Evolving topics in data assimilation

Ross N Hoffman (AER)
AER, Lexington MA and NOAA/AOML, Miami FL

Eugenia Kalnay Symposium, 2.1
Wednesday, January 7, 2015

Eugenia Kalnay ….

- was my postdoc advisor at Goddard
- In Milt Halem's group, which included
  - Bob Atlas
  - Shukla
  - Wayman Baker
  - Joel Susskind, …..
- There are many opportunities for a post doc at GSFC

Postdoc topics

- How can we use ensembles to improve forecasts, skill vs. spread, etc.? (LAF, Ensemble forecasts)
- How can we use the data assimilation cycle to identify and correct model bias? (VarBC)
- How should OSSEs be designed?
- Can we solve variational data assimilation with nonlinear obs operators by using conjugate gradient minimization? (3d-Var)
- Sensitivity to QC decisions? (Robust QC, VarQC, Huber norms)
- What can we do when a feature (a storm) is in the wrong place in the background? (FCA, EnKF)
- What can we do when only half of a feature is observed by a satellite? (EnKF)
- Would it be possible/better to use radiances in data assimilation instead of retrievals? (GSI, IFS)
- Can we determine time continuous solutions to the governing equations that best fit some observation? (Can we solve the 4d-VAR problem?)

30+ years later….

- Many of these topics are still relevant
- I will touch on some of my early attempts in discussing future directions
- In this personal view of where DA is going

For context….

- All the DA methods discussed here depend on optimization:

  $$–\min_X (J);\ \text{where } J = J_b + J_o$$

- Balance the misfits to prior info and current observations.
- Many assumptions, transforms & design choices.
  - Characterization of the obs errors.
  - Choice of control variable X.

40–50+ years later (circa 2025-2035)….

- We will embrace messy data, nonlinear and containing signals from more than one variable and from more than one component of the earth system.
  - More emphasis on wind data
  - Effective use of cloud, precip, hyperspectral data
  - SST couples ocean and atmosphere
  - Ozone couples chemistry and atmosphere
- All data is useful, but even small amounts of very accurate data—GPS/RO, DWL, CLARREO, …—are needed to tie down the DA system.

75+ years later (circa 2060)….

- Large scale quantum computing solves fully nonlinear DA problem

$$-\min_X (J)$$

- by evaluating J for all X.
- For coupled earth system model.
- *Exercise left for the audience: Determine effective way to use quantum computing to evaluate/represent the uncertainty of the solution to use in the next DA cycle.*

