Optimal Spectral Decomposition (OSD) for Ocean Data Assimilation

Peter C. Chu
Naval Postgraduate School
Monterey, CA 93943
pcchu@ns.edu
http://faculty.nps.edu/pcchu

Naval Research Laboratory, Stennis Space Center, Mississippi, 3 April 2014.

I would like to present my preliminary thoughts.

Your comments and critics are highly appreciated.

Outline

- (1) Introduction
- (2) Theoretical Base
- (3) Ocean Modeling (POP Model)
- (4) Twin Experiment
- (5) Error Statistics
- (6) Conclusions

(1) Introduction

Major Ocean Data Assimilation Schemes

- Optimal Interpolation (OI)
- Kalman Filter
- Variational Methods

Assimilated Variable $\mathbf{c}(\mathbf{x}, z, t)$

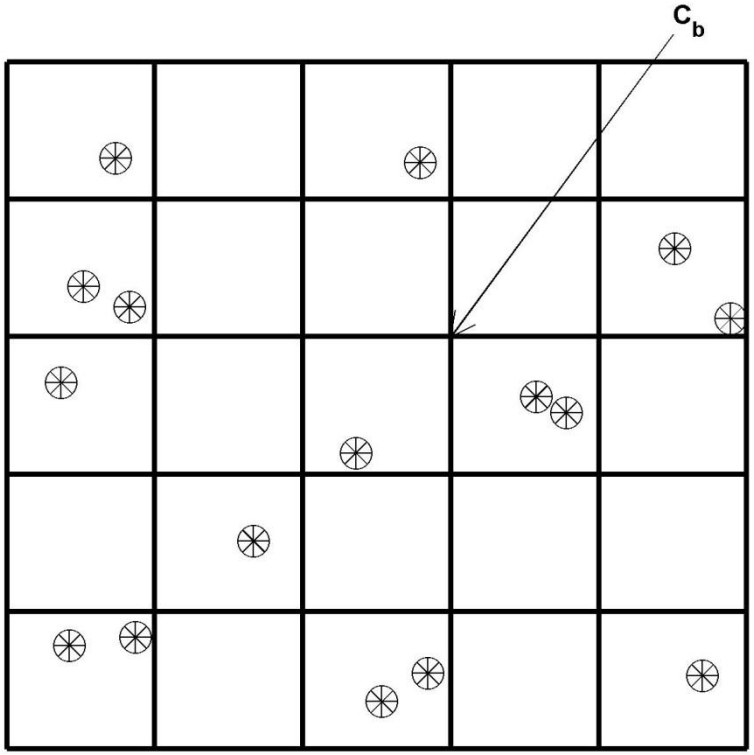
$\mathbf{c}_t \rightarrow$ "true"

$\mathbf{c}_a \rightarrow$ analysis (assimilated)

$\mathbf{c}_b \rightarrow$ background (modeled)

$(\mathbf{c}_t, \mathbf{c}_a, \mathbf{c}_b) \rightarrow$ grid points

$\mathbf{c}_o \rightarrow$ Observation $\rightarrow$ Obs points



Ocean data assimilation with $\boldsymbol{c}_b$ located at the grid points, and $\boldsymbol{c}_o$ located at the points '*'. The ocean data assimilation is to convert the innovation, $\boldsymbol{d} = \boldsymbol{c}_o - \boldsymbol{H}(\boldsymbol{c}_b)$, from the observational points to the grid points.

Data Assimilation

→ Conducted in the Physical Space (**i**, **j**, **k**)

$$\mathbf{c}_a = \mathbf{c}_b + \mathbf{W}\mathbf{d},$$

Innovation → $\mathbf{d} = \mathbf{c}_o - \mathbf{H}(\mathbf{c}_b)$

Various ways → **W** – Weight Matrix

→ Different Data Assimilation Schemes

$\mathbf{W}$ → Gain Matrix (OI)

$$\min\left(\boldsymbol{\varepsilon}_a^T \bullet \boldsymbol{\varepsilon}_a\right) \quad \rightarrow$$

$$\mathbf{W} = \mathbf{B}\mathbf{H}^T\left(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T\right)^{-1}$$

W → Depends on **B**, **R**

B → Background Error Covariance Matrix

R → observational error covariance matrix (usually assumed given)

W – Gain Matrix (3D-Var)

Minimizing Cost Function → $\nabla J(\mathrm{x}) = 0$

$$2J(\mathbf{c}) = (\mathbf{c} - \mathbf{c}_b)^T \mathbf{B}^{-1}(\mathbf{c} - \mathbf{c}_b)$$
$$+[\mathbf{c}_o - \mathbf{H}(\mathbf{c})]^T \mathbf{R}^{-1}[\mathbf{c}_o - \mathbf{H}(\mathbf{c})]$$

$$\mathbf{W} = (\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1}$$

→Depends on **B,** **R** (usually given)

W – Gain Matrix (Kalman Filter)

W – Matrix is called Kalman Gain

$$\mathbf{W}_i = \mathbf{B}^f(t_i)\mathbf{H}^T[\mathbf{R}_i + \mathbf{H}_i\mathbf{B}^f(t_i)\mathbf{H}^T]^{-1}$$

→Depends on forecast error covariance matrix

$\mathbf{B}^f(t_i)$, **R** (known)

It is challenge to determine the matrix B.

Is there any data assimilation method that does not require the matrix B?

All the existing methods were 'borrowed' from atmospheric data assimilation

- Optimal Interpolation (OI)
- Kalman Filter
- Variational Methods

Atmosphere has bottom boundary, but no lateral boundary

Ocean has both bottom and lateral boundaries

Effectively using the ocean topographic characteristics



A new ocean data assimilation method without requiring *a priori* knowledge of matrix B

A New Ocean Data Assimilation Method

$$\mathbf{c}_a = \mathbf{c}_b + F\left[\mathbf{c}_o - \mathbf{H}(\mathbf{c}_b)\right]$$

F → Operator (innovation → updating)

What is "F"?

Optimal Spectral Decomposition (OSD) Operator

(2) Theoretical Base Of OSD

References

- Chu, P.C., L.M. Ivanov, T.P. Korzhova, T.M. Margolina, and O.M. Melnichenko, 2003a: Analysis of sparse and noisy ocean current data using flow decomposition. Part 1: Theory. Journal of Atmospheric and Oceanic Technology, 20 (4), 478-491.

- Chu, P.C., L.M. Ivanov, T.P. Korzhova, T.M. Margolina, and O.M. Melnichenko, 2003b: Analysis of sparse and noisy ocean current data using flow decomposition. Part 2: Application to Eulerian and Lagrangian data. Journal of Atmospheric and Oceanic Technology, 20 (4), 492-512.

