An Overview of Atmospheric Data Assimilation

3rd International Symposium on Integrating CFD and Experiments in Aeronautics

Capt. Paul Homan
Maj. Bob Wacker
USAFA/DFP
20 Jun 07
Report Documentation Page

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

<table>
<thead>
<tr>
<th>1. REPORT DATE</th>
<th>2. REPORT TYPE</th>
<th>3. DATES COVERED</th>
</tr>
</thead>
<tbody>
<tr>
<td>JUN 2007</td>
<td>N/A</td>
<td>-</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>4. TITLE AND SUBTITLE</th>
<th>5a. CONTRACT NUMBER</th>
</tr>
</thead>
<tbody>
<tr>
<td>An Overview of Atmospheric Data Assimilation</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>5b. GRANT NUMBER</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>5c. PROGRAM ELEMENT NUMBER</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>6. AUTHOR(S)</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)</th>
<th>8. PERFORMING ORGANIZATION REPORT NUMBER</th>
</tr>
</thead>
<tbody>
<tr>
<td>U.S. Air Force Academy, USA</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)</th>
<th>10. SPONSOR/MONITOR’S ACRONYM(S)</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>11. SPONSOR/MONITOR’S REPORT NUMBER(S)</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>12. DISTRIBUTION/AVAILABILITY STATEMENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Approved for public release, distribution unlimited</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>13. SUPPLEMENTARY NOTES</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>14. ABSTRACT</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>15. SUBJECT TERMS</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>16. SECURITY CLASSIFICATION OF:</th>
<th>17. LIMITATION OF ABSTRACT</th>
<th>18. NUMBER OF PAGES</th>
<th>19a. NAME OF RESPONSIBLE PERSON</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. REPORT</td>
<td>classified</td>
<td>classified</td>
<td></td>
</tr>
<tr>
<td>b. ABSTRACT</td>
<td>classified</td>
<td>classified</td>
<td></td>
</tr>
<tr>
<td>c. THIS PAGE</td>
<td>classified</td>
<td>classified</td>
<td></td>
</tr>
<tr>
<td></td>
<td>UU</td>
<td>37</td>
<td></td>
</tr>
</tbody>
</table>

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18
Overview

- **Purpose:** Convey how weather forecasting models incorporate diverse observational data

- Governing Equations
- Introduction to Meteorological Observations
- Introduction to Numerical Weather Prediction (NWP)
- Methods of Data Assimilation
  - Optimal Interpolation (OI)
  - 3-Dimensional Variational Assimilation (3DVAR)
  - 4-Dimensional Variational Assimilation (4DVAR)
  - Kalman Filter
Atmospheric Governing Equations

**Conservation of Momentum**

\[
\frac{dU}{dt} = -\frac{1}{\rho} \nabla p + g + F_r - 2\Omega \times U
\]

- Total acceleration of air parcel relative to Earth surface (non-inertial frame due to Earth rotation)
- Pressure gradient acceleration
- Acceleration due to gravity; includes effect of centrifugal force due to Earth rotation
- Viscous acceleration; neglected above planetary boundary layer; parameterized within PBL
- Coriolis acceleration (horizontal component 90° right of velocity in NH)
Atmospheric Governing Equations

 Conservation of Momentum

\[ \frac{dU}{dt} = -\frac{1}{\rho} \nabla p + g + F_r - 2\Omega \times U \]

**Geostrophic Balance:** Coriolis acceleration balances pressure-gradient acceleration in large-scale flow; mass field determines velocity field

\[ \frac{du}{dt} - \frac{uv \tan(\phi)}{a} + \frac{uw}{a} = -\frac{1}{\rho} \left( \frac{\partial p}{\partial x} \right) + F_{rx} + 2\Omega v \sin(\phi) - 2\Omega w \cos(\phi) \]

\[ \frac{dv}{dt} + \frac{u^2 \tan(\phi)}{a} + \frac{vw}{a} = -\frac{1}{\rho} \left( \frac{\partial p}{\partial y} \right) + F_{ry} - 2\Omega u \sin(\phi) \]

\[ \frac{dw}{dt} - \left( \frac{u^2 + v^2}{a} \right) = -\frac{1}{\rho} \left( \frac{\partial p}{\partial z} \right) \quad \frac{g}{g} + F_{rz} + 2\Omega u \cos(\phi) \]

