Principles of data assimilation part II

From linear regression to 4DVAR

Examples of
inverse modelling and
state estimation

AWI
Alfred Wegener Institute
for Polar and Marine Research

unusual methods

- simulated annealing
- genetic algorithms
- micro-genetic algorithm
- neural networks
- particle filter
- sequential importance resampling filter

usual problems

- breakdown of ensemble

  in sequential importance resampling filter
  in ensemble methods (error subspace)

  EnKF, SEEK, SEIK

too much confidence in error of model

usual problems

replacing of variables

$$\psi^{\mathrm{mod}el} = \psi^{measured}$$

nudging

$$\frac{\partial\psi}{\partial t} = \frac{\partial\psi}{\partial t}^{physics} - \gamma(\psi - d^{measured})$$

$$\frac{\partial\nabla^2\psi}{\partial t} = \frac{\partial\nabla^2\psi}{\partial t}^{physics} - \gamma(\psi - d^{measured})$$

fitting a straight line through data

of the form $y = x_1 + x_2 t$

with the cost function j

$$j = 0.5 \sum_{i=1}^{N} \frac{(y_i - y)^2}{\sigma_i^2} = 0.5 \sum_{i=1}^{N} \frac{(y_i - x_1 - x_2 t_i)^2}{\sigma_i^2}$$

$$= 0.5 (\mathbf{A}x - y)^T \mathbf{W} (\mathbf{A}x - y)$$

fitting a straight line through data

of the form $$y = x_1 + x_2 t$$

$$\mathbf{A} = \begin{pmatrix} 1 & t_1 \\ 1 & t_2 \\ \cdots & \cdots \\ 1 & t_N \end{pmatrix} \quad \mathbf{W} = \begin{pmatrix} \sigma_1^{-2} & & & 0 \\ & \sigma_2^{-2} & & \\ & & \cdots & \\ 0 & & & \sigma_N^{-2} \end{pmatrix} \quad \mathrm{y} = \begin{pmatrix} y(t_1) \\ y(t_2) \\ \cdots \\ y(t_N) \end{pmatrix}$$

$$\mathrm{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

fitting a straight line through data

The minimum of j can be found by setting the derivative of j to zero:

.

$$j = 0.5(\mathbf{A}x - y)^T \mathbf{W}(\mathbf{A}x - y)$$

$$\frac{dj}{dx} = \mathbf{A}^T \mathbf{W}(\mathbf{A}x - y) = 0$$

$$(\mathbf{A}^T \mathbf{W} \mathbf{A})x = \mathbf{A}^T \mathbf{W} y$$

$$x = (\mathbf{A}^T \mathbf{W} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{W} y$$

fitting a straight line through data

the error covariance of the solution is:

.

$$= \mathrm{cov}\langle(\mathrm{x}-\hat{\mathrm{x}}),(\mathrm{x}-\hat{\mathrm{x}})^{\mathrm{T}}\rangle = \mathbf{H}^{-1}$$

$$\mathbf{H} = \frac{\partial^2 \mathrm{j}}{\partial \mathrm{x}^2}$$

$$= \frac{\partial}{\partial \mathrm{x}} \mathbf{A}^{\mathrm{T}}\mathbf{W}(\mathbf{A}\mathrm{x}-\mathrm{y})$$

$$= (\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A})$$

fitting a straight line through data

the solution is:

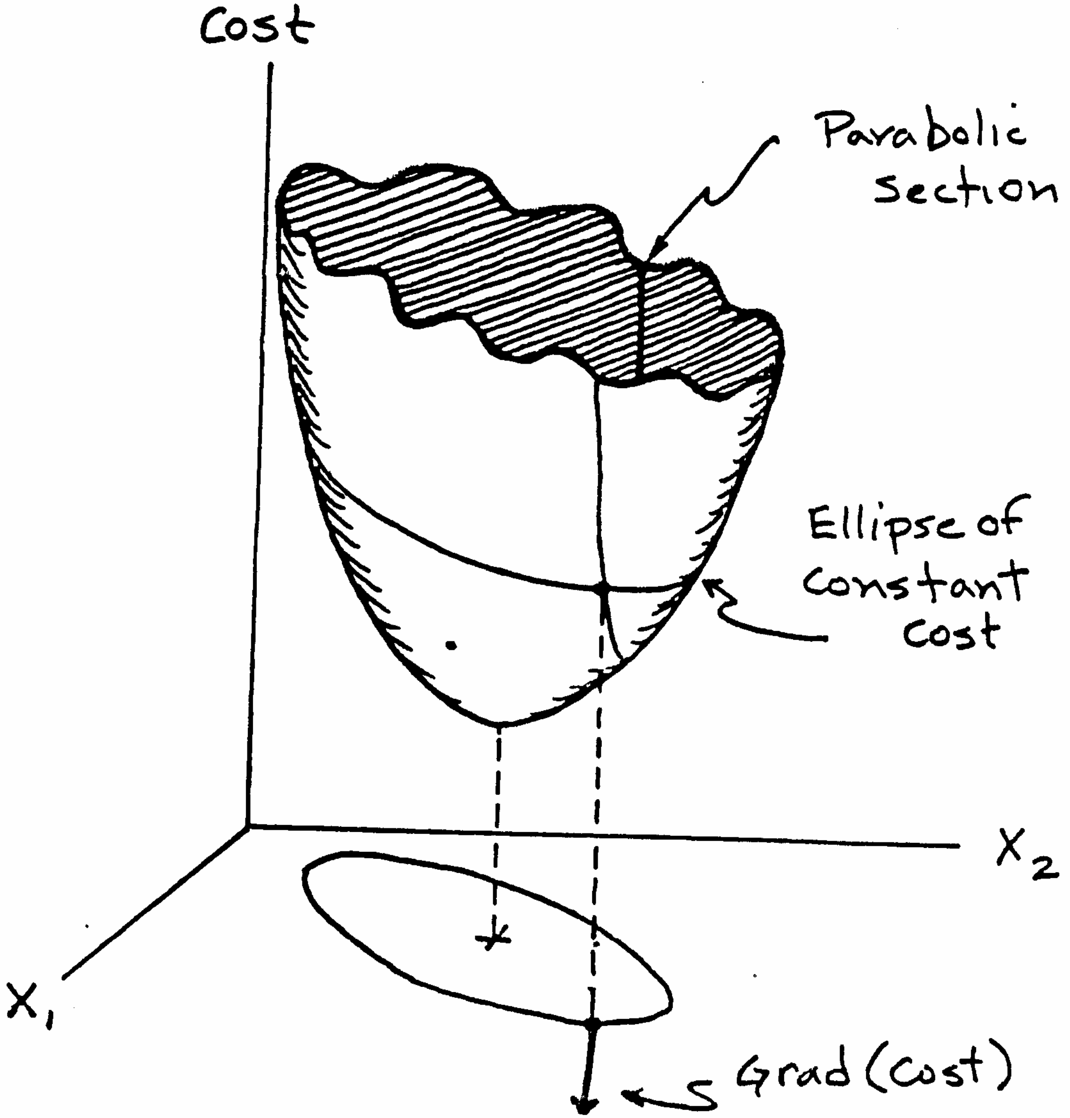
.

$$x = (\mathbf{A}^T\mathbf{W}\mathbf{A})^{-1}\mathbf{A}^T\mathbf{W}y$$

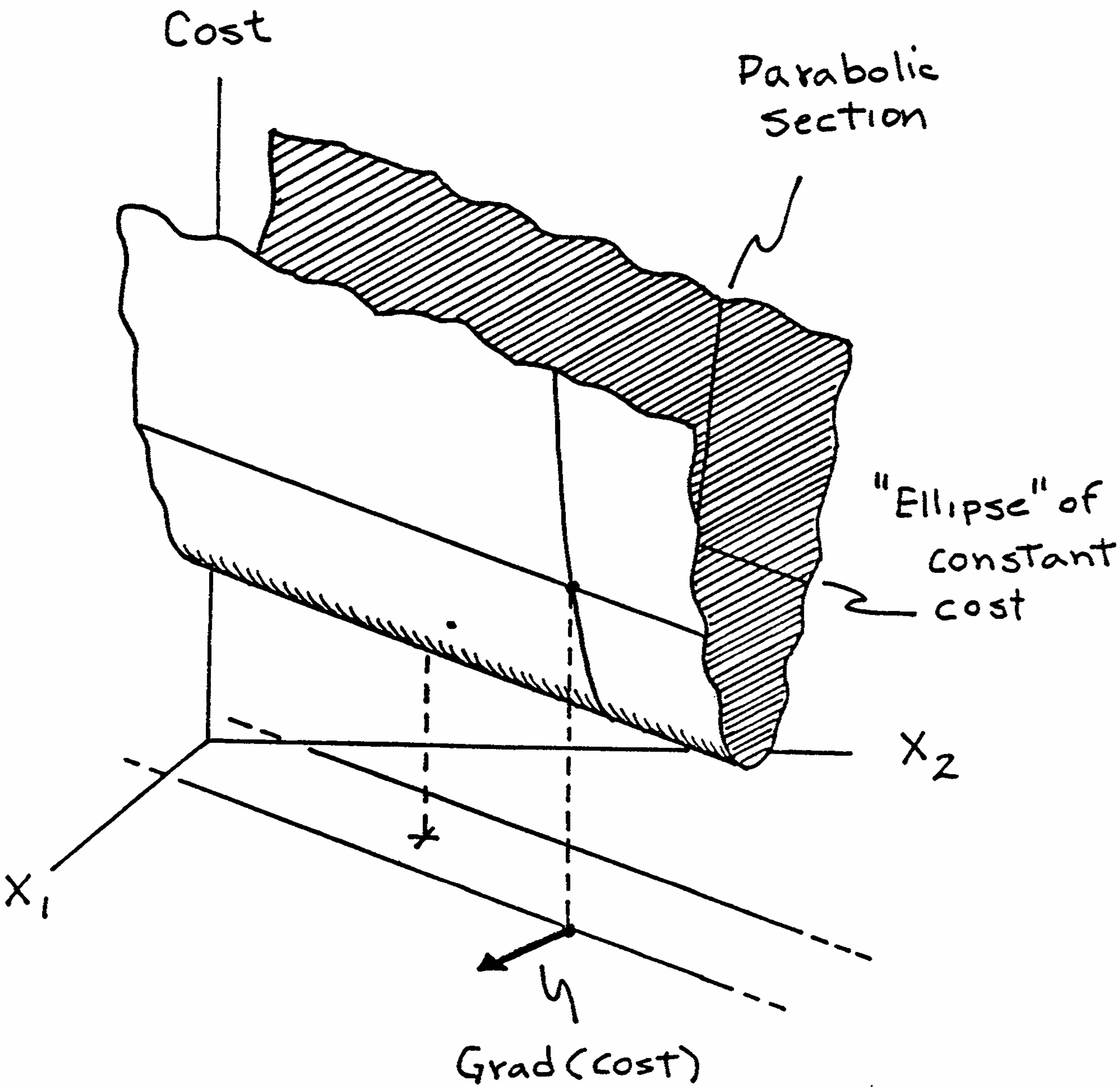
$$x = \mathbf{H}^{-1}\mathbf{A}^T\mathbf{W}y = \mathbf{H}^{-1}\tilde{y}$$

if the inverse of the Hessian
does not exist,
there is no reasonable solution for x

the Hessian matrix describes the curvature of j



the Hessian matrix describes the curvature of j



A singular value decomposition

$$\mathbf{A} = \mathbf{U}\,\mathbf{S}\,\mathbf{V}^{\mathbf{T}}$$

$$\mathbf{U}^{\mathbf{T}}\mathbf{U} = \mathbf{I}$$

$$\mathbf{V}^{\mathbf{T}}\mathbf{V} = \mathbf{I}$$

$$\mathbf{U}\ \mathbf{U}^{\mathbf{T}} \neq \mathbf{I}$$

$$\mathbf{V}\ \mathbf{V}^{\mathbf{T}} \neq \mathbf{I}$$

$$\mathbf{A}^{-1} = \mathbf{U}\,\mathbf{S}^{-1}\,\mathbf{V}^{\mathbf{T}}$$

where U and V contain the singular vectors and the diagonal matrix S the singular values

solving a linear least squares system

The minimum of j can be found by setting the derivative of j to zero:

.

$$j = 0.5(\mathbf{A}x - y)^{T}(\mathbf{A}x - y)$$

$$j = 0.5(\mathbf{U}\,\mathbf{S}\,\mathbf{V}^{\mathbf{T}}x - y)^{T}(\mathbf{U}\,\mathbf{S}\,\mathbf{V}^{\mathbf{T}}x - y)$$

$$\frac{dj}{dx} = (\mathbf{U}\,\mathbf{S}\,\mathbf{V}^{\mathbf{T}})^{T}(\mathbf{U}\,\mathbf{S}\,\mathbf{V}^{\mathbf{T}}x - y) = 0$$

solving a linear least squares system

simple algebra leads to:

$$(\mathbf{V}\,\mathbf{S}\,\mathbf{U}^{\mathbf{T}}\mathbf{U}\,\mathbf{S}\,\mathbf{V}^{\mathbf{T}})x = \mathbf{V}\,\mathbf{S}\,\mathbf{U}^{\mathbf{T}}\,y$$

$$\mathbf{V}\;\mathbf{S}^{2}\,\mathbf{V}^{\mathbf{T}}x = \mathbf{V}\,\mathbf{S}\,\mathbf{U}^{\mathbf{T}}\,y$$

$$x = \mathbf{V}\;\mathbf{S}^{-2}\,\mathbf{V}^{\mathbf{T}}\mathbf{V}\,\mathbf{S}\,\mathbf{U}^{\mathbf{T}}\,y$$

$$x = \mathbf{V}\;\mathbf{S}^{-1}\;\mathbf{U}^{\mathbf{T}}\,y$$

solving a linear least squares system

.

$$x = \mathbf{V}\ \mathbf{S}^{-1}\ \mathbf{U}^{\mathbf{T}}\ y$$

$$x = \sum_{s_i \neq 0} \frac{u_i y}{s_i}\ v_i$$

the solution x is the sum of the singular vectors V, weighted by the inverse of the eigenvalues s_i times the projection of the observations via the singular vectors UT

resolution

.

$$x = \sum_{s_i \neq 0} \frac{u_i y}{s_i} v_i$$

$$\mathbf{V}\,\mathbf{V}^{\mathbf{T}} \neq \mathbf{I}$$

the resolution matrix is non diagonal for rank deficient solutions

resolution

.

the error given by the inverse Hessian matrix for rank deficient solutions does not describe the whole picture.

$$x^{solution} = \mathbf{V}\,\mathbf{V}^{\mathbf{T}} x^{true}$$

there is a null space remaining!

tapered least sqares

.

$$x = \sum_{s_i \neq 0} \frac{s_i}{s_i^{\;2} + \varepsilon} \; u_i y \;\; v_i$$

has full resolution but is a kind of cheating

stabilized solution:

$$x = \sum_{s_i \neq 0} \frac{1}{s_i \;\; + \varepsilon} \; u_i y \;\; v_i$$

use of *prior* information

.

$$j = 0.5(\mathbf{A}x - y)^T(\mathbf{A}x - y) + (x - \hat{x})^T\mathbf{B}(x - \hat{x})$$

$$\frac{dj}{dx} = \mathbf{A}^T\mathbf{A}(x - y) + \mathbf{B}(x - \hat{x}) = 0$$

$$(\mathbf{A}^T\mathbf{A} + \mathbf{B})x = \mathbf{A}^T\mathbf{A}y + \mathbf{B}\hat{x}$$

$$x = (\mathbf{A}^T\mathbf{A} + \mathbf{B})^{-1}(\mathbf{A}^T\mathbf{A}y + \mathbf{B}\hat{x})$$