- Chu, P.C., L.M. Ivanov, and T.M. Margolina, 2004: Rotation method for reconstructing process and field from imperfect data. International Journal of Bifurcation and Chaos, 14(8), 2991-2997.

References

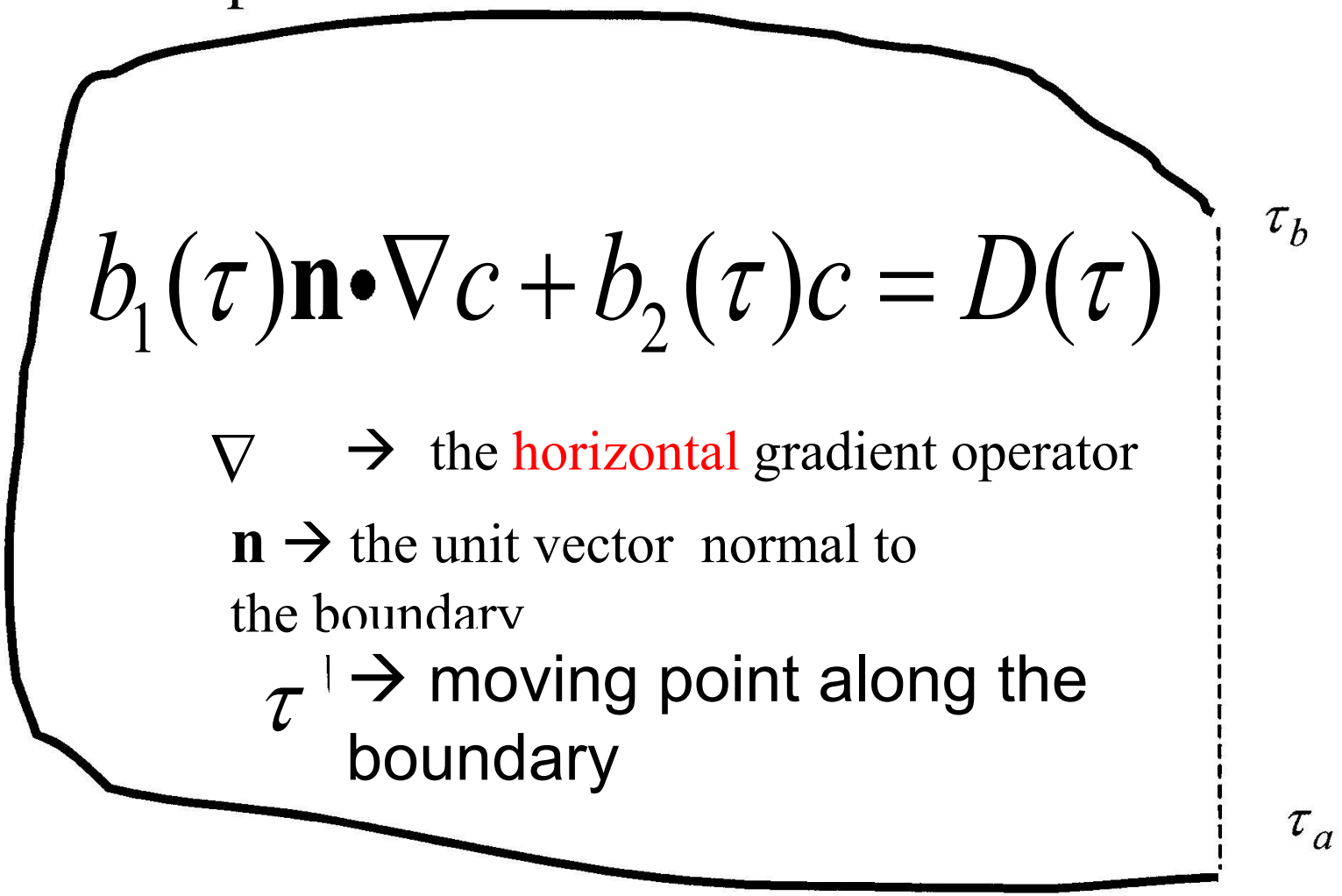
- Chu, P.C., L.M. Ivanov, and O.M. Melnichenko, 2005: Fall-winter current reversals on the Texas-Lousiana continental shelf. Journal of Physical Oceanography, 35, 902-910

- Chu, P.C., L.M. Ivanov, O.M. Melnichenko, and N.C. Wells, 2007: On long baroclinic Rossby Waves in the tropical North Atlantic observed from profiling floats. Journal of Geophysical Research, **112**, C05032, doi:10.1029/2006JC003698..

- Galanis, G., G. Emmanouil, P.C. Chu, and G. Kallos, 2009: A new methodology for the extension of the impact of data assimilation on ocean wave prediction. *Ocean Dynamics*, **59**, 523-535

- These papers can be downloaded from:
- http://faculty.nps.edu/pcchu

Piecewise Lateral Boundary Conditions (General)



Piecewise Lateral Boundary Conditions (LBCs)

$D(\tau) \neq 0$ →Inhomogeneous LBC

$D(\tau) = 0$ →Homogeneous LBC

Homogeneous →

Dirichlet boundary condition when $b_1 = 0$

Neumann boundary conditions when $b_2 = 0$

Assimilated Variable $c(\mathbf{x}, z, \mathrm{t})$

Horizontal coordinates → **x**

Vertical coordinate → z

Time → t

$$b_1(\tau)\mathbf{n}\bullet\nabla c + b_2(\tau)c = D(\tau) \qquad (1)$$

Two Parts of $c(\mathbf{x}, z, t)$

$$c(\mathbf{x},z,t) = \hat{c}(\mathbf{x},z,t) + S(\mathbf{x},z,t),$$

$$\nabla^2 S = 0, \quad k_1(\tau)\mathbf{n}\bullet\nabla S + k_2(\tau)S = D(\tau) \quad \text{at} \quad \Gamma$$

S can be pre-determined before data assimilation.

$$\hat{c}(\mathbf{x},z,t) = c(\mathbf{x},z,t) - S(\mathbf{x},z,t), \qquad (2)$$

Homogeneous Boundary Condition →

$$b_1(\tau)\mathbf{n}\bullet\nabla\hat{c} + b_2(\tau)\hat{c} = 0 \qquad (3)$$

Basis Functions

$$\nabla^2\phi_k = -\lambda_k\phi_{\mathrm{k},} \quad \left[b_1(\tau)\mathbf{n}\bullet\nabla\phi_k + b_2(\tau)\phi_k\right]|_\Gamma = 0, \quad k = 1,\ldots,\infty$$

(4)

ϕ_k → The eigen functions of the 2D Laplacian Operator

satisfaction of the same homogeneous boundary condition (3) of the assimilated variable

