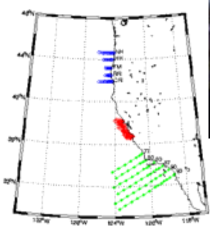
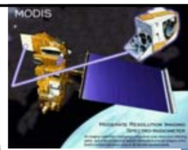
Lecture 5:
Observation Impact & Observation Sensitivity

Outline

- Observation impacts
- Adjoint 4D-Var: $(4D\text{-}Var)^T$
- Observation sensitivity
- Error covariance estimates from $(4D\text{-}Var)^T$

Observation Impacts

(Useful references: Langland & Baker, 2004; Gelaro and Zhu, 2009; Tremolet, 2008)


Given the plethora of different observation platforms, what impact does each have on the 4D-Var analysis?

Observation Impacts

Consider a scalar function of the ocean state vector:

$$I = I(\mathbf{x})$$

Prior	*Posterior*	*Increment*
$I_{\mathrm{b}} = I(\mathbf{x}_{\mathrm{b}})$	$I_{\mathrm{a}} = I(\mathbf{x}_{\mathrm{a}})$	$\Delta I = I_{\mathrm{a}} - I_{\mathrm{b}}$

$$\mathbf{x}_{\mathrm{a}}(\mathrm{t}) = \mathbf{x}_{\mathrm{b}}(\mathrm{t}) + \delta\mathbf{x}(\mathrm{t})$$

$$I_{\mathrm{a}} = I(\mathbf{x}_{\mathrm{b}} + \delta\mathbf{x}) \simeq I_{\mathrm{b}} + \delta\mathbf{x}^{\mathrm{T}} * (\partial I/\partial\mathbf{x})$$

$$\Delta I \simeq \delta\mathbf{x}^{\mathrm{T}} * (\partial I/\partial\mathbf{x}) \quad \textbf{but} \quad \delta\mathbf{x}(t) = \mathbf{M}_{\mathrm{b}}(t,t_0)\tilde{\mathbf{K}}\mathbf{d}$$

$$\boxed{\Delta I \simeq \mathbf{d}^{\mathrm{T}}\tilde{\mathbf{K}}^{\mathrm{T}}\mathbf{M}_{\mathrm{b}}^{\mathrm{T}} * (\partial I/\partial\mathbf{x})}$$

($\mathbf{M}_{\mathrm{b}}^{\mathrm{T}} * (\partial I/\partial\mathbf{x})$ denotes a time convolution)

Observation Impacts

Consider a scalar function of the ocean state vector:

$$I = I(\mathbf{x})$$

Prior	*Posterior*	*Increment*
$I_{\mathrm{b}} = I(\mathbf{x}_{\mathrm{b}})$	$I_{\mathrm{a}} = I(\mathbf{x}_{\mathrm{a}})$	$\Delta I = I_{\mathrm{a}} - I_{\mathrm{b}}$

$$\boxed{\Delta I \simeq \mathbf{d}^{\mathrm{T}}\tilde{\mathbf{K}}^{\mathrm{T}}\mathbf{M}^{\mathrm{T}} * (\partial I/\partial\mathbf{x})}$$

Innovations ($\mathbf{d}^{\mathrm{T}}$)

Adjoint of gain matrix ($\tilde{\mathbf{K}}^{\mathrm{T}}$)

Adjoint model ($\mathbf{M}^{\mathrm{T}}$)

Observation Impacts

Recall the dual form of the gain matrix:

$$\mathbf{K} \approx \tilde{\mathbf{K}}_k = \mathbf{D}\mathbf{G}^{\mathrm{T}}\mathbf{R}^{-1/2}\mathbf{V}_k\mathbf{T}_k^{-1}\mathbf{V}_k^{\mathrm{T}}\mathbf{R}^{-1/2}$$

