, and various physical constraints and prior physical and historical information. The idea is to obtain a sequence of state vectors which is 'close' to the observations, close to a trajectory satisfying the equations of motion, and simultaneously respects the other information available. The state vector may be as big as 10^7 , and the observation vector 10^5 or 10^6 , leading to some interesting implementation questions. Interesting non-standard statistical issues abound.

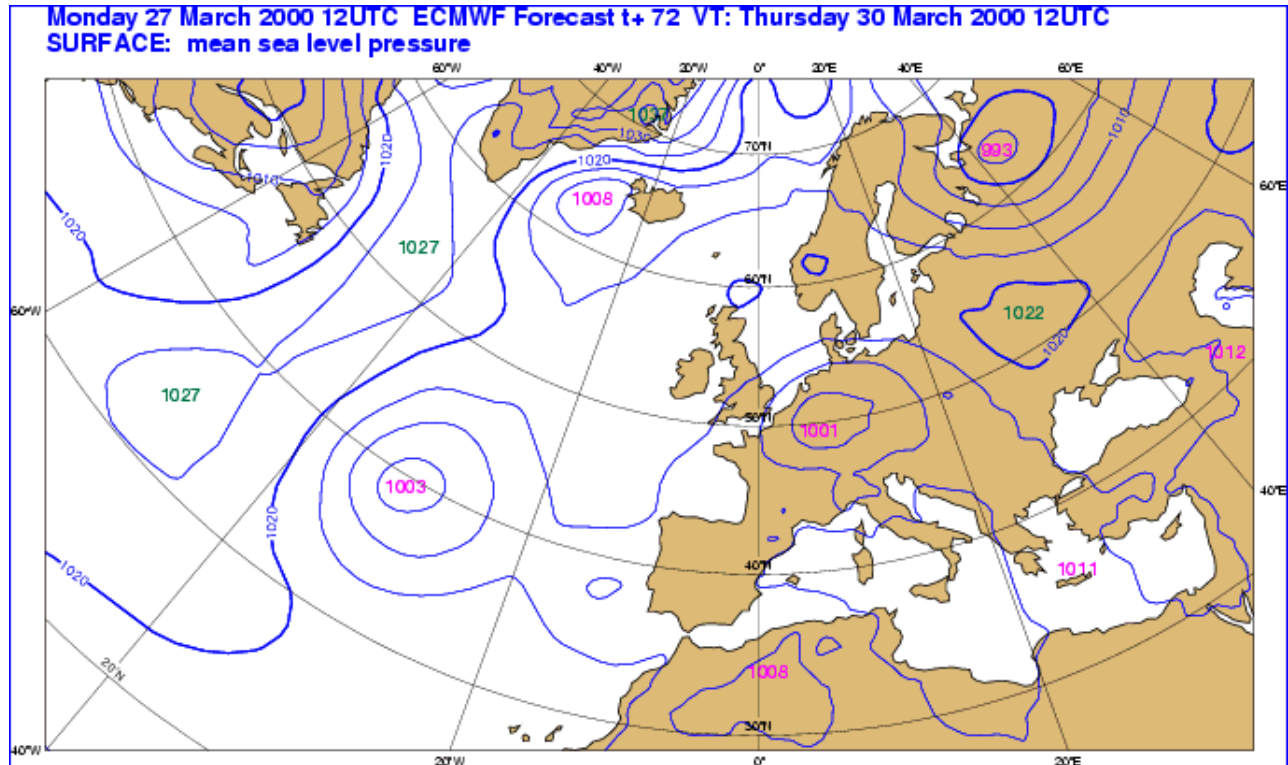
Outline

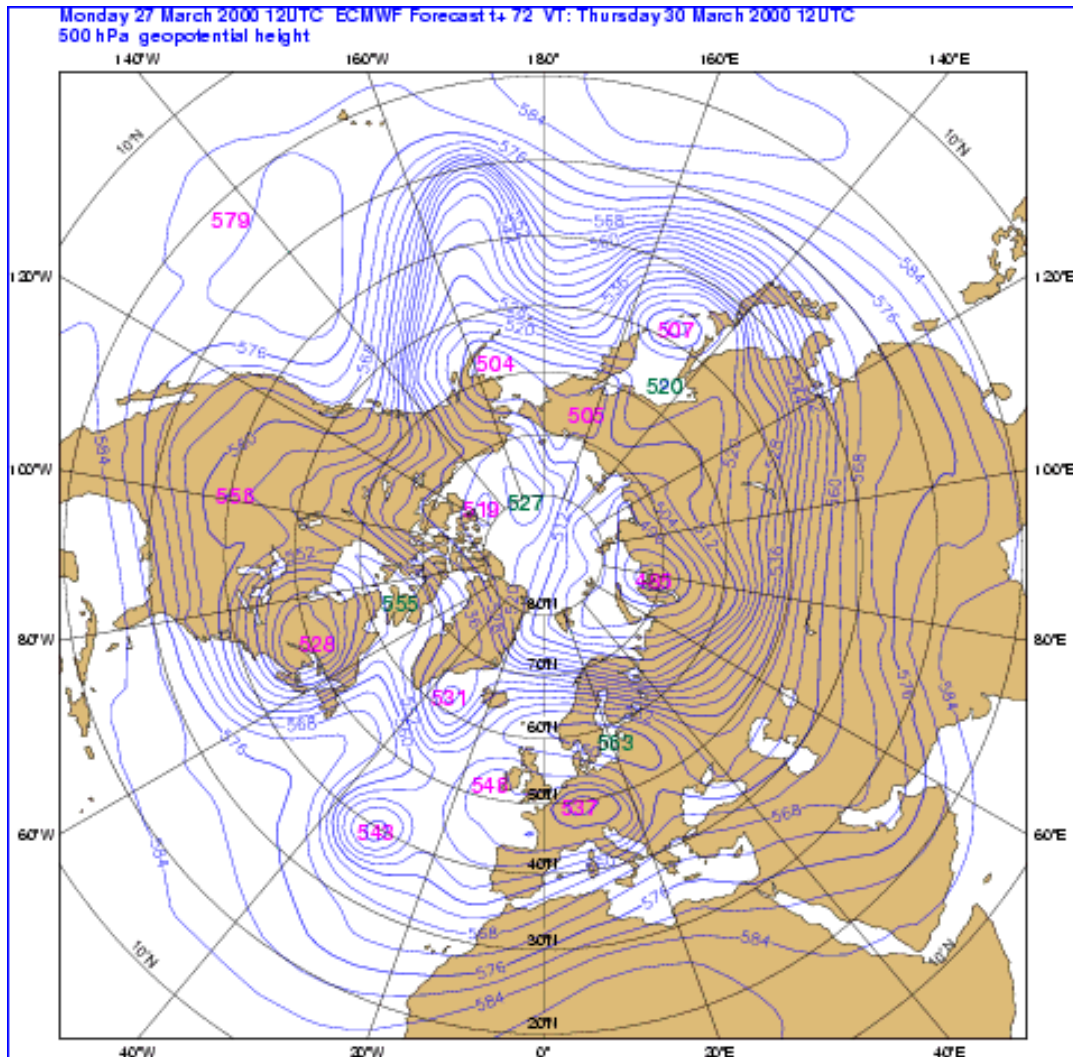
1. What is numerical weather prediction?
2. 3-D VAR, (Three Dimensional Variational Analysis) 4-D VAR (Four Dimensional Variational Analysis)
