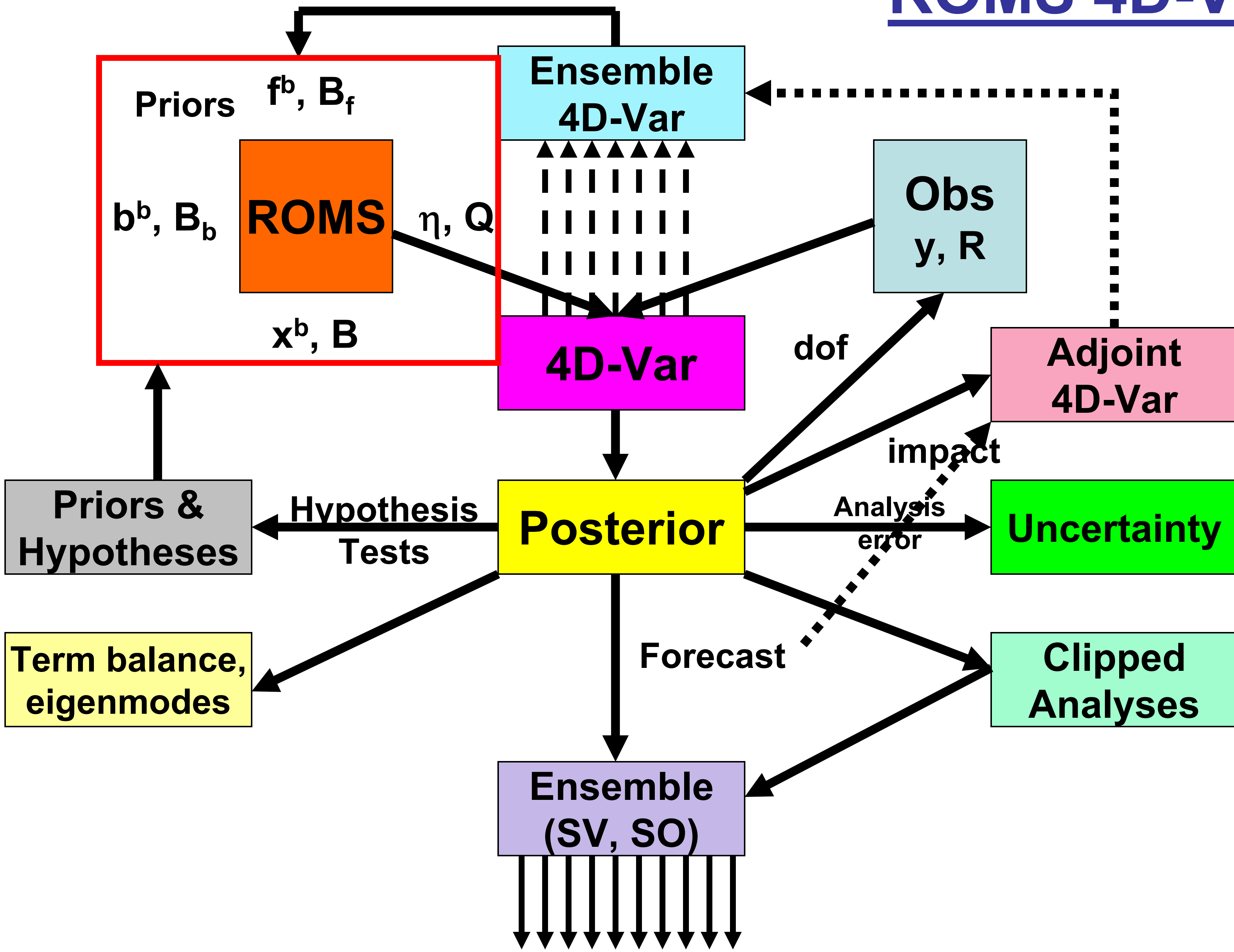
# Lecture 1:
# Primal 4D-Var

# Outline

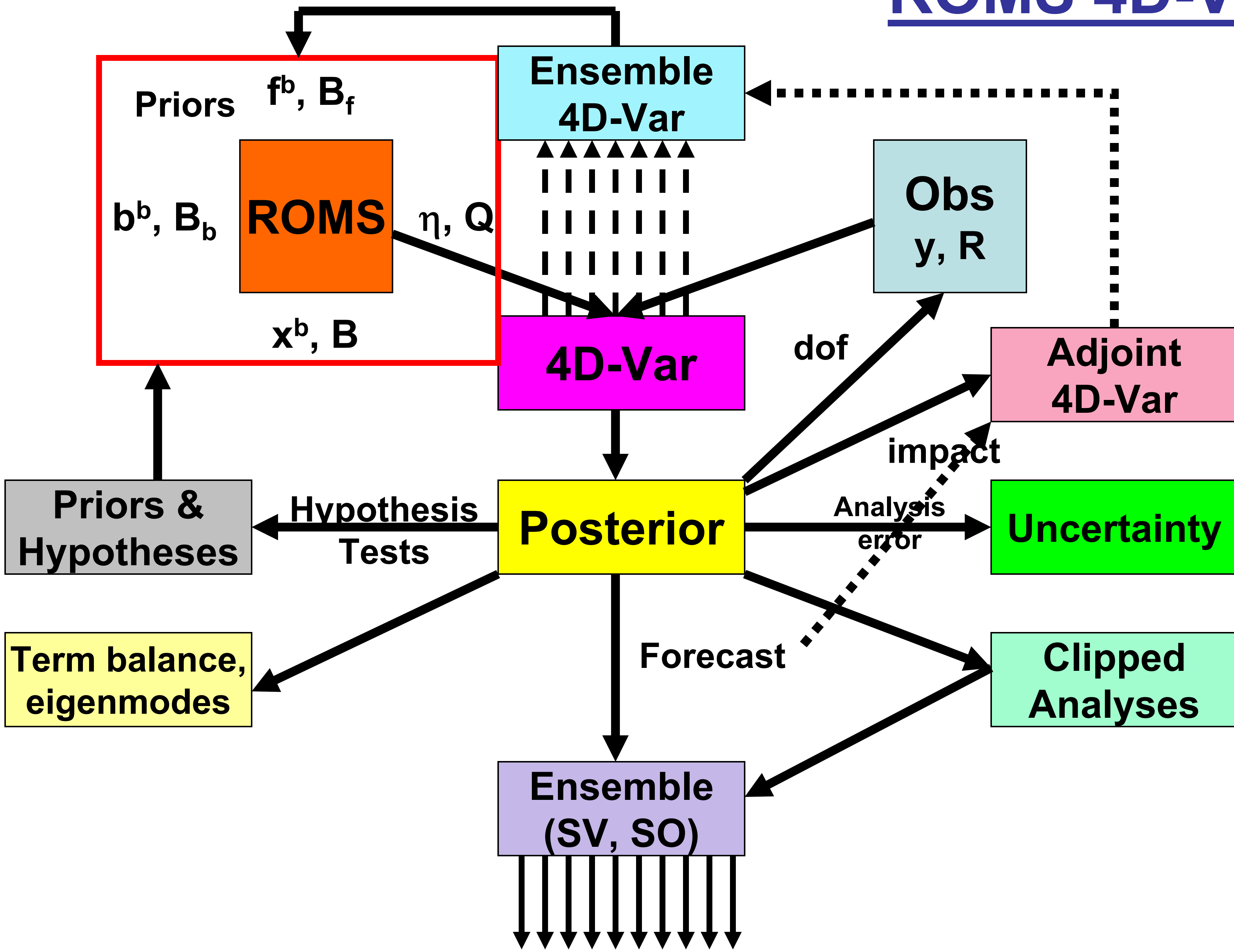
- ROMS 4D-Var overview
- 4D-Var concepts
- Primal formulation of 4D-Var
- Incremental approach used in ROMS
- The ROMS I4D-Var algorithm

# ROMS 4D-Var



Priors
$f^b$, $B_f$
Ensemble 4D-Var
$b^b$, $B_b$
ROMS
η, Q
Obs
y, R
$x^b$, B
4D-Var
dof
Adjoint 4D-Var
impact
Priors & Hypotheses
Hypothesis Tests
Posterior
Analysis error
Uncertainty
Term balance, eigenmodes
Forecast
Clipped Analyses
Ensemble (SV, SO)

# ROMS 4D-Var



Priors
$f^b$, $B_f$
Ensemble 4D-Var
$b^b$, $B_b$
ROMS
$\eta$, Q
Obs
y, R
$x^b$, B
4D-Var
dof
Adjoint 4D-Var
impact
Priors & Hypotheses
Hypothesis Tests
Posterior
Analysis error
Uncertainty
Term balance, eigenmodes
Forecast
Clipped Analyses
Ensemble (SV, SO)