**Hydrostatic Balance:** gravity balances vertical pressure-gradient in large-scale flow; temperature determines mass field

Atmospheric Governing Equations

- Conservation of Mass: Continuity Equation

\[
\frac{\partial \rho}{\partial t} = -\nabla \cdot (\rho \mathbf{V})
\]

- Conservation of Energy: Thermodynamic Equation

\[
\dot{Q} = \frac{d}{dt} \left[ \frac{u^2 + v^2 + w^2}{2} \right] + \Phi + c_v T + p\alpha \right] - \alpha \frac{dp}{dt} - \mathbf{V} \cdot \mathbf{F}_r
\]

- Rate of diabatic heating per mass
- Total energy of air parcel
- Adiabatic heating due to vertical motion
- KE per mass
- PE per mass
- Internal energy per mass due to temperature
- Internal energy per mass due to pressure
- Work done by friction
Atmospheric Governing Equations

- Equation of State

\[ p = \rho RT \]

- Conservation of water vapor mixing ratio

\[ \frac{\partial(\rho q)}{\partial t} = -\nabla \cdot (\rho \mathbf{V} q) + \rho (E - C) \]

*Total amount of water vapor in a parcel is conserved as the parcel moves except where there are sources (evaporation E) and sinks (condensation C)*
Equations Summary

- Seven Equations with seven unknowns
  - \( u, \ v, \ w, \ T, \ p, \ \rho, \ q \)

\[
\begin{align*}
\frac{du}{dt} - \frac{uv \tan(\phi)}{a} + \frac{uw}{a} &= -\frac{1}{\rho} \frac{\partial p}{\partial x} + F_{rx} + 2\Omega v \sin(\phi) - 2\Omega w \cos(\phi) \\
\frac{dv}{dt} + \frac{u^2 \tan(\phi)}{a} + \frac{vw}{a} &= -\frac{1}{\rho} \frac{\partial p}{\partial y} + F_{ry} - 2\Omega u \sin(\phi) \\
\frac{dw}{dt} - \frac{u^2 + v^2}{a} &= -\frac{1}{\rho} \frac{\partial p}{\partial z} - g + F_{rz} + 2\Omega u \cos(\phi) \\
\frac{\partial p}{\partial t} &= -\nabla \cdot (\rho \mathbf{V}) \\
\dot{Q} &= \frac{d}{dt} \left[ \frac{u^2 + v^2 + w^2}{2} + \Phi + c_vT + p\alpha \right] - \alpha \frac{dp}{dt} - \mathbf{V} \cdot \mathbf{F}_r \\
p &= \rho RT \\
\frac{\partial \rho q}{\partial t} &= -\nabla \cdot (\rho \mathbf{V} q) + \rho (E - C)
\end{align*}
\]
Introduction to Meteorological Observations

- **In situ**
  - Surface
  - Upper-air
  - Buoys
  - Aircraft
  - Ships

- **Remotely sensed**
  - Radar
  - Wind Profiler
  - Satellite soundings
  - Satellite winds
  - May not be meteorological variables (i.e. radiance, reflectivity, etc)
Weather Data Observing Platforms

- Geostationary satellites
- Polar orbiting satellites
- Surface instruments
- Aircraft
- Doppler radar
- Ships
- Buoys
- Radiosondes
Global Surface Ocean Observation Coverage
Global Aircraft Coverage – Typical 6 hour period
Global Satellite Wind Coverage
Polar Orbiting Satellite Temperature Sounding Coverage
Introduction to Numerical Weather Prediction

- Global models- Spectral
  - Spherical harmonic representation of data
  - Equations of motion in amplitude space

- ECMWF- European Center for Medium-Range Weather Forecasts
  - T799L91- Horizontal grid length of 25km and 91 vertical levels

- GFS- Run by NCEP (NOAA)
  - T382L64- Equivalent to about 40km resolution with 64 vertical levels

- NOGAPS- Fleet Numerical (Navy’s Global Model)
  - T239L30- 55km with 30 vertical levels

http://www.ecmwf.int/products/forecasts/d/charts/medium/deterministic/
Introduction to Numerical Weather Prediction

- Limited Area (Regional) Models - Finite Difference
  - Nested within larger global or regional models
  - Boundary conditions via global model or bigger regional model
  - WRF - As run by NCEP (NOAA)
    - 12km resolution at 60 vertical levels
  - MM5 - As run by AFWA
    - Nested Grid 45km, 15km resolution at 42 vertical levels
    - 5km output over selected regions

https://weather.afwa.af.mil/data_links/MWORLDLOCA.GIF
Anomaly Correlation of 500-hPa height for 3-, 5-, and 7-day forecasts for the ECMWF operational model as a function of year. Top and bottom of each band correspond to Northern and Southern Hemispheres, respectively.

