Weight Selection when Combining Models with Data

Jodi L. Mead

Department of Mathematics
Center for Geophysical Investigation of the Shallow Subsurface (CGISS)
Boise State University

Bridging-the-Gap Workshop at Princeton University Oct. 1-2, 2012

Why am I here?

To work across boundaries...

- as a mathematician work on problems in near subsurface science with BSU collaborators in geophysics and hydrology, and other collaborators in earth sciences.
- combine probabilistic/statistical models with mathematical models to take advantage of both.
- the intersection of inversion, data assimilation and uncertainty.

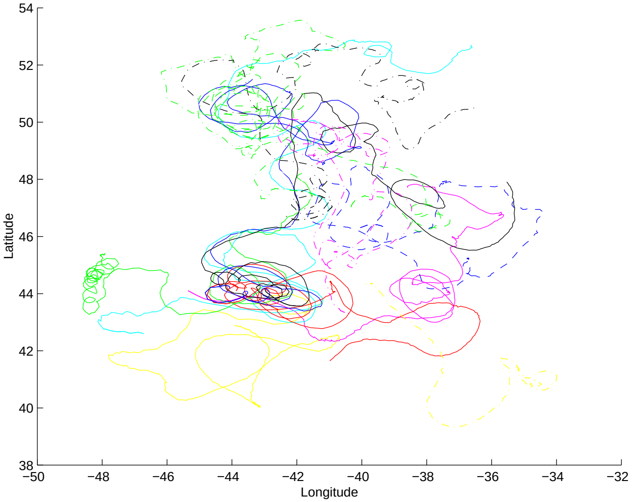
- Accurate prior knowledge is critical for UQ success, and can be estimated by
  - additional parameters in parameter estimation problem (Bayesian)
  - covariance models (Statistics)
  - regularization parameters (Inversion)
  - weights in a deterministic setting (Data Assimilation)

Data Assimilation

- Strengths include the ability to incorporate different types of information into complex models.
- Kalman filter, variational assimilation use least squares
  - assumes Gaussian distributed error,
  - smoothes solutions
- Alternatively, use least squares without Gaussian assumption
  - Covariance matrix weights give piecewise smooth solutions

- The focus is typically on Regularization whereby we make an ill-posed problem well-posed.
- The problem is ill-posed because we don't have enough information to get a solution.
- Regularization can be viewed as adding information to the problem.
    - Finding a regularization parameter amounts to finding weights or uncertainties for the added information.

Oceanographic float data in the North Atlantic



RAFOS sub-surface float

Data: Lagrangian float data

$$\mathbf{d} = [\lambda\ \theta\ h]^T + \boldsymbol{\epsilon}$$

Mathematical model: Lagrangian shallow water equations

$$\frac{\partial^2\lambda}{\partial t^2} = 2\tan\theta \frac{\partial\lambda}{\partial t}\frac{\partial\theta}{\partial t} + \frac{f}{\cos\theta}\frac{\partial\theta}{\partial t} - \frac{g}{r^2\cos^2\theta}J^{-1}\left(\frac{\partial h}{\partial\alpha}\frac{\partial\theta}{\partial\beta} - \frac{\partial\theta}{\partial\alpha}\frac{\partial h}{\partial\beta}\right) + \text{friction}_\lambda + f_\lambda$$

$$\frac{\partial^2\theta}{\partial t^2} = -\sin\theta\cos\theta\left(\frac{\partial\lambda}{\partial t}\right)^2 - f\cos\theta\frac{\partial\lambda}{\partial t} - \frac{g}{r^2}J^{-1}\left(\frac{\partial\lambda}{\partial\alpha}\frac{\partial h}{\partial\beta} - \frac{\partial h}{\partial\alpha}\frac{\partial\lambda}{\partial\beta}\right) + \text{friction}_\theta + f_\theta$$

$$\frac{\partial}{\partial t}(h\cos\theta J) = f_h \quad J \equiv \frac{\partial\lambda}{\partial\alpha}\frac{\partial\theta}{\partial\beta} - \frac{\partial\theta}{\partial\alpha}\frac{\partial\lambda}{\partial\beta}$$

$$\lambda(0) = \alpha + i_\lambda, \quad \theta(0) = \beta + i_\theta, \quad h(0) = h_0 + i_h$$

Variational Formulation

Find dynamics that fit data within specified errors

$$(\hat{\lambda}, \hat{\theta}, \hat{h}) = \underset{(\lambda,\theta,h)}{\operatorname{argmin}} \mathcal{J}$$

$$\begin{aligned}\mathcal{J}(\lambda,\theta,h) \quad = \quad & \iiint f_\lambda C_{f_\lambda}^{-1} f_\lambda + \iiint f_\theta C_{f_\theta}^{-1} f_\theta + \iiint f_h C_{f_h}^{-1} f_h \\ & + \iint i_\lambda C_{i_\lambda}^{-1} i_\lambda + \iint i_\theta C_{i_\theta}^{-1} i_\theta + \iint i_h C_{i_h}^{-1} i_h \\ & + \boldsymbol{\epsilon}^T \mathbf{C}_\epsilon^{-1} \boldsymbol{\epsilon}\end{aligned}$$