Gauss Markov solution

$$\mathbf{B} = \varepsilon\ \mathbf{I}$$

use of *prior* information

the best estimate for our *prior* information is expressed by the ‘background’ or weighting matrix B

we weight by the inverse of the error covariance matrix

$$cov\langle (x-\hat{x}),(x-\hat{x})^T \rangle = \mathbf{B}^{-1}$$

$$\mathbf{B} = \mathbf{H}$$

Example

Describe the stationary barotropic flow of a global ocean circulation model with

$$\mathbf{A}\begin{pmatrix}\psi \\ \zeta\end{pmatrix} - \mathbf{r} = 0$$

And assimilate SSH ζ (altimetry)

and *prior* streamfunction ψ

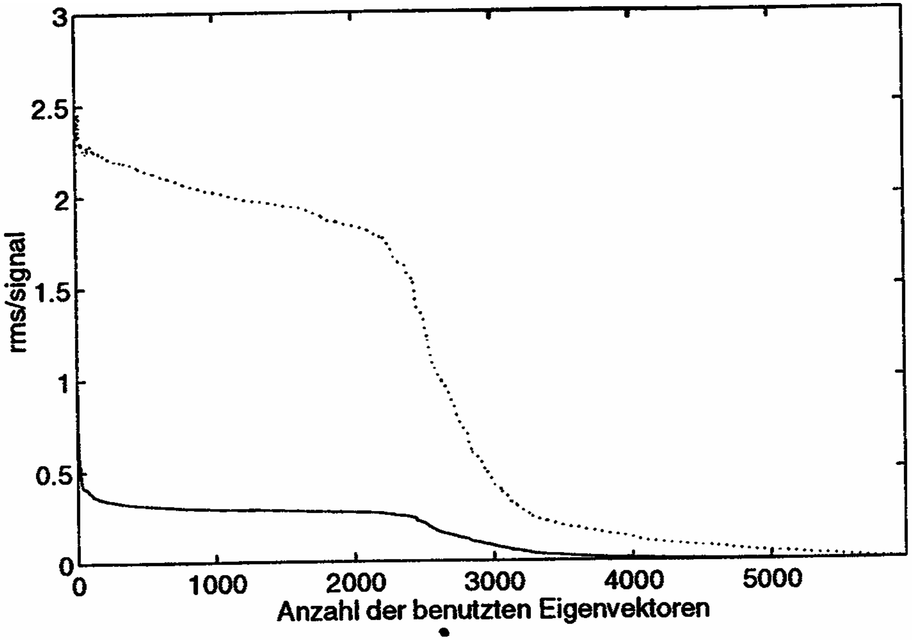
Example

the cost function j is

$$
\begin{aligned}
j = \left(\mathbf{A}\begin{pmatrix}\psi \\ \zeta\end{pmatrix} - \mathrm{r}\right)^{\mathrm{T}} \mathbf{W}_1 \left(\mathbf{A}\begin{pmatrix}\psi \\ \zeta\end{pmatrix} - \mathrm{r}\right) \\
+ (\psi - \psi^{\mathrm{prior}})^{\mathrm{T}} \mathbf{B} \, (\psi - \psi^{\mathrm{prior}}) \\
+ (\zeta - \zeta^{\mathrm{altimeter}})^{\mathrm{T}} \mathbf{W}_2 (\zeta - \zeta^{\mathrm{altimeter}})
\end{aligned}
$$

And solve by a SVD decomposition

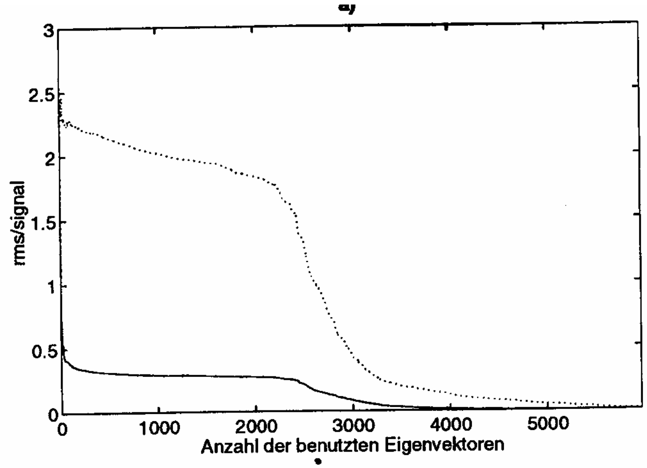
Normalized error of solution



full line altimetry

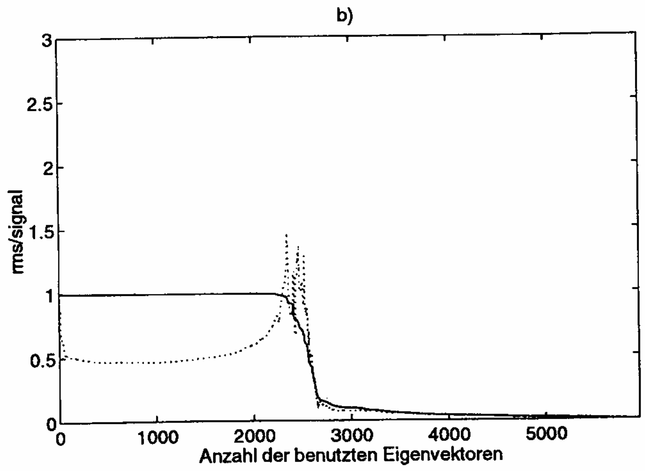
dotted line streamfunction

Normalized error of solution



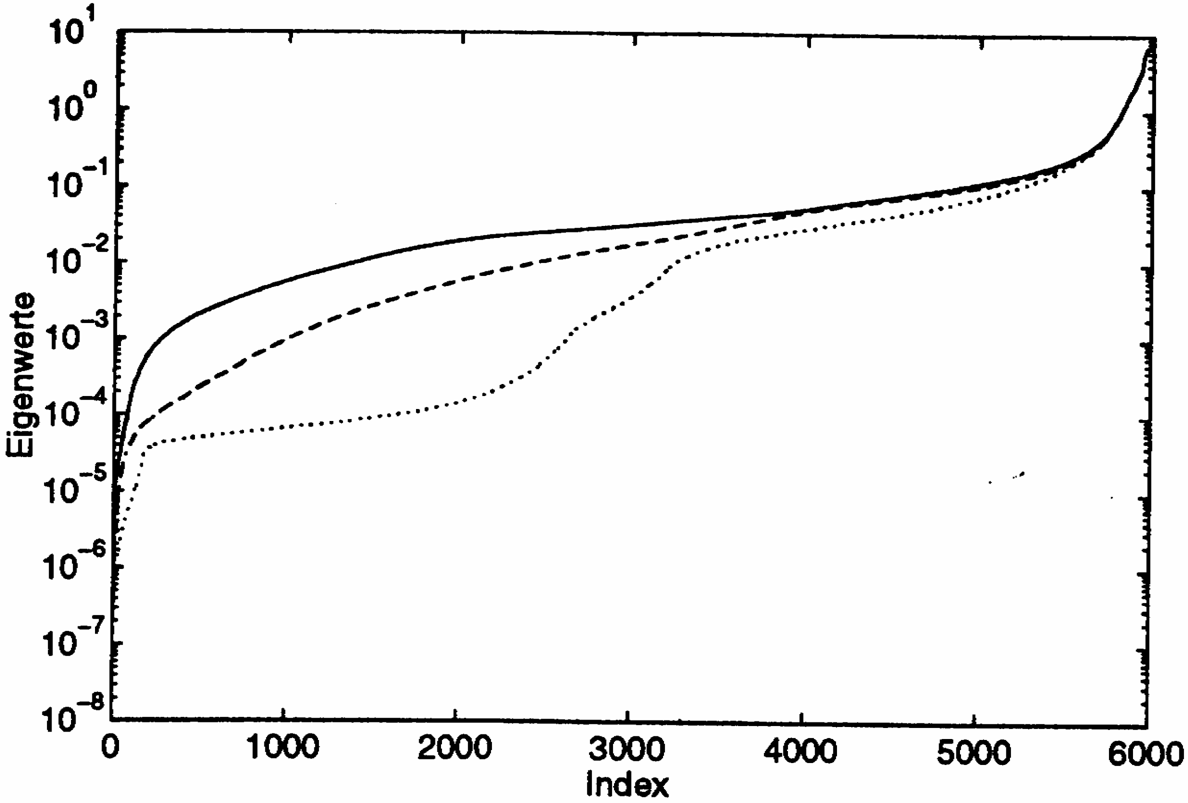
full line altimetry
dotted line streamfunction

scaling depth 5000m



scaling depth 300m

spectra of inverted matrix



dotted: model matrix A

dashed: A + altimetry

full: A + altimetry + prior on streamfunction

fitting a straight line through data

of the form $$y = a + bx$$

with the cost function j

$$j = 0.5 \sum_{i=1}^{N} \frac{(y_i - y)^2}{\sigma_i^2}$$

subject to

$$a + bx - y = 0$$

fitting a straight line through data

construct the Lagrangian function

$$L = j + \lambda(a + bx - y)$$

solve for a stationary point
and of L

using a Lagrangian function

$$\frac{\partial L}{\partial \lambda} = 0 = (a + bx - y)$$

$$\frac{\partial L}{\partial y} = 0 = \lambda + \frac{\partial j}{\partial y} = \lambda + \sum_{i=1}^{N} \frac{(y_i - y)}{\sigma_i^2}$$

$$\frac{\partial L}{\partial a} = 0 = \lambda$$

$$\frac{\partial L}{\partial b} = 0 = \lambda x$$

FEOM model equations

- equation of state $\rho = \rho(T, S, P)$
- steady state primitive equations

  dynamical part $(u, v, w, \zeta) = \Psi(\rho, \tau)$

  strong constraint

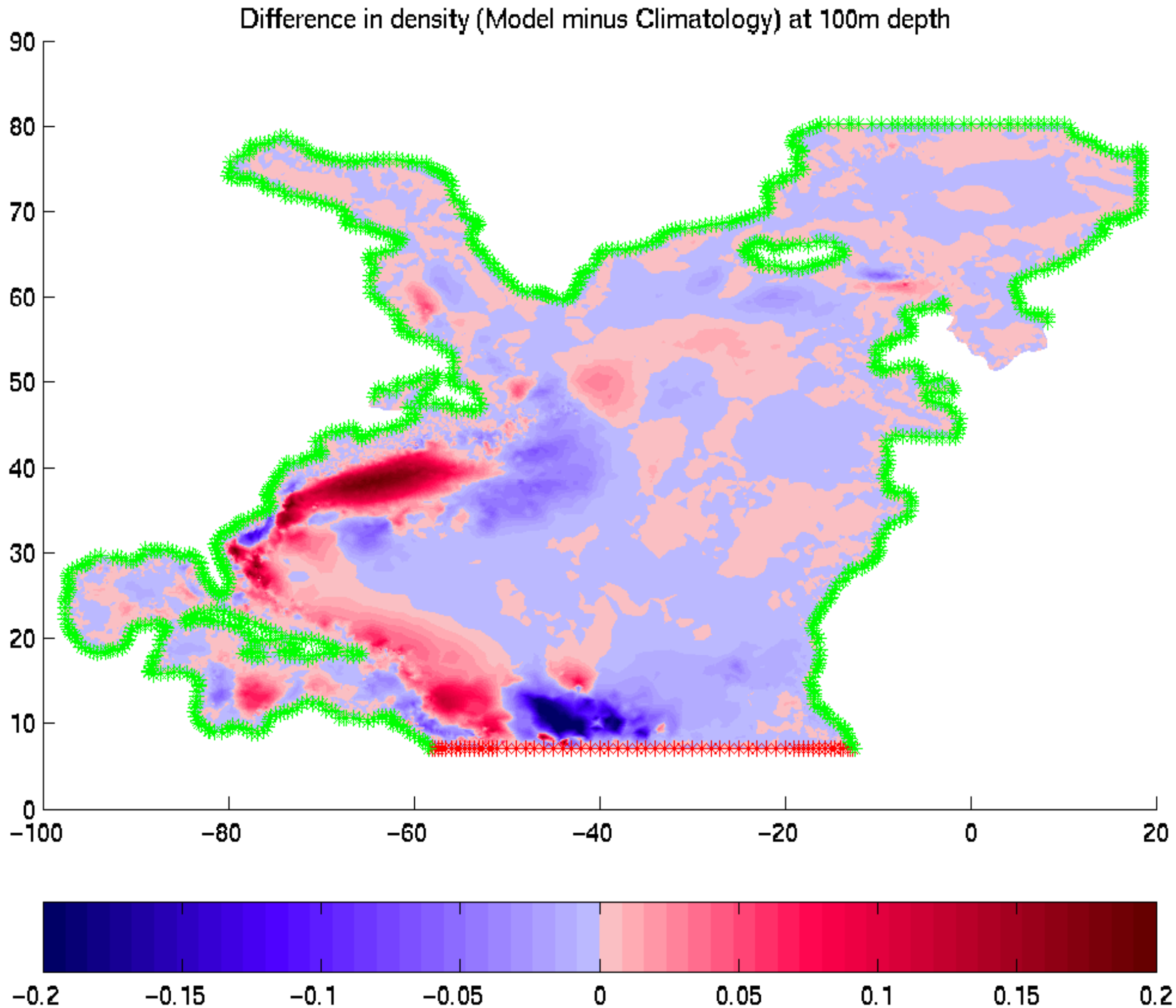
  advection-diffusion equation for density

  $$\nabla \cdot (\vec{u}\rho) - \nabla \cdot (\mathrm{K}\nabla\rho) = F_{\rho}$$

  weak constraint

Difference in density
(optimized minus climatology)
density is the only control parameter

latitude



100m depth

longitude

$$\frac{kg}{m^3}$$

adjoint equations

solution of a constrained optimization problem by transforming it to an unconstrained problem

$$\min j, \quad \text{subject to} \quad E(u,s,x,t)$$

$$L(u,s,\lambda) = j + \int_X \int_{t_0}^{t_1} \lambda(x,t)\, E(u,s,x,t)\, dx\, dt$$

$$\frac{\partial L}{\partial \lambda} = 0$$

$$\frac{\partial L}{\partial s} = 0$$

$$\frac{\partial L}{\partial u} = 0$$

provided E=0 the L and j coincide identically

and we get for the gradient of the (implicit) costfunction:

$$
\begin{aligned}
\nabla_u j &= \frac{dj}{du} \\
&= \frac{dL}{du} \\
&= \frac{\partial L}{\partial u} + \underbrace{\frac{\partial L}{\partial \lambda}}_{=0} \frac{\partial \lambda}{\partial u} + \underbrace{\frac{\partial L}{\partial s}}_{=0} \frac{\partial s}{\partial u} = 0
\end{aligned}
$$

Example for u = initial conditions

we get for the gradient of the (implicit) costfunction:

$$\nabla_u j = \frac{\partial j}{\partial u} + \int_X \int_{t_0}^{t_1} \lambda(x,t) \frac{\partial E(u,s,x,t)}{\partial u} dx\, dt$$

$$\nabla_u j = \left.\frac{\partial j}{\partial u}\right|_{t_0} + \lambda(t_0) \quad \textit{for initial conditions}$$

Example

shallow water equations with a passive tracer

$$\frac{\partial v}{\partial t} + v\frac{\partial v}{\partial x} - A_M \frac{\partial^2 v}{\partial x^2} + \Phi \frac{\partial \phi}{\partial x} = 0$$

$$\frac{\partial \phi}{\partial t} + \frac{\partial (v\phi)}{\partial x} = 0$$

$$\frac{\partial T}{\partial t} + v\frac{\partial T}{\partial x} - A_T \frac{\partial^2 T}{\partial x^2} + \varepsilon T = 0$$

Example

setup of Lagrange function

$$L = \lambda_v \int_X \int_{t_0}^{t_1} \left(\frac{\partial v}{\partial t} + v \frac{\partial v}{\partial x} - A_M \frac{\partial^2 v}{\partial x^2} + \Phi \frac{\partial \phi}{\partial x} \right) dx\, dt$$

$$+ \lambda_\phi \int_X \int_{t_0}^{t_1} \left(\frac{\partial \phi}{\partial t} + \frac{\partial (v\phi)}{\partial x} \right) dx\, dt$$

$$+ \lambda_T \int_X \int_{t_0}^{t_1} \left(\frac{\partial T}{\partial t} + v \frac{\partial T}{\partial x} - A_T \frac{\partial^2 T}{\partial x^2} + \varepsilon T \right) dx\, dt$$

$$+ j$$

Example

partial integration of Lagrange function:

$$\frac{\partial L}{\partial s} - \frac{\partial}{\partial t}\left(\frac{\partial L}{\partial s_t}\right) - \frac{\partial}{\partial x}\left(\frac{\partial L}{\partial s_x}\right) + \frac{\partial^2}{\partial x^2}\left(\frac{\partial L}{\partial s_{xx}}\right) = 0$$