Three Features of the Basis Functions

(a) Satisfaction of the same homogeneous boundary condition of the assimilated variable

(b) Independence on the assimilated variable

(c) Orthonormal

$$\iint_{R(z)} \phi_k(\mathbf{x}, z)\phi_{k'}(\mathbf{x}, z)d\mathbf{x} = \delta_{kk'}$$

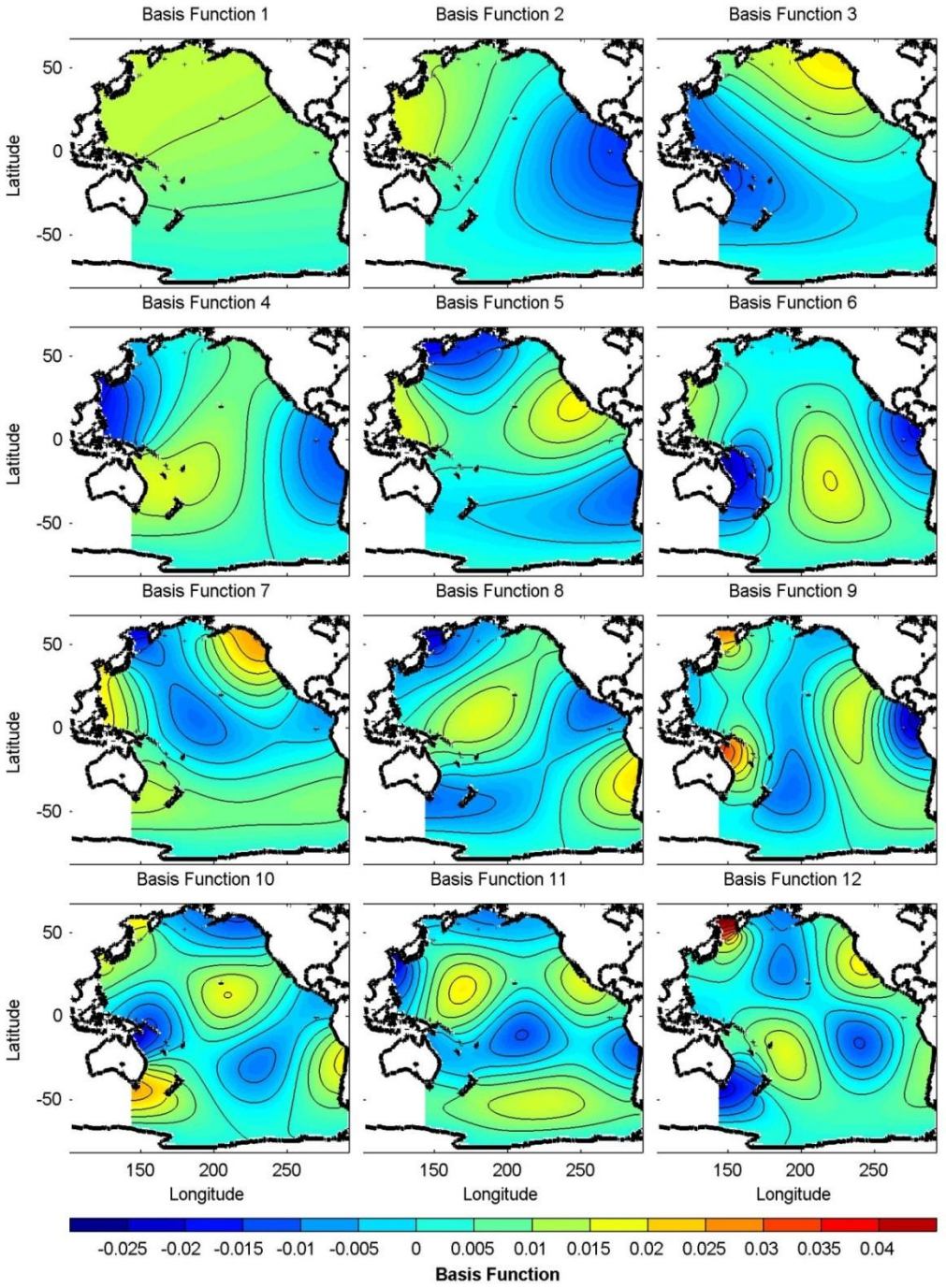
$$\delta_{kk'} = \begin{cases} 0 & \text{if } k \neq k' \\ 1 & \text{if } k = k' \end{cases}$$

First 12 basis functions for the Pacific Ocean at the surface.

DBDB5

Dirichlet boundary condition at the southern boundary (Antarctic),

Newmann boundary condition elsewhere



"True" $c_t(\mathbf{x}, z, t)$

Background (modeled) $c_b(\mathbf{x}, z, t)$

Satisfy the Boundary Condition (1) →

$$c_t(\mathbf{x},z,t) - c_b(\mathbf{x},z,t) = \hat{c}_t(\mathbf{x},z,t) - \hat{c}_b(\mathbf{x},z,t)$$

$$\left[b_1(\tau)\mathbf{n}\bullet\nabla(c_t - c_b) + b_2(\tau)(c_t - c_b)\right]_\Gamma = 0 \qquad (5)$$

Generalized Fourier Series Expansion

$$c_t(\mathbf{x},z,t) - c_b(\mathbf{x},z,t) = \sum_{k=1}^{K} a_k(z,t)\ \phi_k(\mathbf{x},z) - \varepsilon \qquad (6)$$

Given spectral coefficients $\{a_1(z,t),\ a_2(z,t),\ \ldots,\ a_K(z,t)\}$ →

OSD Data Assimilation Equation → Whole Domain

$$c_a(\mathbf{x},z,t) - c_b(\mathbf{x},z,t) = \sum_{k=1}^{K} a_k(z,t)\ \phi_k(\mathbf{x},z) \qquad (7)$$

$$\varepsilon = c_a(\mathbf{x},z,t) - c_t(\mathbf{x},z,t) \qquad (8)$$

Two Issues

(1) Obtaining the spectral coefficients at given K

(2) Determine the optimal truncation

$K = K_{opt}$

(1) Obtaining the spectral coefficients at given K

M observational profiles collected at time t, and horizontal locations $(\mathbf{x}_1, \mathbf{x}_2, .., \mathbf{x}_M)$ with various depth z.

$$[c_o(\mathbf{x}_1, z, t),\ c_o(\mathbf{x}_2, z, t),\ ...,\ c_o(\mathbf{x}_M, z, t)]$$

