Quantifying the Impact of Data and Model Errors in DDDAS*

Adrian Sandu  
Vishwas Hebbur Venkata Rao  
*(with other students and postdocs of the Computational Science Laboratory)*  
Virginia Polytechnic Institute and State University

Dr. Frederica Darema
Data assimilation fuses information from prior, model, and observations, to reduce uncertainty about system background state.

### Chemical kinetics

### Aerosols

### Emissions

### Transport
  - Meteorology

### Model

### Observations

### Data Assimilation

### Configure sensors

### Analysis state
Assimilation of TES ozone column observations from Aug. 2006. IONS-6 ozonesonde data for validation.

TES is one of four instruments on the NASA EOS Aura platform, launched July 14 2004

(Singh, S., et al)
Quality of TES ozone column assimilation results for several DA methods (August 1-15, 2006)

[Singh, Sandu et. al., 2010]
DA is formulated in Bayesian framework, which produces a complete solution to inversion.

- The analysis (posterior) probability density $\mathcal{P}^a(x)$:

  \[
  \text{Bayes: } \mathcal{P}^a(x) = \mathcal{P}(x|y) = \frac{\mathcal{P}(y|x) \cdot \mathcal{P}^b(x)}{\mathcal{P}(y)}.
  \]

Note: In the linear, Gaussian case, Bayes admits closed form solution (the Kalman filter).
The analysis, a best estimate which encapsulates our enhanced knowledge, can be defined in several ways.

**Practical methods:**

1. MAP: 4D-Var
2. MinVar: EnKF
MAP estimate: 4D-Var formulates DA as a single, model-constrained optimization problem

\[ J(x_0) = -\log \mathcal{P}^a(x|y) = -\log \mathcal{P}^b(x) - \log \mathcal{P}(y|x) + \text{const} \]

\[ \approx \frac{1}{2} \left\| x_0 - x_0^b \right\|_B^{-1} + \frac{1}{2} \sum_{i=1}^{N} \left\| \mathcal{H}(x_i) - y_i \right\|_{R_i}^{-1} \]

\[ x_0^a = \arg\min_{x_0} J(x_0) \quad \text{subject to:} \quad x_i = \mathcal{M}_{t_0 \rightarrow t_i}(x_0), \ i = 1, \ldots, N \]
Problem: how do data, model errors impact results?

Reality → Observations

Measurements → Modeling & Discretization

Computer Model → Observation Targeting

Data Assimilation

Inference
**DDDAS/Inverse problem: mathematical formulation**

\[ \theta_{\text{opt}} = \arg \min \ J = \int_{t_0}^{t_F} r(x(t), \theta) \, dt + w(x(t_F), \theta) \]

Subject to forward model:

\[ x' = f(t, x, \theta), \quad x(t_0) = x_0(\theta) \]

Adjoint equation:

\[ \lambda' = -r_x^T(x(t), \theta) - f_x^T(t, x, \theta) \cdot \lambda, \quad \lambda(t_F) = w_x^T(x(t_F), \theta) \]

Optimality condition:

\[ \mu' = -r_\theta^T(x(t), \theta) - f_\theta^T(t, x, \theta) \cdot \lambda, \quad \mu(t_F) = w_\theta(x(t_F), \theta) \]
\[ 0 = \mu(t_0) + x_\theta^T(t_0) \cdot \lambda(t_0) \]
Quantify error in an aspect of interest of the optimal solution (a.k.a., the goal of the inversion)

The aspect of interest of the optimal solution:

\[ E(\theta) \]

Error in forward model:

\[ x' = (f + \Delta f)(t, x, \theta), \quad x(t_0) = x_0(\theta) \]

Error in data/cost function:

\[ J + \Delta J = \int_{t_0}^{t_F} (r + \Delta r)(x(t), \theta) \, dt + (w + \Delta w)(x(t_F), \theta) \]

Error functional changes:

\[ (E + \Delta E)(\theta) \quad \text{with} \quad \Delta E = \Delta E_{adj} + \Delta E_{fwd} + \Delta E_{opt} \]
Aposteriori estimation algorithm for the error in the DDDAS inference solution

1. Solve the optimality equation for $\nu_{\theta}$:

$$j_{\theta,\theta} \cdot \nu_{\theta} = \left( \frac{d^2}{d\theta^2} J(x(\theta), \theta) \right)_{\theta_{\text{opt}}} \cdot \nu_{\theta} = \mathbf{E}_{\theta}$$

2. Solve the tangent linear model for $\nu_{x}$:

$$-\nu'_{x} + f_{x} \cdot \nu_{x} + f_{\theta} \cdot \nu_{\theta} = 0, \quad \nu_{x}(t_0) = x_{\theta}(t_0) \cdot \nu_{\theta}(t_0)$$

