Assimilation of current measurements into a circulation model of Lake Michigan

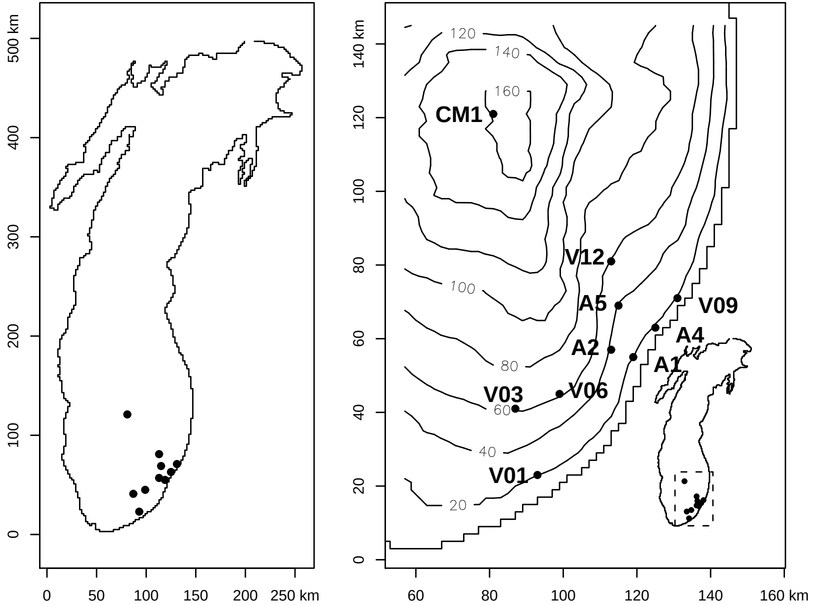
Zepu Zhang, Michael L. Stein
Department of Statistics, University of Chicago

Dmitry Beletsky, David J. Schwab
Great Lake Environmental Research Laboratory, NOAA

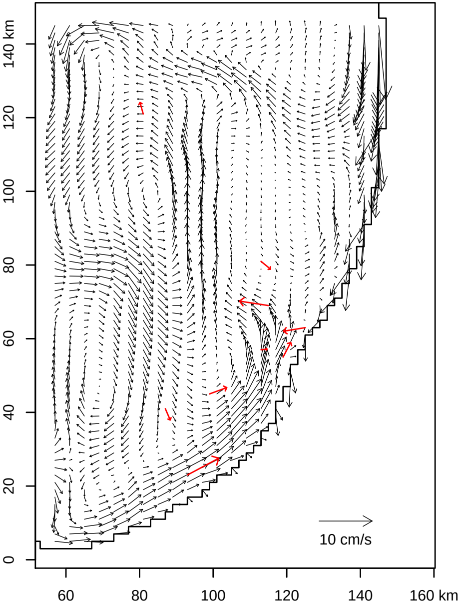
revised May 14, 2007

sponsored by EPA

Lake Michigan and observational sites



A typical view of model output vs observations



Background

Hydrodynamics: Princeton Ocean Model

Forcing: observed meteorology (winds)

Prediction: $\mathbf{u}$, hourly, 2-D (horizontal) velocity

Observations: $\hat{\mathbf{u}}$ at 10 sites, hourly, day 1–60, 1998

Our task:

"Assimilate" data into model output
➜ to increase agreement of prediction with validating data

Constraint 1: physical

- no measurement of surface level
- sparse data sites
- incompressible water

$\Downarrow \quad \Downarrow \quad \Downarrow$

Velocity-updates $\mathbf{\Delta}$ should not change model-predicted water depth h:

$$\left.\frac{\partial h}{\partial t}\right|_{\text{caused by }\mathbf{\Delta}} = (h\Delta_1)_x + (h\Delta_2)_y = 0 \tag{1}$$

$\Rightarrow$ $\mathbf{\Delta}$ can be represented by a "stream function", ψ, as

$$\mathbf{\Delta} = \left(\frac{1}{h}\psi_y, \ -\frac{1}{h}\psi_x\right). \tag{2}$$

The scalar field ψ should be properly differentiable.