Representer solution

Bennett, “Inverse Modeling of the Ocean and Atmosphere”, Cambridge University Press, 2002

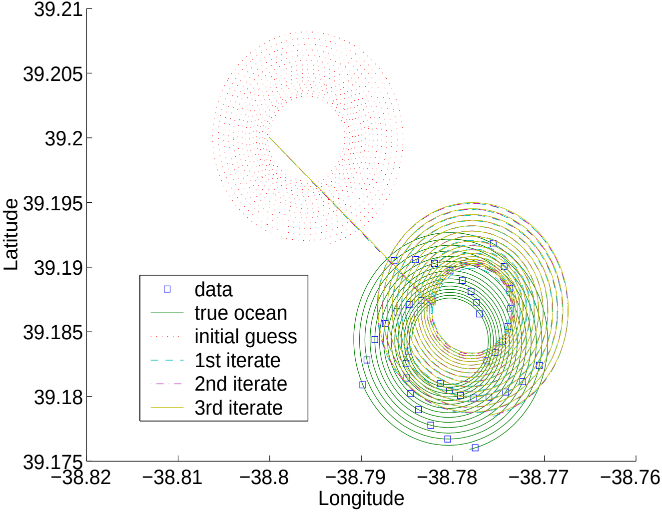
Choice of error variances

	Std. dev. as % of range of reasonable values		
	dynamics *acceleration*	initial conditions *position and velocity*	data *depth and domain*
1	40 %	10%	10 %
2	0.1 %	0.01 %	20 %
3	1 %	1.0 %	10 %

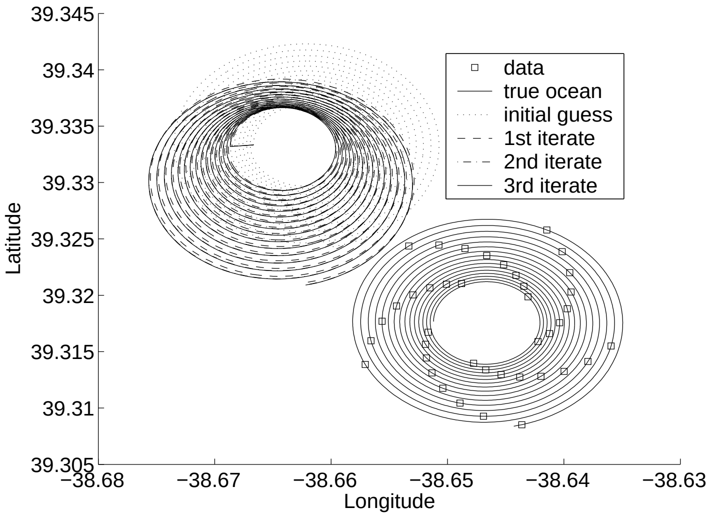
- Experiment 1 heavily weights data.
- Experiment 2 heavily weights the dynamics.
- Experiment 3 doesn't heavily weight either.

The solution will go wherever we place weight

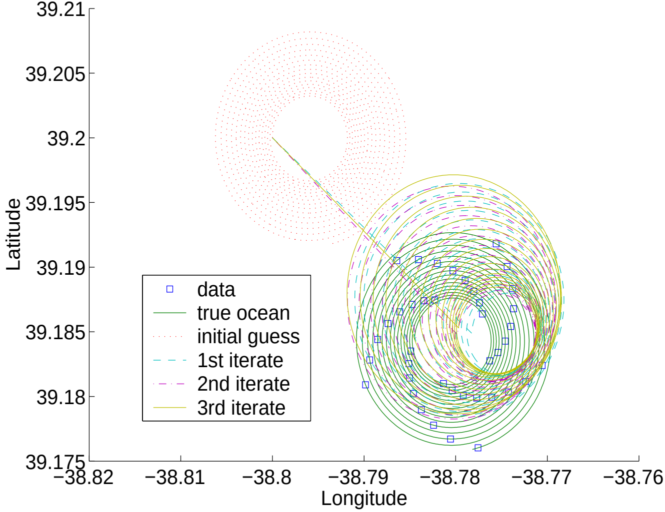
Assimilation Results from Experiment 1



Assimilation Results from Experiment 2



Assimilation Results from Experiment 3



χ^2 Tests

- How do we know if our uncertainties/weights, are correct?

$$\chi^2 \text{ test}: \quad \mathcal{J}(\hat{\mathbf{x}}) \approx \text{number of data}$$

Applies to non-Gaussian errors

- Alternatively, use test to find uncertainties/weights
  - Discrepancy principle (Regularization)

$$\mathcal{J}_{\text{data}}(\hat{\mathbf{x}}) \approx \text{number of data}$$

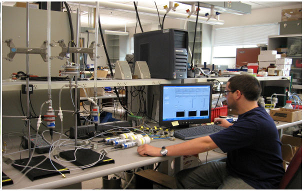
  - χ^2 Method (*Mead and Renaut 2009, Inverse Problems*)