aer
Atmospheric and Environmental Research

LAF :: lagged average forecasting

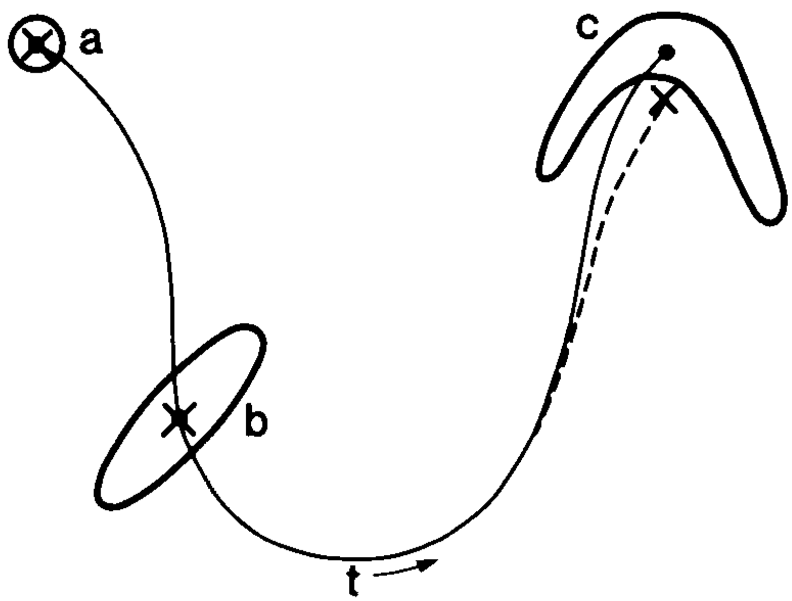
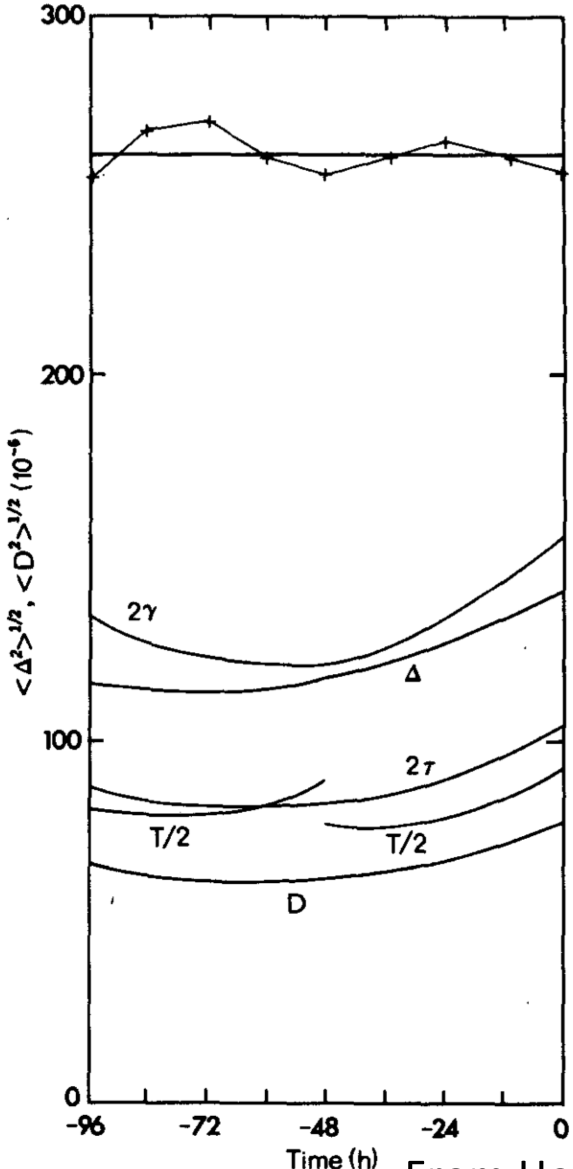


Fig. 1. Schematic phase space description of the central (•) and ensemble (×) forecasts. At the initial time (a) the central forecast is the center of the spherical ensemble. For $t < \tau_{NL}$ (b) the central and ensemble forecasts coincide; the ensemble is ellipsoidal. At later times, $t \gtrsim \tau_{NL}$ (c), the central forecast and the ensemble forecast diverge; the ensemble loses its symmetry.

From Hoffman and Kalnay (1983, Tellus)

4d-Var prototype

- Toy model
  - PE Nature
  - QG forecast
- Finite difference gradients to minimize J_o
- Works, but at the end of the interval, errors project on growing modes



Obs error

Analysis error

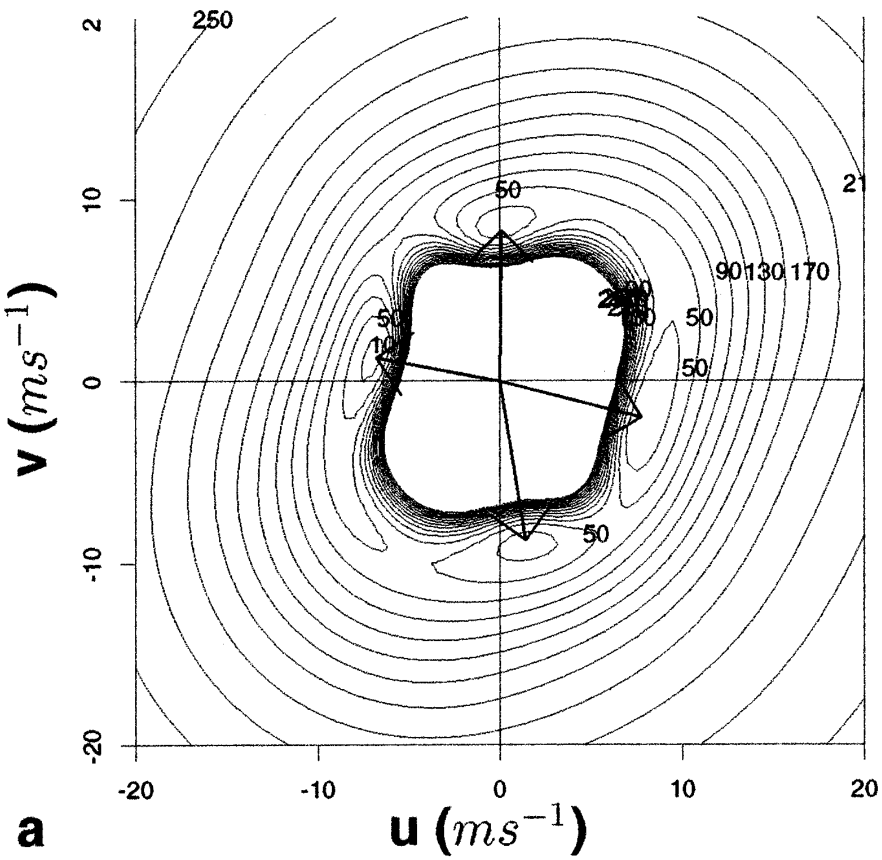
From Hoffman (1986, MWR)

