A strategy to optimize the use of retrievals in data assimilation

R. Hoffman¹, K. Cady-Pereira¹, J. Eluszkiewicz¹, D. Gombos¹, J.-L. Monceľ¹, T. Nehrkorn¹, S. Greybush², K. Ide², E. Kalnay², M. J. Hoffman³, and R. J. Wilson⁴

¹ AER, ² U. Maryland, ³ John Hopkins, ⁴ GFDL

Friday, 6 May, 2011
Ross N. Hoffman, Vice President, R&D
Atmospheric and Environmental Research, www.aer.com
Introduction

• Is it better to use radiances or retrievals in data assimilation?
• Using retrievals reduces complexity in the $H$-operator, reduces data volume, allows arbitrary cloud clearing and retrieval methods and makes the assimilation system more modular.
• Radiiances are available sooner, have uncorrelated errors, and are independent of the prior.
• Using EOFs from the retrieval scheme can reduce data volume and reduces vertical interpolation errors.
• The averaging kernel (AK) concept allows removing the influence of the prior (allowing for interactive retrievals) and rotating to a representation where the obs. errors are uncorrelated.
Examples

• Work in progress from on-going Mars data assimilation experiments using
  – GFDL Mars GCM
  – TES and MCS radiances and retrievals
  – LETKF ensemble Kalman filter method

• Examples from TES retrievals
  – OSS forward model
  – Rodgers type iterative physical retrieval scheme
    • Inverse Methods for Atmospheric Sounding: Theory and Practice
  – Temperature profiles only [for now]
Retrievals, $\hat{x}$ (221 of 8817 total retrievals)

221 retrievals sampled from 6 hour period
Mars Retrieval Locations and Topography

TES data at the northern fall equinox (Ls=180) of Mars Year 24 (MY24)
Forward model vertical grid

- Retrieval data pressure levels (linear)
- Retrieval data pressure levels (log)
- Temperature ($x_a$) profile (linear)

21 pressure levels
Temperature EOFs

From the prior covariance $S_a$
Averaging kernel method (after Rodgers)

\[ \hat{x} = Ax + (I - A)x_a + G_y \epsilon_y. \]

- Here \( \hat{x} \) is the retrieval, \( x \) is the true state vector, and \( x_a \) is the prior. \( A \) is the averaging kernel. \( \epsilon_y \) includes measurement error and forward model error which is unbiased with covariance matrix, \( S_\epsilon \). \( G_y \) is the sensitivity of retrieval to the radiances

\[ G_y = \frac{\partial \hat{x}}{\partial y} = \hat{S}K^T S_\epsilon^{-1} = S_a K^T (K S_a K^T + S_\epsilon)^{-1} \]

Here \( S_a \) is the prior covariance and the retrieval error covariance is determined from

\[ \hat{S}^{-1} = K^T S_\epsilon^{-1} K + S_a^{-1} \]

- And \( K = \frac{\partial F}{\partial x} \) is the Jacobian of the forward problem, \( F \), evaluated at \( \hat{x} \) the solution. Finally

\[ A = G_y K = \hat{S} \left( K^T S_\epsilon^{-1} K \right) \]
Interface: required data for assimilation

- Inputs to the retrieval
  - Prior mean and covariance and radiance error covariance
    \[ x_a, S_a, \text{ and } S_\varepsilon \]
- Outputs from the retrieval
  - Retrieval and Jacobian
    \[ \hat{x} \text{ and } K \]
- OR, if the channel set is large
  - Prior mean
    \[ x_a \]
  - Retrieval, posterior covariance, and averaging kernel
    \[ \hat{x}, \hat{S}, \text{ and } A \]
Jacobian of the forward problem $F$

$K = \frac{\partial F}{\partial x}$

Wavenumber (cm$^{-1}$) vs Pressure (hPa)

Jacobian #1374, $P_0 = 10.111$ hPa
Posterior correlations from $\hat{S}$

Correlations for $\hat{S}$ #1374

Pressure (hPa)

Pressure (hPa)
Data assimilation interface

- Within the data assimilation, define the new observation as

\[ \hat{y}_A = \hat{x} - (I - A) x_a, \]

and the new observation operator by

\[ y_A = Ax. \]

We will now be comparing observed and simulated quantities with the same degree of smoothing. Now the observation increments

\[ \hat{y}_A - y_A = G_y \varepsilon_y, \]

are unbiased if the \( \varepsilon_y \) are unbiased, and have covariance

\[ S_m = G_y S_\varepsilon G_y^T = A\hat{S} \]
Averaging kernel, $A$

Rows of $A_R$ (eq 11) for # 1374, $P_0 = 10.111$ hPa
Old vs. New observations

\[ \hat{y}_A = \hat{x} - (I - A) \hat{x}_a \]

Retrievals, \( \hat{x} \)

\[ \tilde{y}_{\text{eq}}(\hat{x}) \]

Pressure (hPa)

Temperature (K)
Forecast as prior (interactive retrievals)

- Since the observation increment

\[ \hat{y}_A - y_A = G_y \varepsilon_y , \]

does not depend on the prior, we may use the ensemble background mean as \( x_a \) and the ensemble sample covariance as \( S_a \).