- Difference between hemispheres has nearly disappeared in the past few years due to the successful assimilation of satellite data.
Anomaly Correlation Coefficient for ECMWF forecasts for different levels over the Northern Hemisphere in 2004.

http://www.ecmwf.int/products/forecasts/guide/The_data_assimilation_and_analysis_system.html
Introduction to Numerical Weather Prediction

- Sensitivity to parameterization of sub-grid processes
  - Introduction of chaos and error into models

- Model has $O(10^7)$ values to calculate (degrees of freedom) and observational quantities are $O(10^5)$

- Problem statement: *NWP is an under-determined initial value problem; how do we obtain the most realistic representation of the initial condition?*
We deal with three atmospheres

- Real atmosphere
  - Unknown
- Observed atmosphere
  - Data coverage gaps
    - vertical, horizontal, temporal
  - Observation error
    - random and systematic
- Analysis/model atmosphere
  - Observation limitations
  - Data assimilation system limitations
  - Model limitations
Three Sources of Information

- Imperfect Observations
- Imperfect Forecast Background
- Error Statistics Describing Imperfections

- Sampled Reality
- Model Physics
- Weighting

Best Linear Unbiased Estimate of Atmospheric State
The forecast model provides the background estimate of the current atmospheric state.

The analysis provides the initial conditions for the next forecast.

These corrections are added to the background to form the analysis.

The data assimilation algorithm produces corrections to the model fields.

Differences between observations and background are called departures or innovations.
Methods of Data Assimilation

- Optimal Interpolation (OI)
- 3-Dimensional Variational Assimilation (3DVAR)
- 4-Dimensional Variational Assimilation (4DVAR)
- Extended Kalman Filter
- Ensemble Kalman Filter
**Methods of Data Assimilation**

- **Optimal Interpolation**

  one observation

  \[ T_a = T_b + W(T_o - T_b) \]

  one background value

  \[ x_a = x_b + W[y_o - H(x_b)] \]

  multiple variables

  \[ x_a = x_b + Wd \]

  interpolate each grid point to the observation location first

  then weight each innovation

  \[
  \begin{bmatrix}
  x_1 \\
  \vdots \\
  x_n \\
  \end{bmatrix}
  \quad \begin{bmatrix}
  d_1 \\
  \vdots \\
  d_p \\
  \end{bmatrix}
  \quad \begin{bmatrix}
  w_1 & \cdots & w_{p1} \\
  \vdots & \ddots & \vdots \\
  w_{1n} & \cdots & w_{np} \\
  \end{bmatrix}
  \]

  number of model variables

  number of observations

  \[ n \times p \]
Methods of Data Assimilation

- Optimal Interpolation
  - Optimal analysis is found by minimizing the analysis error variance (A) by finding the optimal weights of observation increments through a least squares approach

\[
\frac{1}{\sigma_a^2} = \frac{1}{\sigma_o^2} + \frac{1}{\sigma_b^2}
\]

\[
W = \frac{\sigma_b^2}{\sigma_o^2 + \sigma_b^2}
\]

\[
W = B H^T (R + H B H^T)^{-1}
\]

\[
A = (I - WH) B
\]

- \(\sigma_a^2\) = analysis error variance
- \(\sigma_o^2\) = observational error variance
- \(\sigma_b^2\) = background error variance
- \(W\) = optimal weight
- \(W\) = weight matrix
- \(B\) = background error covariance matrix
- \(R\) = observations error covariance matrix
- \(H\) = linear observation operator matrix
- \(I\) = identity matrix
- \(A\) = analysis error covariance matrix
Methods of Data Assimilation

3DVAR Introduction - Cost Function

Why use the term Cost Function?

\[ J(x) = \frac{1}{2} \left[ \frac{(y - x_a)^2}{\sigma_o^2} + \frac{(x_b - x_a)^2}{\sigma_b^2} \right] \]

- \( x_a \) = analyzed value
- \( y \) = observed value
- \( x_b \) = background value