3. Wind, Divergence and Vorticity, Spherical Harmonics.
4. The NCEP (National Centers for Environmental Prediction), and ECMWF (European Center for Medium Range Weather Prediction)
Global Scale NWP (Numerical Weather Weather Prediction Models).
5. Model variables, ζ , D , T , P_s , q . Analysis variables. 'Balance'.

6. The analysis state vector.
7. The variational problem to be solved in 3D-VAR models. Weighting, smoothing and tuning parameters.
8. 4D-VAR Models
9. A toy experiment to examine the feasibility of tuning 4D-VAR models via Generalized Cross Validation.

Today's forecast

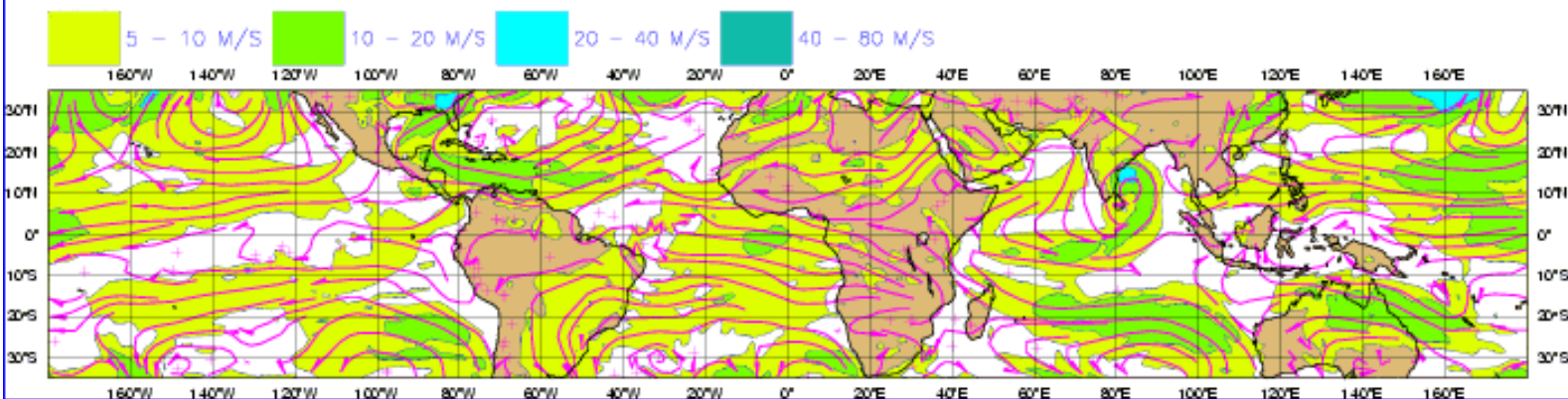
A copy of the surface temperature and pressure forecast plot from a daily newspaper goes here.