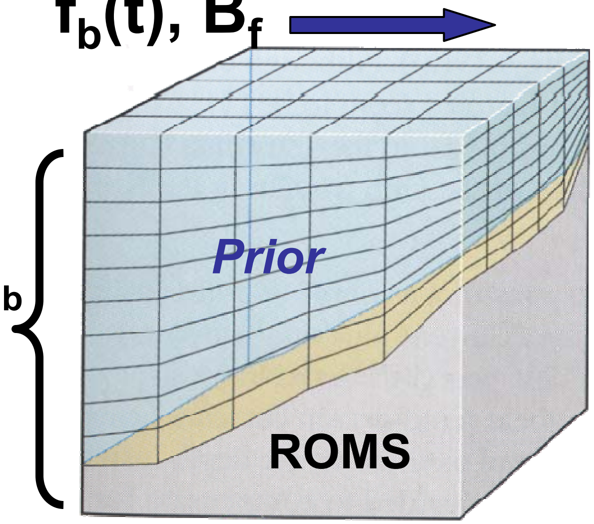
# ROMS 4D-Var Applications

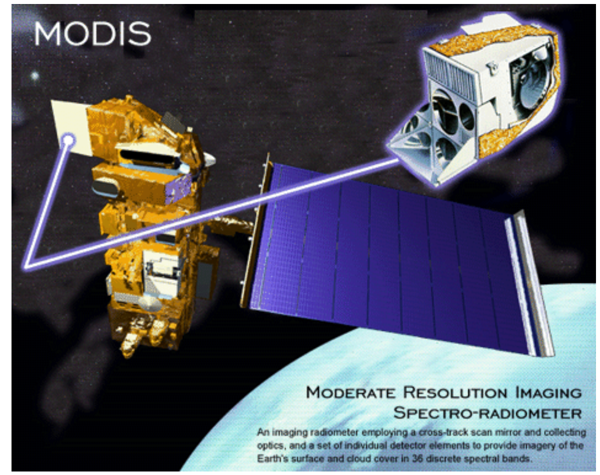

- Broquet, G., C.A. Edwards, A.M. Moore, B.S. Powell, M. Veneziani and J.D. Doyle, 2009: Application of 4D-variational data assimilation to the California Current System. *Dyn. Atmos. Oceans*, **48**, 69-91.
- Broquet, G., A.M. Moore, H.G. Arango, C.A. Edwards and B.S. Powell, 2009: Ocean state and surface forcing correction using the ROMS-IS4DVAR data assimilation system. *Mercator Ocean Quarterly Newsletter*, #**34**. 5-13.
- Broquet, G., A.M. Moore, H.G. Arango, and C.A. Edwards, 2010: Corrections to ocean surface forcing in the California Current System using 4D-variational data assimilation. *Ocean Modelling*, Under review.
- Di Lorenzo, E., A.M. Moore, H.G. Arango, B.D. Cornuelle, A.J. Miller, B. Powell, B.S. Chua and A.F. Bennett, 2007: Weak and strong constraint data assimilation in the inverse Regional Ocean Modeling System (ROMS): development and application for a baroclinic coastal upwelling system. *Ocean Modeling*, **16**, 160-187.
- Powell, B.S., H.G. Arango, A.M. Moore, E. DiLorenzo, R.F. Milliff and D. Foley, 2008: 4DVAR Data Assimilation in the Intra-Americas Sea with the Regional Ocean Modeling System (ROMS). *Ocean Modelling*, **23**, 130-145.
- Powell, B.S., A.M.Moore, H.G. Arango, E. Di Lorenzo, R.F. Milliff and R.R. Leben, 2009: Near real-time assimilation and prediction in the Intra-Americas Sea with the Regional Ocean Modeling System (ROMS). *Dyn. Atmos. Oceans*, **48**, 46-68.
- Zhang, W.G., J.L. Wilkin, and J.C. Levin, 2010: Towards an integrated observation and modeling system in the New York Bight using variational methods, Part I. *Ocean Modelling,* Under review.
- Zhang, W.G., J.L. Wilkin, and J.C. Levin, 2010: Towards an integrated observation and modeling system in the New York Bight using variational methods, Part II. *Ocean Modelling,* Under review

# Data Assimilation

$f_b(t)$, $B_f$

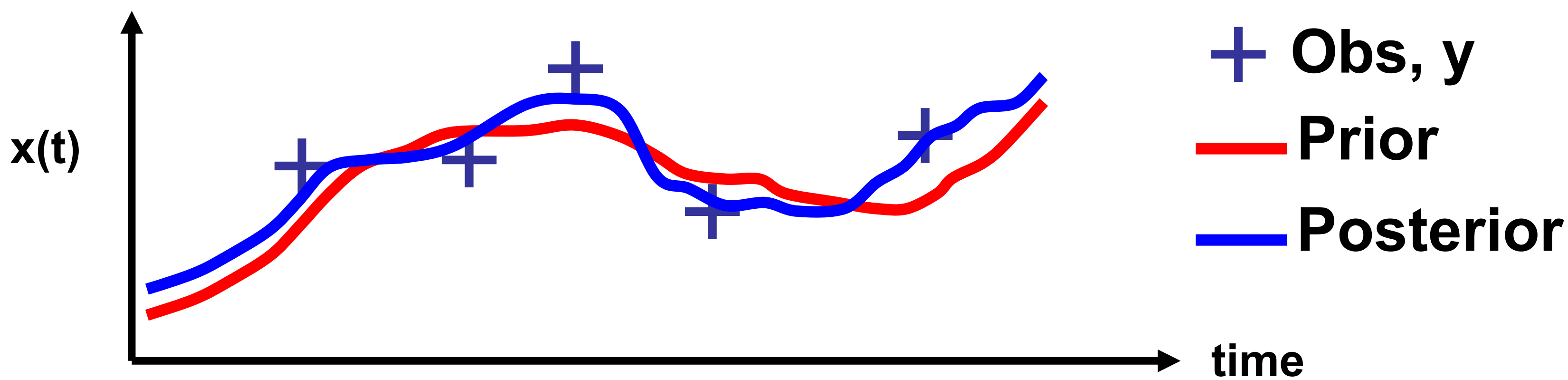
$b_b(t)$, $B_b$




$x_b(0)$, $B_x$

$\mathbf{y}, \mathbf{R}$




Model solutions depends on $x_b(0)$, $f_b(t)$, $b_b(t)$, $\eta(t)$

# Notation & Nomenclature

$$\mathbf{x} = \begin{bmatrix} \mathbf{T} \\ \mathbf{S} \\ \boldsymbol{\zeta} \\ \mathbf{u} \\ \mathbf{v} \end{bmatrix}$$

State vector

$$\mathbf{z} = \begin{bmatrix} \mathbf{x}(0) \\ \mathbf{f}(t) \\ \mathbf{b}(t) \\ \boldsymbol{\eta}(t) \end{bmatrix}$$

Control vector

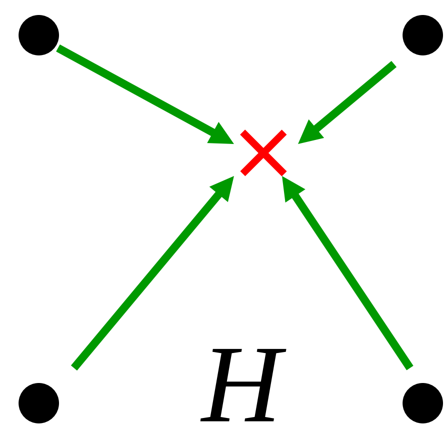
$$\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ \vdots \\ y_N \end{bmatrix}$$

Observation vector

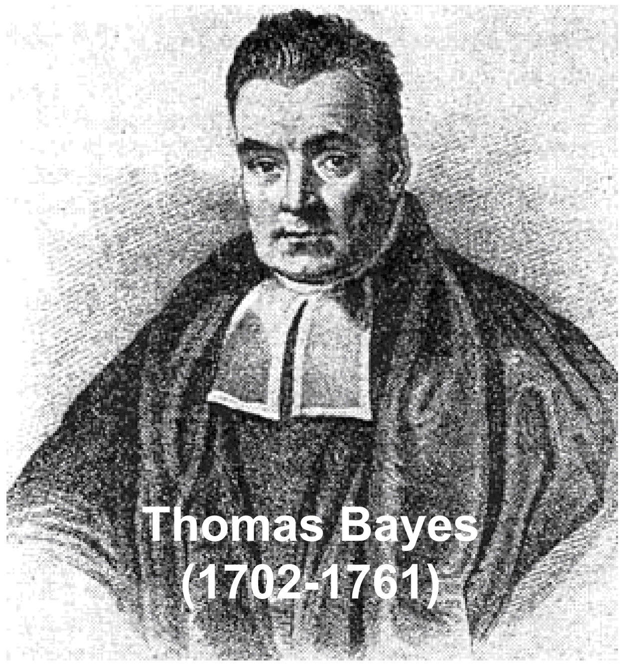
Prior

$$\mathbf{d} = \left(\mathbf{y} - H(\mathbf{x}_\mathrm{b})\right)$$

Innovation vector




Observation operator


Thomas Bayes
(1702-1761)