It is reasonable to assume that the "true" field takes the observational values at the observational points, i.e.,

$$c_t(\mathbf{x}_m, z, t) = c_o(\mathbf{x}_m, z, t)$$

Definition of Errors (8) →

$$\varepsilon = c_a(\mathbf{x}, z, t) - c_t(\mathbf{x}, z, t)$$

At observational points →

$$c_t(\mathbf{x}_m, z, t) = c_o(\mathbf{x}_m, z, t)$$

→

$$\varepsilon_m = c_a(\mathbf{x}_m, z, t) - c_o(\mathbf{x}_m, z, t), \qquad m = 1, 2, \ldots, M$$

Apply (6) at the observational points →

$$c_o(\mathbf{x}_m,z,t)-c_b(\mathbf{x}_m,z,t)=\sum_{k=1}^{K}a_k(z,t)\ \phi_k(\mathbf{x}_m,z)-\varepsilon_m,\quad m=1,2,...,M \qquad (9)$$

Variables with the red ink are given →

$$\mathbf{Da}=\mathbf{g}+\boldsymbol{\varepsilon} \qquad (10)$$

$$\mathbf{D}=[d_{mk}],\quad d_{mk}=\phi_k(\mathbf{x}_m,z,t)$$

$$\mathbf{g}=[g_1,...,g_M]^T,\quad g_m=c_o(\mathbf{x}_m,z,t)-c_b(\mathbf{x}_m,z,t)$$

$$\boldsymbol{\varepsilon}=[\varepsilon_1,\varepsilon_2,...,\varepsilon_M]^T$$

Training error J_{tr} → $J_{tr}(a_1,...,a_K,K,M)=\sqrt{\frac{1}{M}\sum_{m=1}^{M}\varepsilon_m^2(z,t)}$

$$\frac{\partial J_{tr}^2}{\partial a_k}=\frac{1}{M}\frac{\partial}{\partial a_k}\left\{\sum_{m=1}^{M}\left[\sum_{k=1}^{K}d_{mk}a_k(z,t)-g_m(z,t)\right]^2\right\}=0$$

$$\sum_{l=1}^{K}\sum_{m=1}^{M}d_{ml}d_{mk}a_l(z,t)=\sum_{m=1}^{M}d_{mk}g_m(z,t)$$

OSD Assimilation Equation → for the whole domain

$$c_a(\mathbf{x},z,t)-c_b(\mathbf{x},z,t)=\sum_{k=1}^{K}a_k(z,t)\,\phi_k(\mathbf{x},z) \qquad (7)$$

(2) Determine the optimal truncation $K = Kopt$

Using the Vapnik-Chervonenkis Dimension
(Statistical Learning Theory)

For a given mode truncation K, the error →

$$\varepsilon_K = c_b(\mathbf{x}, z, t) + \sum_{k=1}^{K} a_k(z,t)\ \phi_k(\mathbf{x}, z) - c_t(\mathbf{x}, z, t) \qquad (11)$$

$ca(\mathbf{x}, z, t)$

$c_t(\mathbf{x}, z, t)$ → Unknown →

ε_K estimated probabilistically

$c_t(\mathbf{x}_m, z, t) = c_o(\mathbf{x}_m, z, t)$ ← Observational points

$J^2 \rightarrow$ ensemble average of ε_K^2

$$P\left\{ J \le J_{tr} + J_* \sqrt{\frac{K\left[\ln(2M/K)+1\right] - \ln(\eta/4)}{M}} \right\} = 1-\eta \qquad (12)$$

(Vapnik 2000; Chu t al. 2003a)

$J \rightarrow$ Test Error

$J_* \rightarrow$ the upper bound of J_{tr}

Minimization of the VC cost function (J_K)

$$J_K = J_{tr} + \mu(K, M, \eta),$$

$$\mu(K, M, \eta) = J_* \sqrt{\frac{K\left[\ln(2M / K) + 1\right] - \ln(\eta / M)}{M}}, \quad K = 1, 2, \ldots, \infty$$

$$\min_K (J_K) = J_{K_{opt}}$$

(3) Ocean Model

- The Parallel Ocean Program (POP) model is used to show the feasibility of the OSD data assimilation.

- As the Bryan-Cox-Semtner class of models, the POP (Dukowicz and Smith 1994) was officially adopted as the ocean component of the CESM based at NCAR in 2001.

Corse Resolution

Horizontal: 3 degree Time Step: 2 hours

Depths of vertical levels in the POP model.

Level	Depth (m)	Level	Depth (m)	Level	Depth (m)	Level	Depth (m)	Level	Depth (m)
1	5	13	125	25	268	37	708	49	2649
2	15	14	135	26	285	38	787	50	2890
3	25	15	145	27	305	39	879	51	3133
4	35	16	155	28	328	40	985	52	3380
5	45	17	165	29	351	41	1106	53	3628
6	55	18	175	30	378	42	1245	54	3876
7	65	19	186	31	409	43	1401	55	4126
8	75	20	198	32	443	44	1574	56	4375
9	85	21	210	33	483	45	1764	57	4625
10	95	22	223	34	528	46	1969	58	4875
11	105	23	236	35	579	47	2186	59	5125
12	115	24	251	36	639	48	2414	60	5375

Atmospheric Forcing

- Annually varying climatology derived from the surface Coordinated Ocean Research Experiments (CORE) version 2 (Large and Yeager 2009).
- All fluxes are computed over the 23 years from 1984 to 2006.
- https://climatedataguide.ucar.edu/climate-data/large-yeager-air-sea-surface-flux-corev2-1949-2006
- The forcing is interpolated to the time step of the model.

"Truth Field"