Variational analysis method (VAM)

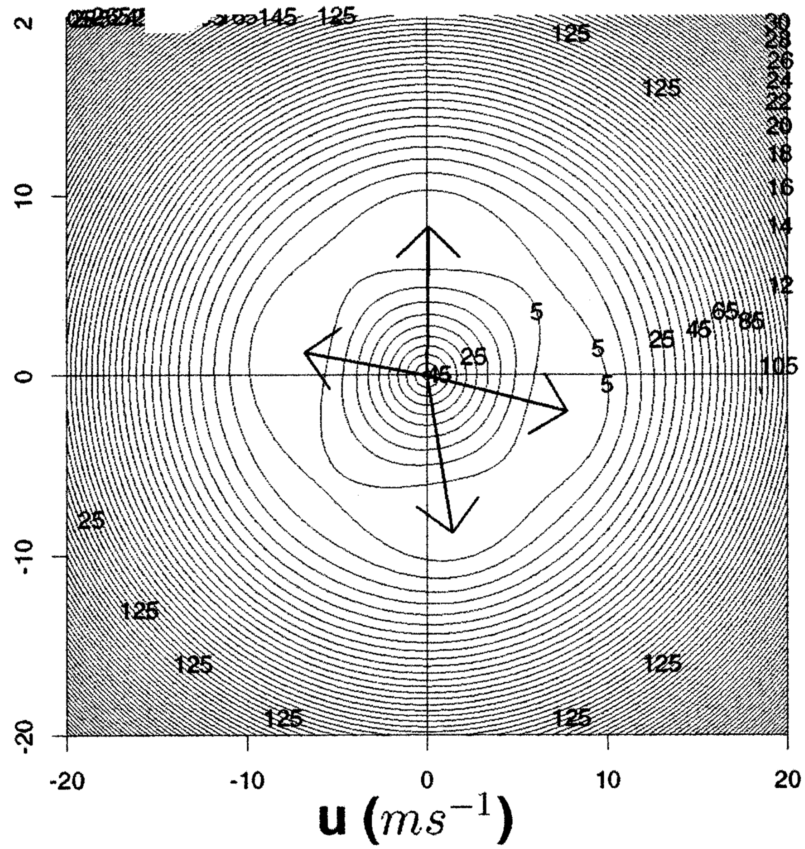
- 2d-Var based on the idea of smoothing splines to analyze scatterometer data
- Initial work on the QE2 storm of 1978
  – Large scale minimization
  – Very nonlinear
    - ambiguity removal, dynamical constraint
  – Half a storm is observed
  – Background is too weak
- VAM is now used to produce the CCMP ocean surface wind data product

Scatterometer obs functions

Sigma0 backscatter values



Ambiguous wind vectors



From Hoffman et al. (2003, JTech)