But \( G_y \) does depend on \( S_a \), the prior covariance.
Rotated averaging kernel method

If we scale by $S_m^{-1/2}$ we effectively rotate to a space where the observation errors are unbiased, uncorrelated, and have unit variance. The new observation is

$$\hat{y}_A = S_m^{-1/2} \left( \hat{x} - (I - A)x_a \right),$$

and the new observation operator by

$$y_A = S_m^{-1/2} Ax = A_R x,$$

where

$$A_R = S_m^{-1/2} A.$$
Rotated averaging kernel $A_R$

Rows of $A_R$ (eq 11) for # 1374, $P_0 = 10.111$ hPa

$$A_R = S_m^{-1/2} A$$

© 2011 Atmospheric and Environmental Research (AER). All rights reserved.
New vs. rotated observations

\[ \hat{y}_A = \hat{x} - (I - A)x_a \]
EOF analysis of the retrieval method

- Subtract \( \bar{x} = A\bar{x} + (I - A)x \) from \( \hat{x} = Ax + (I - A)x_a + G_y \varepsilon_y \) to get a version in terms of anomalies,

\[
\hat{x}' = Ax' + (I - A)x'_a + G_y \varepsilon_y.
\]

- We now project this version into EOF coefficient space.
  - For justification, the filtering step is normally done within the retrieval algorithm initially and at each step of iteration.

- If we define \( \tilde{A} = E^T A E \) then we have

\[
\hat{\alpha} = \tilde{A}\alpha + \left( I - \tilde{A} \right) \alpha_a + \varepsilon_\alpha
\]

with

\[
\varepsilon_\alpha = E^T G_y \varepsilon_y \quad \hat{\alpha} = E^T (\hat{x} - \bar{x}) \quad \alpha_a = E^T (x_a - \bar{x})
\]
Rotated EOF averaging kernel method

- $\varepsilon_a$ has covariance given by $S_\alpha = E^T S_m E$. We scale by $S_\alpha^{-1/2}$ to rotate to a space where the observation errors are unbiased, uncorrelated, and have unit variance.

- The new observation as $\hat{y}_A = S_\alpha^{-1/2} \left( \hat{\alpha} - (I - \tilde{A}) \alpha_a \right)$

- The new observation operator by $y_A = \tilde{A}_R \left( x - \bar{x} \right)$,

where we have defined the “rotated” EOF-space AK, $\tilde{A}_R = S_\alpha^{-1/2} \tilde{A} E^T$

Or work in EOF coefficient space:

$\tilde{A}_R = S_\alpha^{-1/2} \tilde{A} \quad y_A = \tilde{A}_R \alpha$
EOF projected (rotated) averaging kernel

\[\tilde{A}^T \]

\(A^T\) for #1374, \(P_0 = 10.111\) hPa

Rows of \(\tilde{A}_R\) (eq 28) for #1374, \(P_0 = 10.111\) hPa

\[\tilde{A}_R = S_{\alpha}^{-1/2} \tilde{A}^E\]
Old vs. EOF-projected observations

\[ \hat{y}_A = S_{\alpha}^{-1/2} \left( \hat{\alpha} - (I - \tilde{A}) \alpha_a \right) \]
Summary

• Convert standard retrievals into "observations" with expected errors that should be zero mean, uncorrelated, and unit variance, and independent of the background or prior.

• Define a corresponding obs-function (or H-operator) that is a weighted sum of the temperatures on the radiative transfer model vertical grid.
  • No changes to the assimilation method are needed, except to interpolate to the radiative transfer model vertical grid and to calculate the weighted sum.

• Projecting onto EOFs used by the retrieval can reduce the number of observations.

• Based on ideas in Rodgers' book, "Inverse Methods for Atmospheric Sounding: Theory and Practice".

• Next step: Try it out in our Mars LETKF data assimilation system.

For details visit http://arxiv.org/abs/1009.1561
Vertical interpolation

- We want to determine the EOF coefficients that best fit the temperatures from the dynamical model.
- Define an interpolation $V$ that transforms $x$ from the retrieval grid to the dynamical model grid.
- Then the best fit solution is

$$E^T V^T V E \alpha = E^T V^T (x_b - V \bar{x})$$