March 27 72 hour forecast 500 hPa geopotential height (in 10's of meters), from ECMWF.

Monday 27 March 2000 12UTC ECMWF Forecast t+ 72 VT: Thursday 30 March 2000 12UTC
850 hPa Winds



March 27 72 hour forecast 850 hPa Winds, from
ECMWF.

Wind, Divergence and Vorticity

P = point on the sphere

P = (latitude, longitude) = (λ, ϕ)

$\lambda \in (0, 2\pi), \phi \in (-\frac{\pi}{2}, \frac{\pi}{2})$

$u(P)$ = East Wind at P

$v(P)$ = North Wind at P

vorticity = ζ

divergence = D

$$\zeta = \frac{1}{a \cos \phi} \left[-\frac{\partial}{\partial \phi}(u \cos \phi) + \frac{\partial v}{\partial \lambda} \right]$$

$$D = \frac{1}{a \cos \phi} \left[-\frac{\partial u}{\partial \lambda} + \frac{\partial}{\partial \phi}(v \cos \phi) \right]$$

Spherical Harmonics

$$P = (\lambda, \phi)$$

$$Y_{\ell s} = \begin{cases} \theta_{\ell s} \cos(s\lambda) P_{\ell s}(\sin \phi) & 0 \leq s \leq \ell \\ \theta_{\ell s} \sin(s\lambda) P_{\ell|s|}(\sin \phi) & -\ell \leq s < 0 \end{cases}$$

$\ell = 0, 1, 2, \dots$, $P_{\ell s} =$ Legendre Polynomials

The spherical harmonics are the eigenfunctions of the (horizontal) Laplacian Δ on the sphere:

$$\Delta f = \frac{1}{a^2} \left[\frac{1}{\cos^2 \phi} f_{\lambda\lambda} + \frac{1}{\cos \phi} (\cos \phi f_{\phi})_{\phi} \right]$$

$$\Delta Y_{\ell s} = -\ell(\ell + 1)Y_{\ell s}$$

and play the same role on the sphere as sines and cosines on the circle.

Spherical Harmonics (con't)

$$f \in \mathcal{L}_2(\text{Sphere})$$
$$f \sim \sum_{l=0}^{\infty} \sum_{s=-l}^l f_{ls} Y_{ls}$$

where

$$f_{ls} = \int_{\text{Sphere}} f(P) Y_{ls}(P) dP.$$

(Given a constant) its easy to go back and forth from (u, v) to (ζ, D) in spherical harmonic coordinates.

Model and Analysis Variables

The model variables are ζ, D, T, p_s, q , respectively, vorticity, divergence, temperature, surface pressure and humidity, as a function of $P = (lat, long)$, and a vertical coordinate. The vertical coordinate is typically pressure divided by p_s rescaled to $[0, 1]$ (1 at the ‘top’ of the atmosphere) ($\sim 10hPa$). $[0, 1]$ is then discretized to N discrete levels. The model variables are expanded in spherical harmonics $Y_{\ell,s}, s = -\ell, \dots, \ell; \ell = 0, \dots, L$, at each level. Thus, there will be $[4N + 1] \times [(L + 1)/2]^2$ coefficients in the model state vector. The model state vector at time $t - 1$ is updated via the equations of motion of the atmosphere to get a predicted model state vector at time t . Then the model state vector is transformed to analysis variables to get the forecast analysis state vector $x_f(t)$ which contains the analysis variables: $\zeta, D_u, (T, p_s)_u, q$, (next slide), expanded in spherical harmonics, at N levels (for ECMWF). (NCEP similar). The NCEP model has $N = 28$ discrete levels in the vertical, with $L = 126$. The ECMWF model has $N = 31$ discrete levels and L about twice as big. (latest info I have, may not be current, other caveats)