# Bayes Theorem

**Conditional probability:**

**(Wikle and Berliner, 2007)**

$$p(\mathbf{z}\,|\,\mathbf{y}) = p(\mathbf{y}\,|\,\mathbf{z})\,p(\mathbf{z})/p(\mathbf{y})$$

**Posterior distribution** | **Data distribution** | **Prior** | **Marginal**

**("likelihood")**

$$= c\ \exp\left(-1/2\left(\mathbf{y}-H(\mathbf{x})\right)^{\mathrm{T}}\mathbf{R}^{-1}\left(\mathbf{y}-H(\mathbf{x})\right)\right)$$

$$\times\ \exp\left(-1/2\left(\mathbf{z}-\mathbf{z}_{\mathbf{b}}\right)^{\mathrm{T}}\mathbf{D}^{-1}\left(\mathbf{z}-\mathbf{z}_{\mathbf{b}}\right)\right)$$

**Maximum likelihood estimate: identify the minimum of**

$$J_{NL}(\mathbf{z}) = \frac{1}{2}(\mathbf{z}-\mathbf{z}_{\mathbf{b}})^{\mathrm{T}}\mathbf{D}^{-1}(\mathbf{z}-\mathbf{z}_{\mathbf{b}}) + \frac{1}{2}(\mathbf{y}-H(\mathbf{x}))^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{y}-H(\mathbf{x}))$$

**which maximizes** $p(\mathbf{z}|\mathbf{y})$.

# Variational Data Assimilation

**Conditional Probability:** $P(\mathbf{z}\,|\,\mathbf{y}) \propto \exp\left(-J_{NL}\right)$

$$J_{NL}(\mathbf{z}) = \frac{1}{2}(\mathbf{z}-\mathbf{z_b})^{\mathrm{T}}\mathbf{D}^{-1}(\mathbf{z}-\mathbf{z_b}) + \frac{1}{2}(\mathbf{y}-H(\mathbf{x}))^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{y}-H(\mathbf{x}))$$

**Observation error covariance**

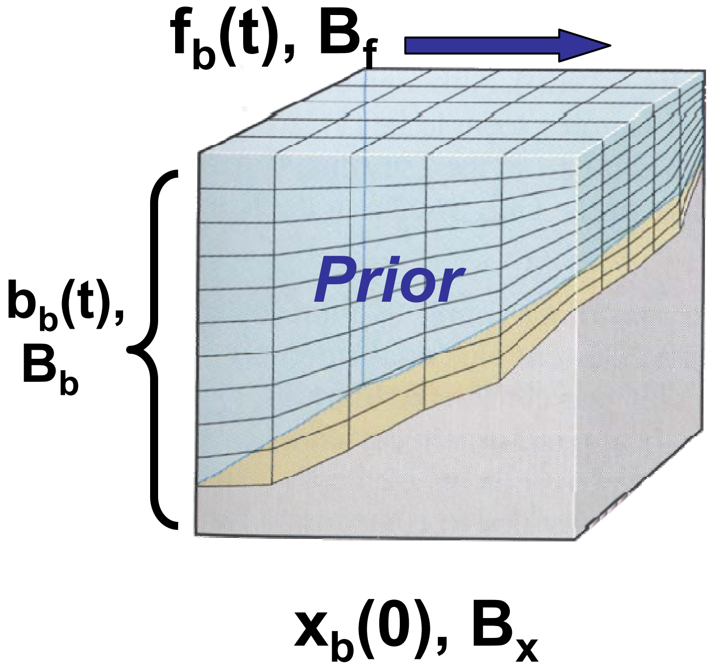
$$\mathbf{D} = \underbrace{\mathrm{diag}(\mathbf{B_x}, \mathbf{B_b}, \mathbf{B_f}, \mathbf{Q})}_{\text{Background error covariance}}$$

**Background error covariance**

$J_{NL}$ is called the "cost" or "penalty" function.

**<u>Problem</u>: Find z=$z_a$ that minimizes *J (i.e. maximizes P)* using principles of variational calculus.**
**$z_a$ is the "maximum likelihood" or "minimum variance" estimate.**

# Incremental Formulation




$$\delta \mathbf{z} = (\delta \mathbf{x}^T(0), \boldsymbol{\varepsilon}_b^T(t), \boldsymbol{\varepsilon}_f^T(t), \boldsymbol{\eta}(t))^T$$

- $\delta \mathbf{x}^T(0)$: initial condition increment
- $\boldsymbol{\varepsilon}_b^T(t)$: boundary condition increment
- $\boldsymbol{\varepsilon}_f^T(t)$: forcing increment
- $\boldsymbol{\eta}(t)$: corrections for model error

(Courtier *et al.*, 1994)

$$J = \frac{1}{2}\delta \mathbf{z}^T \mathbf{D}^{-1} \delta \mathbf{z} + \frac{1}{2}\left(\mathbf{G}\delta \mathbf{z} - \mathbf{d}\right)^T \mathbf{R}^{-1}\left(\mathbf{G}\delta \mathbf{z} - \mathbf{d}\right)$$

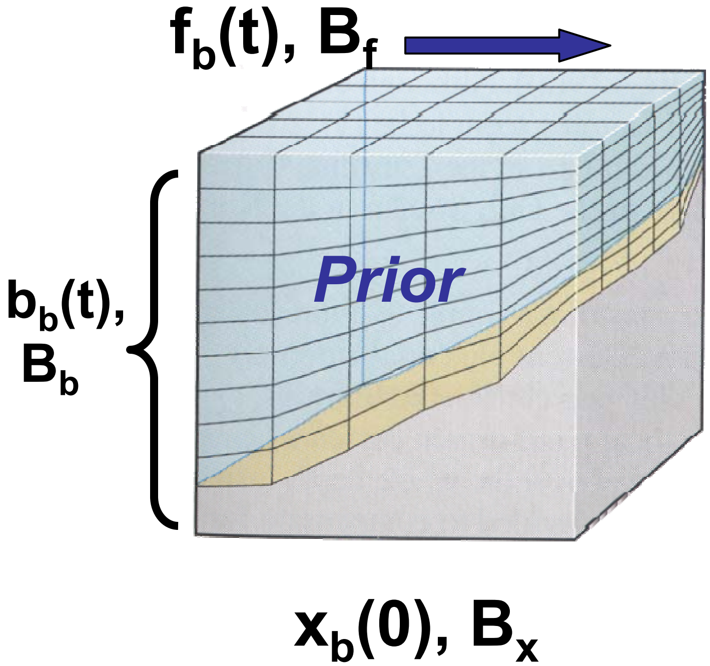
$$\mathbf{D} = \mathrm{diag}(\mathbf{B}_x, \mathbf{B}_b, \mathbf{B}_f, \mathbf{Q})$$

- $\mathbf{D}$: Prior (background) error covariance
- $\mathbf{G}$: Tangent Linear Model sampled at obs points
- $\mathbf{R}$: Obs Error Cov.
- $\mathbf{d}$: Innovation

$$\mathbf{d} = \mathbf{y} - H(\mathbf{x}_b)$$

(H=linearized obs operator)