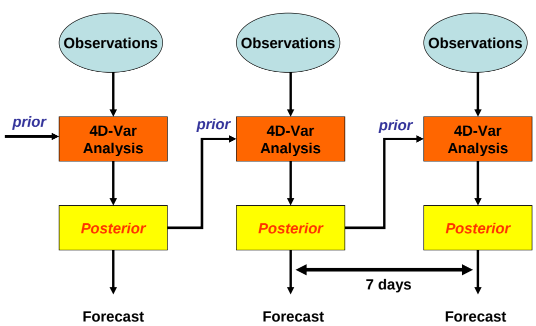
So:

$$\tilde{\mathbf{K}}_k^{\mathrm{T}} = \mathbf{R}^{-1/2}\mathbf{V}_k\mathbf{T}_k^{-1}\mathbf{V}_k^{\mathrm{T}}\mathbf{R}^{-1/2}\mathbf{G}\mathbf{D}$$

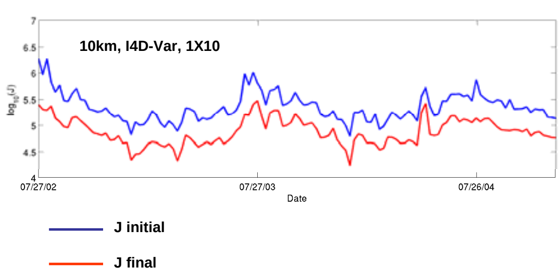
Therefore:

$$\Delta I \simeq \mathbf{d}^{\mathrm{T}}\mathbf{R}^{-1/2}\mathbf{V}_k\mathbf{T}_k^{-1}\mathbf{V}_k^{\mathrm{T}}\mathbf{R}^{-1/2}\mathbf{G}\mathbf{D}\mathbf{M}^{\mathrm{T}} * (\partial I/\partial \mathbf{x})$$

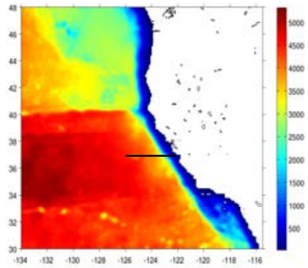
Sequential 4D-Var with 10km CCS ROMS



Sequential 4D-Var CCS ROMS



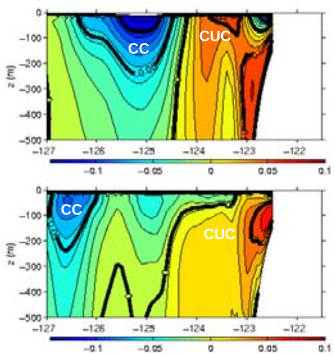
Example: 37N Transport



10km, CCS ROMS

Example: 37N Transport

No assim



JAS time mean alongshore Flow (10km, 42 lev)

Primal Strong

CC = California Current
CUC = California Under Current

37N Transport Observation Impacts

The time average 37N transport can be written as:

$$I_{37N} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{h}^{\mathrm{T}}\mathbf{x}_i$$

where: $\mathbf{x}_i \equiv \mathbf{x}(i\Delta t) = \mathbf{x}(t)$

Model timestep

therefore:

$$\Delta I_{37N} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{h}^{\mathrm{T}}\left((\mathbf{x_a})_i - (\mathbf{x_b})_i\right) \qquad \mathbf{M}_\mathbf{b}^{\mathrm{T}} * (\partial I/\partial \mathbf{x})$$

$$\simeq \frac{1}{N}\sum_{i=1}^{N} \mathbf{h}^{\mathrm{T}}(\mathbf{M_b})_i\tilde{\mathbf{K}}\mathbf{d} = \mathbf{d}^{\mathrm{T}}\tilde{\mathbf{K}}^{\mathrm{T}}\boxed{\sum_{i=1}^{N}\frac{1}{N}(\mathbf{M_b})_i^{\mathrm{T}}\mathbf{h}}$$

where: $(\mathbf{M_b})_i \equiv \mathbf{M}(t_0 + i\Delta t, t_0) = \mathbf{M}(t, t_0)$