The (Unbalanced) Analysis Variables

The analysis variables are: $\zeta, D_{\mathbf{u}}, (T, p_s)_{\mathbf{u}}, q$, where $D_{\mathbf{u}}, (T, p_s)_{\mathbf{u}}$ are the unbalanced part of D and (T, p_s) . What this means is the following: There are approximate physical relationships between these variables, generally called 'balance' which means that physical forces must approximately balance each other out - in particular the pressure gradient along pressure surfaces and the Coriolis force (acceleration due to the rotation of the earth) approximately balance each other out. Then the model variables (ζ, D, T, p_s) are divided into two parts, a balanced part which can essentially be computed via a balance equation from the vorticity, and the rest, which are called the unbalanced part. Later this allows for control of the size of the unbalanced part, which is known a priori to be 'not too big'.

The forecast vector $x_f(t)$.

Also known as the 'background' vector $x_b(t)$ (about which everything will be expanded).

$$\zeta(P, z_k, t) = \sum_{\ell, s} a_{\ell sk}(t) Y_{\ell s}(P)$$

$$D_{\mathbf{u}}(P, z_k, t) = \sum_{\ell, s} b_{\ell sk}(t) Y_{\ell s}(P)$$

$$T_{\mathbf{u}}(P, z_k, t) = \sum_{\ell, s} c_{\ell sk}(t) Y_{\ell s}(P)$$

$$P_{\text{su}}(P, t) = \sum_{\ell, s} d_{\ell s}(t) Y_{\ell s}(P)$$

$$q(P, z_k, t) = \sum_{\ell, s} e_{\ell sk}(t) Y_{\ell s}(P)$$

$$x_f \equiv x_b = \begin{pmatrix} a \\ b \\ c \\ d \\ e \end{pmatrix}$$

Next, to combine x_b with observations to get an updated estimate of the present state (the analysis).

Observations

There are many kinds of observations, most of which are ‘direct’ observations on u, v, T, p_s, q , some (from satellites) being indirect observations. The satellites observe upwelling radiances at different wavelengths, which are related to integrals of the vertical temperature distribution over different (vertical) regions of the atmosphere. After linearization the observation vector y can be considered as linear functionals on a (mythical) state vector x_{true} , $y = Hx_{true} + \epsilon_o$, where ϵ_o includes instrumental error of all kinds as well as ‘errors of representativeness’ essentially due to the fact that a finite model cannot represent the atmosphere exactly. From another point of view, direct observations are at a point, but the model resolution has finite extent (about 60 km for ECMWF).

Some Observation Types: Observation types listed in <http://sgi62.wwb.noaa.gov:8080/RTPUB/research/jhtml/ssi2.html> dated March 2, 2000 go here.

The Variational Problem for the Present State

Let $x = x_{true}$,

$$y = Hx + \epsilon_o, \quad \epsilon_o \sim N(0, R) \quad \text{obs'n error}$$

$$x_b = x + \epsilon_b, \quad \epsilon_b \sim N(0, B) \quad \text{forecast error}$$

$$x \sim N(\mu, \Sigma) \quad \text{prior information}$$

To estimate x , find x to minimize

$$(y - Hx)'R^{-1}(y - Hx) + (x - x_b)'B^{-1}(x - x_b) + \\ (x - \mu)'\Sigma^{-1}(x - \mu).$$

Let $\delta = x - x_b$, $d = y - Hx_b$, results in: find δ to minimize

$$(d - H\delta)'R^{-1}(d - H\delta) + \delta'B^{-1}\delta + J(\delta, \text{other stuff}).$$

Then \hat{x} , a.k.a. the analysis vector, is given by

$$\hat{x} = x_b + \delta.$$

$$B \sim 10^6 \times 10^6, R \sim 10^5 \times 10^5.$$

B : Instrumental error, errors of representativeness

R : Historical data, NMC method, empirical method

J : Dynamical constraints: Gravity waves, etc.

4D-VAR

We briefly describe the main idea behind 4D-VAR and then describe a ‘toy’ experiment designed to examine the possibility of adaptive tuning of 4D-VAR models via Generalized Cross Validation and related methods.