# Incremental Formulation




$$\delta\mathbf{z} = (\delta\mathbf{x}^T(0), \boldsymbol{\varepsilon}_\mathbf{b}^T(t), \boldsymbol{\varepsilon}_\mathbf{f}^T(t), \boldsymbol{\eta}(t))^T$$

initial condition increment

boundary condition increment

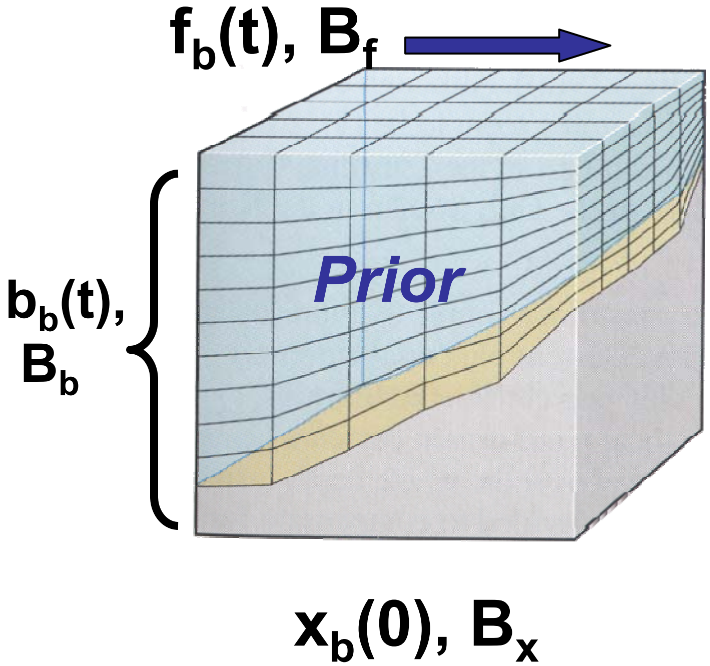
forcing increment

corrections for model error

$$J = \frac{1}{2}\delta\mathbf{z}^T\mathbf{D}^{-1}\delta\mathbf{z} + \frac{1}{2}\left(\mathbf{G}\delta\mathbf{z} - \mathbf{d}\right)^T \mathbf{R}^{-1}\left(\mathbf{G}\delta\mathbf{z} - \mathbf{d}\right)$$

**The minimum of *J* is identified iteratively by searching for $\partial J/\partial\delta z=0$**

# Incremental Formulation




$$\delta\mathbf{z} = (\delta\mathbf{x}^T(0), \boldsymbol{\varepsilon}_{\mathbf{b}}^T(t), \boldsymbol{\varepsilon}_{\mathbf{f}}^T(t), \boldsymbol{\eta}(t))^T$$

initial condition increment

boundary condition increment

forcing increment

corrections for model error

$$\mathbf{z} = \mathbf{z}_{\mathbf{b}} + \delta\mathbf{z}$$

**Assumptions:**

(i) $\delta\mathbf{z} \ll \mathbf{z}_b$

(ii) $\mathbf{x}(t)=\mathbf{x}_b(t)+\delta\mathbf{x}(t)$

(iii) $\delta\mathbf{x}(t) \simeq \mathbf{M}\delta\mathbf{z}$

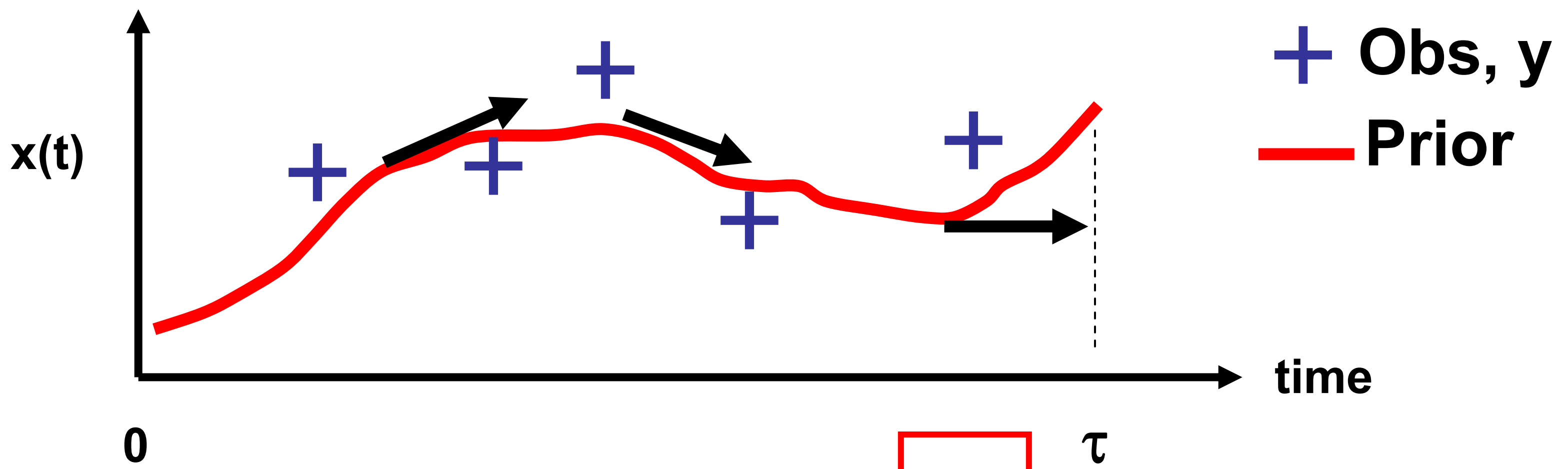
(iv) $\mathbf{H} * \delta\mathbf{x}(t) \simeq \mathbf{H} * \mathbf{M}\delta\mathbf{z}=\mathbf{G}\delta\mathbf{z}$

$\mathbf{M}$ = Tangent Linear Model

$\mathbf{H}$ = Tangent Linear $H$

# The Tangent Linear Model

**(TLROMS)**




**Prior is solution of model:** $\mathbf{x}_b(t_i) = \mathcal{M}(\mathbf{x}_b(t_{i-1}), \mathbf{f}_b(t_i), \mathbf{b}_b(t_i))$

**Nonlinear model**

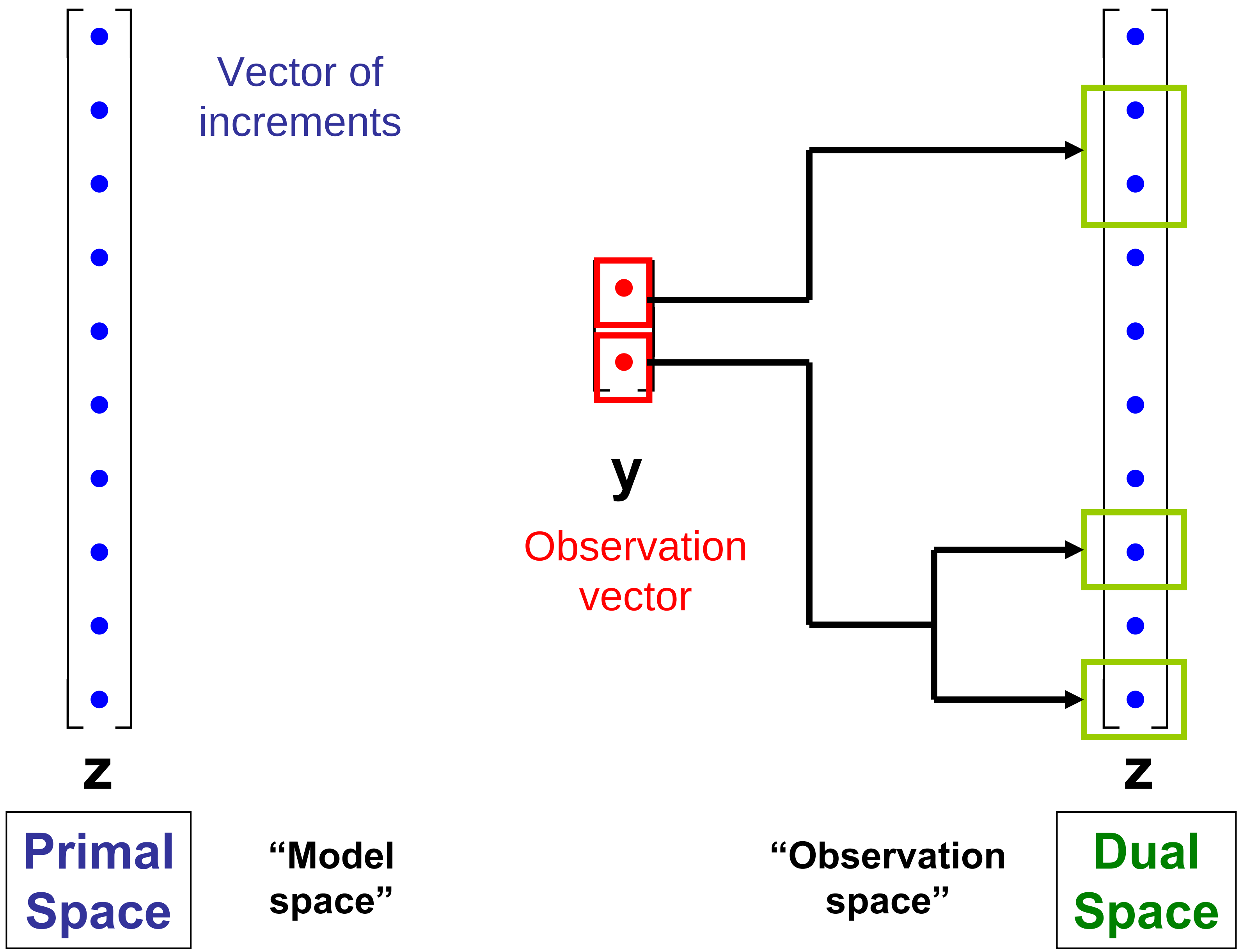
**Increment:** $\delta\mathbf{x}(t) \ll \mathbf{x}_b(t);\ \ \delta\mathbf{f}(t) \ll \mathbf{f}_b(t);\ \text{etc}$

$$\mathbf{x}(t_i) = \mathcal{M}(\mathbf{x}_b(t_{i-1}) + \delta\mathbf{x}(t_{i-1}), \cdots)$$

$$\simeq \mathcal{M}(\mathbf{x}_b(t_{i-1}), \cdots) + \mathbf{M}_{\mathbf{x}_b}\,\delta\mathbf{z}$$