Example

partial integration of Lagrange function

yields the adjoint equations

$$-\frac{\partial\lambda_v}{\partial t}-v\frac{\partial\lambda_v}{\partial x}-\phi\frac{\partial\lambda_\phi}{\partial x}-\lambda_T\frac{\partial T}{\partial x}-A_M\frac{\partial^2\lambda_v}{\partial x^2}=-\frac{\partial j}{\partial v}$$

$$-\frac{\partial\lambda_\phi}{\partial t}-v\frac{\partial\lambda_\phi}{\partial x}-\Phi\frac{\partial\lambda_v}{\partial x}=-\frac{\partial j}{\partial\phi}$$

$$-\frac{\partial\lambda_T}{\partial t}-v\frac{\partial(v\lambda_T)}{\partial x}-A_T\frac{\partial^2\lambda_T}{\partial x^2}+\varepsilon\lambda_T=-\frac{\partial j}{\partial T}$$

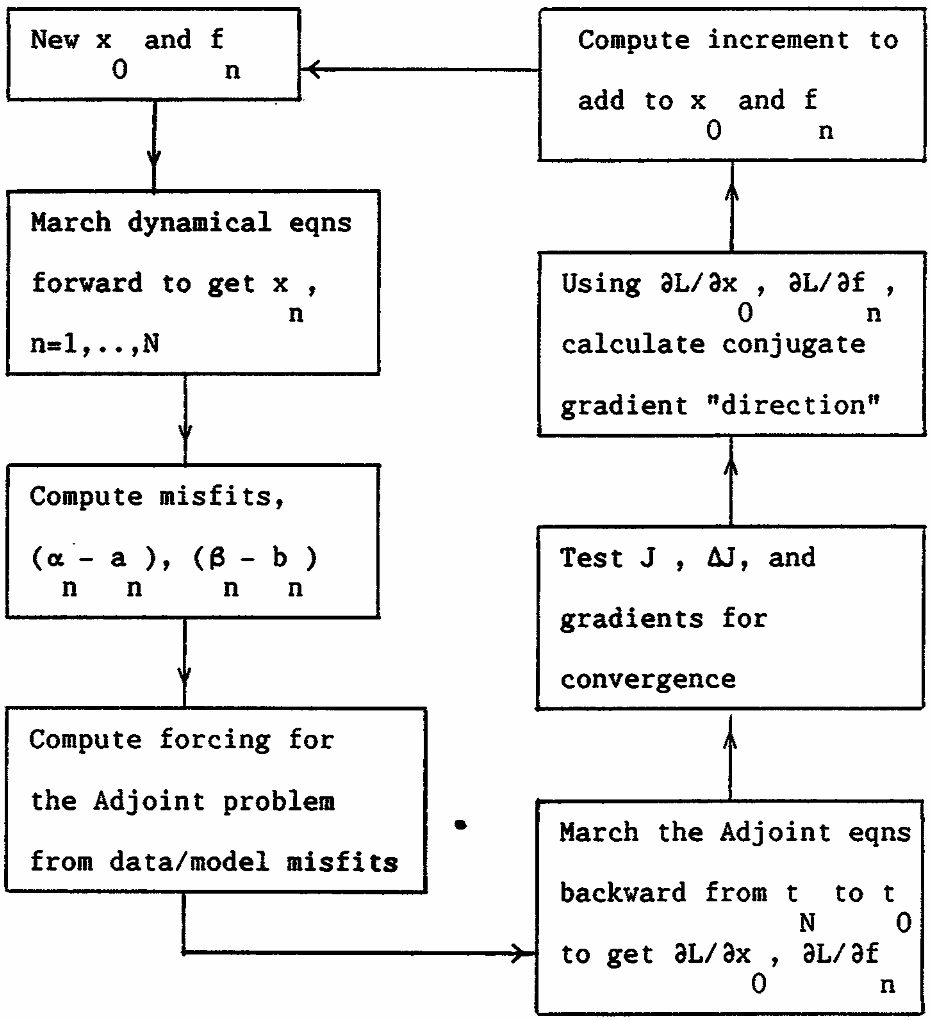
Example

the adjoint equations integrated backwards in time yield any gradient information we can think of.

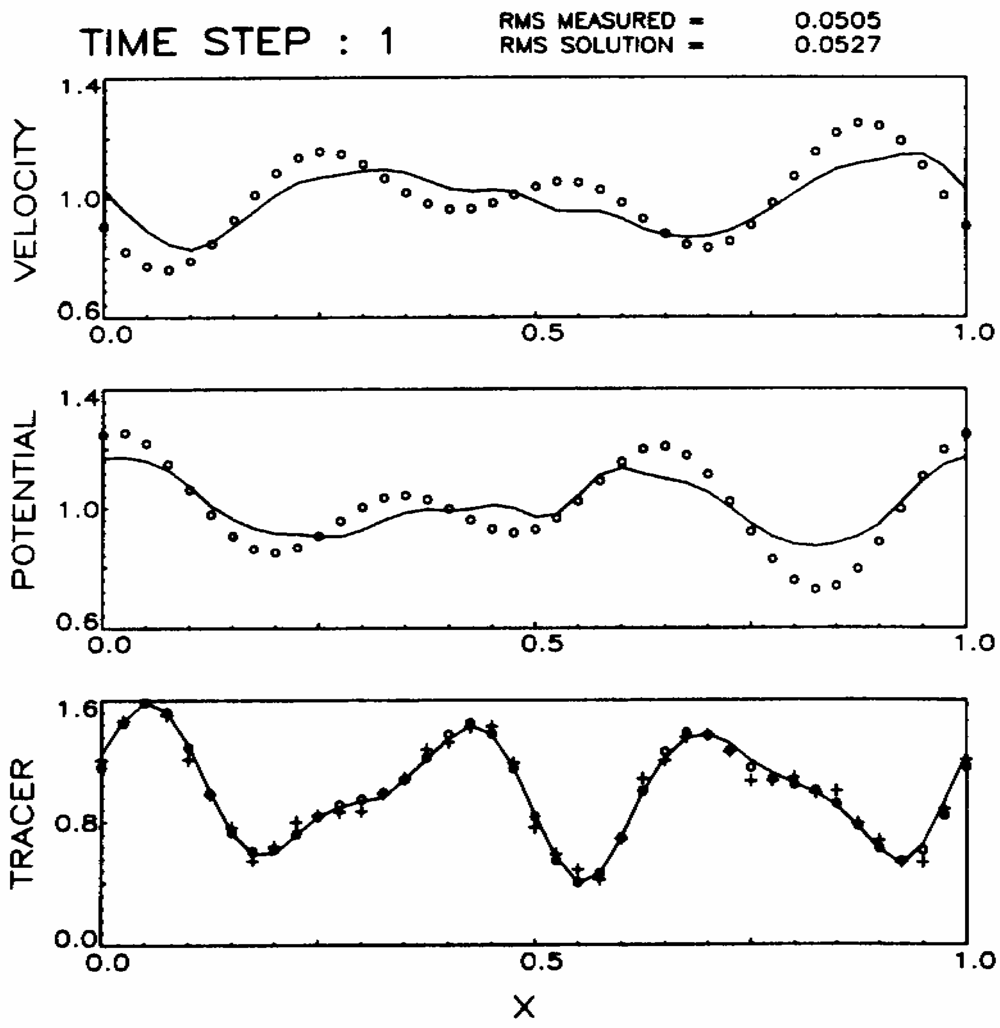
(however only for one specific cost function. A model forward integration perturbed in one component of u yields the sensitivity of any model variable and any cost function for this specific disturbance)

FIGURE 1: Solution by Conjugate Gradient Descent

Schematic:



inverting a tracer field for velocities



use of control variables

other than initial conditions

$$-\frac{\partial \lambda_T}{\partial t} - v\frac{\partial (v\lambda_T)}{\partial x} - A_T \frac{\partial^2 \lambda_T}{\partial x^2} + \varepsilon \lambda_T = -\frac{\partial j}{\partial T}$$

$$\nabla_u j = \frac{\partial j}{\partial u} + \int_X \int_{t_0}^{t_1} \lambda(x,t) \frac{\partial E(u,s,x,t)}{\partial u} \, dx \, dt$$

$$= - \int_X \int_{t_0}^{t_1} \lambda_T \frac{\partial^2 T}{\partial x^2} \, dx \, dt \qquad \text{for } u = A_T$$

use of control variables

other than initial conditions

$$-\frac{\partial \lambda_T}{\partial t} - v\frac{\partial (v\lambda_T)}{\partial x} - A_T\frac{\partial^2 \lambda_T}{\partial x^2} + \varepsilon\lambda_T = -\frac{\partial j}{\partial T}$$

$$\nabla_u j = \frac{\partial j}{\partial u} + \int_X \int_{t_0}^{t_1} \lambda(x,t)\frac{\partial E(u,s,x,t)}{\partial u}\,dx\,dt$$

$$= - \int_X \int_{t_0}^{t_1} \lambda_T\, T\,dx\,dt \qquad \text{for } u = \varepsilon$$

use of control variables

other than initial conditions

$$-\frac{\partial \lambda_T}{\partial t} - v\frac{\partial (v\lambda_T)}{\partial x} - A_T\frac{\partial^2 \lambda_T}{\partial x^2} + \varepsilon\lambda_T = -\frac{\partial j}{\partial T}$$

$$\nabla_u j = \frac{\partial j}{\partial u} + \int_X \int_{t_0}^{t_1} \lambda(x,t)\frac{\partial E(u,s,x,t)}{\partial u} dx\, dt$$

$$= - \int_X \int_{t_0}^{t_1} \lambda_T\, T\, dx\, dt \qquad \text{for } u = \varepsilon$$

TASK: Ocean state during the 1993-2001

- Model: mass conserving *(2° x 2°, 23 layers)*
- Nine years (1993-2001) T/P altimeter referenced to GRACE + Reynolds surface temperatures + oceanic measurements are assimilated into the model
- Method: 4D VAR data assimilation
- As control parameters we use the model initial state and the model forcing (the first guess taken from NCEP)

cost function for global model

$$J = J_{misfit} + W_{cycle} J_{cycle} + W_{bogus} J_{bogus}$$

$$+ W_{SSH} J_{SSH} + W_{hmv} J_{hmv} + W_{atl} J_{atl}$$

$$J_{misfit} = \sum \frac{R_{i,j,k}}{\sigma^2(T)_k} (T_{i,j,k,n} - T^*_{i,j,k,n})^2$$

$$+ \sum \frac{R_{i,j,k}}{\sigma^2(S)_k} (S_{i,j,k,n} - S^*_{i,j,k,n})^2$$

cost function for global model

$$for\ \vec{x} = \{T, S, u, v\}$$

$$J_{cycle} = \sum \frac{1}{\sigma^2 (\Delta \vec{x})_k} (\vec{x}_{i,j,k,n} - \vec{x}_{i,j,k,n-ntyear})^2$$

$$J_{SSH} = \sum \frac{1}{\sigma^2{}_{SSM}} (<\zeta> - <\zeta^*>)^2$$

$$+ \sum \frac{1}{\sigma^2{}_{SSA}} [\zeta - <\zeta> - (\zeta^* - <\zeta^*>)]^2$$

cost function for global model

$$J_{\mathrm{hmw}} = \sum_{sec\,\mathrm{t}} \frac{1}{\sigma^2(\mathrm{T})} (< T > - T^*)^2$$

T denotes transports

$$J_{\mathrm{atl}} = \sum_{sec\,\mathrm{t}} \frac{1}{\sigma^2(\mathrm{atl})} (\Psi_{\mathrm{m}} - \Psi_{\mathrm{m}}^*)^2$$

Ψ_{m} denotes meridional streamfunction

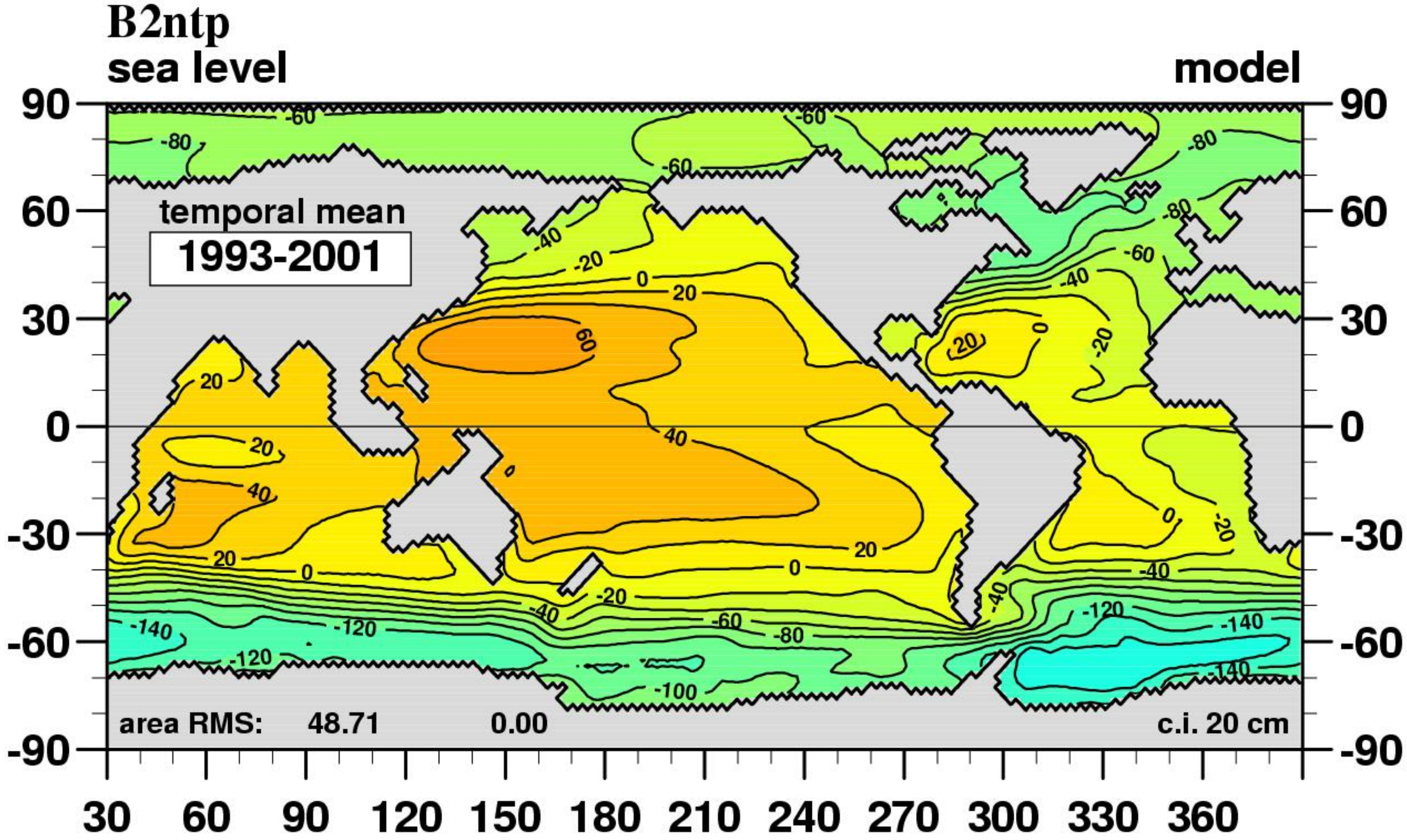
cost function for global model

$$J = J_{misfit} + W_{cycle} J_{cycle} + W_{bogus} J_{bogus}$$
$$+ W_{SSH} J_{SSH} + W_{hmv} J_{hmv} + W_{atl} J_{atl}$$