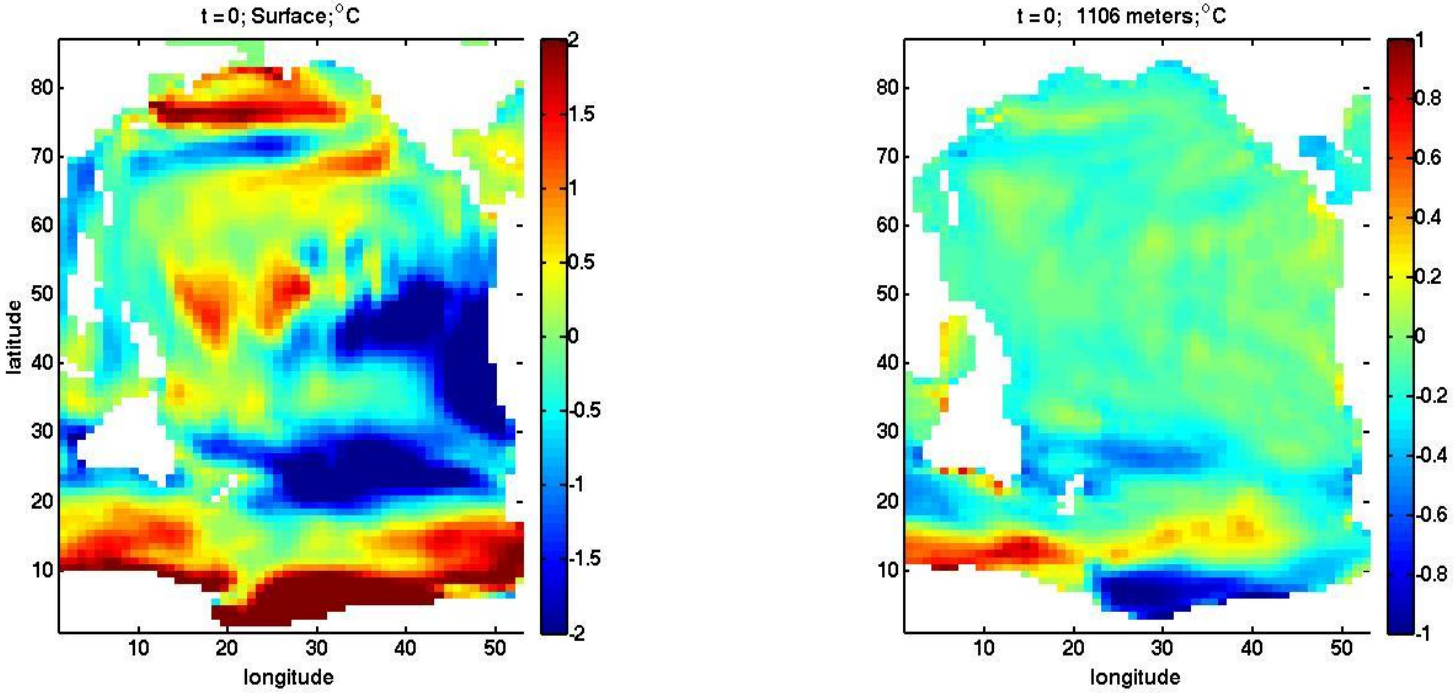
- The POP model has been spun up from rest and climatological annual mean (T, S) with the daily climatological surface forcing from the CORE version 2 (Large and Yeager 2009) and integrated for a period of over 300 simulation years. The model output data of the year-210 (c_{210}) is treated as the "truth field",

$$c_t(\mathbf{x}, z, t)$$

Initial Error

- The model is integrated from 1 March of year-210 but uses the fields from 1 March of year-300. The initial error (the variable c denoting temperature) is

$$\varepsilon(x,z,t_0) = c_{210}(\mathbf{x},z,t_0)\text{-}c_{300}(\mathbf{x},z,t_0)$$



Initial errors in temperature (°C) at: (a) the sea surface, and (b) the depth of 1106 m (at the level 41)

(4) Twin Experiment

- (a) Model Integration with no Data Assimilation

- (b) Model Integration with Data Assimilation

Exp-1→ No Data Assimilation

The model is integrated without data assimilation from 1 March with the initial condition,

$$c_b(\mathbf{x}, z, t_0) = c_{300}(\mathbf{x}, z, t_0)$$

Use daily surface forcing for 20 days, and get

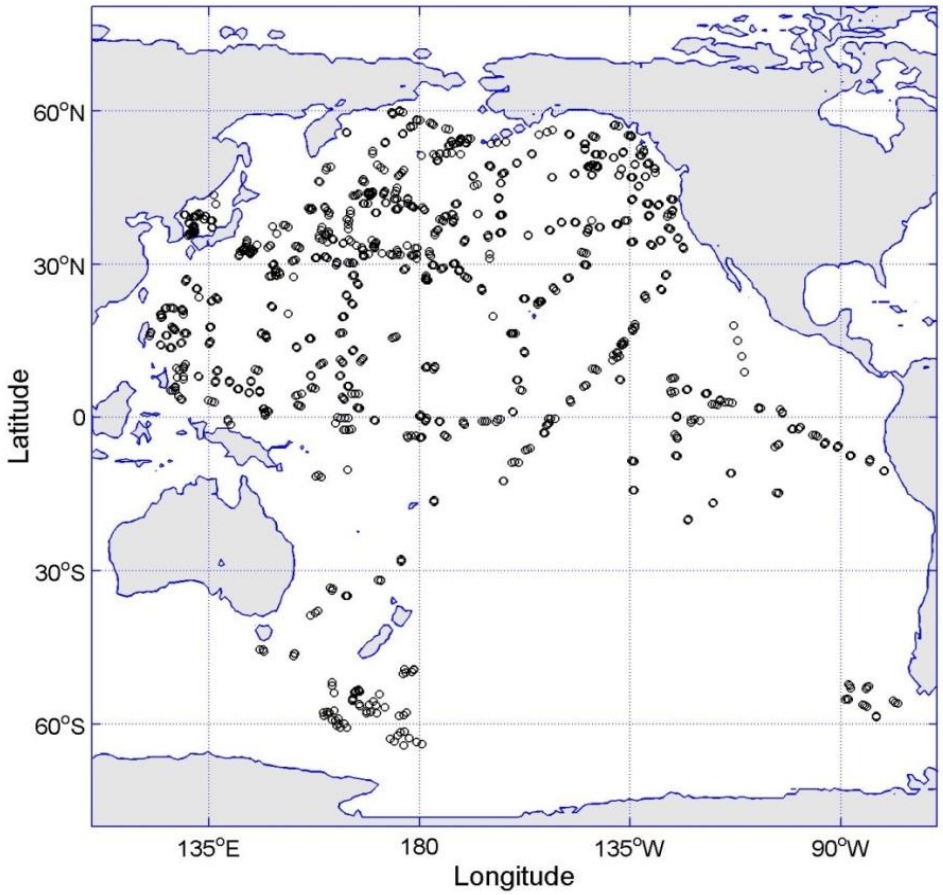
$$c_m(\mathbf{x}, z, t)$$

Exp-2 → OSD Data Assimilation

$$c_a(\mathbf{x},z,t)-c_b(\mathbf{x},z,t)=\sum_{k=1}^{K}a_k\left(z,t\right)\,\phi_k\left(\mathbf{x},z\right)$$

(a) The assimilation model is then run forward in time for 24 hours with the model field saved at the end of 24 hours, which is the background field for the day = (t+1), $c_b(\mathbf{x},z,t+1\text{ day})$

(b) At each assimilation time (1 day), the optimal mode truncation (K_{opt}) is re-calculated.