**Tangent linear model**

# Primal vs Dual Formulation



# The Solution

**Analysis:** $$\mathbf{z_a} = \mathbf{z_b} + \mathbf{Kd}$$

**Gain matrix (dual form):**

$$\mathbf{K} = \mathbf{DG}^{\mathrm{T}}(\mathbf{GDG}^{\mathrm{T}} + \mathbf{R})^{-1}$$

**Gain matrix (primal form):**

$$\mathbf{K} = (\mathbf{D}^{-1} + \mathbf{G}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{G})^{-1}\mathbf{G}^{\mathrm{T}}\mathbf{R}^{-1}$$

# Two Spaces

**Gain (dual):**

$$\mathbf{K} = \mathbf{D}\mathbf{G}^{\mathrm{T}}\underbrace{\left(\mathbf{G}\mathbf{D}\mathbf{G}^{\mathrm{T}} + \mathbf{R}\right)}_{N_{\mathrm{obs}} \times N_{\mathrm{obs}}}^{-1}$$

**Gain (primal):**

$$\mathbf{K} = \underbrace{\left(\mathbf{D}^{-1} + \mathbf{G}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{G}\right)}_{N_{\mathrm{model}} \times N_{\mathrm{model}}}^{-1}\mathbf{G}^{\mathrm{T}}\mathbf{R}^{-1}$$

$$N_{\mathrm{obs}} \ll N_{\mathrm{model}}$$

# Two Spaces

**G** maps from model space
to observation space

$\mathbf{G}^T$ maps from observation space
to model space

# Iterative Solution of Primal Formulation

**(define IS4DVAR, is4dvar_ocean.h)**

**Recall the incremental cost function:**

$$J = \underbrace{\frac{1}{2}\delta\mathbf{z}^T\mathbf{D}^{-1}\delta\mathbf{z}}_{J_b} + \underbrace{\frac{1}{2}\left(\mathbf{G}\delta\mathbf{z}-\mathbf{d}\right)^{\mathrm{T}}\mathbf{R}^{-1}\left(\mathbf{G}\delta\mathbf{z}-\mathbf{d}\right)}_{J_o}$$

**At the minimum of *J* we have** $\partial J/\partial\delta\mathbf{z} = \mathbf{0}$

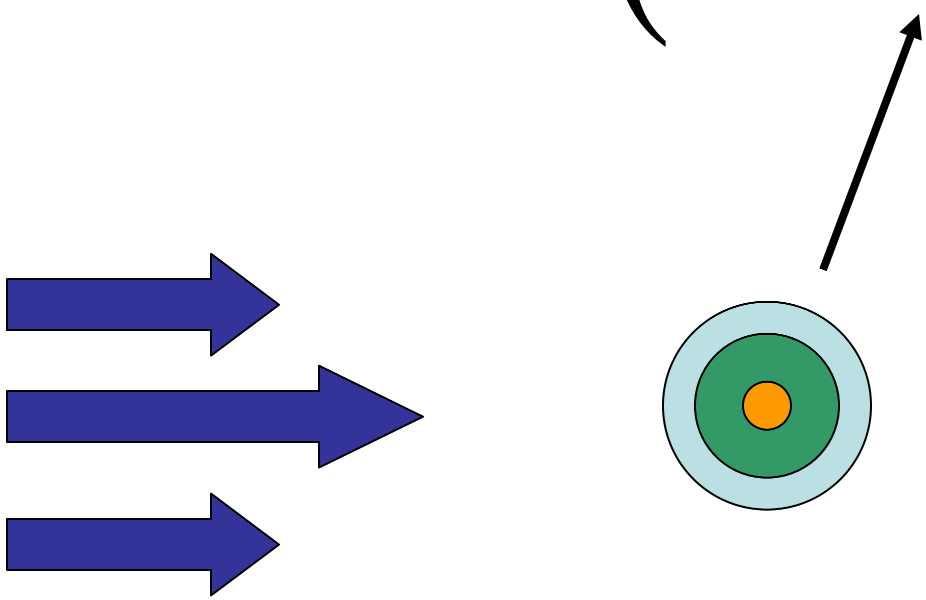
$$\partial J/\partial\delta\mathbf{z} = \mathbf{D}^{-1}\delta\mathbf{z} + \mathbf{G}^{\mathrm{T}}\mathbf{R}^{-1}\left(\mathbf{G}\delta\mathbf{z}-\mathbf{d}\right)$$

**Given *J* and $\partial J/\partial\delta\mathbf{z}$, we can identify the $\delta\mathbf{z}$ that minimizes *J***

# Matrix-less Operations

**There are no matrix multiplications!**

$$\mathbf{G}^{\mathrm{T}}\mathbf{R}^{-1}\left(\mathbf{G}\delta\mathbf{z}-\mathbf{d}\right)$$



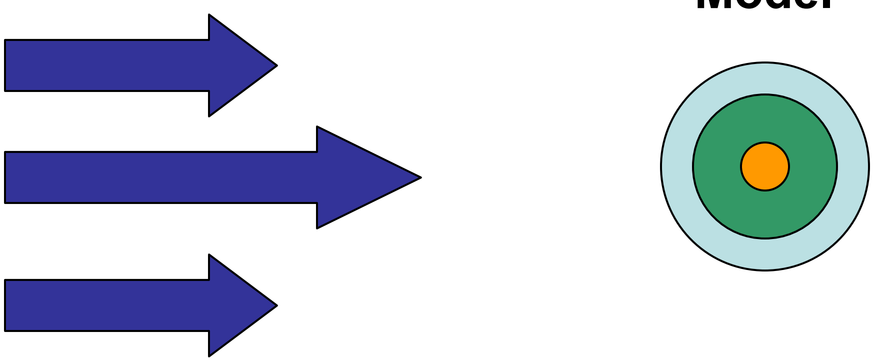
Zonal shear flow

# Matrix-less Operations

**There are no matrix multiplications!**

$$\mathbf{G}^{T}\mathbf{R}^{-1}\left(\mathbf{G}\delta\mathbf{z}-\mathbf{d}\right)$$

↑ Tangent Linear Model



Zonal shear flow

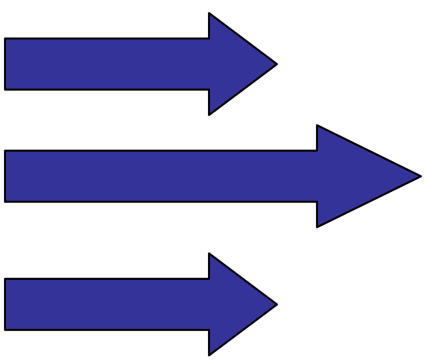
# Matrix-less Operations

**There are no matrix multiplications!**

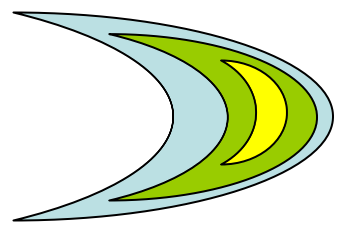
$$\mathbf{G}^{\mathrm{T}}\mathbf{R}^{-1}\left(\mathbf{G}\delta\mathbf{z}-\mathbf{d}\right)$$

↑

**Tangent Linear Model**



Zonal shear flow

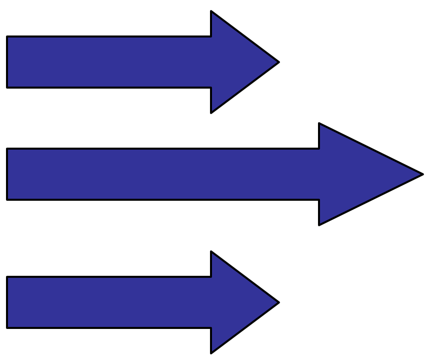


# Matrix-less Operations

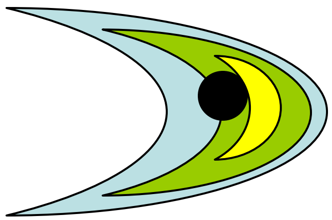
**There are no matrix multiplications!**

$$\mathbf{G}^{\mathrm{T}}\mathbf{R}^{-1}\underbrace{\left(\mathbf{G}\delta\mathbf{z}-\mathbf{d}\right)}$$

