Ocean Data Assimilation without Background Error Covariance Matrix

Peter C. Chu, C.W. Fan, and R.T. Tokmakian
Naval Postgraduate School
Monterey, CA 93943
pcchu@ns.edu
http://faculty.nps.edu/pcchu

47th International Liege Colloquium on Ocean Dynamics
Marine Environmental Monitoring, Modelling and Prediction
4 -8 May 2015

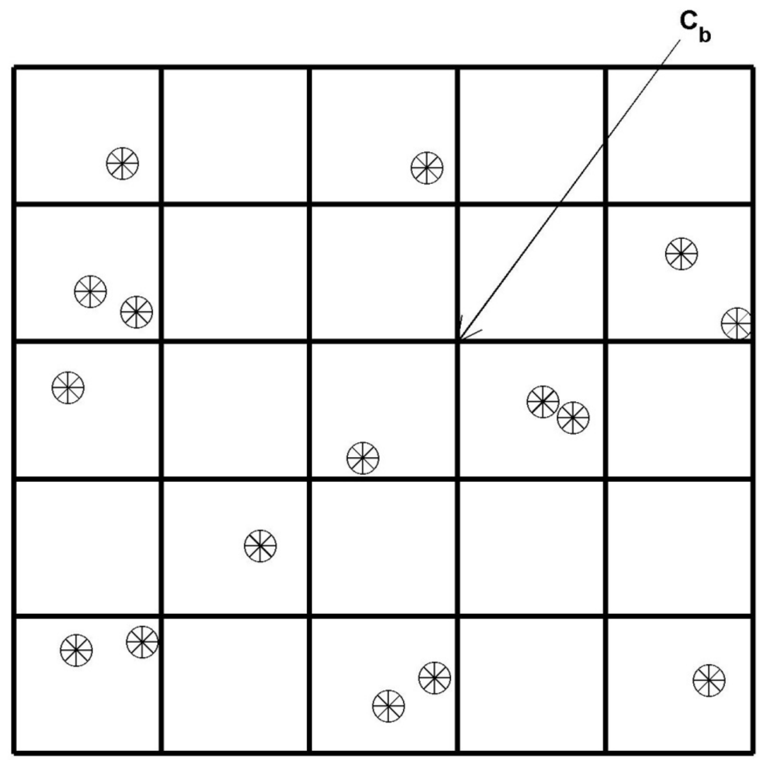
Outline

- (1) Introduction
- (2) OSD Method
- (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



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, $\boldsymbol{d} = \boldsymbol{c}_o - \mathbf{H}\boldsymbol{c}_b$, from the observational points to the grid points.

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

$\mathbf{c}_t$ → "true"

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

$\mathbf{c}_b$ → background (modeled)

$(\mathbf{c}_t, \mathbf{c}_a, \mathbf{c}_b)$ → grid points → $\mathbf{r}_n$ → total N

$\mathbf{c}_o$ → Observation → Obs points → $\mathbf{r}^{(m)}$

→ total M

Data Assimilation

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

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

Various ways → **W** – Weight Matrix
→ Different Data Assimilation Schemes

$\mathbf{H} = [h_{mn}]$ → the $M \times N$ linear observation operator matrix

W → Depends on **B**, **R**

B → Background Error Covariance Matrix

R → observational error covariance matrix (usually assumed given)

Background Error Covariance Matrix B

$$\boldsymbol{\varepsilon}_a = \mathbf{c}_a - \mathbf{c}_t, \quad \boldsymbol{\varepsilon}_o \equiv \mathbf{H}^T \mathbf{c}_o - \mathbf{c}_t$$

$$E^2 = \left\langle \boldsymbol{\varepsilon}_a^T \boldsymbol{\varepsilon}_a \right\rangle \rightarrow \min \qquad \boxed{\partial E^2 / \partial w_{nm} = 0}$$

- Optimal Interpolation (OI) $\quad \mathbf{W} = \mathbf{B}\mathbf{H}^T(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1}$
- Kalman Filter (KF) $\quad \mathbf{W}_i = \mathbf{B}^f(t_i)\mathbf{H}^T[\mathbf{R}_i + \mathbf{H}_i\mathbf{B}^f(t_i)\mathbf{H}^T]^{-1}$
- Variational Method $\quad \mathbf{W} = (\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1}$
- Data Assimilation Equation (OI/KF) →

$$\boxed{\mathbf{c}_a = \mathbf{c}_b + \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{d}}$$

Effectively using the ocean topographic characteristics

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

(2) OSD Method

Basis Functions

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

ϕ_k → The eigen functions of the 2D Laplacian Operator

satisfaction of the same homogeneous boundary condition of the assimilated variable anomaly