Constraint 2: coastal

- no cross-shore velocity component
- model-prediction is (assumed) parallel to the coast

$$\Downarrow \quad \Downarrow \quad \Downarrow$$

Updates $\mathbf{\Delta}$ are parallel to the coast.

$\Rightarrow$ the coastline is a contour of ψ.

$\Rightarrow$ enforced by "pseudo data" along the coast:

$$\psi(\mathbf{c}) = 0 \quad \text{for coastal point } \mathbf{c}$$

Constraint 3: numerical

Ignoring data error, then the update at data sites should follow the observations:

- $\mathbf{\Delta} = \hat{\mathbf{u}} - \mathbf{u}$

$$\Downarrow \quad \Downarrow \quad \Downarrow$$

Define

$$\hat{\boldsymbol{\xi}} \triangleq h(\hat{\mathbf{u}} - \mathbf{u})$$
$$\boldsymbol{\xi} = h\mathbf{\Delta} = \left(\psi_y,\ -\psi_x\right)$$

$\Rightarrow$ at any data site,

$$\hat{\mathbf{\Delta}} = \hat{\boldsymbol{\xi}}/h$$

Data vector

$$\mathbf{z} = \left[\psi(\mathbf{p}_1^{(c)}), \ldots, \psi(\mathbf{p}_m^{(c)}),\right.$$
$$\xi_1(\mathbf{p}_1^{(d)}), \ldots, \xi_1(\mathbf{p}_n^{(d)}),$$
$$\left.\xi_2(\mathbf{p}_1^{(d)}), \ldots, \xi_2(\mathbf{p}_n^{(d)})\right]^T$$

Contains information about ψ, ψ_x, ψ_y, $\boldsymbol{\epsilon}$, and h.

Pseudo data $\hat{\psi}(\mathbf{p}^{(c)}) = 0$ maintain parallel-to-the-coast velocities near the coast.

$\hat{\xi}_1(\mathbf{p}^{(d)})$ and $\hat{\xi}_2(\mathbf{p}^{(d)})$ steer $\mathbf{u}$ at $\mathbf{p}^{(d)}$ to be the observed $\hat{\mathbf{u}}$.

Kriging interpolation

Estimate $\mathbf{\Delta}$ everywhere by spatial interpolation using data $\mathbf{z}$:

$$\hat{\Delta}_1(\mathbf{x}) = \mathbf{k}^T \boldsymbol{\lambda}$$

where

$\mathbf{k} = \text{cov}\big(\Delta_1(\mathbf{x}), \mathbf{z}\big)$: unknown-to-data covariance vector

$\boldsymbol{\lambda}$: interpolation coefficients

$$\boldsymbol{\lambda} = \mathbb{K}^{-1}\mathbf{z}$$

where $\mathbb{K} = \text{cov}(\mathbf{z})$ is data-to-data covariance matrix.

Therefore

$$\hat{\Delta}_1 = \mathbf{k}^T \mathbb{K}^{-1} \hat{\mathbf{z}}$$

Similar for Δ_2.

Need to know covariances between things...

k & $\mathbb{K}$ involve covariances between ψ, ψ_x, and ψ_y, which are all functions of the covariance function of ψ, which we assume to be spatially stationary:

$$\mathrm{cov}\big(\psi(\mathbf{p}_1), \psi(\mathbf{p}_2)\big) \triangleq K(\boldsymbol{\ell}), \quad \text{where } \boldsymbol{\ell} = \mathbf{p}_2 - \mathbf{p}_1.$$