Radiances or retrievals

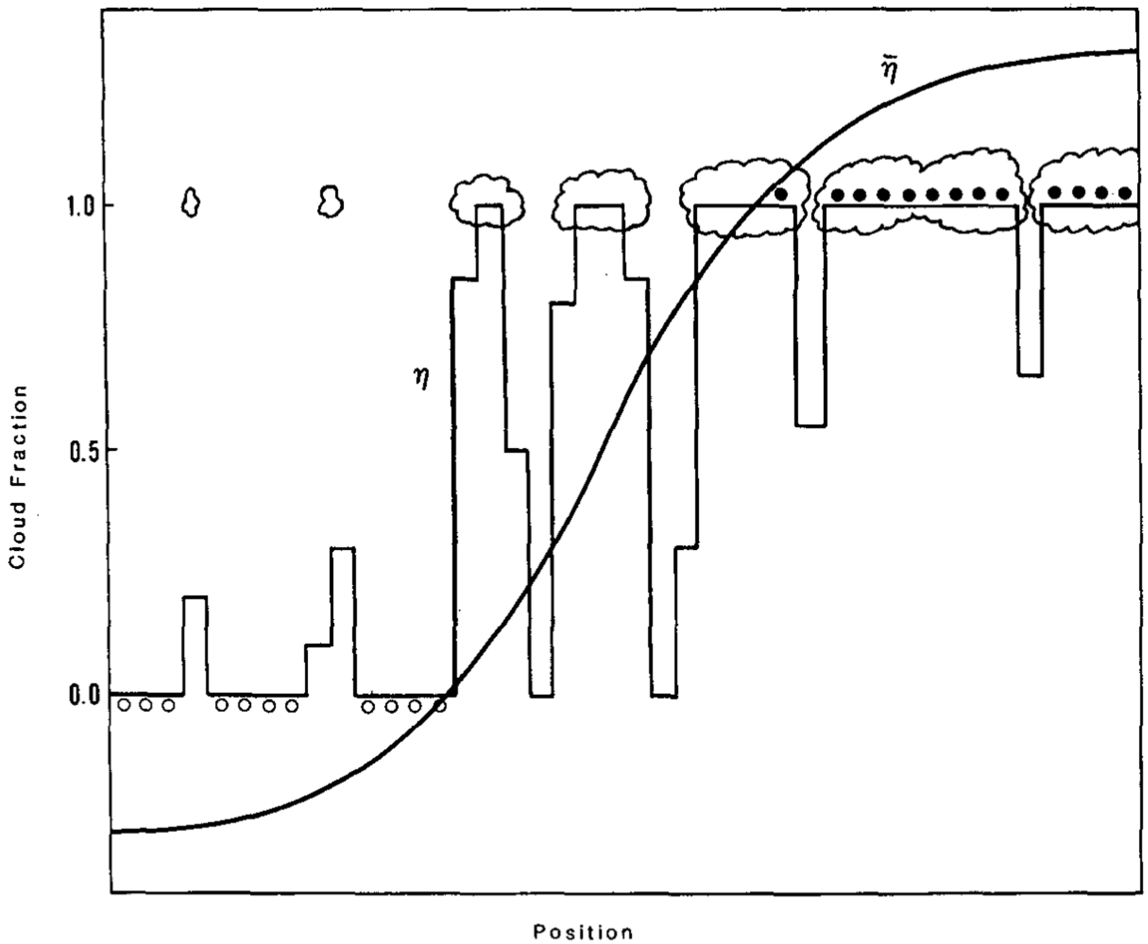


FIG. 1. Cloud fraction representation. The modeled cloud fraction η is the cropped sum of a smooth cloud fraction $\bar{\eta}$ and a ragged cloud fraction η' (not shown). See text for discussion.

From Hoffman and Nehrkorn (1989, MWR)

Radiances or retrievals

- Linearization

$$x_R = Ax_T + (I - A)x_a$$

- Provide retrievals x_R, covariance S_R along with prior information, x_a and S_a to the DA