Φ Matrix →

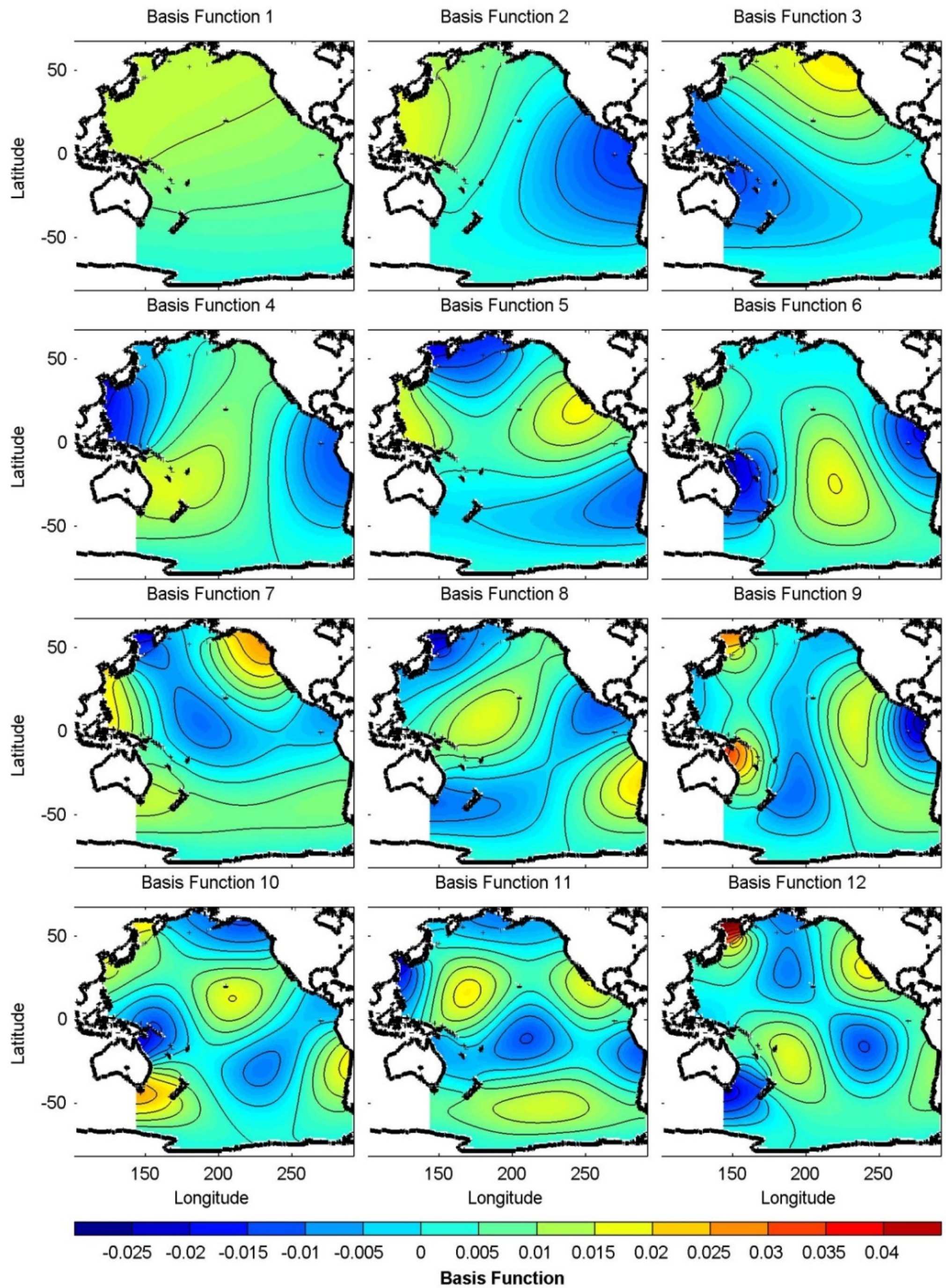
$$\mathbf{\Phi} = \{\phi_{kn}\} = \begin{bmatrix} \phi_1(\mathbf{r}_1) & \phi_2(\mathbf{r}_1) & \ldots & \phi_K(\mathbf{r}_1) \\ \phi_1(\mathbf{r}_2) & \phi_2(\mathbf{r}_2) & \ldots & \phi_K(\mathbf{r}_2) \\ \ldots & \ldots & \ldots & \ldots \\ \phi\,(\mathbf{r}_N) & \phi_2(\mathbf{r}_N) & \ldots & \phi_K(\mathbf{r}_N) \end{bmatrix}$$

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

DBDB5

Dirichlet boundary condition at the southern boundary (Antarctic),

Newmann boundary condition elsewhere



Spectral Ocean Data Assimilation

$$\mathbf{c}_a = \mathbf{c}_b + f_n \mathbf{s}^{(K)}, \quad s_K(\mathbf{r}_n) \equiv \sum_{k=1}^{K} a_k \, \phi_k(\mathbf{r}_n), \qquad f_n \equiv \sum_{m=1}^{M} h_{nm}$$