data are SST, SSH, T,S

section analysis, model free run

Mean sea level



local sea level changes due to:

$$\frac{\partial}{\partial t}\zeta = P - E$$

$$+ \quad \nabla . \int_{-H}^{\zeta} \vec{v}\, dz$$

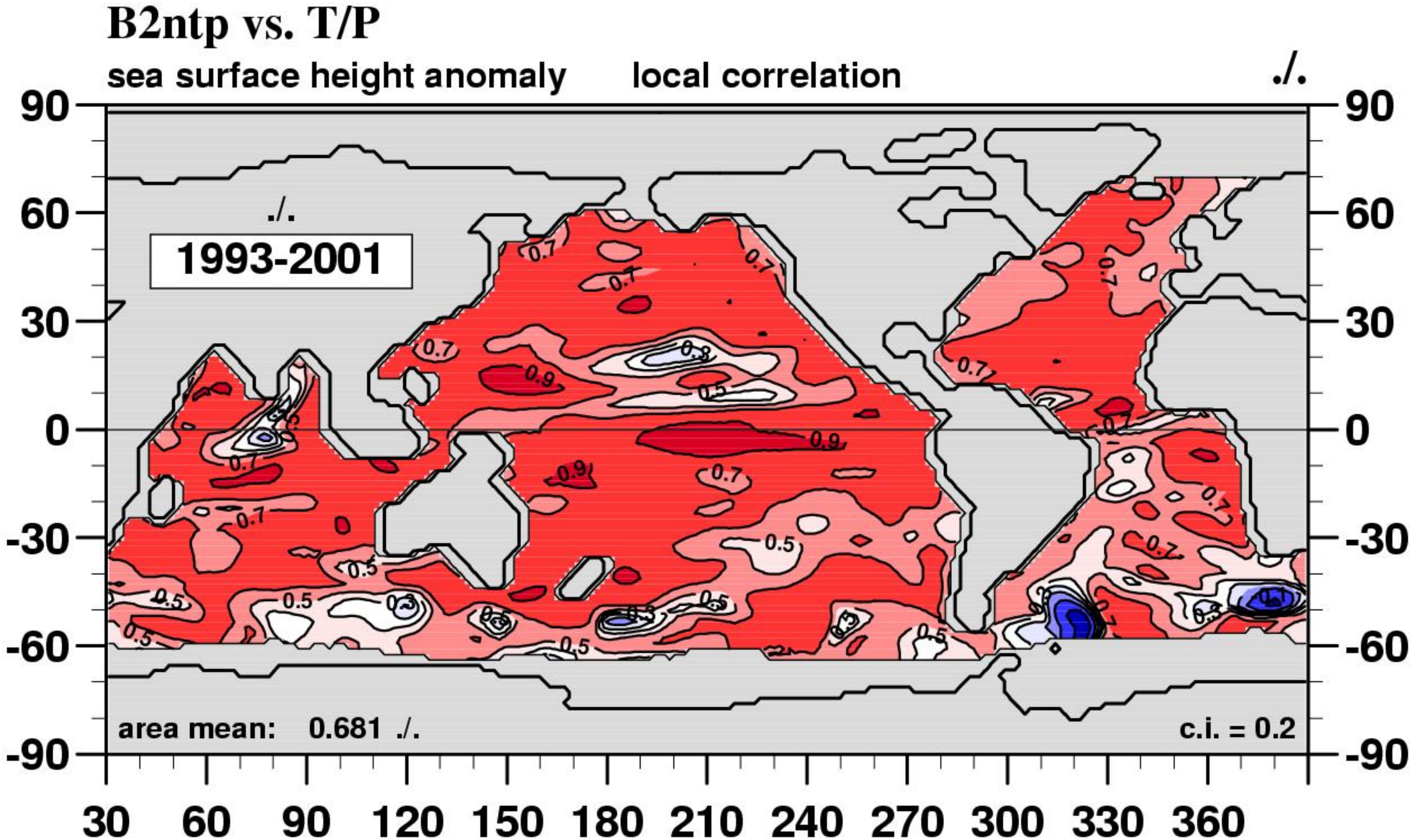
$$+ \quad A_h \Delta\zeta$$

$$+ \quad \int_{-H}^{\zeta} \frac{1}{\alpha} \frac{\partial \alpha}{\partial T}\bigg|_{S,p} \frac{\partial}{\partial t} T\, dz$$

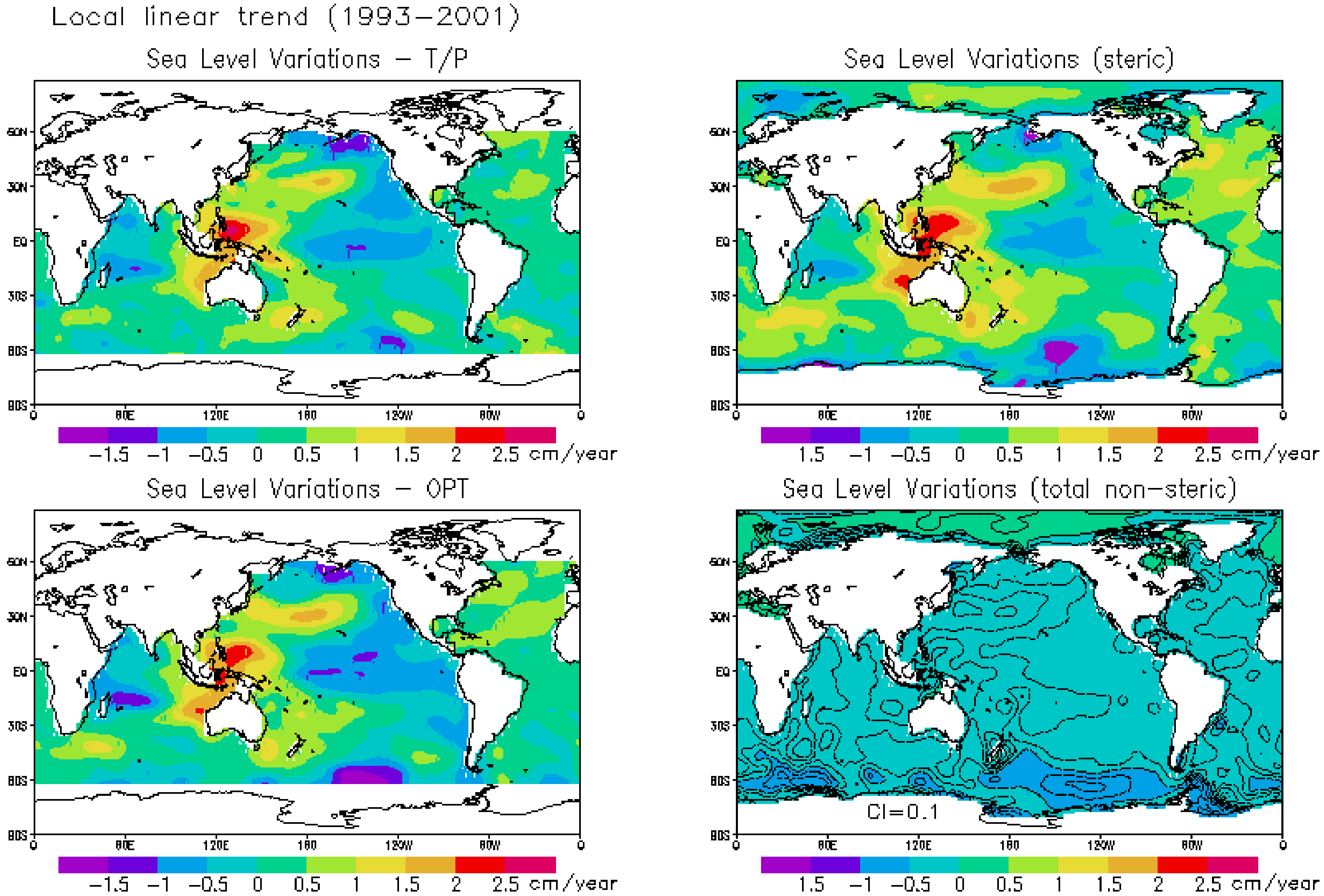
$$+ \quad \int_{-H}^{\zeta} \frac{1}{\alpha} \frac{\partial \alpha}{\partial S}\bigg|_{T,p} \frac{\partial}{\partial t} S\, dz$$

- freshwater flux
- divergence
- sub grid gravity waves
- thermosteric
- halosteric

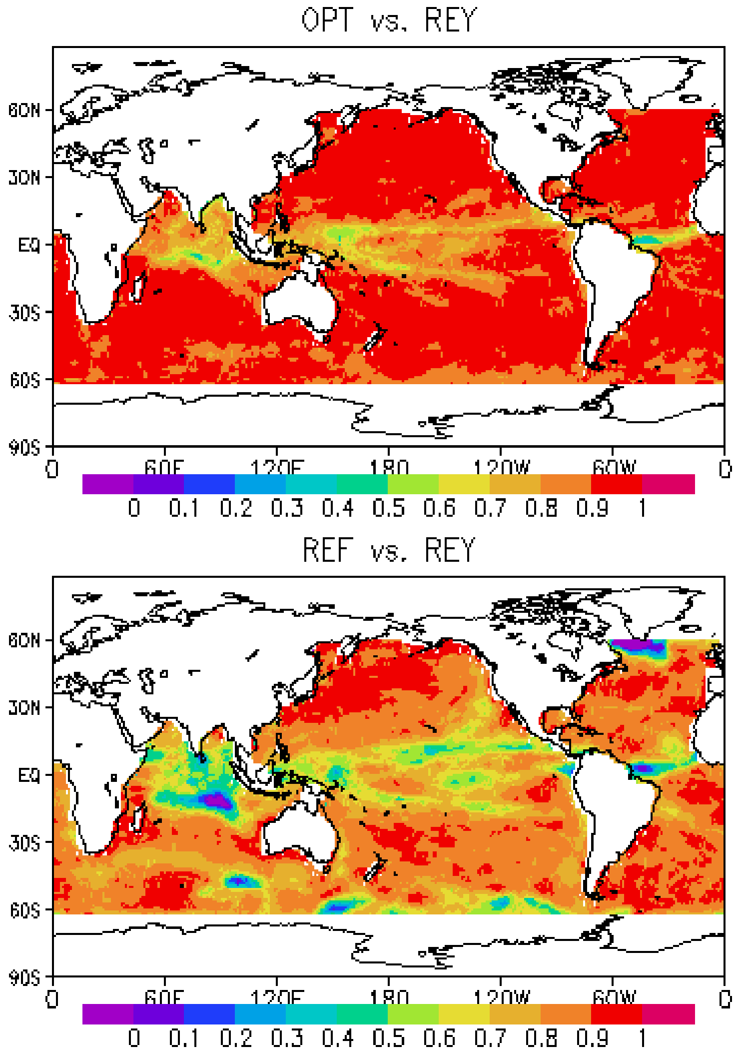
SLA correlation



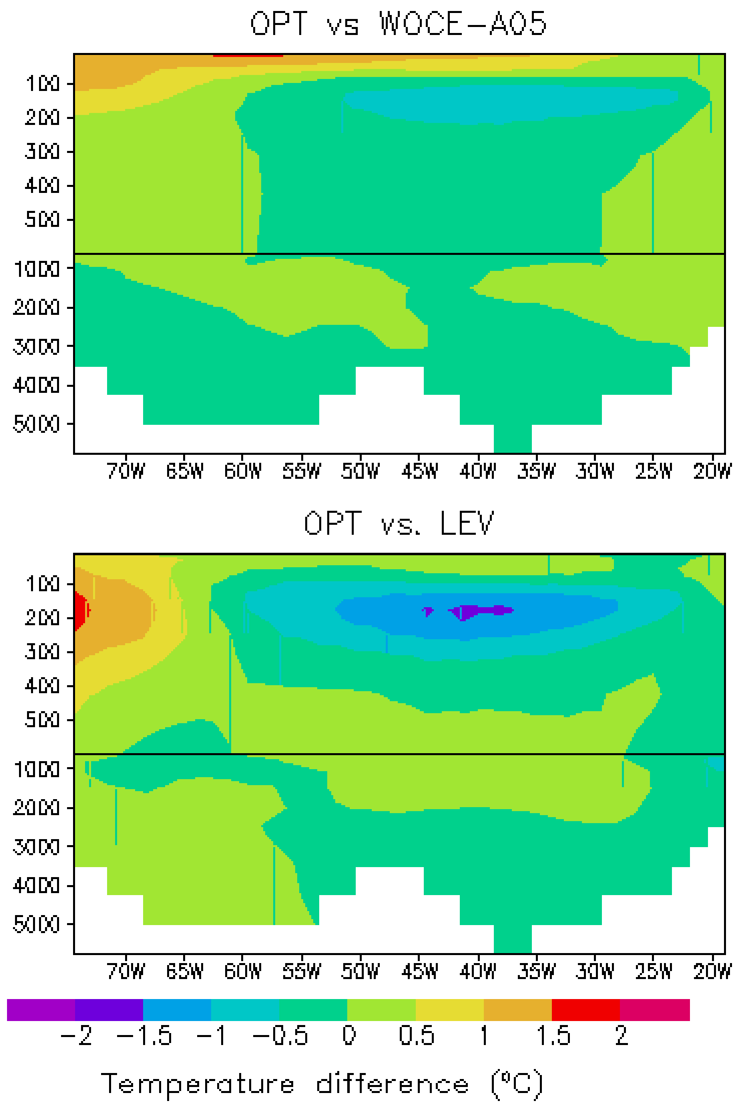
Sea level trends



Correlation between Reynolds and model SST



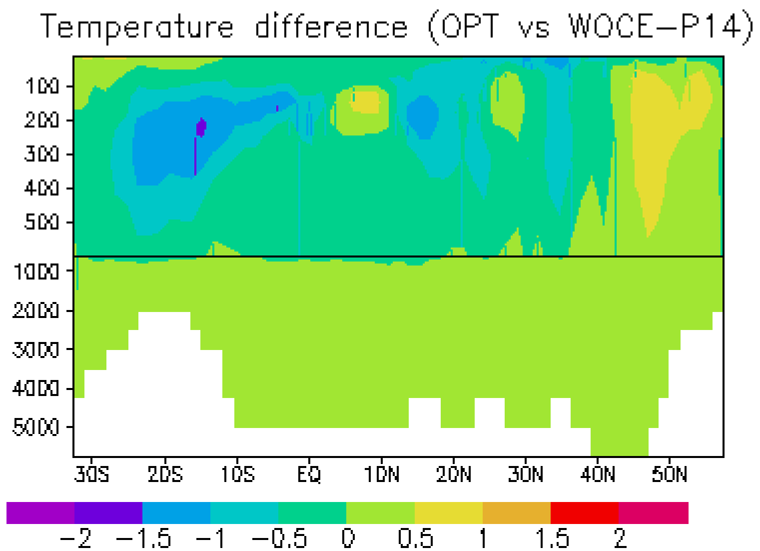
Temperature Difference - Atlantic Section (24.5 °N)