The Model: Let $t = 1, \dots, T$ denote discrete time, $x_t, t = 1, \dots, T$ be a sequence of model state vectors representing (some part of) nature which evolves according to

$$x_{t+1} = M_t x_t + N_t + \xi_t, \quad t = 1, \dots, T - 1 \quad (1)$$

where $M_t = M_t(\theta)$ is the model (one step in time) evolution operator, possibly depending on some parameters θ , N_t is a forcing function and the ξ_t are what's unexplained - finite, approximate model, approximate forcing, numerical approximations. x_b is the forecast for $t = 1$, assumed to satisfy

$$x_b = x_1 + \epsilon_b.$$

The observations y_t are assumed to satisfy

$$y_t = H_t x_t + \epsilon_t, \quad t \in \Lambda$$

where H_t is a map from state vector space to observation space at time $t \in \Lambda$ and Λ is the subset of $\{1, \dots, T\}$ where there are observations. After making numerous unjustified independence assumptions, one obtains the 4D-VAR variational problem:

Find $x = (x_1, \dots, x_T)$ to min

$$\begin{aligned} \sum_{t \in \Lambda} \|y_t - H_t x_t\|_{R_t}^2 &+ \alpha \sum_{t=1}^{T-1} \|x_{t+1} - M_t x_t - N_t\|_{Q_t}^2 \\ &+ \gamma \|x_b - x_1\|_B^2 + \eta \|x_T\|_{\Sigma}^2, \end{aligned}$$

where $\|\delta\|_B^2 \equiv \delta' B^{-1} \delta$. The four terms represent closeness to the data, closeness to the model, closeness to the forecast, and (prior) physical information concerning physical (approximate) constraints.

Letting $\hat{y}_t = H_t \hat{x}_t$, $\hat{y} = (\hat{y}_1', \dots, \hat{y}_T')'$, then, (more assumptions)) there exists a matrix $A(\lambda)$, ($\lambda = \alpha, \gamma, \eta, \theta, \dots$) known as the influence matrix, such that

$$\hat{y} = A(\lambda)y + \text{quantities independent of } y.$$

The *GCV* (generalized cross validation) estimate of θ is the minimizer of $V(\lambda)$ where

$$V(\lambda) = \frac{\frac{1}{n_{dat}} RSS(\lambda)}{\left[\frac{1}{n_{dat}} \text{tr}(I - A(\lambda)) \right]^2}$$

where n_{dat} is the number of data points (dimension of y) and $RSS(\lambda) = \|y - \hat{y}\|^2$. Letting $x_{t'true}$, be the 'true' but unknown x_t , $V(\lambda)$ is, under suitable assumptions, a proxy for the predictive mean square error (*PMSE*), given by $R(\lambda)$ where

$$R(\lambda) = \frac{1}{n_{dat}} \sum_{t \in \Lambda} \|H_t x_{t'true} - H_t \hat{x}_t(\lambda)\|^2,$$

in the sense that the minimizer $V(\lambda)$ is a good estimates of the minimizer of $R(\lambda)$. $\text{tr}A$ can be estimated via the randomized trace technique, so that it can be feasible even in very large problems.

Toy Example: The Equivalent Barotropic Vorticity Equation on a Latitude Circle

From Gong, Wahba, Johnson and Tribbia, *Mon. Wea. Rev.* (1998).

$$\frac{\partial}{\partial t}(x_{ss} - \lambda^2 x) + U(\theta; s)x_{sss} + \beta x_s = -f_0 U_s,$$

where

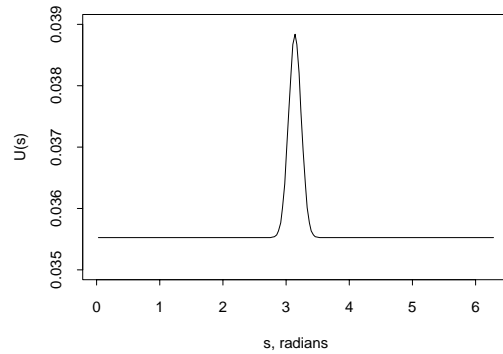
s is the space variable on the 45° latitude circle.
 x is the streamfunction, $dx/ds = u(s)$ is the wind.
 λ^2, β, f_0 are physical constants depending on Coriolis, gravity, etc. $U(\theta, s) = U_0(1 + \delta g(s))$ where $\theta = (U_0, \delta)$ are 'distributed' parameters to be estimated, and $g(x)$ is a fixed perturbation function. $U(\theta, s)$ affects the speed of the (wind) wave.