Calculation of $\mathbb{K}$ and **k** is based on the following relations:

$$\begin{aligned}
\mathrm{cov}\big(\psi(\mathbf{p}_1), \psi_x(\mathbf{p}_2)\big) &= K_x(\mathbf{p}_2 - \mathbf{p}_1) \\
\mathrm{cov}\big(\psi(\mathbf{p}_1), \psi_y(\mathbf{p}_2)\big) &= K_y(\mathbf{p}_2 - \mathbf{p}_1) \\
\mathrm{cov}\big(\psi_x(\mathbf{p}_1), \psi_x(\mathbf{p}_2)\big) &= -K_{xx}(\mathbf{p}_2 - \mathbf{p}_1) \\
\mathrm{cov}\big(\psi_y(\mathbf{p}_1), \psi_y(\mathbf{p}_2)\big) &= -K_{yy}(\mathbf{p}_2 - \mathbf{p}_1) \\
\mathrm{cov}\big(\psi_x(\mathbf{p}_1), \psi_y(\mathbf{p}_2)\big) &= -K_{xy}(\mathbf{p}_2 - \mathbf{p}_1) \\
\mathrm{cov}\big(\psi_y(\mathbf{p}_1), \psi_x(\mathbf{p}_2)\big) &= -K_{xy}(\mathbf{p}_2 - \mathbf{p}_1)
\end{aligned}$$

But we don't know K!...

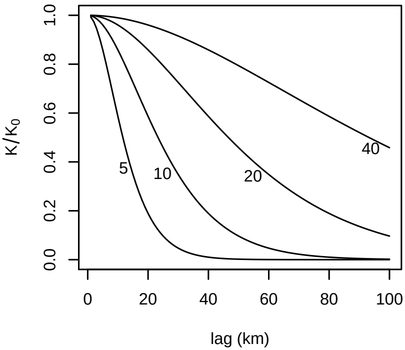
— Let's assume one:

$$K(\boldsymbol{\ell}) = K(\ell) = K_0 e^{-\ell/\alpha}(1 + \ell/\alpha + \ell^2/(3\alpha^2))$$

where

$K_0 = K(0)$

α: "range"
(Larger α corresponds to covariance structures with longer ranges of influence.)



Then all elements of $\mathbb{K}$ and $\mathbf{k}$ are derived as functions of this K.

Define a measure to quantify the impact of assimilation

Define an indicator of the "similarity" between two vectors:

$$\rho(\mathbf{u}_1, \mathbf{u}_2) = 1 - \frac{\|\mathbf{u}_1 - \mathbf{u}_2\|}{b + \max(\|\mathbf{u}_1\|, \|\mathbf{u}_2\|)}$$

where empirical parameter b, $b > 0$, helps to maintain continuity of ρ and de-emphasize pairs of small vectors.

- $\rho \in (-1, 1]$
- $\rho \to -1$ when $\mathbf{u}_1 = -\mathbf{u}_2$ (and $\|\mathbf{u}_1\| \gg b$, $\|\mathbf{u}_2\| \gg b$)
- $\rho = 1$ when $\mathbf{u}_1 = \mathbf{u}_2$