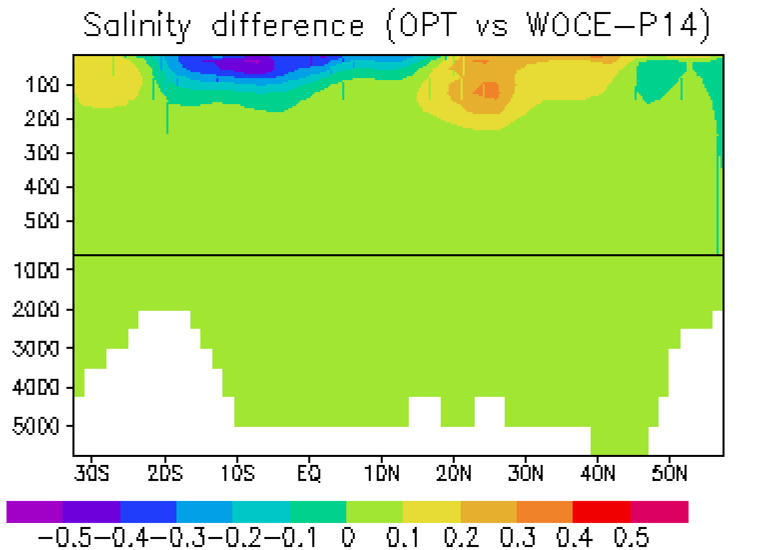


After assimilation

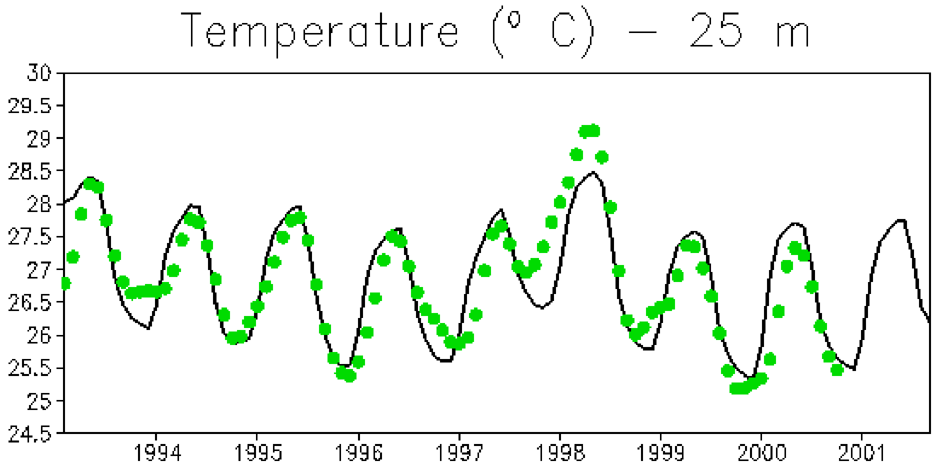
Before assimilation

OPT vs. WOCE- Pacific Section (179 °W)

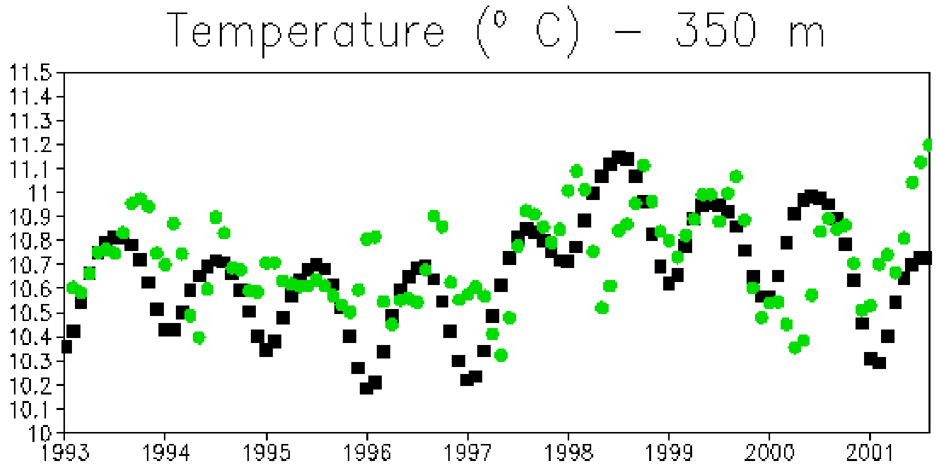




Temperature (OPT vs. TAO Data)

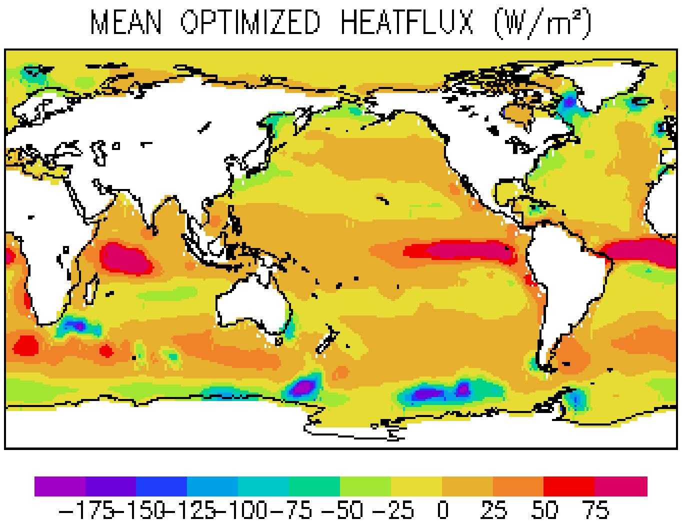


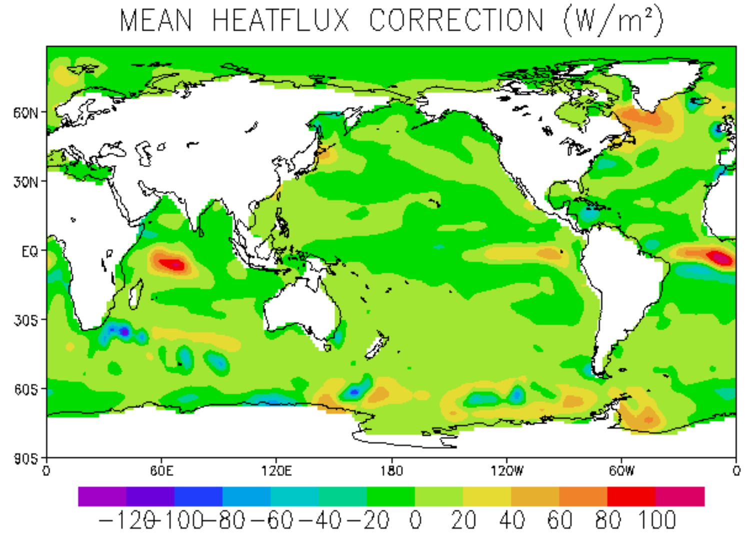
(235°, –8°)

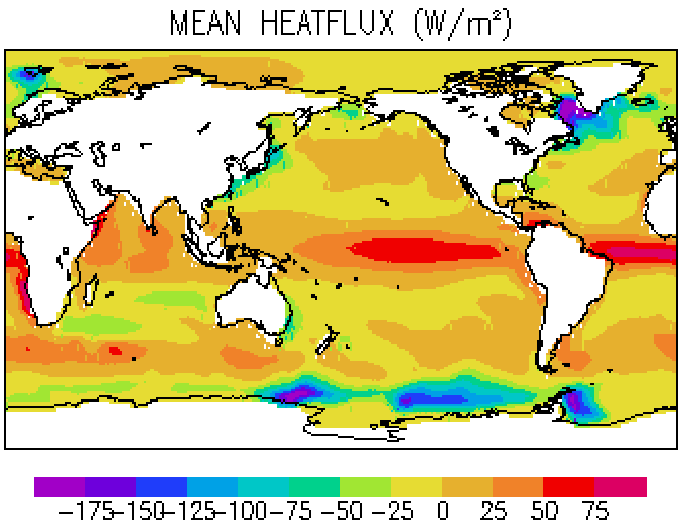


(235°, –8°)