37N Transport Observation Impacts

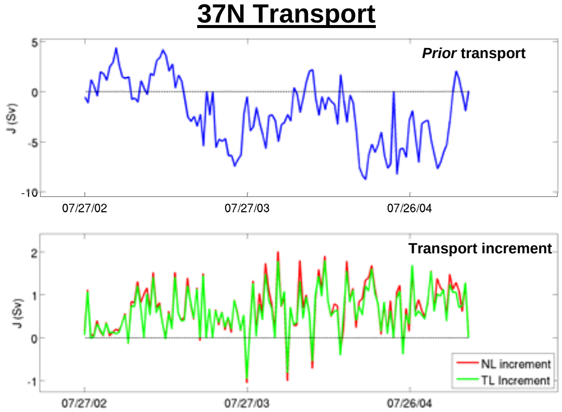
37N time averaged transport increment:

$$\Delta I_{37N} \simeq \frac{1}{N}\mathbf{d}^{\mathrm{T}}\tilde{\mathbf{K}}^{\mathrm{T}}\underbrace{\sum_{i=1}^{N}(\mathbf{M}_{\mathrm{b}})_i^{\mathrm{T}}\mathbf{h}}_{\text{ADROMS forced by h}}$$

$$\tilde{\mathbf{K}}_k^{\mathrm{T}} = \mathbf{R}^{-1/2}\underbrace{\mathbf{V}_k\mathbf{T}_k^{-1}\mathbf{V}_k^{\mathrm{T}}}_{\text{Dual space Lanczos vectors}}\mathbf{R}^{-1/2}\mathbf{G}\mathbf{D}$$

↑ **G**: TLROMS sampled at observation points

37N Transport



Control Vector Impacts

37N time averaged transport increment:

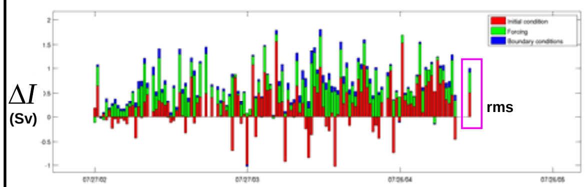
$$\Delta I_{37N} \simeq \frac{1}{N}\mathbf{d}^{\mathrm{T}}\tilde{\mathbf{K}}^{\mathrm{T}}\sum_{i=1}^{N}(\mathbf{M}_{\mathrm{b}})_i^{\mathrm{T}}\mathbf{h}$$

$$= \mathbf{d}^{\mathrm{T}}\mathbf{g} = \mathbf{d}^{\mathrm{T}}\left(\mathbf{g}_{\mathrm{x}} + \mathbf{g}_{\mathrm{f}} + \mathbf{g}_{\mathrm{b}}\right)$$

where: $\mathbf{g} \simeq \frac{1}{N}\tilde{\mathbf{K}}\sum_{i=1}^{N}(\mathbf{M}_{\mathrm{b}})_i^{\mathrm{T}}\mathbf{h}$

$\mathbf{g}_{\mathrm{x}}$ **- contribution from initial condition increments**
$\mathbf{g}_{\mathrm{f}}$ **- contribution from surface forcing increments**
$\mathbf{g}_{\mathrm{b}}$ **- contribution from open boundary increments**

37N Transport Control Vector Impacts



Observation Impacts

37N time averaged transport increment:

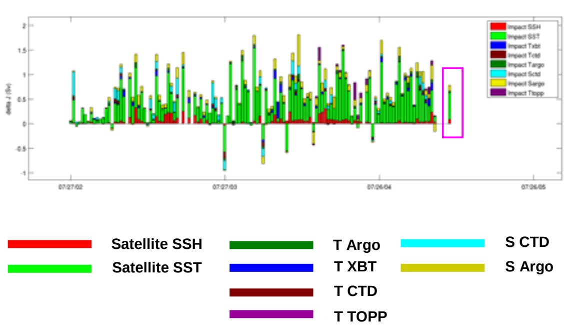
$$\Delta I_{37N} \simeq \frac{1}{N}\mathbf{d}^{\mathrm{T}}\tilde{\mathbf{K}}^{\mathrm{T}}\sum_{i=1}^{N}(\mathbf{M}_{\mathbf{b}})_i^{\mathrm{T}}\mathbf{h}$$