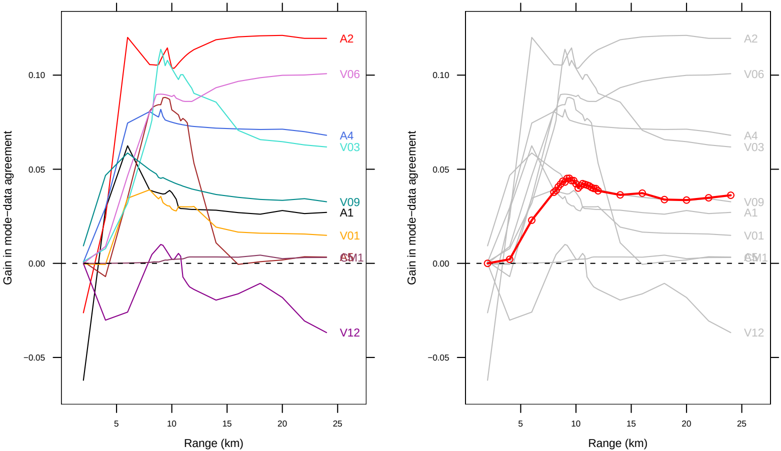
Now with this tool,

- $\rho(\mathbf{u}, \hat{\mathbf{u}})$ large $\longrightarrow$ original prediction is good
- $\rho(\mathbf{u} + \hat{\boldsymbol{\Delta}}, \mathbf{u})$ small $\longrightarrow$ assimilation has big impact
- $\rho(\mathbf{u} + \hat{\boldsymbol{\Delta}}, \hat{\mathbf{u}})$ large $\longrightarrow$ assimilation result is good

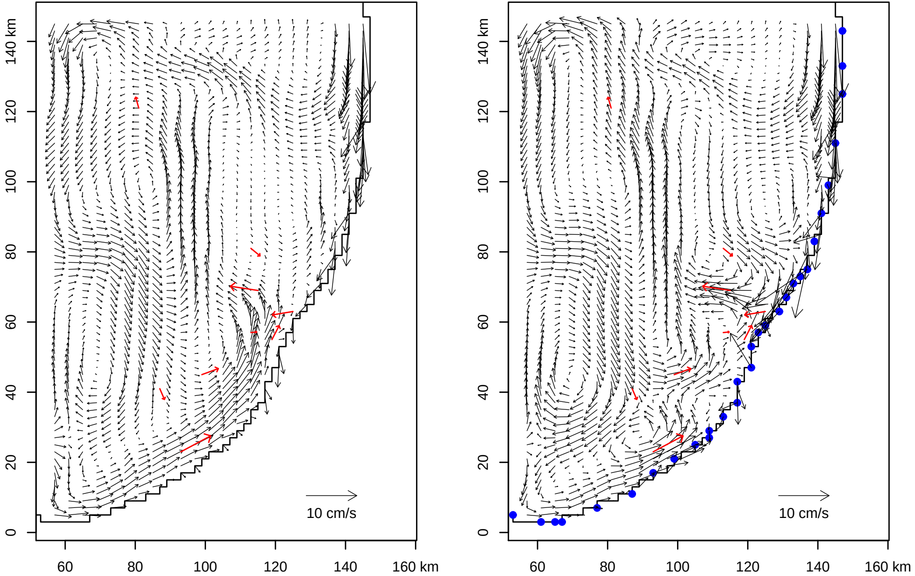
Estimating range α

Leave one site out, assimilate other sites, find α that makes greatest improvement to the site left out.