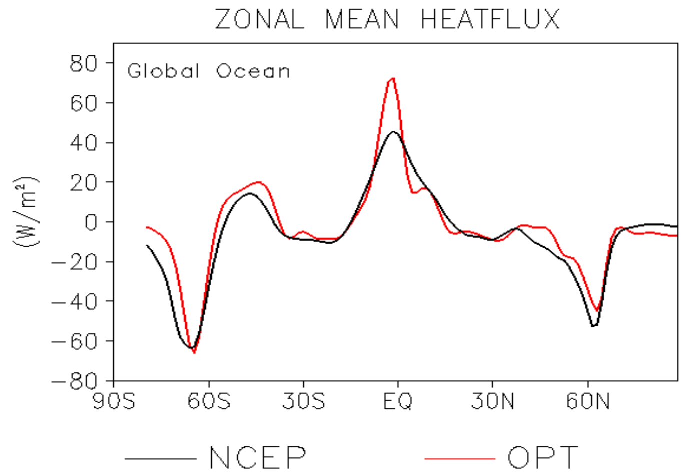
Heat flux



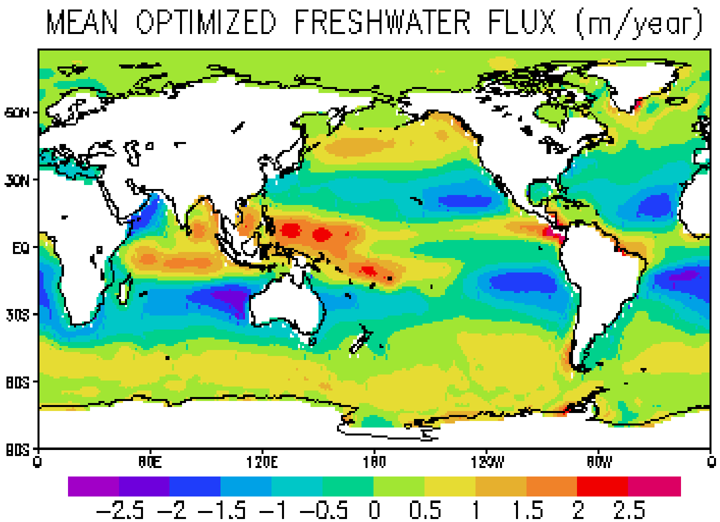


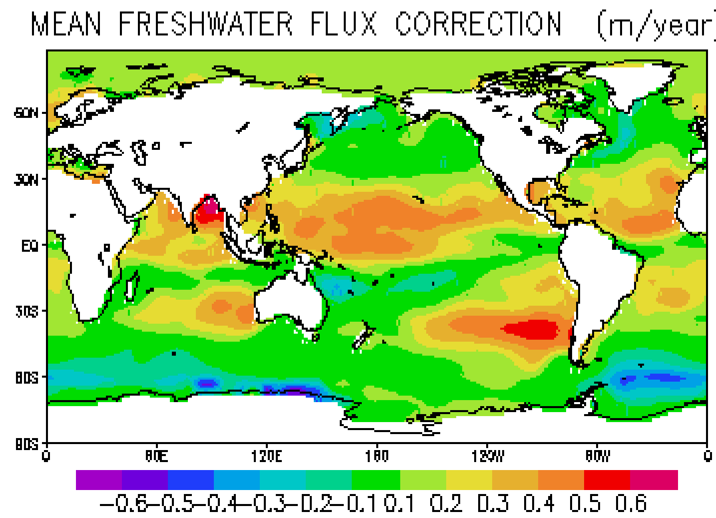


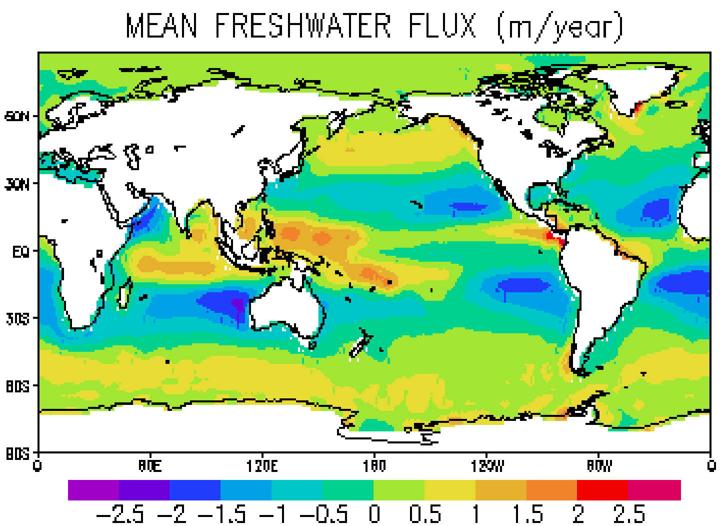
milation



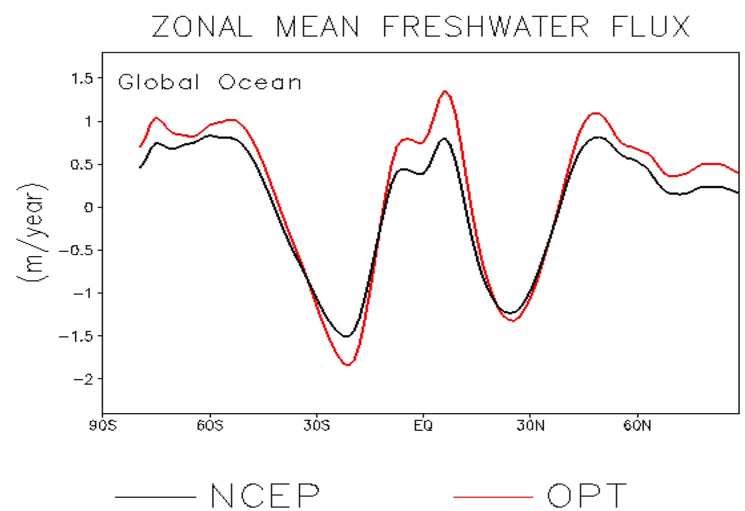
Freshwater flux



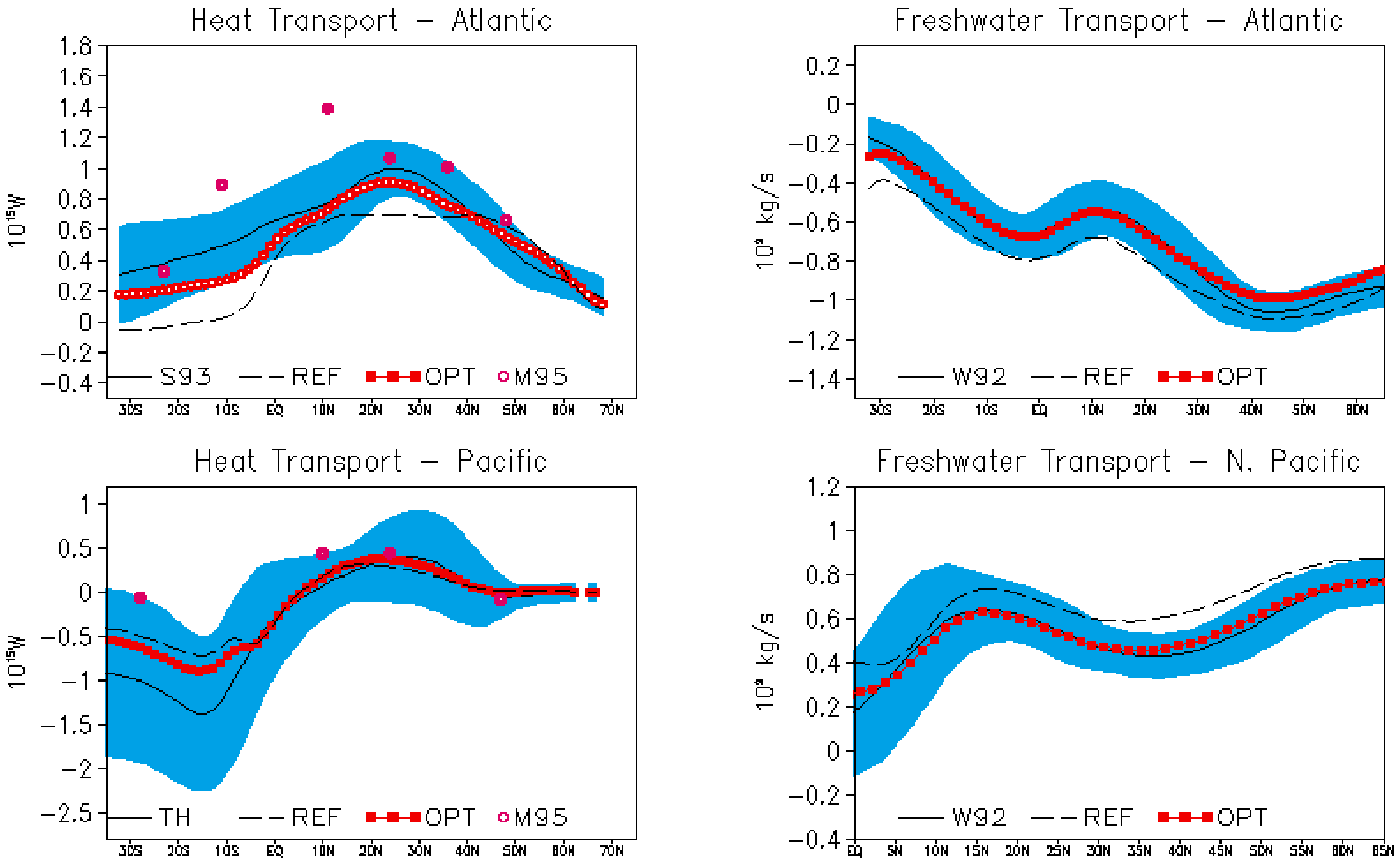




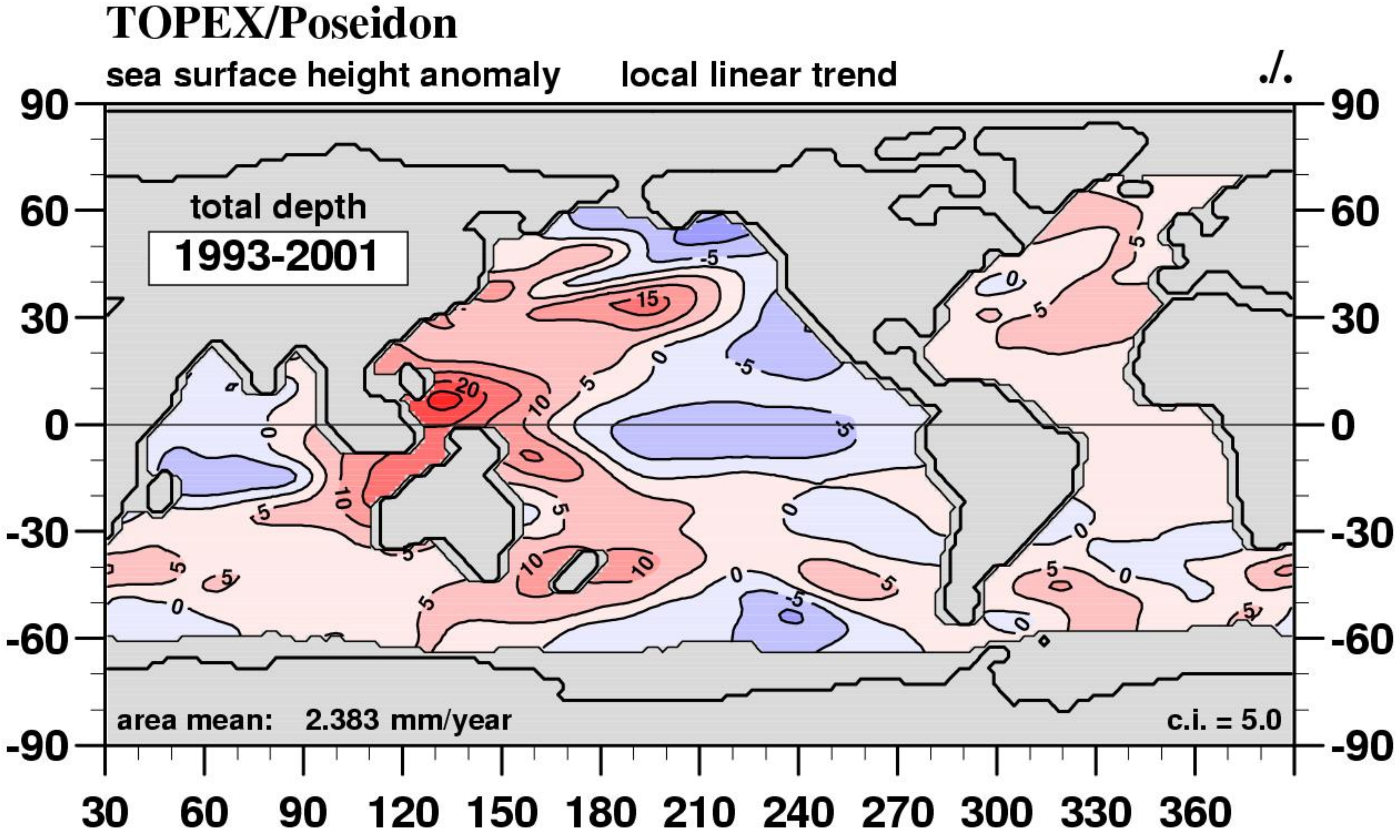
ssimilat



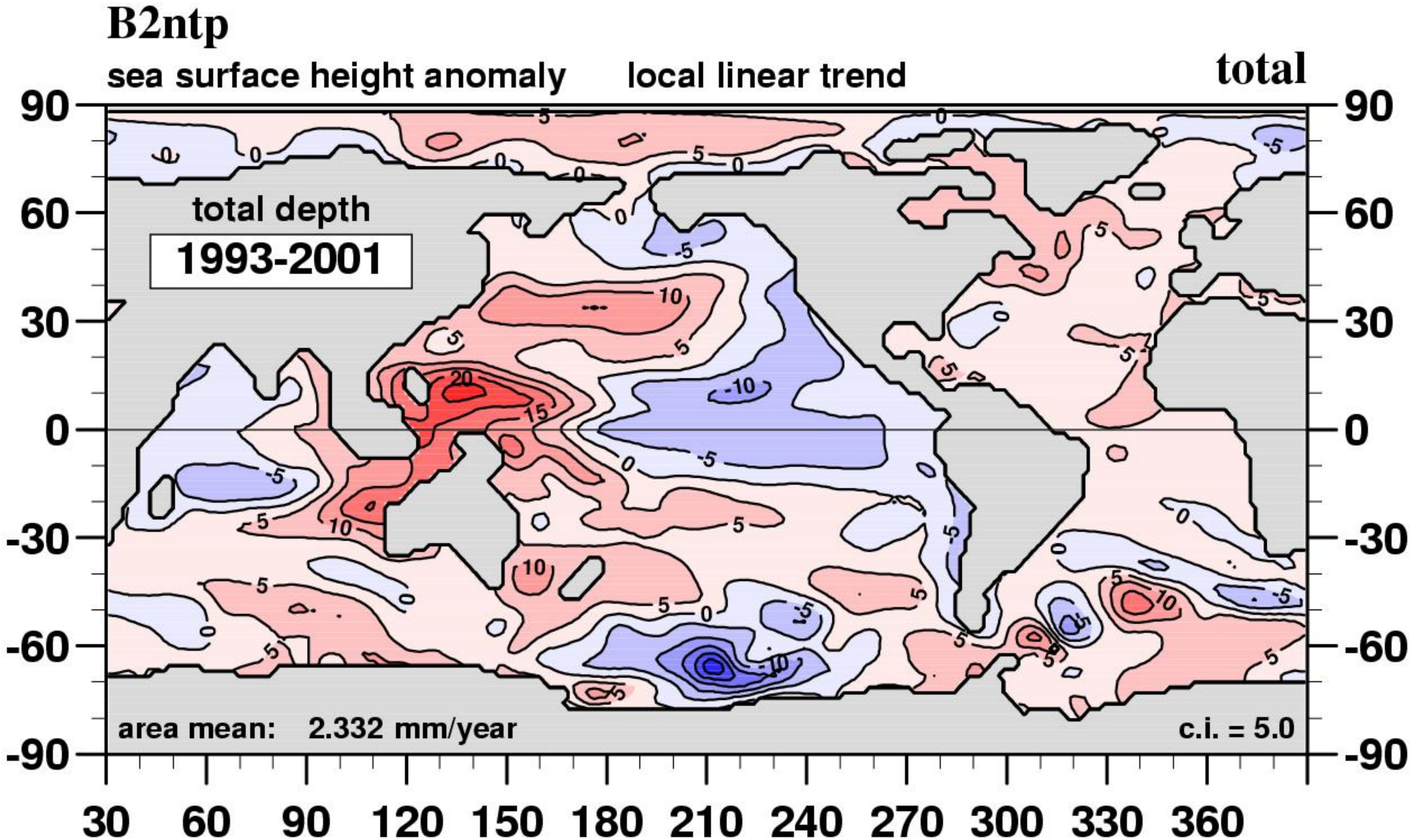
Heat and freshwater transport



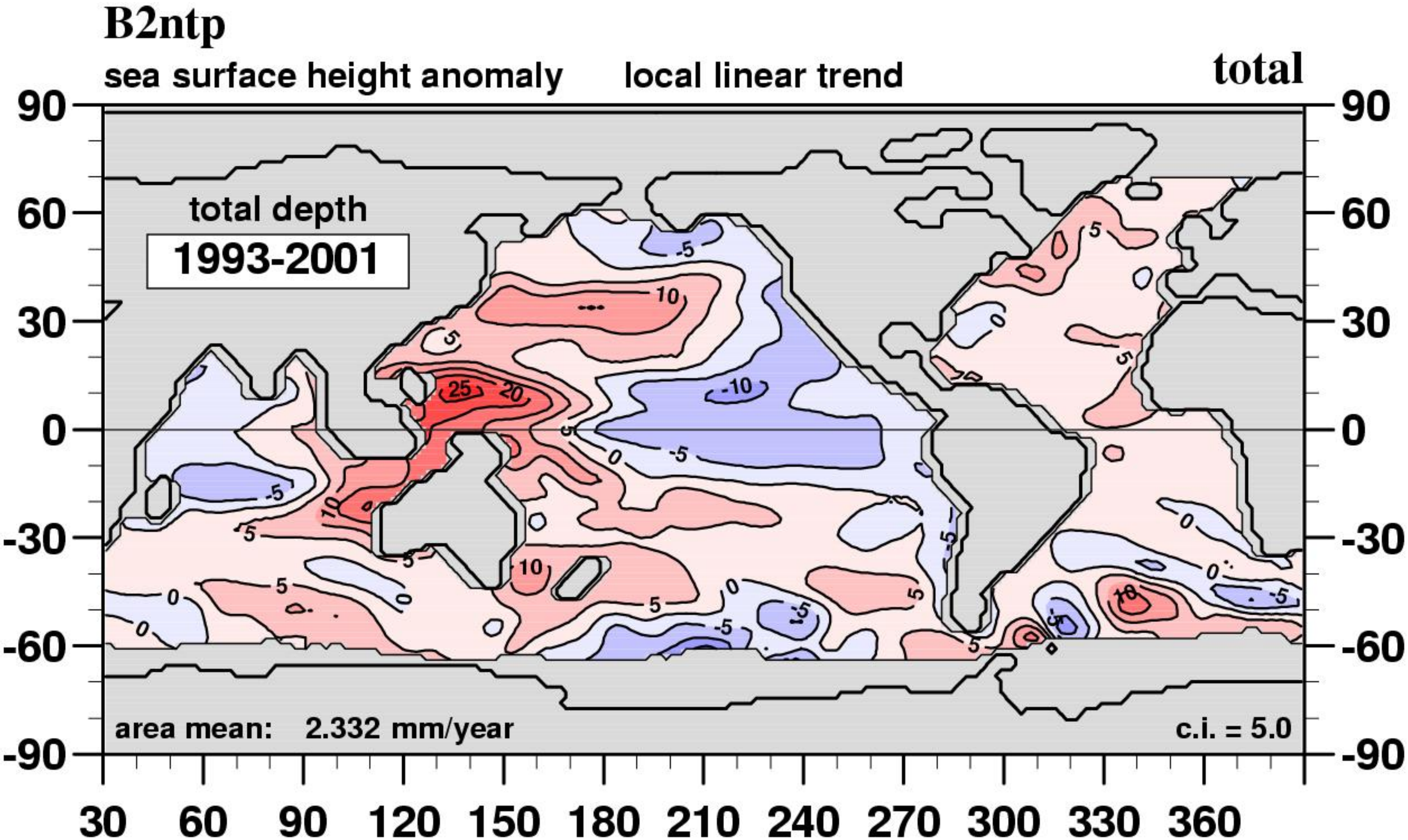
Observed sea level trend

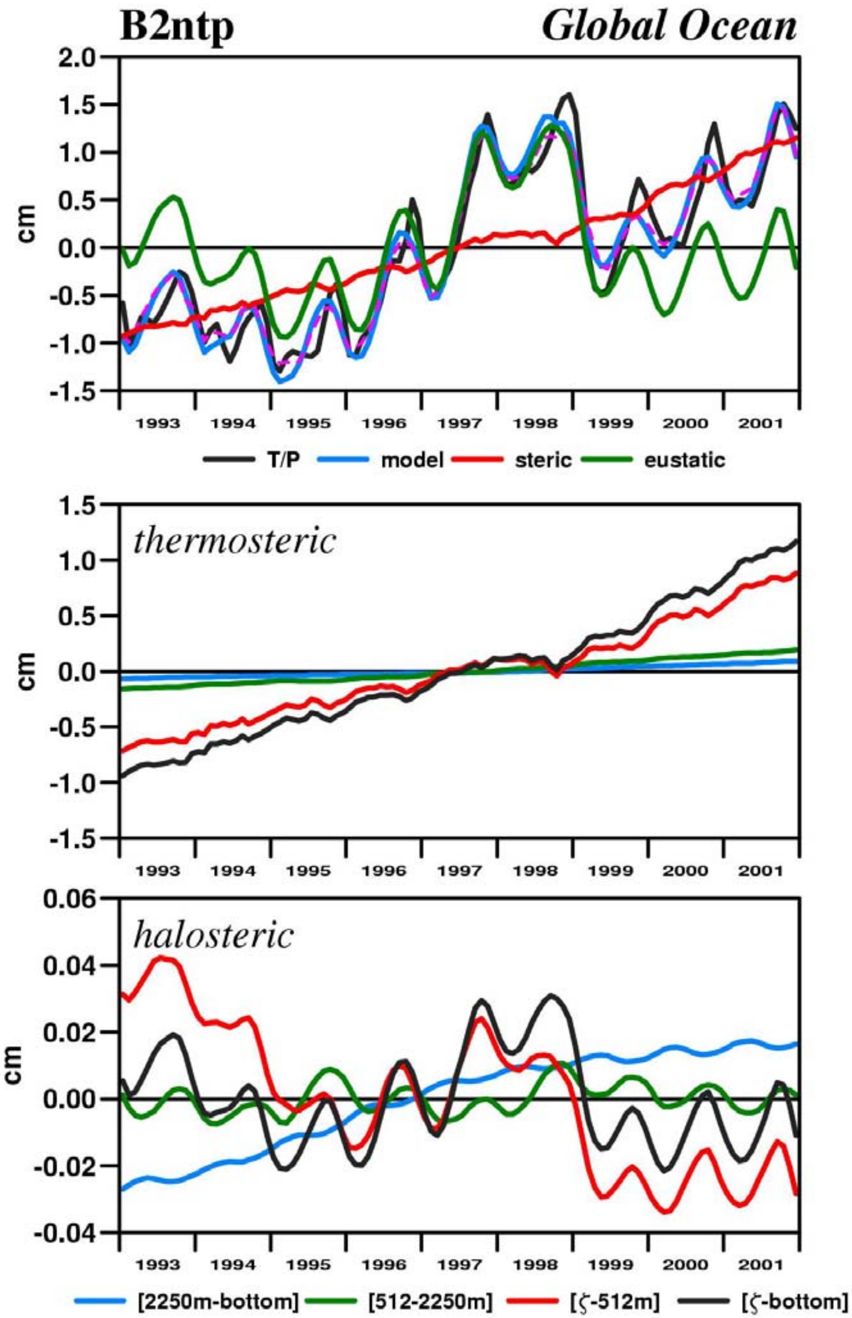


modelled sea level trend



modelled sea level trend