$$= \mathbf{d}^{\mathrm{T}}\mathbf{g} = \sum_{i=1}^{N_{obs}} d_i g_i$$

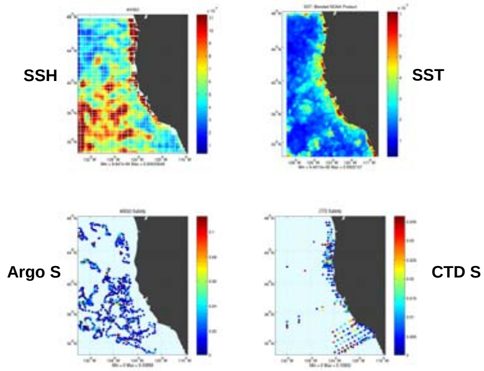
$$= \sum_{i=1}^{N_{obs}} \underbrace{\left(y_i - H_i\left(\mathbf{x}_{\mathbf{b}}(t)\right)\right) g_i}_{}$$

Contribution of each observation to ΔI

37N Transport Observation Impacts



37N Transport Observation Impacts



Two Spaces: Obs Impact

$\mathbf{K}$ maps from observation (dual) space
to model (primal) space

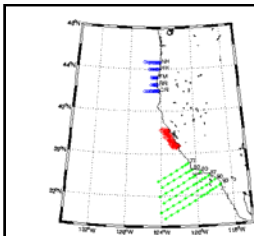
$\mathbf{K}^{T}$ maps from model (primal) space
to observation (dual) space

Identifies the part of model space that controls 37N transport ***<u>and</u> that is activated by the observations***

Observation Impacts: ROMS Implementation

- **Primal (I4D-Var) and dual (4D-PSAS & R4D-Var) forms available:**
  - **define IS4DVAR_SENSITIVITY**
    Drivers/obs_sen_is4dvar.h
  - **define W4DPSAS_SENSITIVITY**
    define OBS_IMPACT
    define OBS_IMPACT_SPLIT
    Drivers/obs_sen_w4dpsas.h
  - **define W4DVAR_SENSITIVITY**
    define OBS_IMPACT
    define OBS_IMPACT_SPLIT
    Drivers/obs_sen_w4dvar.h

Adjoint 4D-Var & Observation Sensitivity



How will the circulation analysis change if some of the observations or the observation array change?

Adjoint 4D-Var & Observation Sensitivity

The analysis increments are a nonlinear function of the innovation vector d:

4D-Var

$$\mathbf{z}_a = \mathbf{z}_b + \boxed{\mathcal{K}(\mathbf{d})}$$

where:

$$\mathbf{d} = \mathbf{y} - H\left(\mathbf{z}_b(t)\right)$$

Consider variations in the observation vector δy:

$$\delta\mathbf{d} = \delta\mathbf{y}; \quad \mathbf{z}_a + \delta\mathbf{z}_a = \mathbf{z}_b + \mathcal{K}(\mathbf{d} + \delta\mathbf{d})$$

$$\approx \mathbf{z}_b + \mathcal{K}(\mathbf{d}) + \left(\partial\mathcal{K}/\partial\mathbf{y}\right)\delta\mathbf{y}$$

$$\delta\mathbf{z}_a \approx \boxed{\frac{\partial\mathcal{K}}{\partial\mathbf{y}}}\delta\mathbf{y}$$

Tangent linearization of 4D-Var

Adjoint 4D-Var & Observation Sensitivity

Consider a scalar function of the *posterior* control vector z_a:

$$I_{\mathbf{a}} = I(\mathbf{z}_{\mathbf{a}}) = I(\mathbf{z}_{\mathbf{b}} + \mathcal{K}(\mathbf{d}))$$

A change $\delta\mathrm{y}$ in the observations yields a change in ΔI_{a} :

$$I_{\mathbf{a}} + \Delta I_{\mathbf{a}} = I(\mathbf{z}_{\mathbf{b}} + \mathcal{K}(\mathbf{d} + \delta\mathbf{y}))$$

$$\approx I\left(\mathbf{z}_{\mathbf{b}} + \mathcal{K}(\mathbf{d}) + (\partial\mathcal{K}/\partial\mathbf{y})\,\delta\mathbf{y}\right)$$

$$\approx I(\mathbf{z}_{\mathbf{a}}) + \left((\partial\mathcal{K}/\partial\mathbf{y})\,\delta\mathbf{y}\right)^{\mathrm{T}}\left(\partial I/\partial\mathbf{z}\right)$$