**Consider a single Observation**



**Zonal shear flow**



# Matrix-less Operations

**There are no matrix multiplications!**

$$\mathbf{G}^{T}\mathbf{R}^{-1}\underbrace{\left(\mathbf{G}\delta\mathbf{z}-\mathbf{d}\right)}$$

**Consider a single Observation**

**Zonal shear flow**

# Matrix-less Operations

**There are no matrix multiplications!**

$$\mathbf{G}^{T}\mathbf{R}^{-1}\left(\mathbf{G}\delta\mathbf{z}-\mathbf{d}\right)$$

**Inverse Obs Error Covariance**

**Zonal shear flow**

# Matrix-less Operations

**There are no matrix multiplications!**

$$\mathbf{G}^{\mathrm{T}}\mathbf{R}^{-1}\left(\mathbf{G}\delta\mathbf{z}-\mathbf{d}\right)$$

**Inverse Obs Error Covariance**

**Zonal shear flow**

# Matrix-less Operations

**There are no matrix multiplications!**

$$\mathbf{G}^{T}\mathbf{R}^{-1}\left(\mathbf{G}\delta\mathbf{z}-\mathbf{d}\right)$$

↑

**Adjoint Model**

**Zonal shear flow**

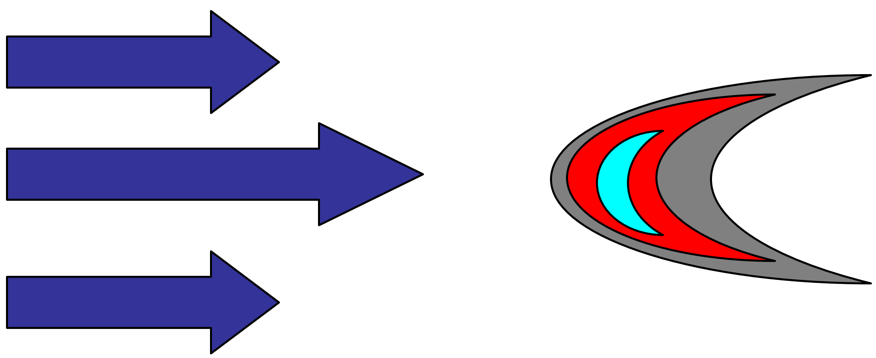
# Matrix-less Operations

**There are no matrix multiplications!**

$$\mathbf{G}^{T}\mathbf{R}^{-1}\left(\mathbf{G}\delta\mathbf{z}-\mathbf{d}\right)$$

↑

**Adjoint Model**

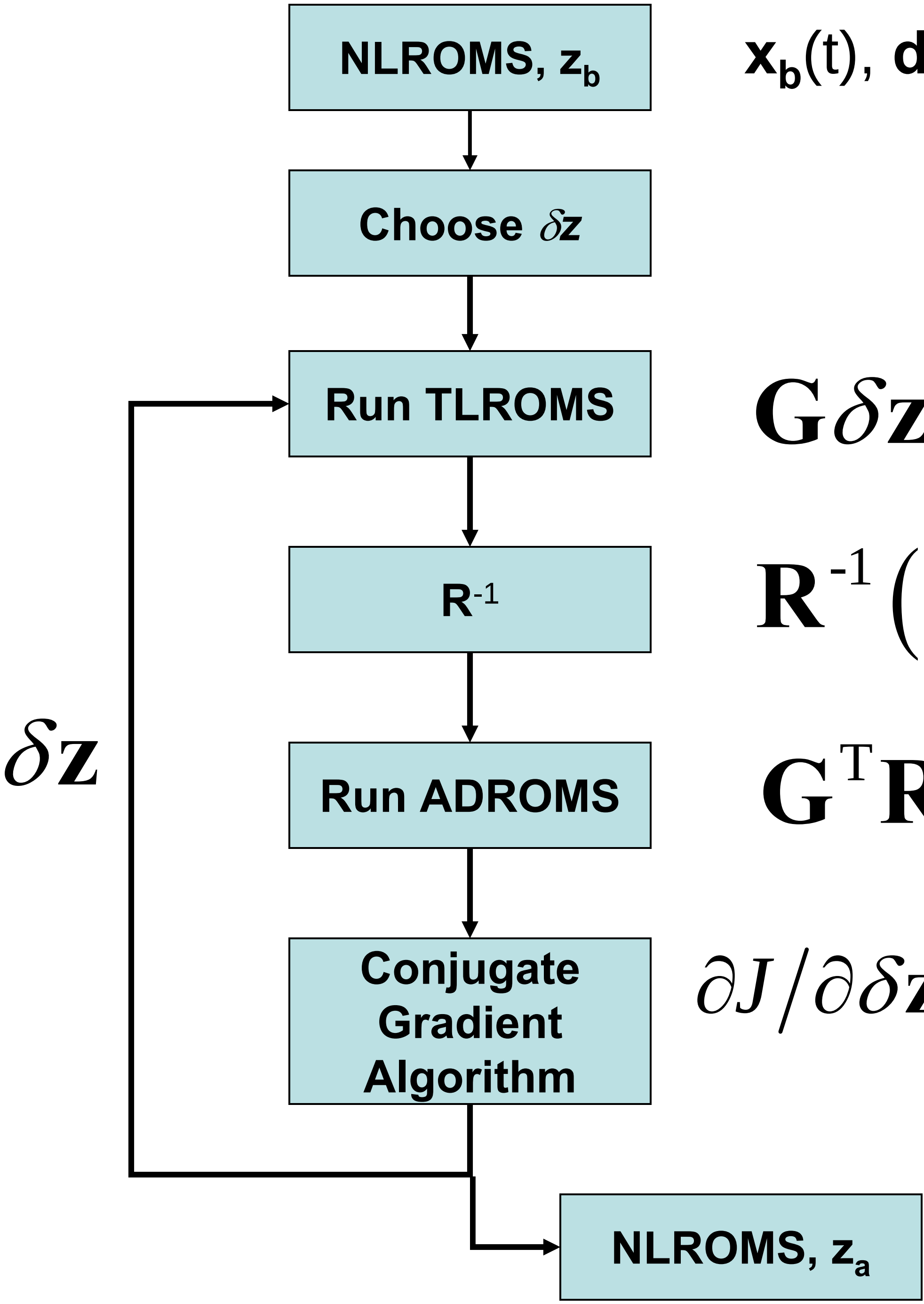


**Zonal shear flow**

**Green's Function**

$$\partial J_{o}/\partial\delta\mathbf{z}$$

# Primal 4D-Var Algorithm (I4D-Var)




$\mathbf{x}_b(t)$, $\mathbf{d}$

$$\mathbf{G}\delta\mathbf{z}, \quad (\mathbf{G}\delta\mathbf{z}-\mathbf{d}), \quad J$$