Think of the function as defining how far away the analysis value is away from the input values. You will be “charged” for not fitting the observation. You will also be “charged” for not fitting the background. You choose the analysis to minimize your “cost.”
3DVAR Introduction- Vector Form of Cost Function

Vector Form:

\[ J(x) = \frac{1}{2} \left[ (y - x_a)^T R^{-1} (y - x_a) + (x_b - x_a)^T B^{-1} (x_b - x_a) \right] \]

Minimizing \( J(x) \) with respect to \( x_a \) yields:

\[ x_a = x_b + B (B + R)^{-1} (y - x_b) \]
Methods of Data Assimilation

- **3DVAR**
  - Minimizes cost function
  - Performs analysis at set times
    - Only observational data from set times are included or observations in a +/- time window are lumped together and given essentially an equivalent time
  - Employs a forward operator or observation operator: H
    - Interpolates observations spatially to grid points
    - Converts observed quantities (i.e. satellite radiances, radar reflectivities) into model variables (i.e. temperatures, humidities)

\[
J(x) = \min \frac{1}{2} \left[ (x_b - x_a)B^{-1}(x_b - x_a) + (y - Hx_a)R^{-1}(y - Hx_a) \right]
\]

Distance to forecast
Distance to observations
At analysis time
Methods of Data Assimilation

- **4DVAR**
  - Minimizes cost function

  - Cost function is weighted difference between model forecasts during an assimilation window and coincident observations

    \[
    J(x) = \min \frac{1}{2} \left[ (x(t_0) - x_a(t_0))B^{-1}(x_b(t_0) - x_a(t_0)) + \sum_{i=0}^{N} (y_i - Hx_i)R^{-1}(y_i - Hx_a) \right]
    \]

    - Distance to background at initial time
    - Distance to observations in a time window interval

  - 4D-Var seeks initial conditions such that the forecast best fits the observations within the assimilation interval

  - Computes the increments at observation times during a forward integration using the forecast model and then integrates these weighted increments back into the initial time using the adjoint model \((L^T)\)
Methods of Data Assimilation

4DVAR data assimilation window

assimilation window

3z 6z 9z 12z 16z time

x_b J_b x_a J_o obs obs obs J_o J_o analysis corrected forecast previous forecast
Methods of Data Assimilation

- Extended Kalman Filter
  - Same weighted approach as OI, but using a background error covariance matrix (P) that evolves with the forecast rather than a constant background covariance matrix (B)
  - Forecast error covariance is obtained using a tangent linear model (L) to transform the perturbation from the initial time to the final time and the adjoint model (L^T) to "advance" the perturbation backwards from the final to initial time to optimize the initial conditions

Forecast step:

\[ x_n^f = M_n(x_{n-1}^q) \]
\[ P_n = L_n A_{n-1} L_n^T + Q_n \]

Analysis step:

\[ x_n^q = x_n^f + K_n (y_n - H x_n^f) \]

The optimal weight matrix:

\[ K_n = P_n (R + H P_n H^T)^{-1} \]

The analysis error covariance:

\[ A_n = (I - K_n H)_n \]
“Extended Kalman filter is gold standard of data assimilation” (Kalnay, 2006)

A poor initial guess can be transitioned through time to provide the best linear unbiased estimate of the state of the atmosphere and its error covariance provided observations are frequent and system is stable (Kalnay, 2006)

Very computationally expensive

So can it be done more cheaply?
Methods of Data Assimilation

- **Ensemble Kalman Filter**
  - Estimates forecast error covariance matrix from ensemble of forecasts initialized by random perturbations added to the same sets of observations

  - Computational cost increased $O(10^2)$ from OI or 3DVAR, but cheap compared to Extended Kalman Filter whose cost is $O(10^7)$

  - Tangent linear or adjoint model not necessary

  - Most promising approach for the future
Limit of predictability is ~2 weeks due to sensitive dependence on initial condition (chaos)

Quality of forecast therefore is critically dependent on the quality of the initial condition

Throughout 50 years of NWP, methods of obtaining that initial condition have evolved with increasing sophistication of NWP models, new remotely-sensed observation types, and increasing computing power
Acknowledgments


- Slides 9-15 and 21-25 were taken directly from course notes and presentations of MR4323, Air Ocean Numerical Modeling, at the Naval Postgraduate School. A special thanks to Commander Rebecca Stone and Prof Mary Jordan for providing this material.
Questions?