Therefore:

$$\Delta I_{\mathbf{a}} \approx \delta\mathbf{y}^{\mathrm{T}}\left(\partial\mathcal{K}/\partial\mathbf{y}\right)^{\mathrm{T}}\left(\partial I/\partial\mathbf{z}\right)$$

Adjoint 4D-Var & Observation Sensitivity

$$\Delta I_{\mathbf{a}} \approx \delta\mathbf{y}^{\mathrm{T}}\boxed{\left(\partial\mathcal{K}/\partial\mathbf{y}\right)^{\mathrm{T}}}\left(\partial I/\partial\mathbf{z}\right)$$

Adjoint of 4D-Var

Observation System Experiments (OSEs)

Suppose that during a particular assimilation cycle the satellite altimeter goes offline.

How would this have impacted the analysis?

We could run 4D-Var again with SSH obs removed.

Or let $\delta y_i = -d_i$ **for all SSH obs.**

The change in the analysis is: $\delta\mathbf{z}_{\mathbf{a}} \approx \left(\partial\mathcal{K}/\partial\mathbf{y}\right)\delta\mathbf{y}$

The change in ΔI_{a} **is:** $\Delta I_{\mathbf{a}} \approx \delta\mathbf{y}^{\mathrm{T}}\left(\partial\mathcal{K}/\partial\mathbf{y}\right)^{\mathrm{T}}\left(\partial I/\partial\mathbf{z}\right)$

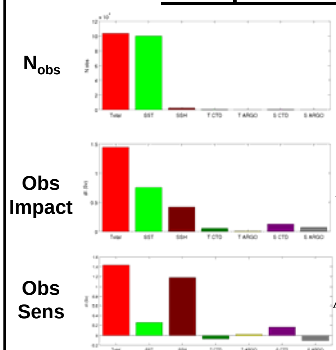
Observation System Experiments (OSEs)

The cost of $(4D\text{-}Var)^T$ = cost of 4D-Var

But ONLY one run of $(4D\text{-}Var)^T$ is needed for ALL OSEs.

Example: 37N transport

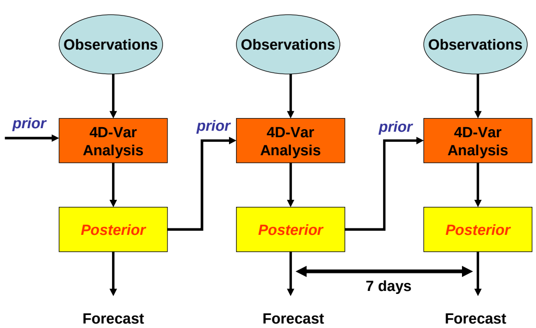
(10km, CCS ROMS)



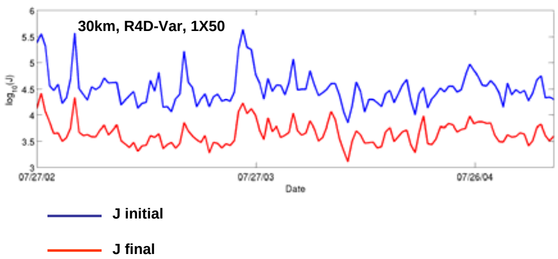
$$\Delta I = \mathbf{d}^T \tilde{\mathbf{K}}^T \sum_{i=1}^{N} \frac{1}{N} (\mathbf{M}_b)_i^T \mathbf{h}_i$$

$$\Delta I = \mathbf{d}^T \left(\frac{\partial K}{\partial \mathbf{y}}\right)^T \sum_{i=1}^{N} \frac{1}{N} (\mathbf{M}_b)_i^T \mathbf{h}_i$$

Sequential 4D-Var with 30km CCS ROMS

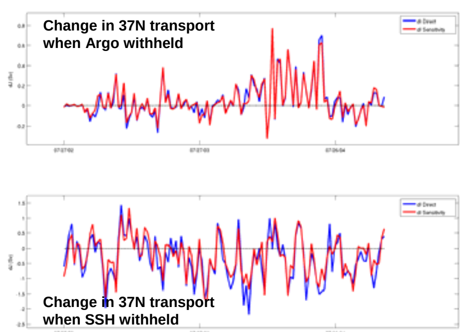


Sequential 4D-Var CCS ROMS



Observing System Experiments (OSEs)