Experimental Design to Test Feasibility of GCV Tuning

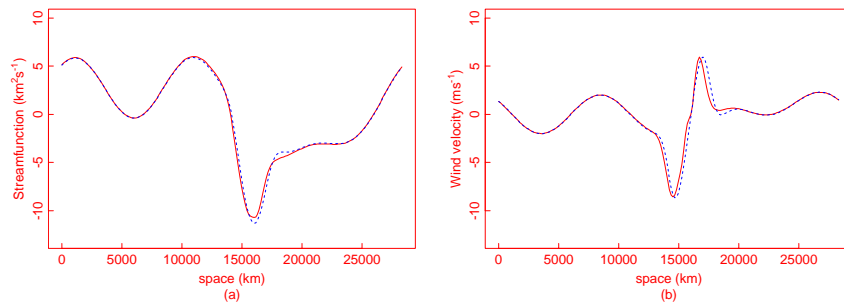
- Nature: A high resolution leapfrog scheme for M_t , time grid 8.7 sec, space grid 21 km.
- The Model: A low resolution first order forward difference scheme for M_t . The Model time grid 4 hours, space grid 146 km (194 points).
- Model time $t = 1, \dots, 13$ where $1 = 0hrs$, $13 = 48hrs$.
- Dimension of Model state vector = $13 \times 194 = 2252$. 164 wind observations at times at $t \in \Lambda = 1, 4, 7, 10, 13$, for a total of $n_{obs} = 5 \times 164 = 820$, and a forecast vector for $t = 1$ of dimension $n_b = 194$.

Steps in the Experiment

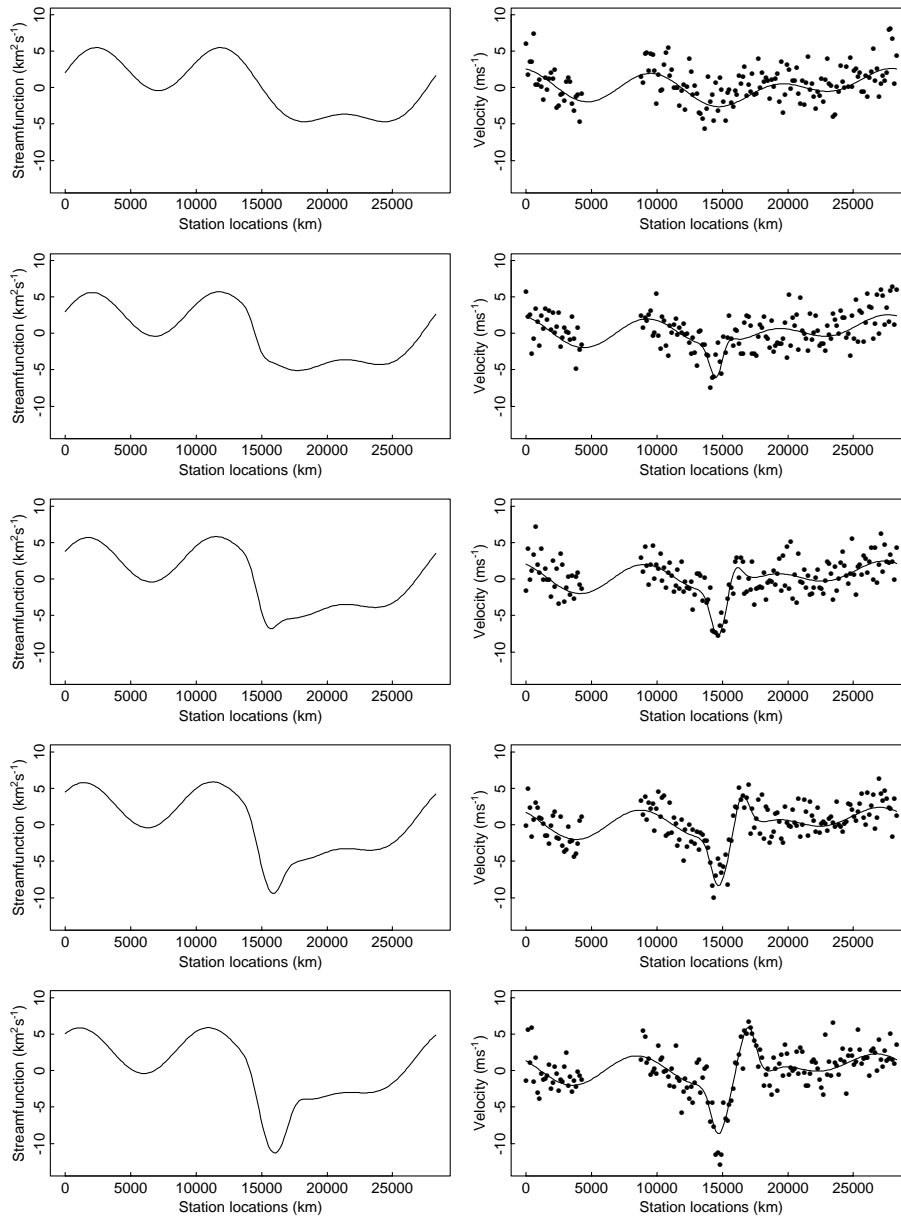
1. Generate initial 'truth' and evolve it forward via Nature
2. Generate forecast by adding ϵ_b to initial truth.
3. Generate wind observations by computing wind 'truth' from Nature and adding observation error.
4. Fix $\lambda = \alpha, \gamma, \eta, \theta$. Solve the variational problem, using the Model. (Packed Cholesky decomposition)
5. Evaluate $V(\lambda)$. Increment $V(\lambda)$ and repeat. Find minimum over λ . (Experimental Global search, but in practice use downhill simplex).