3. Solve the second order adjoint equation for $\nu_{\lambda}$:

$$\nu'_{\lambda} + f_{x}^{T} \cdot \nu_{\lambda} + r_{x,x} \cdot \nu_{x} + (f_{x,x} \cdot \nu_{x})^{T} \cdot \lambda + r_{\theta,x} \cdot \nu_{\theta} + (f_{\theta,x} \cdot \nu_{\theta})^{T} \cdot \lambda = 0$$

$$\nu_{\lambda}(t_F) = w_{\theta,x}^{T} \cdot \nu_{\theta}(t_F) + w_{x,x}^{T} \cdot \nu_{x}(t_F)$$
Aposteriori error estimation (contd.)

\[
\Delta E = \Delta E_{adj} + \Delta E_{fwd} + \Delta E_{opt}
\]

\[
\Delta E_{adj} = \int_{t_0}^{t_F} \nu_x^T \cdot (\Delta r_x + \Delta f_x^T \cdot \lambda) \, dt + \nu_x^T \cdot \Delta w_x \bigg|_{t_F}
\]

\[
\Delta E_{fwd} = \int_{t_0}^{t_F} \nu_\lambda^T \cdot \Delta f \, dt
\]

\[
\Delta E_{opt} = \int_{t_0}^{t_F} \nu_\theta^T \cdot (\Delta r_\theta + \Delta f_\theta^T \cdot \lambda) \, dt
\]
4D-Var with imperfect model and data

\[ x^a_0 = \arg \min_{x_0} J(x_0) \]

Imperfect forward model:

\[ x' = (f + \Delta f)(t, x), \quad x(t_0) = \theta \]

Cost function with imperfect data:

\[ J(x_0) = (x_0 - x^b_0)^T B_0^{-1} (x_0 - x^b_0) \]
\[ + \frac{1}{2} \sum_{i=1}^{N} (\mathcal{H}(x_i) - (y_i + \Delta y_i))^T R_i^{-1} (\mathcal{H}(x_i) - (y_i + \Delta y_i)) \]
Error impact in 4D-Var data assimilation

Error impact via the adjoint equation:

\[ \lambda' = -f_x^T \cdot \lambda - \sum_{i=1}^{N} \mathbf{H}_i^T \mathbf{R}_i^{-1} \left( \mathcal{H}(\mathbf{x}_i) - (\mathbf{y}_i + \Delta \mathbf{y}_i) \right) \delta(t - t_i) \]

\[ \Delta E_{\text{adj}} \approx \sum_{i=1}^{N} \nu_x^T(t_i) \cdot (\mathbf{H}_i^T \mathbf{R}_i^{-1} \Delta \mathbf{y}_i) + \sum_{i=1}^{N} \Delta t_i \Delta f_x^T(t_i) \cdot \lambda(t_i) \]

Error impact is via the forward equation:

\[ \Delta E_{\text{fwd}} = \int_{t_0}^{t_F} \nu_\lambda^T \cdot \Delta f \, dt \approx \sum_{i=1}^{N} \Delta t_i \nu_\lambda^T(t_i) \cdot \Delta f(t_i) \]
Test problem: the heat equation

- Test problem (heat equation)

\[ u_t = u_{xx} + s(t, x) \]
\[ x \in [0, 1], \quad u(t_0) = u_0, \quad u(t, 0) = u(t, 1) = 0. \]

- Convert the above PDE to an ODE by using central finite difference scheme in space, which leads to linear ODE.
- Invert for initial conditions

- Quantity of interest:

\[ E = \frac{1}{N_{\text{grid}}} \sum_{i=1}^{N_{\text{grid}}} u^a(t_0, x_i) \]
The solution of the heat equation in time and space
Solution to the associated adjoint equation
Impact of data errors on fine grid

Error in observations

Contributions to a posteriori error

Aposteriori estimate: 1.01e-2
Actual error: 1.30e-2
Impact of model errors on coarse grid

Artificial model error introduced by adding constants (=1) except for leftmost grid cell

Error contributions of different grid points at separate time instances.

Aposteriori estimate: 10.45e-3
Actual error: 7.88e-3
DDDAS project status after the first year

- Theory complete to characterize impact of data and model errors on DDDAS inference results
- One Ph.D. student (V. Rao) educated on the topic
- Currently working on implementation and demonstration with more realistic test problems: shallow water on the sphere, later WRF
- Error impact on sensor configuration problem
SIAM/ASA Journal on UNCERTAINTY QUANTIFICATION

Publishing research articles presenting significant mathematical, statistical, algorithmic, and application advances in uncertainty quantification.

Senior Editor: Max Gunzburger
Editors-in-Chief: James Berger and Donald Estep
More than 35 other professionals comprising the editorial board.

Continuous electronic publication covering the analysis and quantification of areas as divergent as finance, disaster preparedness, and porous media flows.

The first papers have published online to Volume 1, and are freely available through 2013.


Authors submit for consideration at http://juq.siam.org.