(30km, CCS ROMS)



Two Spaces: Obs Sensitivity

$\partial K/\partial \mathbf{y}$ maps from observation (dual) space to model (primal) space

$(\partial K/\partial \mathbf{y})^{\mathrm{T}}$ maps from model (primal) space to observation (dual) space

Identifies the part of model space that controls 37N transport <u>and</u> that is activated by the observations during 4D-Var

Observation Sensitivity: ROMS Implementation

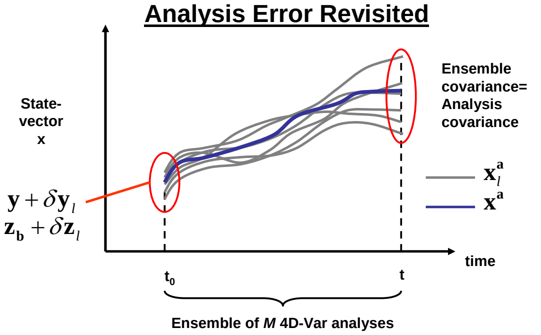
• Dual (4D-PSAS & R4D-Var) forms only available:

- define W4DPSAS_SENSITIVITY
  (define RECOMPUTE_4DVAR)
  Drivers/obs_sen_w4dpsas.h

- define W4DVAR_SENSITIVITY
  (define RECOMPUTE_4DVAR)
  Drivers/obs_sen_w4dvar.h

Error Covariance Estimates from (4D-Var)^T

Analysis Error Revisited



Belo Pereira & Berre (2006): $\mathbf{E}^a=\frac{1}{M}\sum_{l=1}^{M}\left(\mathbf{x}_l^a-\mathbf{x}^a\right)\left(\mathbf{x}_l^a-\mathbf{x}^a\right)^T$
(also Daget et al, 2010)

Analysis Error Revisited

An ensemble of 4D-Var analyses is very expensive!

But <u>one</u> run of (4D-Var)^T yields $(\partial K/\partial \mathbf{d})^{\mathrm{T}}$ **and:**

$$\delta \mathbf{z}_l^{\mathbf{a}} \approx \delta \mathbf{z}_l + (\partial K/\partial \mathbf{d})\,\delta \mathbf{d}_l; \qquad \delta \mathbf{d}_l \approx \delta \mathbf{y}_l + \mathbf{G}\,\delta \mathbf{z}_l$$

and: $\delta \mathbf{x}_l^{\mathbf{a}}(t) \approx \mathcal{M}(t,t_0)\,\delta \mathbf{z}_l^{\mathbf{a}}$

Therefore:

$$\mathbf{E}_{\mathbf{x}}^{\mathbf{a}}(t) = \left\langle \delta \mathbf{x}^{\mathbf{a}}(t) \left(\delta \mathbf{x}^{\mathbf{a}}(t)\right)^{\mathrm{T}} \right\rangle$$

$$= \mathcal{M}\left\{ \left(\mathbf{I} - \left(\frac{\partial K}{\partial \mathbf{d}}\right)\mathbf{G}\right)\mathbf{D}\left(\mathbf{I} - \left(\frac{\partial K}{\partial \mathbf{d}}\right)\mathbf{G}\right)^{\mathrm{T}} + \left(\frac{\partial K}{\partial \mathbf{d}}\right)\mathbf{R}\left(\frac{\partial K}{\partial \mathbf{d}}\right)^{\mathrm{T}} \right\}\mathcal{M}^{\mathrm{T}}$$

Here, *M* is essentially infinite!