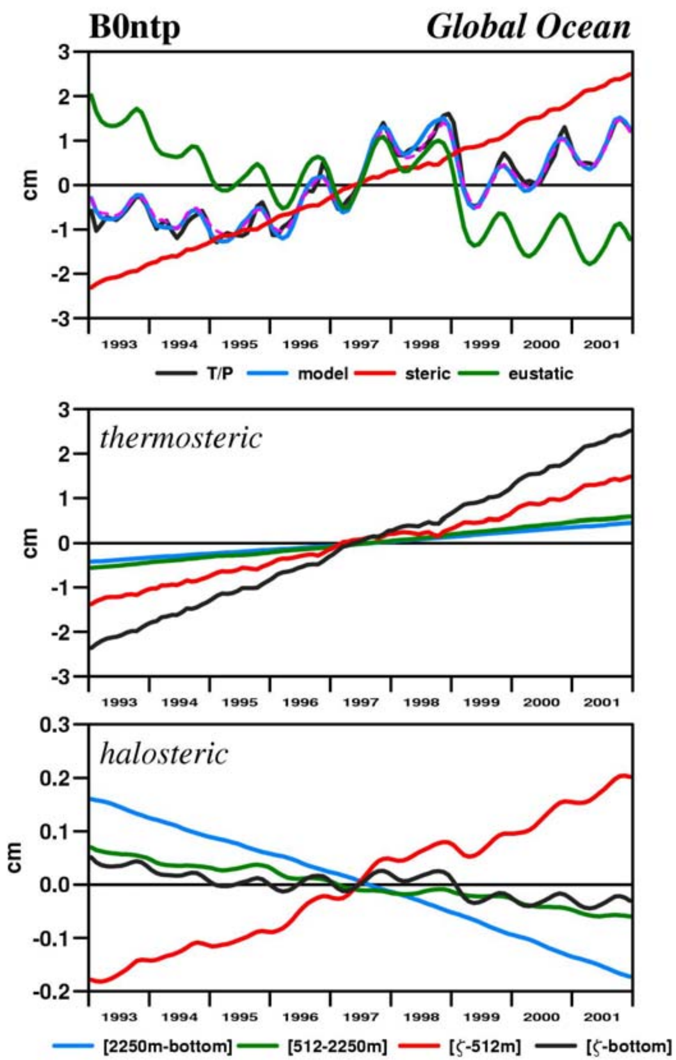


Global Mean Sea Level

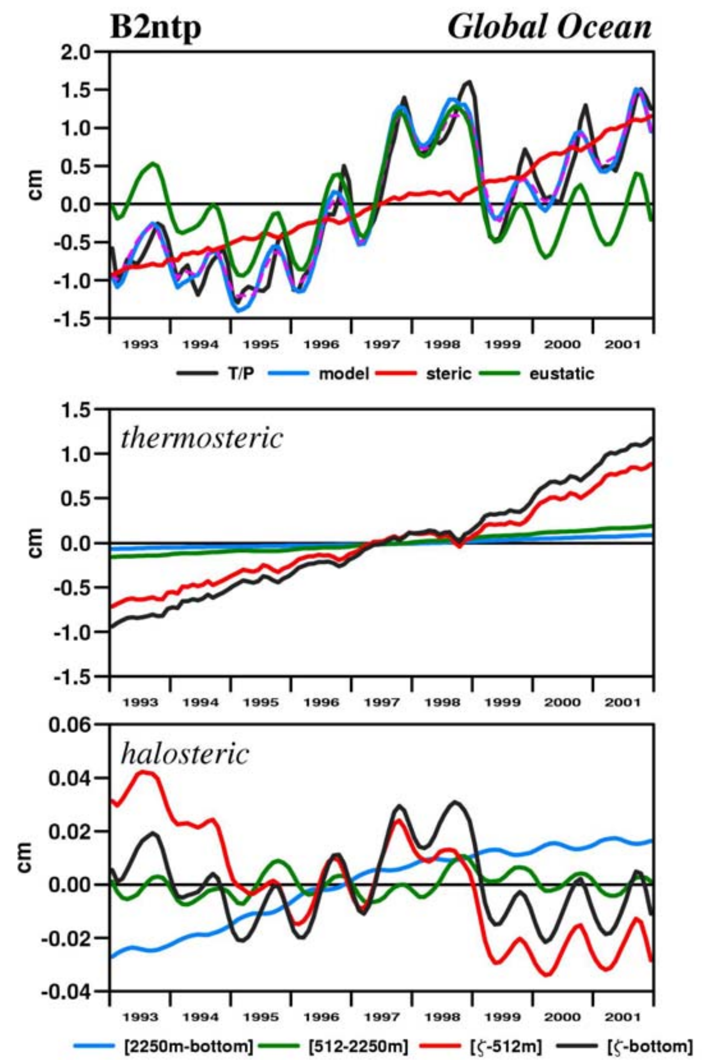
sea level rise is explained by thermal expansion

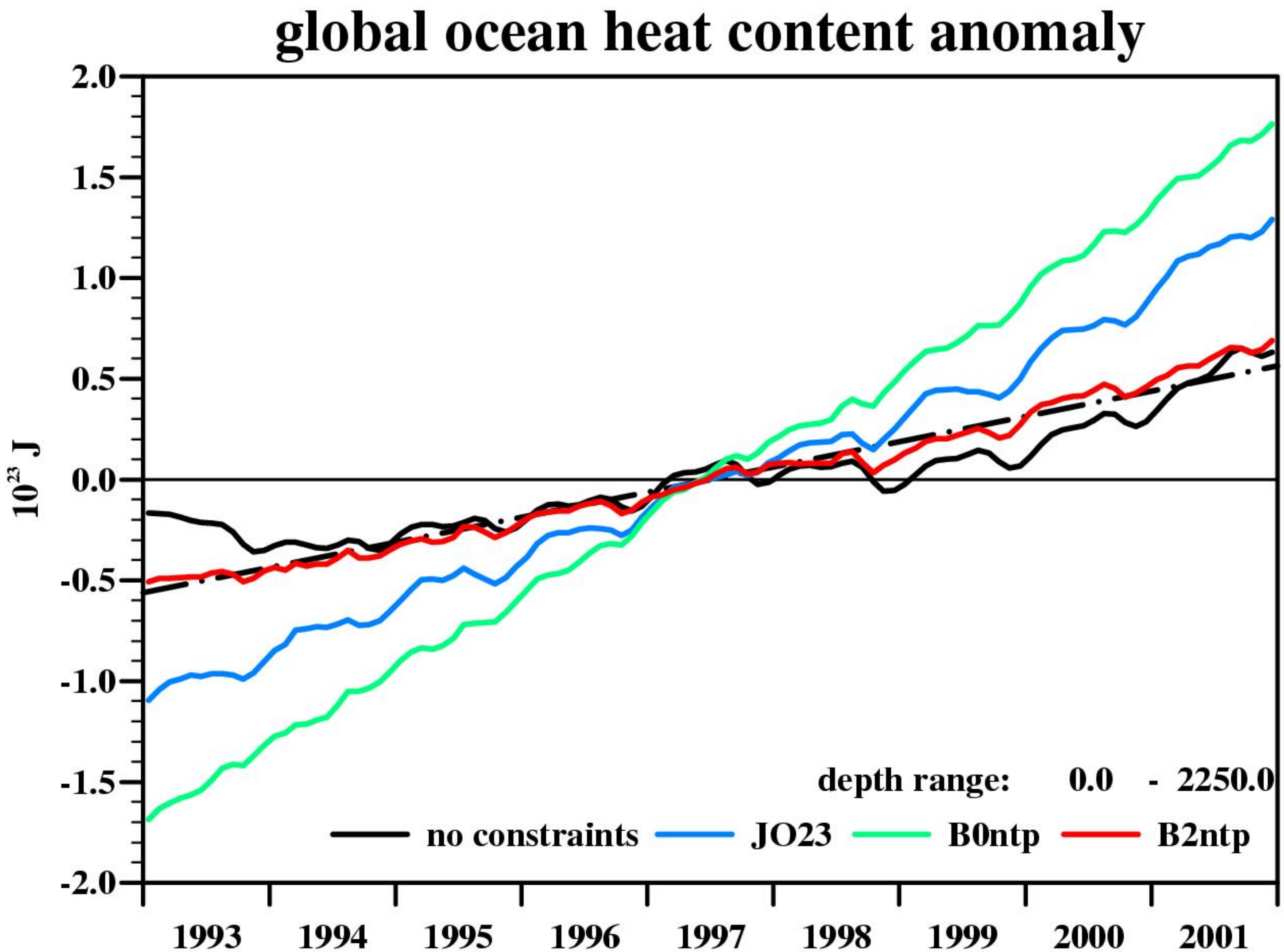
interannual variability and seasonal cycle are mostly eustatic

Global Mean Sea Level



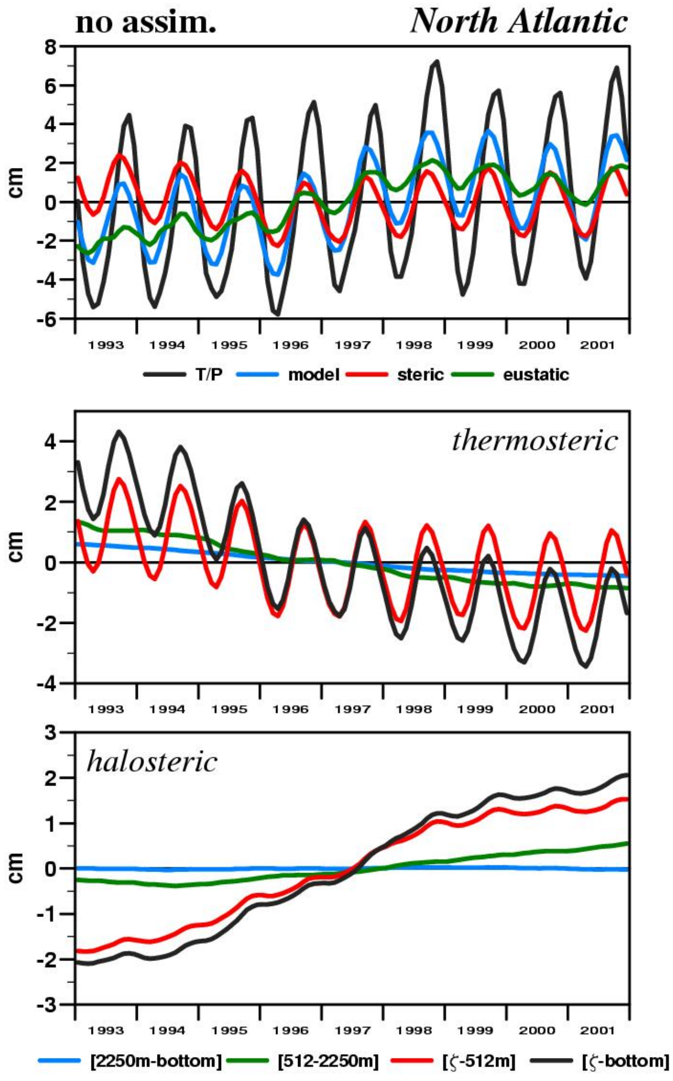
assimilat

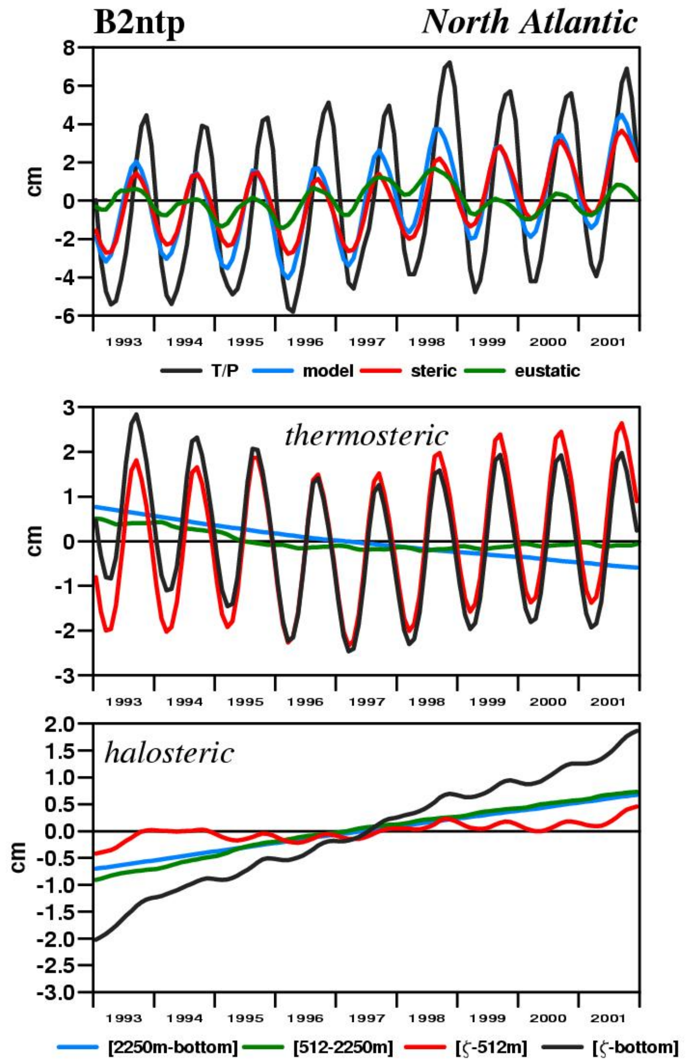




global ocean heat content anomaly
2.0
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
-2.0
10²³ J
depth range: 0.0 - 2250.0
no constraints
JO23
B0ntp
B2ntp
1993
1994
1995
1996
1997
1998
1999
2000
2001