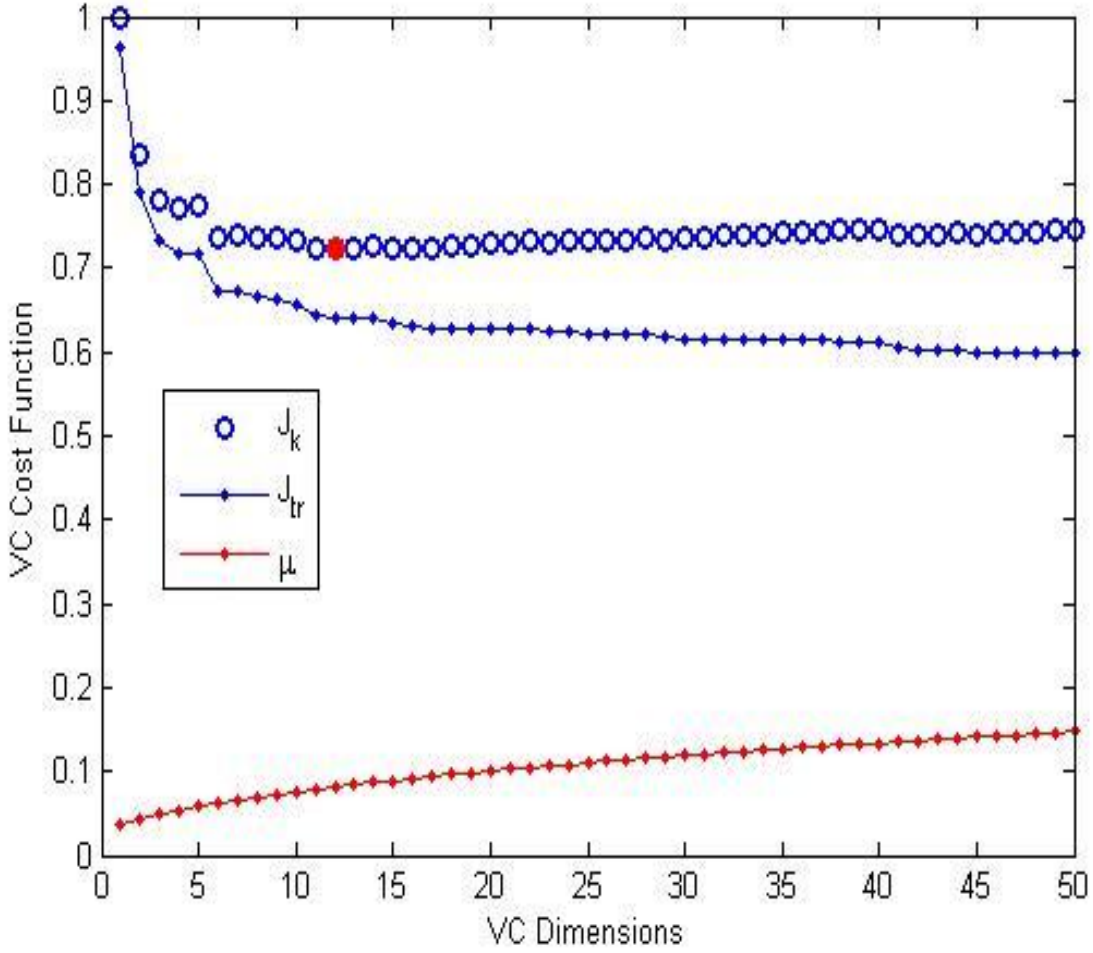


Sampling

$$c_t(\mathbf{x}, z, t)$$

$$= c_{210}(\mathbf{x}, z, t)$$

Daily sampling taking from horizontal distribution of the Argo floats (just using the horizontal locations) in March 2003. It is noted that the "observational" data-rich area is north of 20°S, and the "observational" data-poor area is south of 20°S.

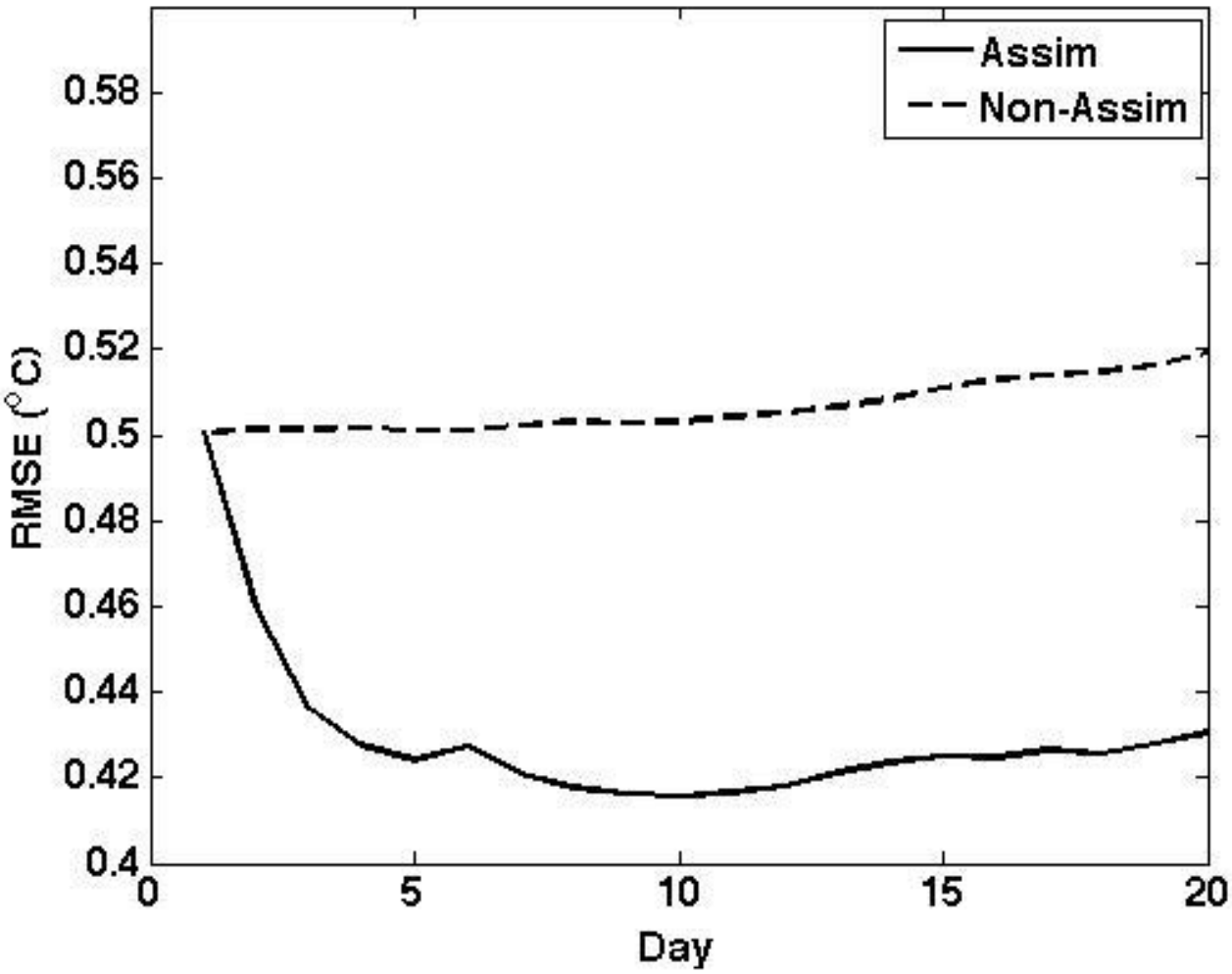


Optimal mode decomposition (K_{opt}) at 125 m depth and day-0 is determined the minimization of the VC cost function (denoted by red square).