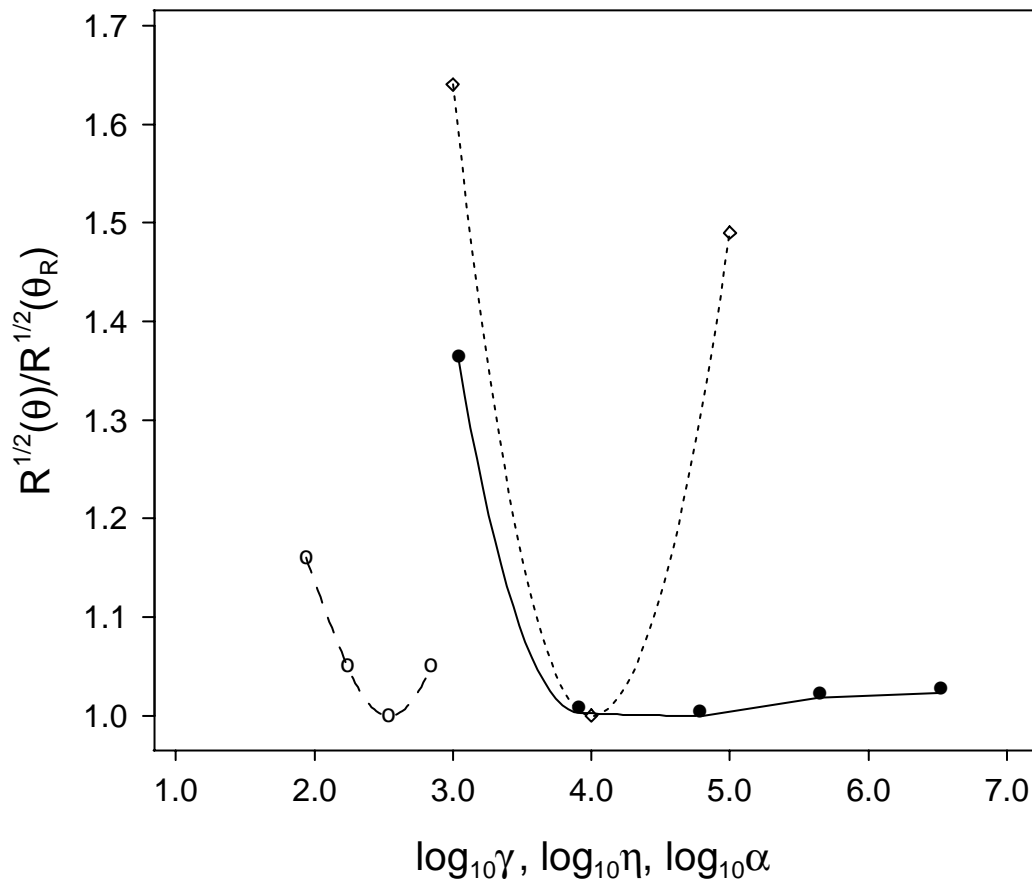
$U(\theta, s)$ the speed coefficient. U_0 controls the constant component, δ the height of the peak.



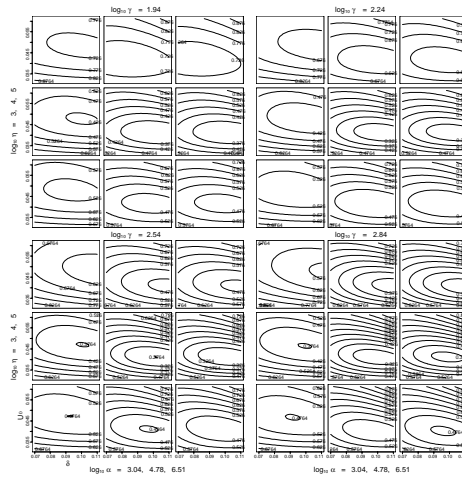
Nature and model streamfunction (left) and nature and model wind (right) at 48 hours.