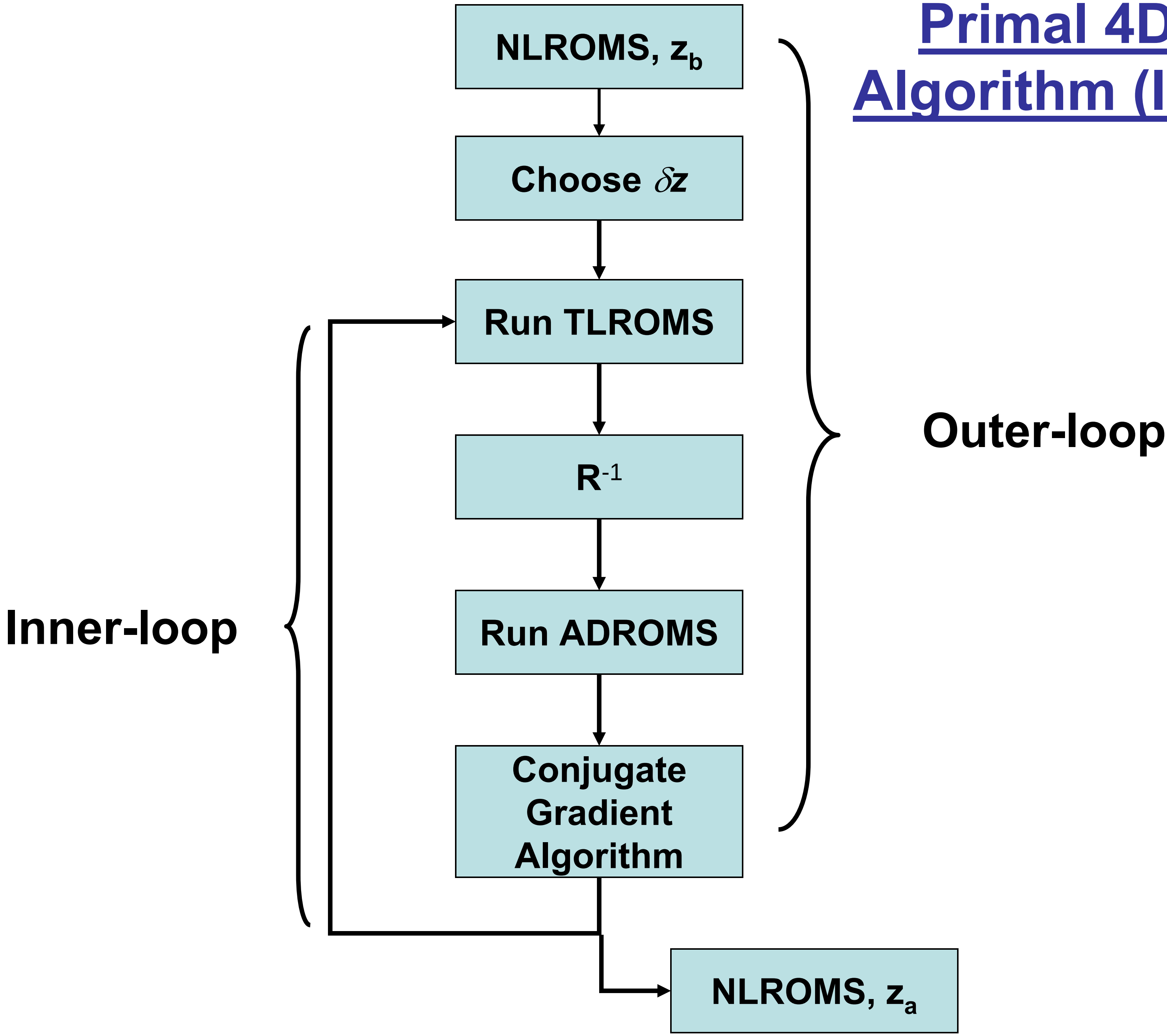
$$\mathbf{R}^{-1}(\mathbf{G}\delta\mathbf{z}-\mathbf{d})$$

$$\mathbf{G}^{T}\mathbf{R}^{-1}(\mathbf{G}\delta\mathbf{z}-\mathbf{d})$$

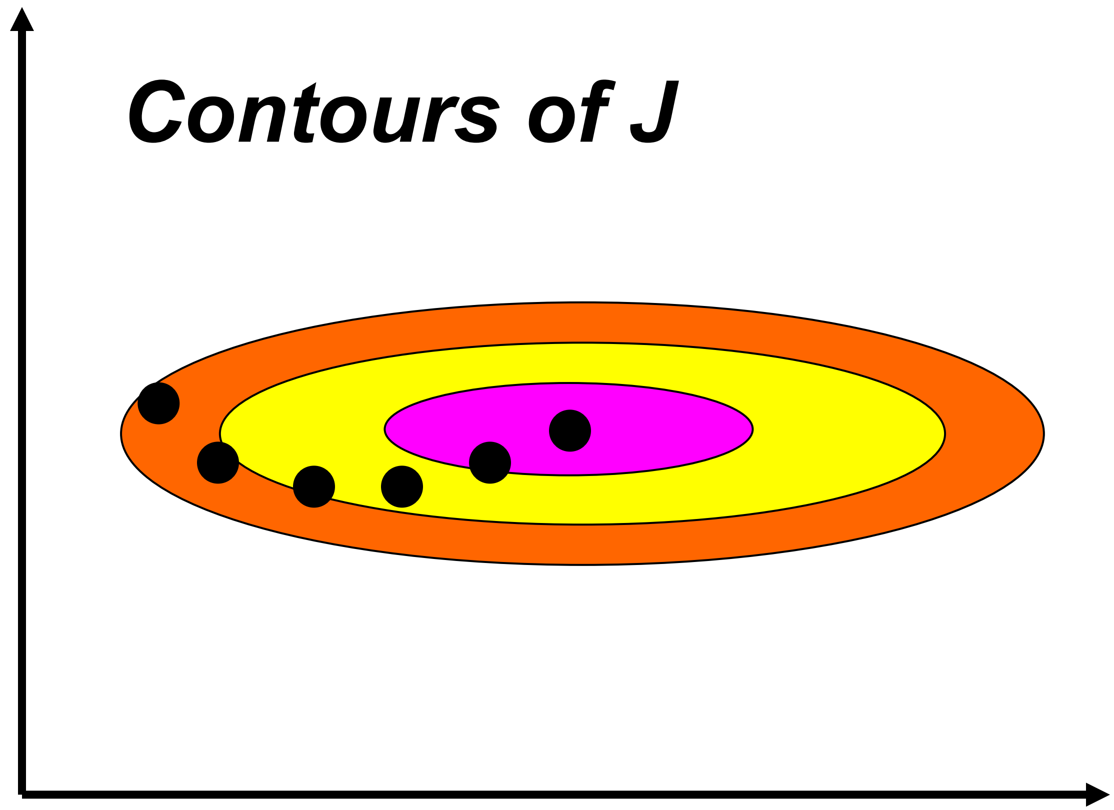
$$\partial J/\partial\delta\mathbf{z}=\mathbf{D}^{-1}\delta\mathbf{z}+\mathbf{G}^{T}\mathbf{R}^{-1}(\mathbf{G}\delta\mathbf{z}-\mathbf{d})$$