$$A = I - S_R\, S_a^{-1}$$

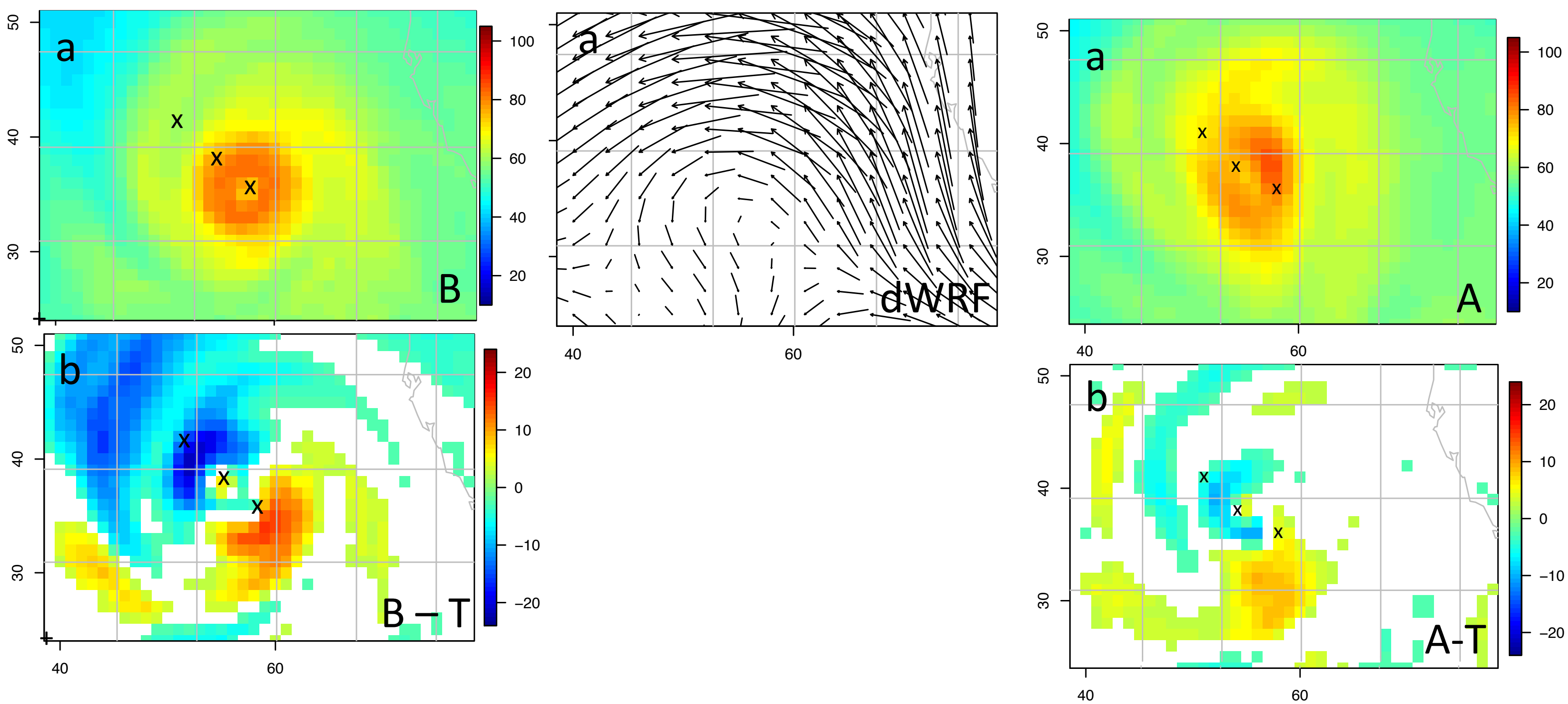
- Can transform to observations with unbiased, uncorrelated, unit variance errors
- x_a and S_a can come from the ensemble in an EnKF setting

From Hoffman (2011, arXiv)

Feature Calibration and Alignment (FCA)

- Position errors of features are both common and problematical
  - Non-Gaussian error statistics
  - Poor convergence of variational analysis schemes
- FCA represents errors (or differences) in terms of errors of alignment and errors of amplitude and “random” errors
  - dWRF developed as a feature alignment pre-processor for WRFDA
  - dWRF uses WRF software and we plan to integrate feature alignment in the WRFDA

dWRF :: feature alignment in WRFDA



Integrated water vapor (IWV) at 12 UTC
28 Aug 2005 for H. Katrina.

From Nehrkorn et al. (2015, MWR)

Lessons learned

- Strong nonlinear signals are valuable, even if the signals come from difference earth system components.
- We do need some highly accurate data to tie down our error statistics. (GPS, DWL, …)
- Knowledge, intuition about error structure is helpful to reduce degrees of freedom
- EnKF vs. 4d-Var? This may not be the right question. Hybrid DA now seems superior if the necessary discipline and effort are available for 4d-Var.
- DA systems relatively insensitive to how we implement the obs functions, but can be very sensitive to data selection and QC.

Closing thoughts

- At best our models only shadow reality.
- Our solutions are never optimal because we make approximations, have inexact knowledge of the error statistics.
  - In the nonlinear regime ad hoc methods may out-perform classical optimization.
  - Define optimal!