Mean Sea Level North Atlantic

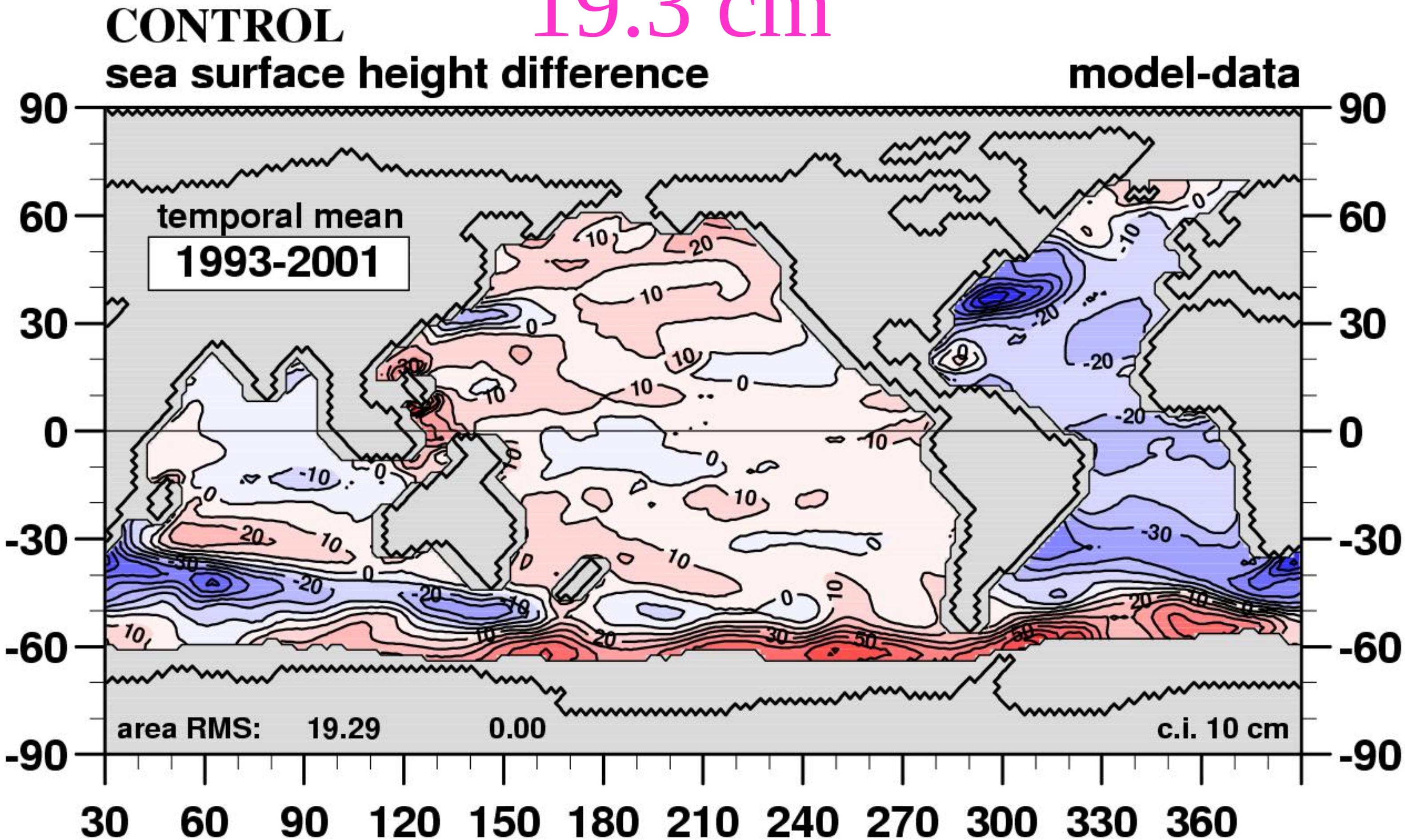




assimilat

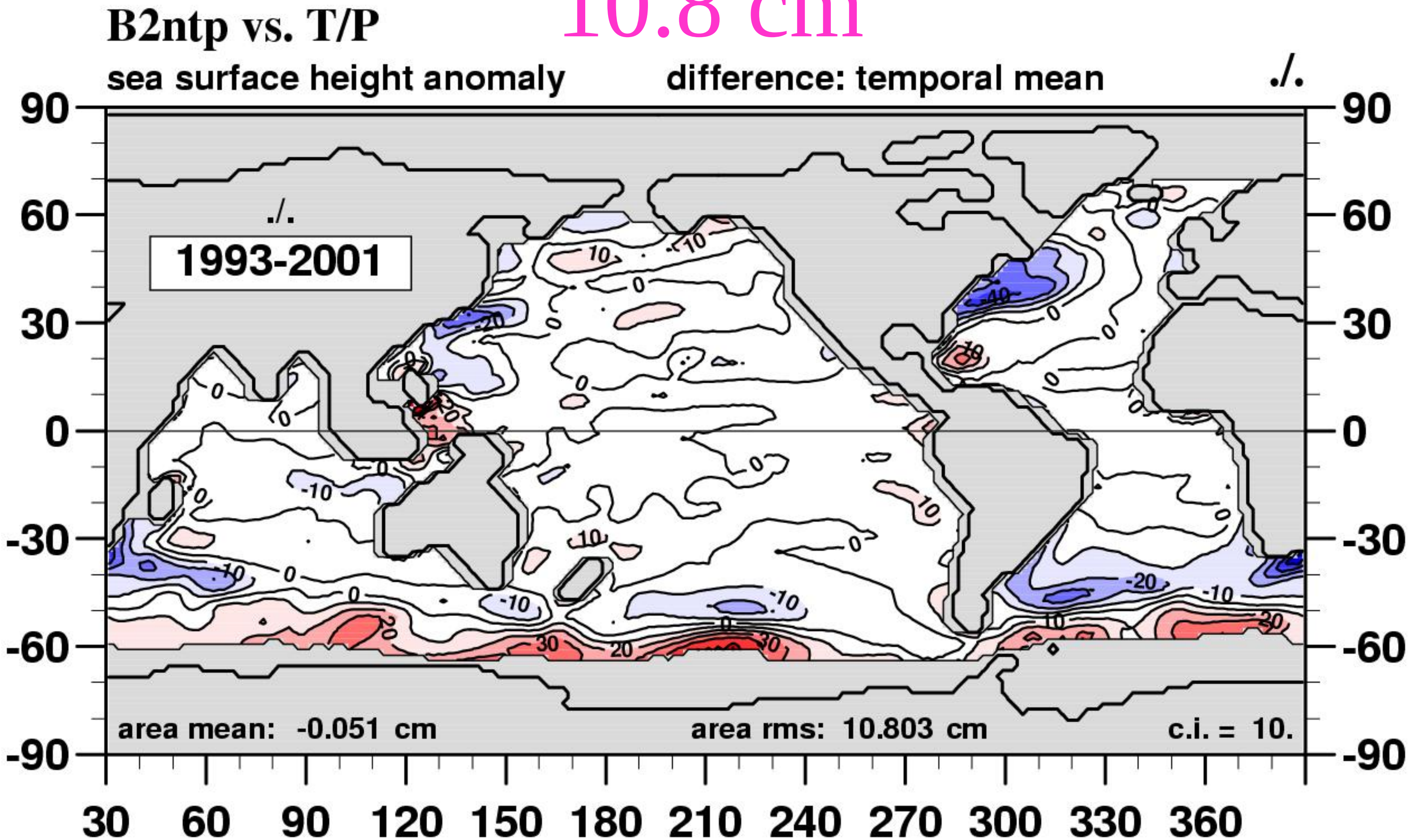
MSL difference (no assimilation)

19.3 cm



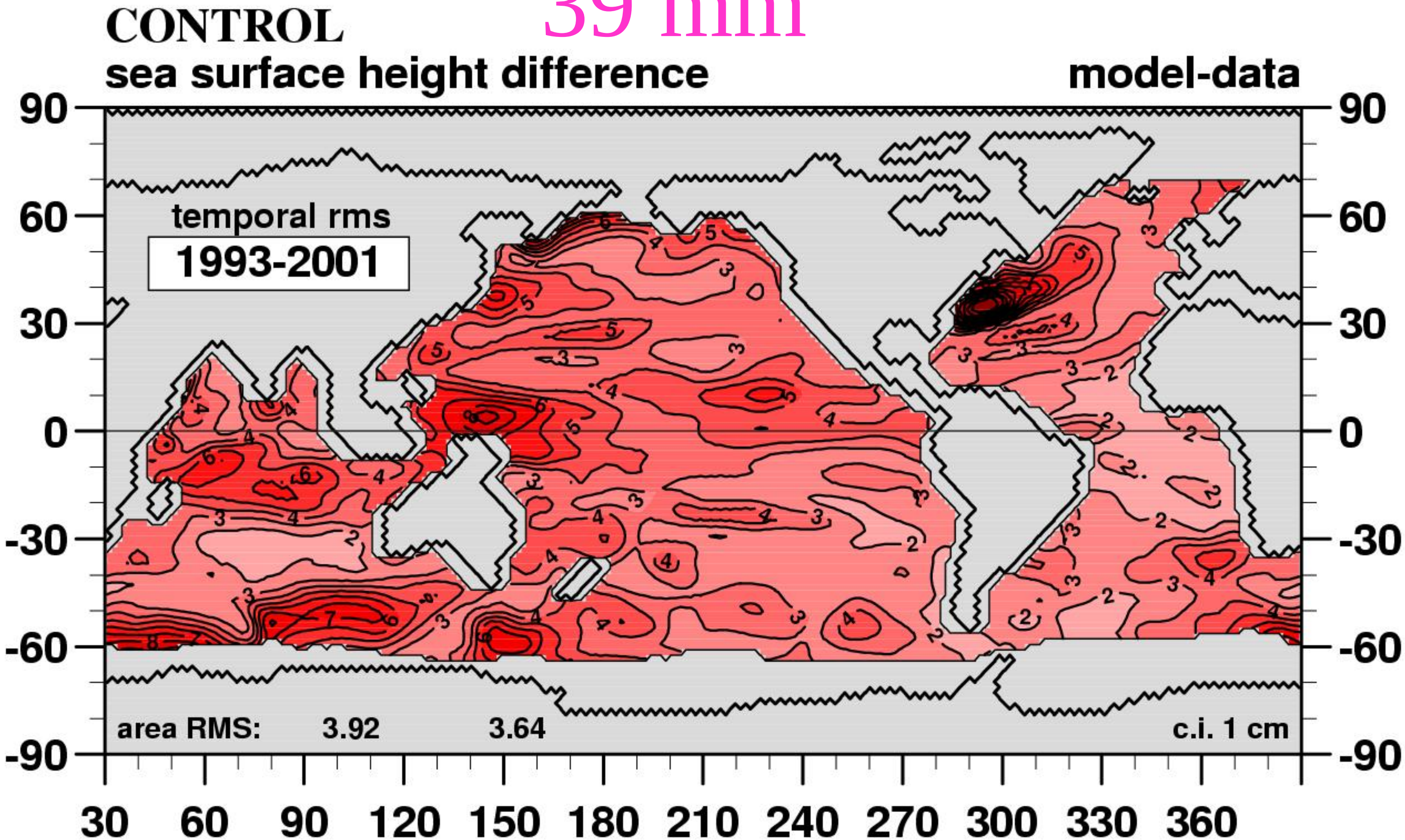
MSL difference (with GRACE)

10.8 cm



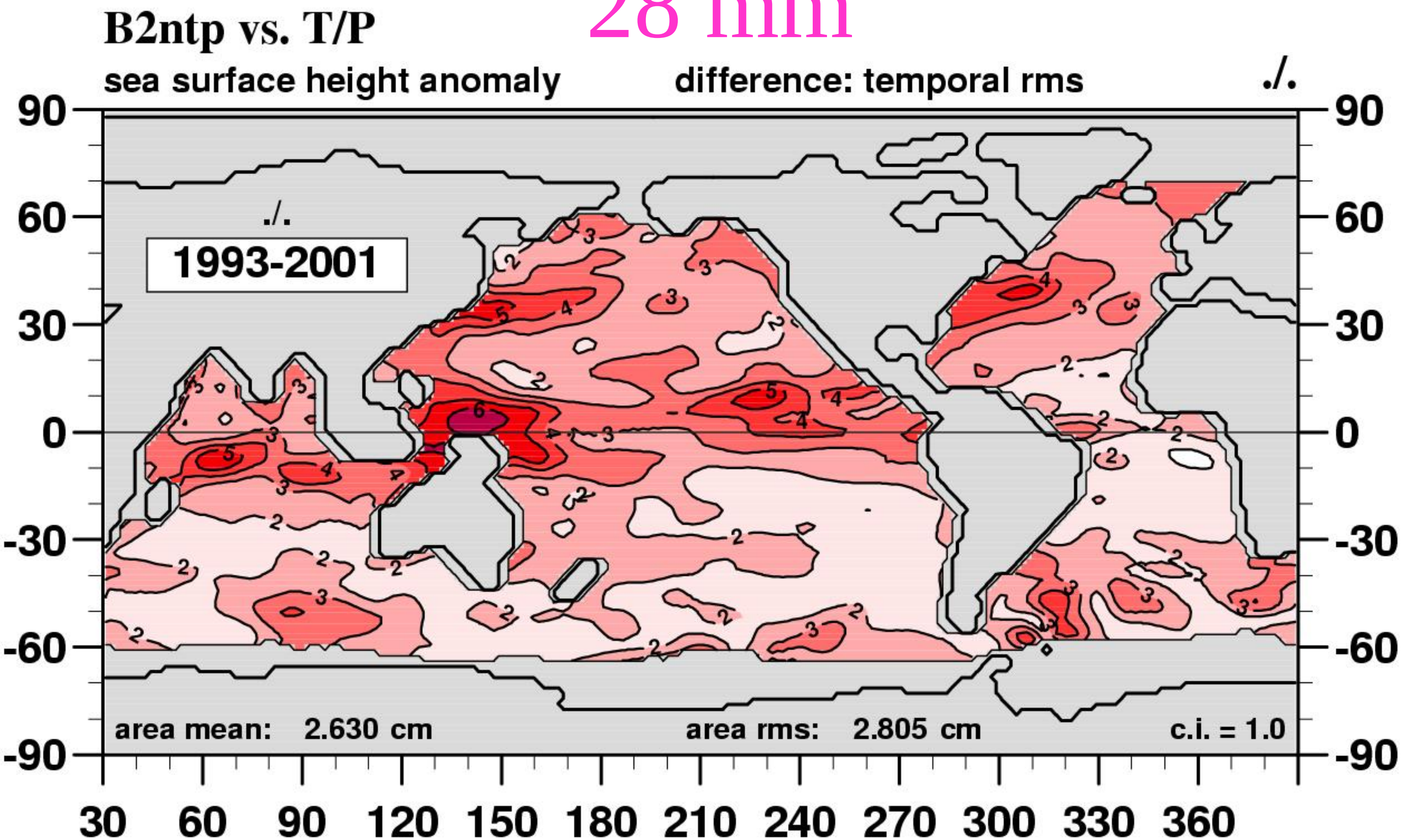
SLA difference (no assimilation)

39 mm

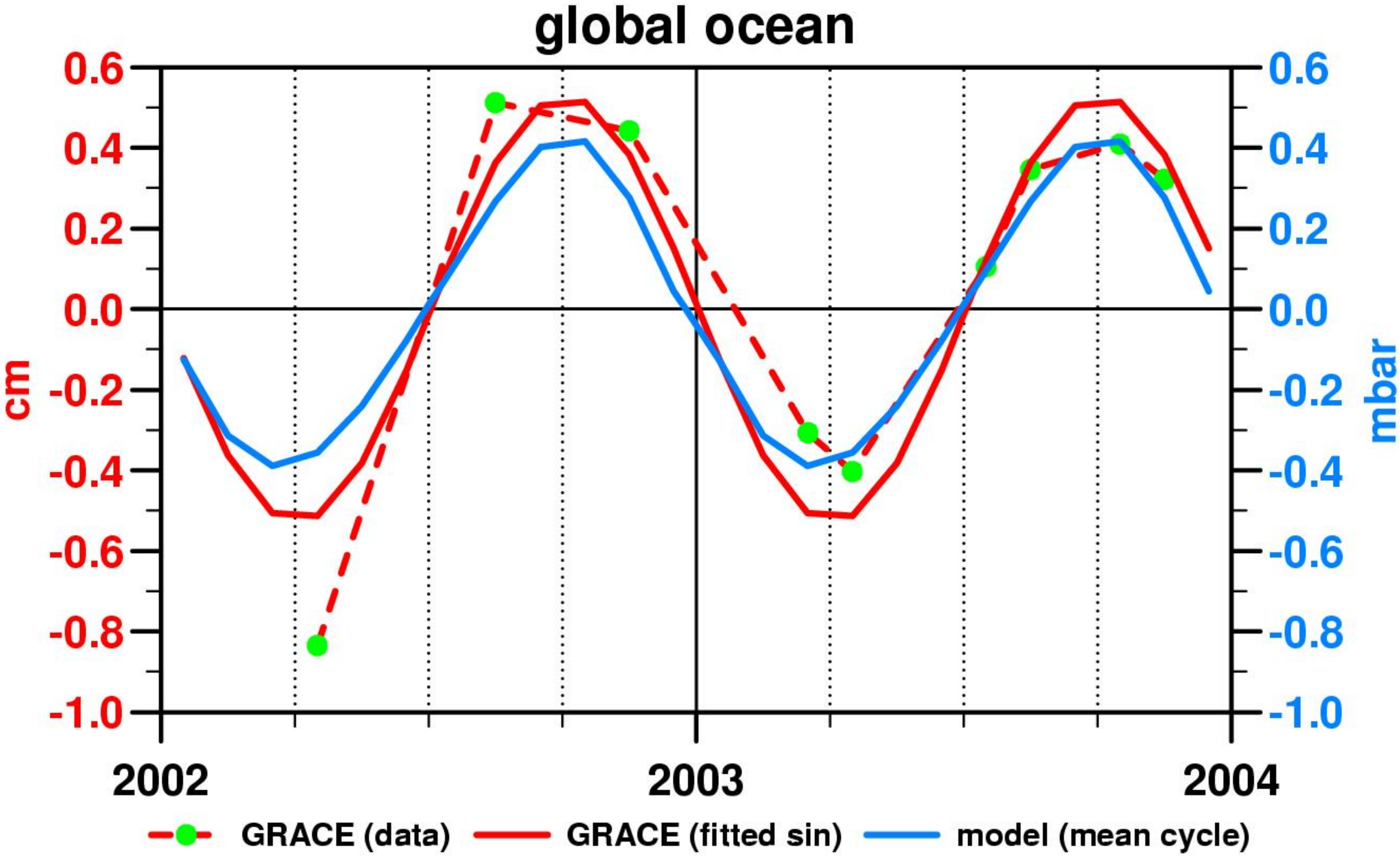


SLA difference (with GRACE)

28 mm



Global Ocean Mass



THE END