Analysis Error Revisited

For linear functions: $I(\mathbf{x}) = \sum_{k=1}^{N} \mathbf{h}_k^{\mathrm{T}} \mathbf{x}_k$

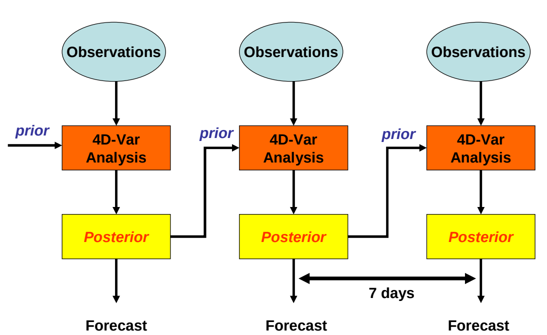
***Posterior*/analysis error variance:**

$$\left(\sigma_I^a\right)^2 = \left(\sum_{k=1}^{N} \mathbf{h}_k^{\mathrm{T}} \mathcal{M}_k\right) \mathbf{E}_{\mathbf{x}}^{\mathbf{a}}(t_0) \left(\sum_{j=1}^{N} \mathcal{M}_j^{\mathrm{T}} \mathbf{h}_j\right)$$

$$= \mathbf{g}^{\mathrm{T}} \mathbf{E}_{\mathbf{x}}^{\mathbf{a}}(t_0)\,\mathbf{g}$$

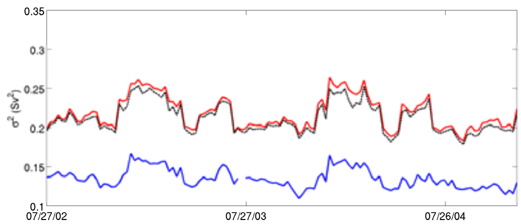
where: $\mathbf{g} = \sum_{j=1}^{N} \mathcal{M}_j^{\mathrm{T}} \mathbf{h}_j$ **(ADROMS forced by h)**

Sequential 4D-Var with 30km CCS ROMS



Example: 37N Transport

30km, ROMS CCS, 7day average transport errors

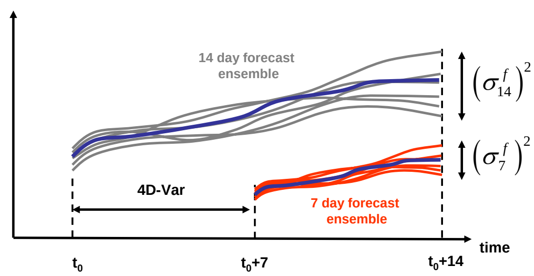


Prior error

Posterior error using (4D-Var)^T

Posterior error using $\tilde{\mathbf{E}}^a$

Predictability



Predictability due to assimilating observations during [t_0,t_0+7]:

$$\left(\sigma_{14}^f\right)^2-\left(\sigma_7^f\right)^2$$

Example: 37N Transport Predictability

$\left(\sigma_{14}^f\right)$ = spread of 14 day forecast ensemble of transport

$\left(\sigma_7^f\right)$ = spread of 7 day forecast ensemble of transport

$$\left(\sigma_{14}^f\right)^2-\left(\sigma_7^f\right)^2=2\mathbf{g}^{\mathrm{T}}\mathbf{GD}\mathcal{M}_{\mathbf{b}}^{\mathrm{T}}\sum_k\left(\mathcal{M}_{14}\right)_k^{\mathrm{T}}\mathbf{h}_k-\mathbf{g}^{\mathrm{T}}\left(\mathbf{GDG}^{\mathrm{T}}+\mathbf{R}\right)\mathbf{g}$$

where: $\mathbf{g}=\left(\partial K/\partial\mathbf{d}\right)^{\mathrm{T}}\mathcal{M}_{\mathbf{b}}^{\mathrm{T}}\sum_k\left(\mathcal{M}_{14}\right)_k^{\mathrm{T}}\mathbf{h}_k$

Seemingly complicated expressions, but really just
TL and AD operators strung together in the right order!