$\mathbf{H} = [h_{mn}]$ → the $M \times N$ linear observation operator matrix

$$\boldsymbol{\varepsilon}_a \equiv \mathbf{c}_a - \mathbf{c}_t = (\mathbf{c}_a - \mathbf{c}_b) + (\mathbf{c}_b - \mathbf{c}_t) = \boldsymbol{\varepsilon}_K + \boldsymbol{\varepsilon}_o$$

$$\boldsymbol{\varepsilon}_K \equiv \left[f_n \mathbf{s}^{(K)} - \mathbf{H}^T (\mathbf{c}_o - \mathbf{H}\mathbf{c}_b) \right], \qquad \boldsymbol{\varepsilon}_o \equiv \mathbf{H}^T \mathbf{c}_o - \mathbf{c}_t$$

$$\left\langle \boldsymbol{\varepsilon}_o^T \boldsymbol{\varepsilon}_K \right\rangle = 0$$

$$E^2 = \left\langle \boldsymbol{\varepsilon}_a^T \boldsymbol{\varepsilon}_a \right\rangle = E_K^2 + E_o^2, \quad E_K^2 \equiv \left\langle \boldsymbol{\varepsilon}_K^T \boldsymbol{\varepsilon}_K \right\rangle, E_o^2 \equiv \left\langle \boldsymbol{\varepsilon}_o^T \boldsymbol{\varepsilon}_o \right\rangle$$

OSD Data Assimilation Equation

$$E^2 \rightarrow \min, \quad \partial E^2 / \partial a_k = \partial E_K^2 / \partial a_k = 0, \quad k = 1, \ldots, K$$

$$E_K^2 = \sum_{n=1}^{N} \left(f_n \sum_{k=1}^{K} a_k \phi_{kn} - D_n \right)^2 \rightarrow \min$$

$$\sum_{k'=1}^{K} \sum_{n=1}^{N} (\phi_{kn} f_n \phi_{nk'}) a_{k'} = \sum_{n=1}^{N} \phi_{kn} f_n D_n, \qquad k = 1, 2, \ldots, K$$

$$\mathbf{\Phi F \Phi}^T \mathbf{A} = \mathbf{\Phi F D}, \qquad \mathbf{A} = \left[\mathbf{\Phi F \Phi}^T \right]^{-1} \mathbf{\Phi F D}$$

OSD → $\boxed{\mathbf{c}_a = \mathbf{c}_b + \mathbf{F \Phi}^T \left[\mathbf{\Phi F \Phi}^T \right]^{-1} \mathbf{\Phi H}^T \mathbf{d}}$

OI/KF → $\boxed{\mathbf{c}_a = \mathbf{c}_b + \mathbf{B H}^T (\mathbf{H B H}^T + \mathbf{R})^{-1} \mathbf{d}}$

Two Issues

(1) Determine the optimal truncation using the steep-descending method (Chu and Fan 2015)

K = Kopt

(2) Obtaining the spectral coefficients at given K from observational data

(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)	Leve l	Depth (m)	Level	Depth (m)	Leve l	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.