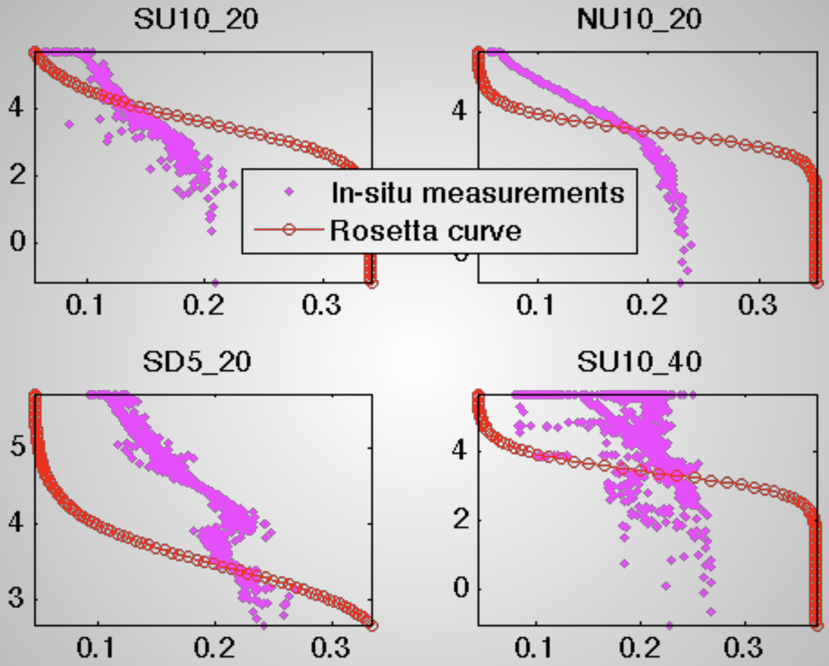
$$\mathcal{J}(\hat{\mathbf{x}}) \approx \text{number of data}$$

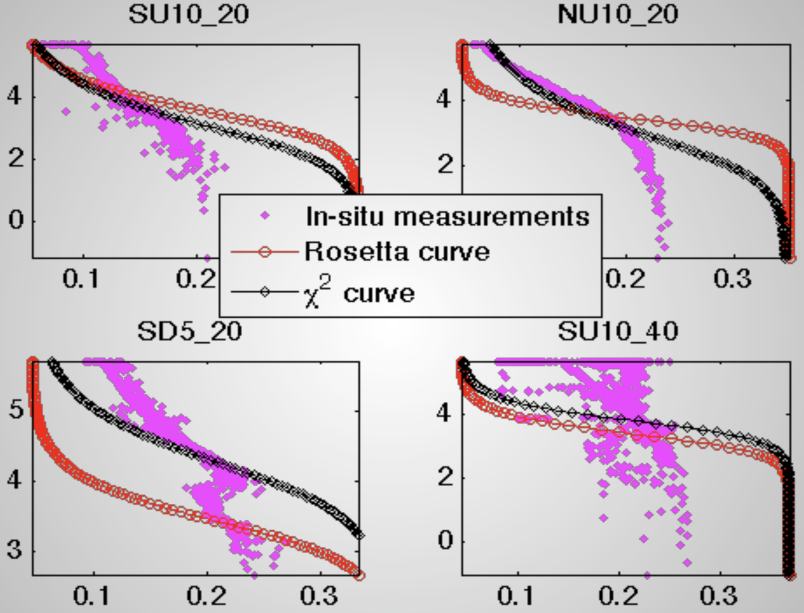
Two different types of data

- In-situ measurements of soil moisture and pressure head.

- Collect soil samples, measure % of sand, silt, clay etc. and input in neural network algorithm (Rosetta).




SU10_20
NU10_20
SD5_20
SU10_40
In-situ measurements
Rosetta curve
0.1
0.2
0.3



SU10_20
NU10_20
SD5_20
SU10_40
In-situ measurements
Rosetta curve
χ2 curve
0.1
0.2
0.3

χ^2 Method

- Uses the well known χ^2 test to "back out" uncertainty information.
  - Computationally efficient, statistically justified with minimal assumptions, but not statistically deep.
- Extensions
  - More dense uncertainty estimates/weights can be estimated with multiple χ^2 tests (*Mead 2012, in revision*)
  - χ^2 tests for nonlinear problems (*Mead and Hammerquist 2012, submitted*)
  - Statistical tests with L_1 functional (*Mead and Nelson, in preparation*)

$$\mathcal{J}(\mathbf{x}) = \|\mathbf{d} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda\|\mathbf{L}(\mathbf{x} - \mathbf{x}_0)\|_1$$

$$\begin{aligned} \mathcal{J}(\hat{\mathbf{x}}) &\approx \text{Number of data} \\ \lambda &\approx \sigma_{Lx}^{-1} \end{aligned}$$