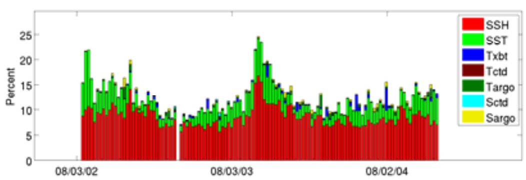
Example: 37N Transport Predictability

$$\left(\sigma_{14}^{f}\right)^2-\left(\sigma_{7}^{f}\right)^2=2\mathbf{g}^{\mathrm{T}}\mathbf{GD}\mathcal{M}_{\mathbf{b}}^{\mathrm{T}}\sum_{k}\left(\mathcal{M}_{14}\right)_k^{\mathrm{T}}\mathbf{h}_k-\mathbf{g}^{\mathrm{T}}\left(\mathbf{GDG}^{\mathrm{T}}+\mathbf{R}\right)\mathbf{g}$$

[red box] change in predictability due to the covariance between errors in the *priors* z_b and errors in the time evolving *prior* circulation $x_b(t)$ evaluated at the observation points.

[blue box] change in predictability associated with the stabilized representer matrix - a combination of the covariance between errors in the time evolving *prior* circulation at the observation points, and the covariance between the observation errors (including errors of representativeness).

Example: 37N Transport Predictability



$$r=100\left\{\left(\sigma_{14}^{f}\right)^2-\left(\sigma_{7}^{f}\right)^2\right\}\Big/\left(\sigma_{14}^{f}\right)^2$$

$r>0$ implies 4D-Var increases predictability

Issues, Things to do, & Coming Soon

- Observation sensitivity <u>only</u> available for dual 4D-Var.
- Observation impact and observation sensitivity calculations are currently restricted to a single outer-loop – multiple outer-loops coming soon.
- Increase the modularity of ROMS drivers so that arbitrary sequences of operators (linear and non-linear) can be formed.

Summary

- Observation impact is based on $\tilde{\mathbf{K}}^{\mathrm{T}}$ and yields the actual contribution of each obs to the circulation increments.
- Observation sensitivity is based on $(4D\text{-}Var)^{T}$ and yields the change in circulation due to changes in obs (or array)
  - useful for efficient generation of OSEs.
- Both obs impact and obs sensitivity were applied in examples during analysis cycle, but can be applied during forecast cycle also (Moore et al, 2010c).
- $(4D\text{-}Var)^{T}$ yields more reliable estimates of $\mathbf{E}^{a}$ and $\mathbf{E}^{f}$ and predictability.

References

- Belo Pereira, M. and L. Berre, 2006: The use of an ensemble approach to study the background error covariances in a global NWP model. *Mon. Wea. Rev*., **134**, 2466-2498.
- Daget, N., A.T. Weaver and M.A. Balmaseda, 2009: Ensemble estimation of background error variances in a three-dimensional variational data assimilation system for the global ocean. *Q. J. R. Meteorol. Soc*., **135**, 1071-1094.
- Langland, R.H. and N.L. Baker, 2004: Estimation of observation impact using the NRL atmospheric variational data assimilation adjoint system. *Tellus*, **56A**, 189-201.
- Gelaro, R. and Y. Zhu, 2009: Examination of observation impacts derives from Observing System Experiments (OSEs) and adjoint models. *Tellus*, **61A**, 179-193.
- Moore, A.M., H.G. Arango, G. Broquet, B.S. Powell, J. Zavala-Garay and A.T. Weaver, 2010a: The Regional Ocean Modeling System (ROMS) 4-dimensional data assimilation systems: Part I – System overview. *Ocean Modelling*, Submitted.

References

- Moore, A.M., H.G. Arango, G. Broquet, C.. Edwards, M. Veneziani, B.S. Powell, D. Foley, J. Doyle, D. Costa and P. Robinson, 2010b: The Regional Ocean Modeling System (ROMS) 4-dimensional data assimilation systems: Part II – Performance and application to the California Current System. *Ocean Modelling*, Submitted.
- Moore, A.M., H.G. Arango, G. Broquet, C.. Edwards, M. Veneziani, B.S. Powell, D. Foley, J. Doyle, D. Costa and P. Robinson, 2010c: The Regional Ocean Modeling System (ROMS) 4-dimensional data assimilation systems: Part III – Observation impact and observation sensitivity the California Current System. *Ocean Modelling*, Submitted.
- Trémolet, Y., 2008: Computation of observation sensitivity and observation impact in incremental variational data assimilation. *Tellus*, **60A**, 964-978.