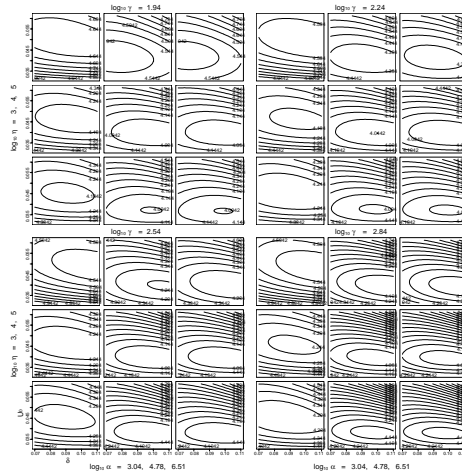
Nature streamfunction (left) and nature wind and wind observations (right), at $t = 1, 4, 7, 10, 13$.



Sensitivity of the Predictive Mean Square Error to α (weight for model), γ (weight for forecast) and η (weight for constraint).



$$R^{1/2}(\gamma, \eta, \alpha, U_0, \delta)$$



$$V^{1/2}(\gamma, \eta, \alpha, U_0, \delta)$$

Four 3×3 blocks. The four blocks correspond to four values of γ . Each block corresponds to 3 values of η by 3 values of α . Each picture within the block is a U_0, δ plot.

Closing Remarks

Operational Numerical Weather Prediction Models solve gigantic penalized least squares problems. However there are many features that do not satisfy the usual statistical assumptions. Observation errors and Model errors are correlated in complicated ways that are not easy to unscramble, and systematic errors abound in observing system calibration and drift, and model inadequacies. Model and observation operators have nonlinearities. Data sets are humongous but some are of notoriously poor quality. A large amount of prior information is available but some of it is highly 'physical' or qualitative rather than statistical, for example, penalties based on the understanding that certain physical quantities are 'not too large'. Operational models are sensitive to numerous tuning parameters, whose effects are interact. In practice, tuning of various kinds goes on continuously. The GCV and related methods may be useful in some of the tuning problems associated with numerical weather prediction.

References

- [CDE⁺93] P. Courtier, J. Derber, R. Errico, J. Louis, and T. Vukicevic. Important literature on the use of adjoint, variational methods and the Kalman filter in meteorology. *Tellus*, 45A:342–357, 1993.
- [Cou97] P. Courtier. Dual formulation of four-dimensional variational assimilation. *Q. J. R. Meteorol. Soc.*, 123:2449–2461, 1997.
- [CTH94] P. Courtier, J. Thépaut, and A. Hollingsworth. A strategy for operational implementation of 4D-Var, using an incremental approach. *Q. J. R. Meteorol. Soc.*, 120:1367–1387, 1994.
- [DB99] J. Derber and F. Bouttier. A reformulation of the background error covariance in the ECMWF global data assimilation system. *Tellus*, 51A:195–221, 1999.
- [GWJT98] J. Gong, G. Wahba, D. Johnson, and J. Tribbia. Adaptive tuning of numerical weather prediction models: simultaneous estimation of weighting, smoothing and physical parameters. *Monthly Weather Review*, 125:210–231, 1998.
- [Lor86] A. Lorenc. Analysis methods for numerical weather prediction. *Quart. J. Roy. Meteor. Soc.*, 112:1177–1194, 1986.
- [PD92] D. Parrish and J. Derber. The National Meteorological Center’s spectral statistical interpolation analysis system. *Mon. Wea. Rev.*, 120:1747–1763, 1992.

- [Wah82] G. Wahba. Variational methods in simultaneous optimum interpolation and initialization. In D. Williamson, editor, *The Interaction Between Objective Analysis and Initialization*, pages 178–185. Atmospheric analysis and Prediction Division, National Center for Atmospheric Research, Boulder, CO, 1982.
- [Wah85] G. Wahba. Design criteria and eigensequence plots for satellite- computed tomography. *J. Atmos. Ocean. Tech.*, 2:125–132, 1985.
- [Wah99] G. Wahba. Adaptive tuning, four dimensional variational data assimilation and representers in rkhs. In ECMWF, editor, *Diagnosis of Data Assimilation Systems*, pages 45–52, Reading England, 1999. European Center for Medium Range Weather Prediction.
- [WJGG95] G. Wahba, D. Johnson, F. Gao, and J. Gong. Adaptive tuning of numerical weather prediction models: randomized GCV in three and four dimensional data assimilation. *Monthly Weather Review*, 23:3358–3369, 1995.