$\mathbf{x}_a(t)$

# Primal 4D-Var Algorithm (I4D-Var)



# Conjugate Gradient (CG) Methods



# An Example: ROMS CCS

COAMPS forcing

$f_b(t), B_f$

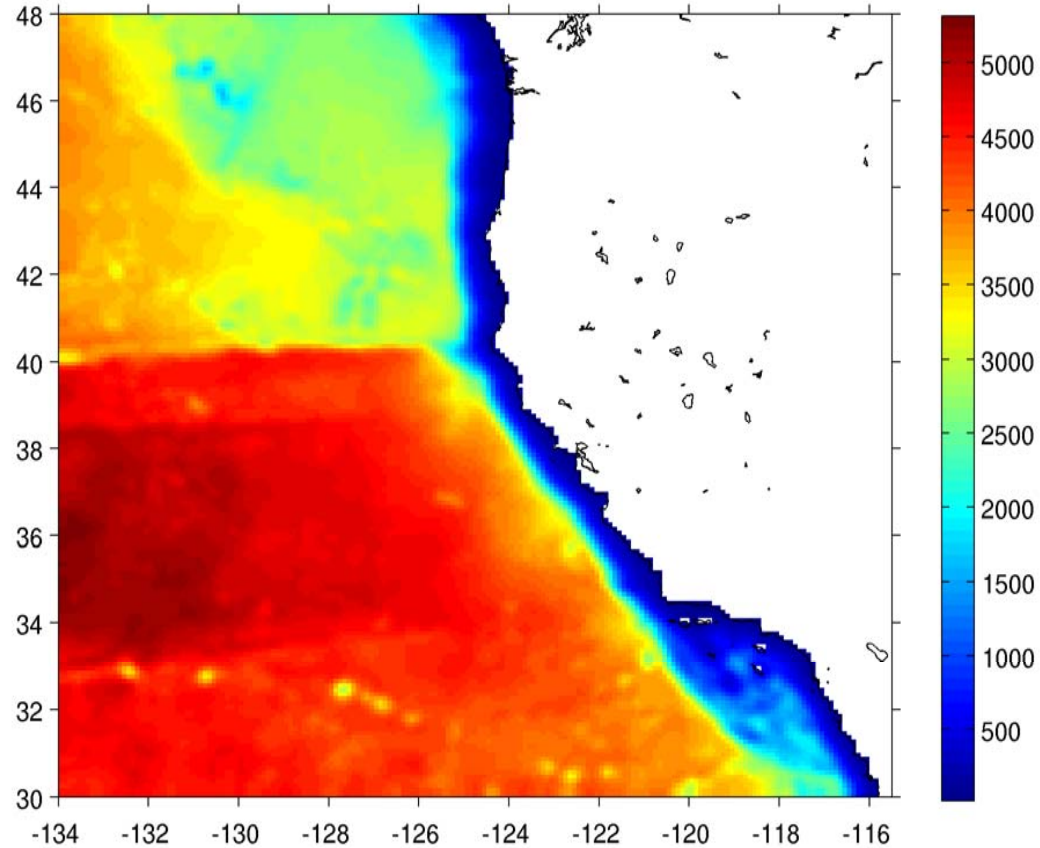
ECCO open boundary conditions

$b_b(t), B_b$

$x_b(0), B_x$

↑

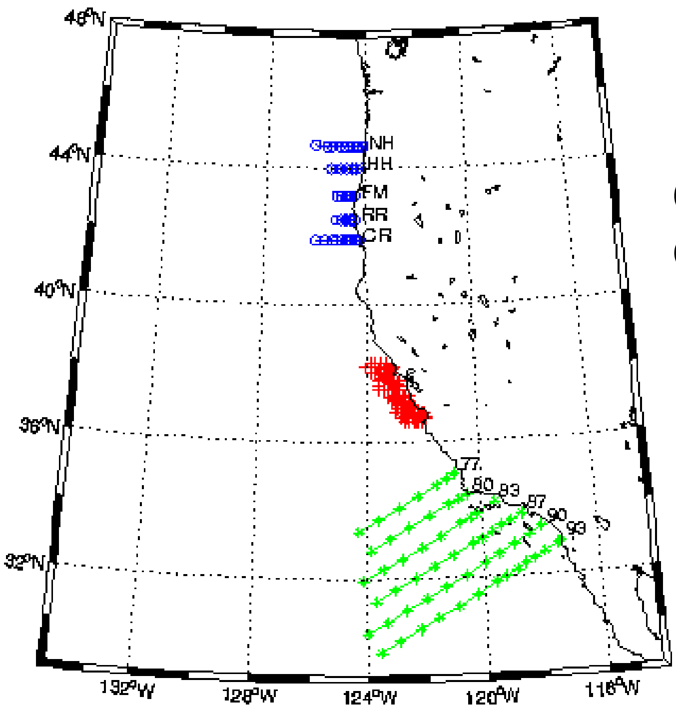
Previous assimilation cycle



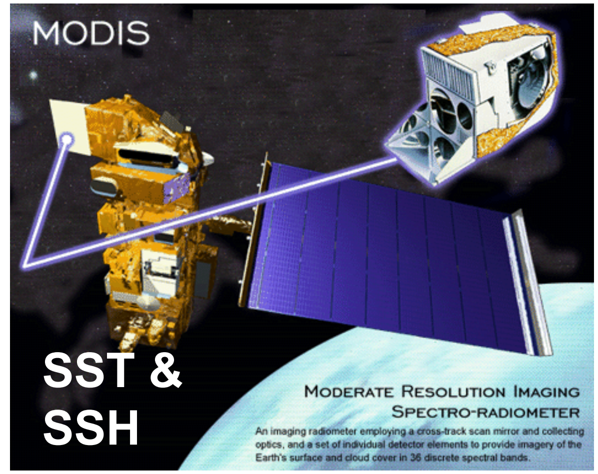

30km, 10 km & 3 km grids, 30- 42 levels

Veneziani et al (2009)
Broquet et al (2009)
Moore et al (2010)

# <u>Observations</u> (y)




**CalCOFI & GLOBEC**



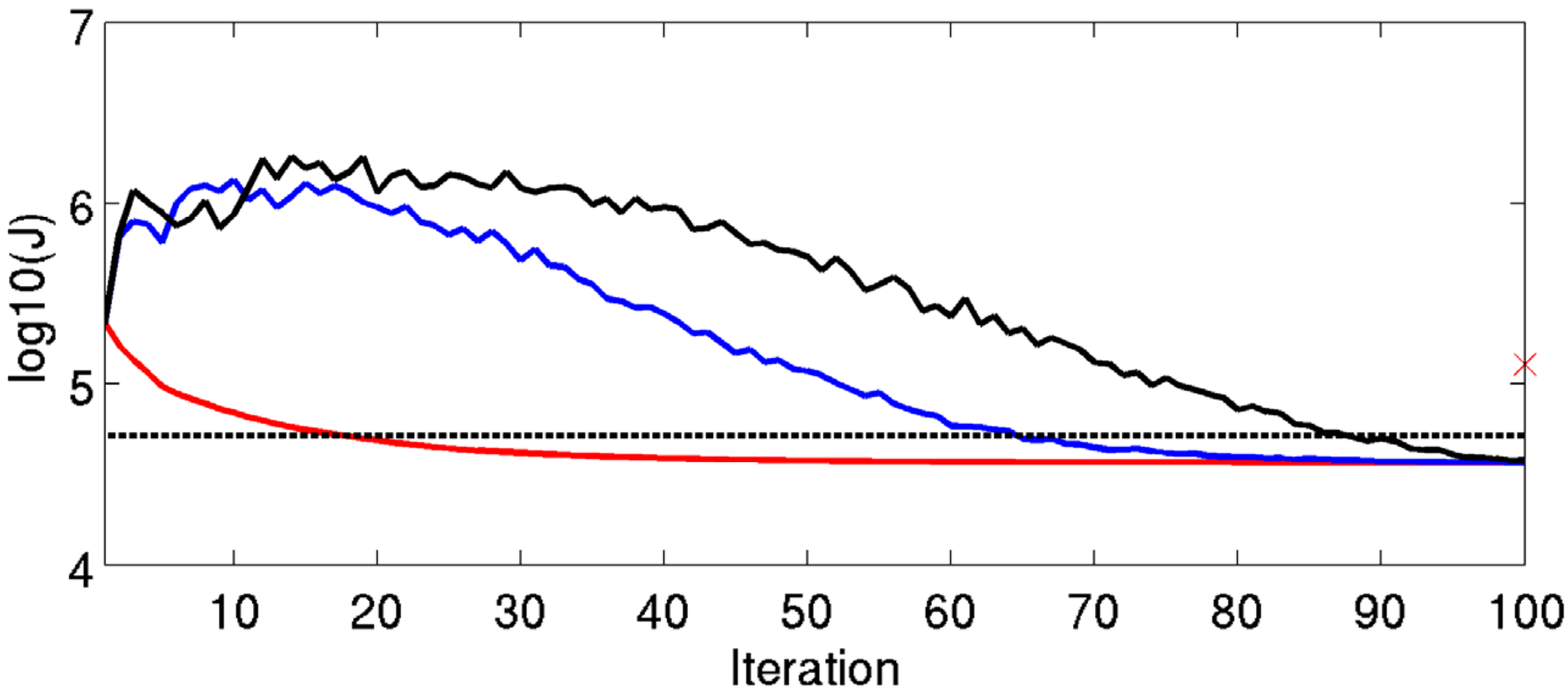


**Ingleby and Huddleston (2007)**

# 4D-Var Configuration

- Case studies for a representative case 3-10 March, 2003.
- 1 outer-loop, 100 inner-loops
- 7 day assimilation window
- *Prior* **D***:* **x** $L_h$=50 km, $L_v$=30m, $\sigma$ from clim
  **f** $L_\tau$=300km, $L_Q$=100km, $\sigma$ from COAMPS
  **b** $L_h$=100 km, $L_v$=30m, $\sigma$ from clim
- Super observations formed
- Obs error **R** (diagonal):
  SSH 2 cm
  SST 0.4 C
  hydrographic 0.1 C, 0.01psu

# 4D-Var Performance




**3-10 March, 2003**
**(10km, 42 levels)**

# Summary

- Strong constraint incremental 4D-Var, primal formulation:

  define IS4DVAR

  Drivers/is4dvar_ocean.h

- Matrix-less iterations to identify cost function minimum using TLROMS and ADROMS

# References


- Broquet, G., C.A. Edwards, A.M. Moore, B.S. Powell, M. Veneziani and J.D. Doyle, 2009: Application of 4D-variational data assimilation to the California Current System. *Dyn. Atmos. Oceans*, **48**, 69-91.
- Courtier, P., J.-N. Thépaut and A. Hollingsworth, 1994: A strategy for operational implemenation of 4D-Var using an incremental approach. *Q. J. R. Meteorol. Soc*., **120**, 1367-1388.
- Ingleby, B. and M. Huddleston, 2007: Quality control of ocean temperature and salinity profiles - historical and real-time data. *J. Mar. Systems*, **65**, 158-175.
- Moore, A.M., H.G. Arango, G. Broquet, B.S. Powell, J. Zavala-Garay and A.T. Weaver, 2010: The Regional Ocean Modeling System (ROMS) 4-dimensional data assimilation systems: Part I. *Ocean Modelling*, Submitted.
- Veneziani, M., C.A. Edwards, J.D. Doyle and D. Foley, 2009: A central California coastal ocean modeling study: 1. Forward model and the influence of realistic versus climatological forcing. *J. Geophys. Res*., **114**, C04015, doi:10.1029/2008JC004774.
- Wikle, C.K. and L.M. Berliner, 2007: A Baysian tutorial for data assimilation. *Physica D*, **230**, 1-16.