$$\hat{\alpha} = 9.4 \text{ km}$$



Results of assimilation, with 100 pseudo coastal points

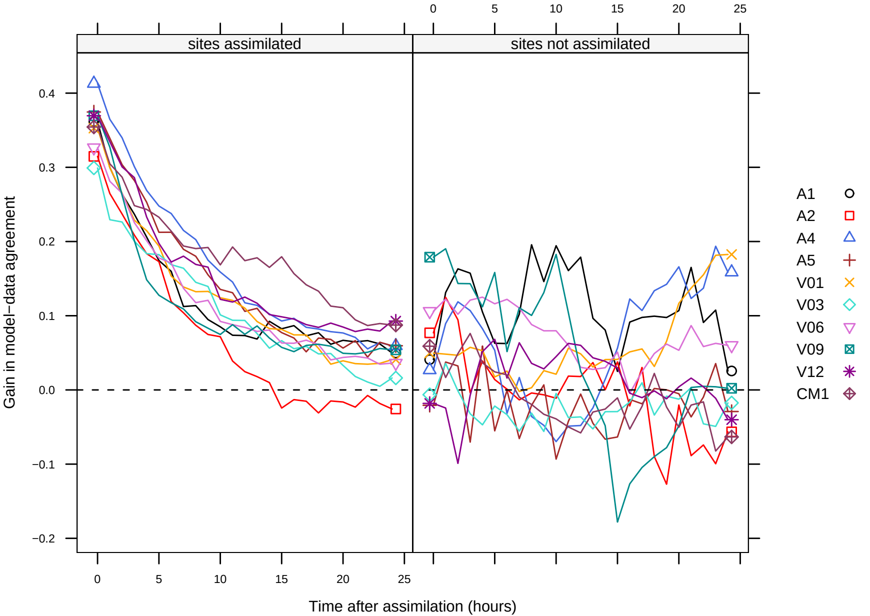


Temporal propagation of the assimilation effect

1. Assimilate 10 hours in a row, leaving out one site and using the other 9;
2. Stop assimilation, let the model continue for 24 hours;
3. Check $\rho(\mathbf{u}, \hat{\mathbf{u}})$ at data sites (assimilated or left out) in these 24 hours, compare with a reference model run.

$\Rightarrow$ shows how much improvement the assimilation makes at assimilated or validating data sites, and how this improvement evolves after the assimilation.

Temporal propagation of the assimilation effect

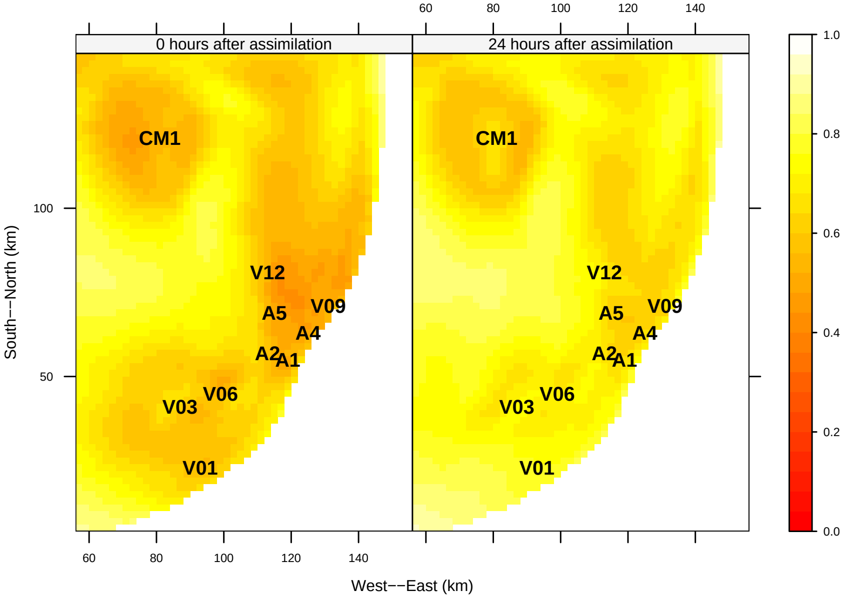


Spatial propagation of the assimilation effect

1. Assimilate 10 hours in a row, leaving out one site and using the other 9;
2. Stop assimilation, let the model continue for 24 hours;
3. Check $\rho(\mathbf{u}, \mathbf{u}^{(r)})$ at every grid point in these 24 hours, where $\mathbf{u}^{(r)}$ is from a reference model run.

$\Rightarrow$ shows how much impact the assimilation makes throughout the lake, and how this impact evolves after the assimilation.

Spatial propagation of the assimilation effect



Summary

- Assimilate observed 2-D currents into model.
- Update is required not to change model-predicted surface level:
  - guarantees mass conservation
  - leads to representation via stream function
- Update is estimated by kriging interpolation.
- Coastal constraint is imposed by incorporating *pseudo data* into the interpolation:
  - transforms a "relation"—velocity is parallel to the coast—to "data"—$\psi(\mathbf{c}) = 0$
  - eliminates the need to construct a complex covariance function K
- Results show temporal and spatial propagation of the assimilation effect.