Basin-wide RMSEs for the assimilation run (E_{ass}) and non-assimilation run (E_{non})

$$E_{ass}(t) = \sqrt{\frac{1}{L}\sum_{l=1}^{L}\left\{\frac{1}{N_l}\sum_{i=1}^{N_l}\left[c_a(\mathbf{x}_i, z_l, t) - c_t(\mathbf{x}_i, z_l, t)\right]^2\right\}}$$

$$E_{non}(t) = \sqrt{\frac{1}{L}\sum_{l=1}^{L}\left\{\frac{1}{N_l}\sum_{i=1}^{N_l}\left[c_m(\mathbf{x}_i, z_l, t) - c_t(\mathbf{x}_i, z_l, t)\right]^2\right\}}$$

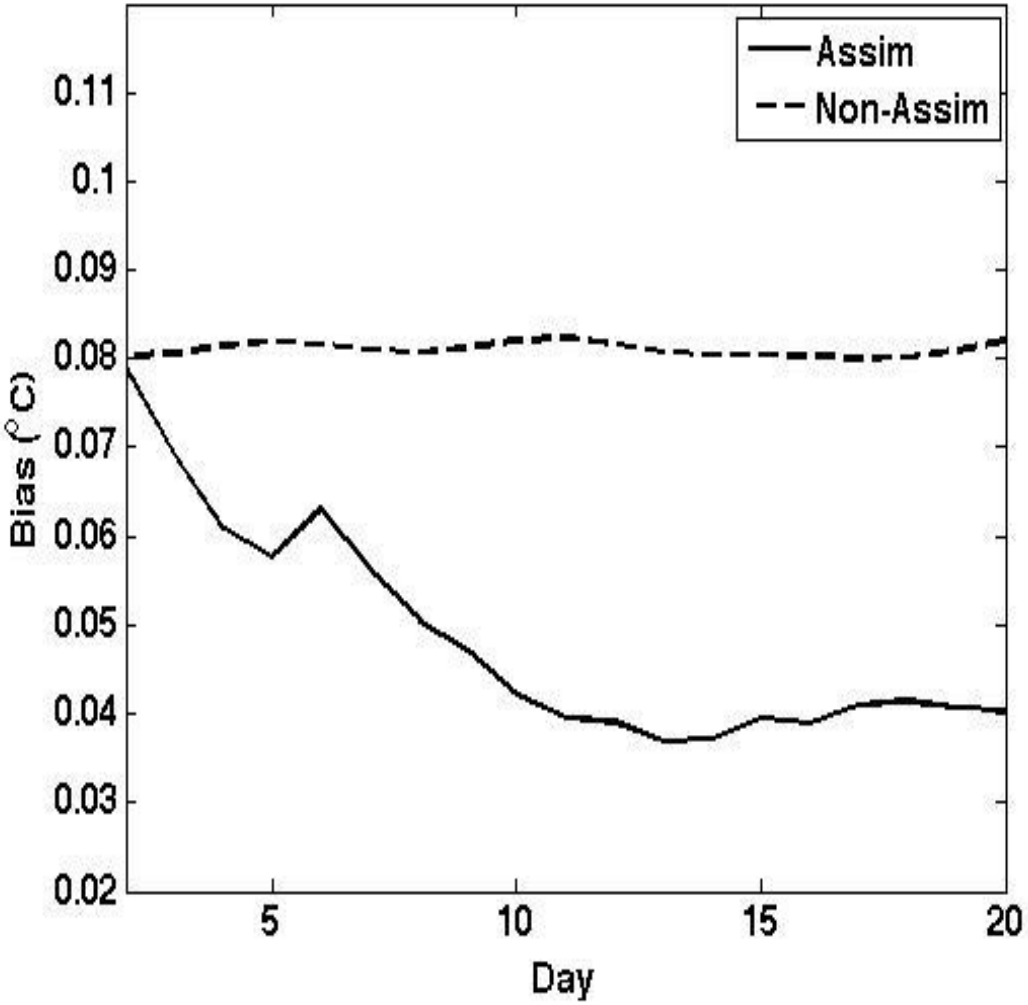


Comparison of temporally varying basin-wide RMSEs for the assimilation and non-assimilation runs

Basin-wide biases for the assimilation run (B_{ass}) and non-assimilation run (B_{non})

$$B_{ass}(t) = \frac{1}{L}\sum_{l=1}^{L}\left\{\frac{1}{N_l}\sum_{i=1}^{N_l}\left[c_a(\mathbf{x}_i, z_l, t) - c_t(\mathbf{x}_i, z_l, t)\right]\right\}$$

$$B_{non}(t) = \frac{1}{L}\sum_{l=1}^{L}\left\{\frac{1}{N_l}\sum_{i=1}^{N_l}\left[c_m(\mathbf{x}_i, z_l, t) - c_t(\mathbf{x}_i, z_l, t)\right]\right\}.$$



Comparison of temporally varying basin-wide biases for the assimilation and non-assimilation runs.

Local RMSEs for the assimilation and non-assimilation runs

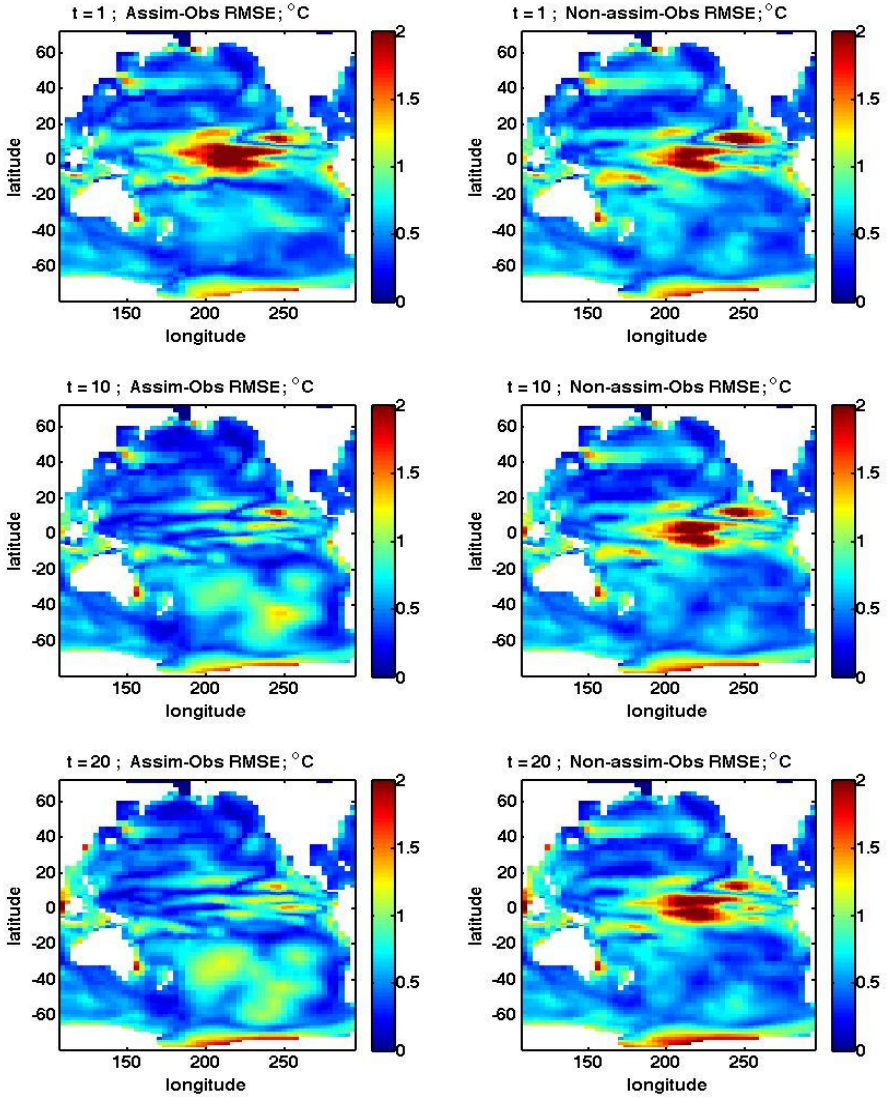
$$E_{ass}(\mathbf{x}_i, t) = \sqrt{\frac{1}{L}\sum_{l=1}^{L}\left[c_a(\mathbf{x}_i, z_l, t) - c_t(\mathbf{x}_i, z_l, t)\right]^2}$$

$$E_{non}(\mathbf{x}_i, t) = \sqrt{\frac{1}{L}\sum_{l=1}^{L}\left[c_m(\mathbf{x}_i, z_l, t) - c_t(\mathbf{x}_i, z_l, t)\right]^2}$$

Day-1

Day-10

Day-20



Left → Assimilation

Right → Non-assimilation

Comparison of temporally varying local RMSEs for the assimilation and non-assimilation runs

Local biases for the assimilation and non-assimilation runs

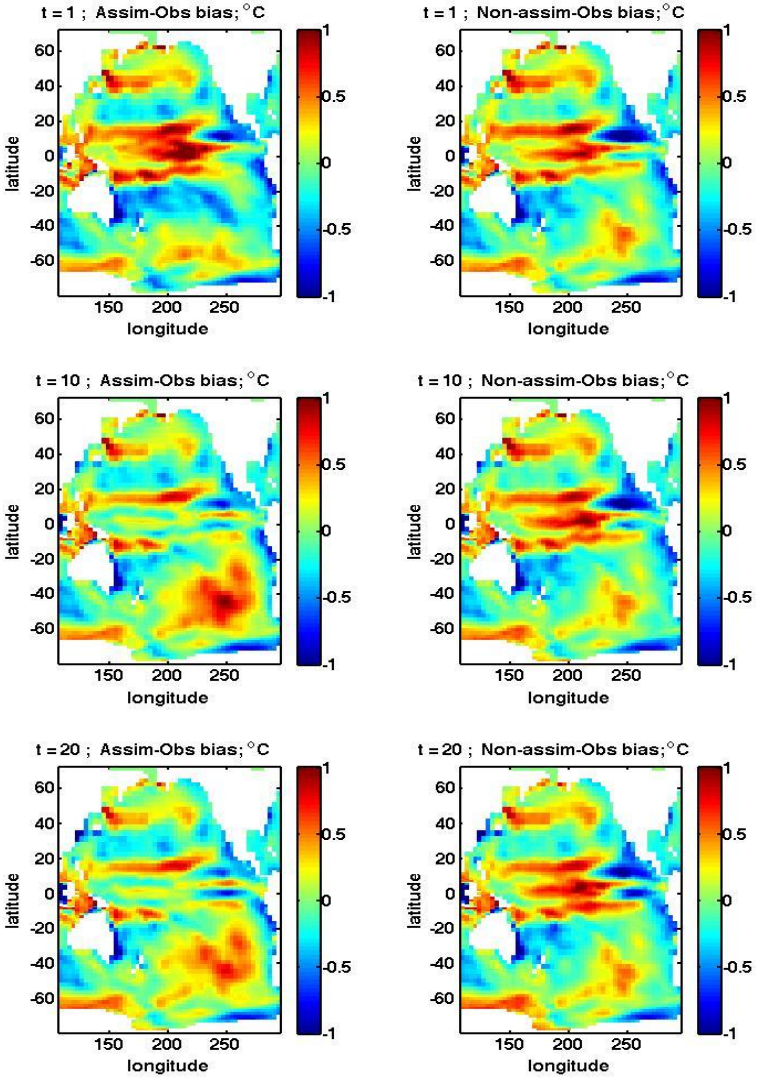
$$E_{ass}(\mathbf{x}_i, t) = \frac{1}{L}\sum_{l=1}^{L}\left[c_a(\mathbf{x}_i, z_l, t) - c_t(\mathbf{x}_i, z_l, t)\right]$$

$$E_{non}(\mathbf{x}_i, t) = \frac{1}{L}\sum_{l=1}^{L}\left[c_m(\mathbf{x}_i, z_l, t) - c_t(\mathbf{x}_i, z_l, t)\right]$$

Day-1

Day-10

Day-20



Left → Assimilation

Right → Non-assimilation

Comparison of temporally varying local biases for the assimilation and non-assimilation runs

Conclusions

- OSD is a useful method especially for ocean data assimilation.
- OSD method effectively uses the ocean topography and lateral boundary conditions.
- The basis functions are orthonormal
- The basis functions are the eigen functions of the Laplacian operator (pre-determined).