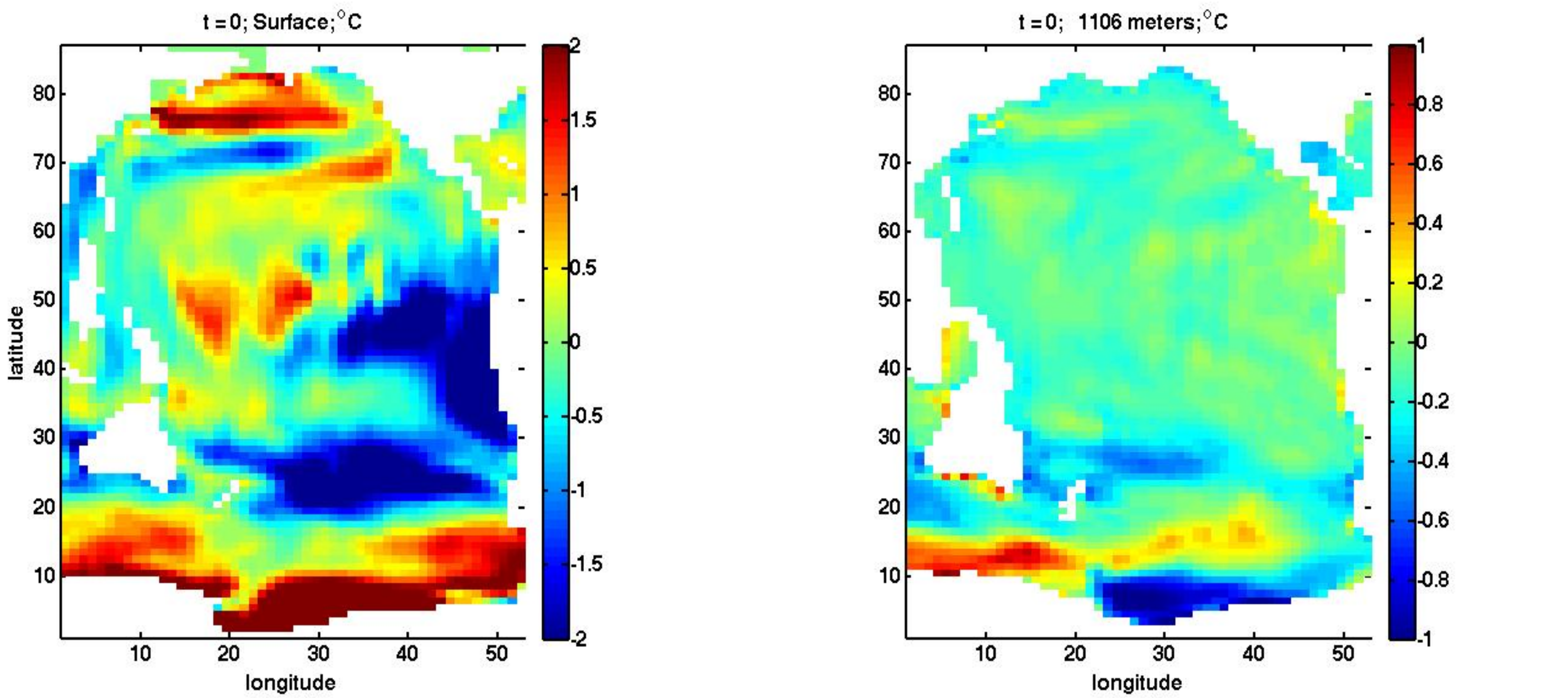
"Truth Field"

- 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

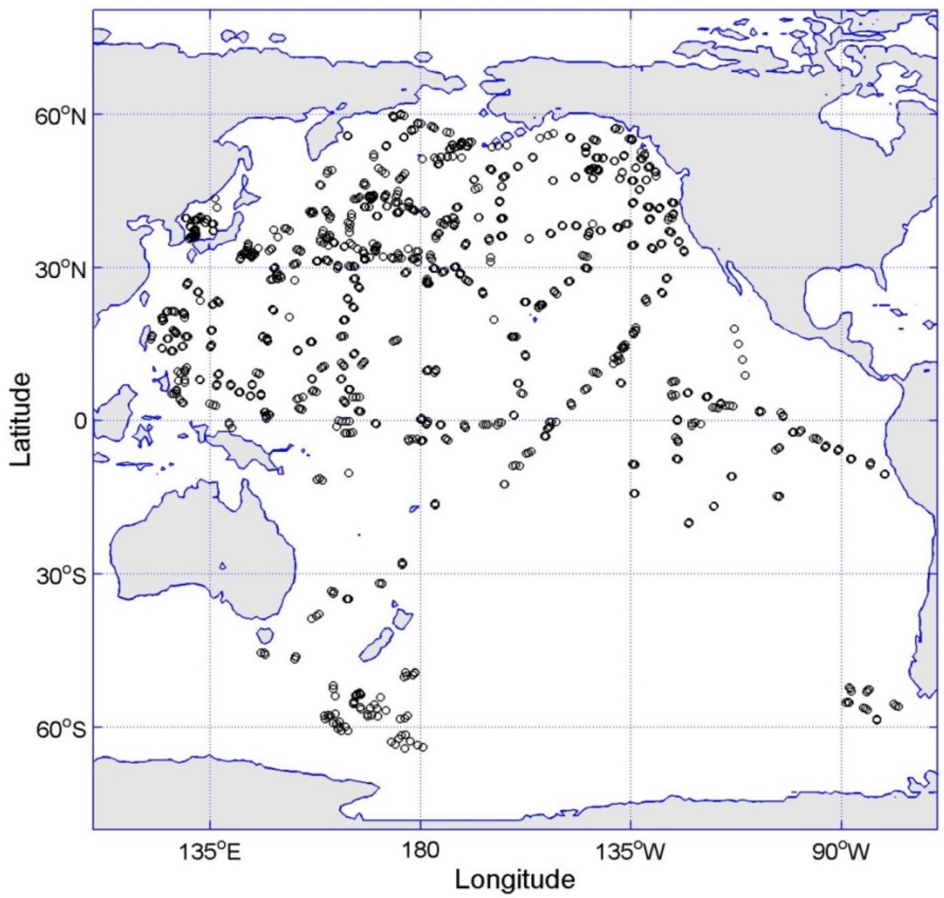
$$c_m(\mathbf{x}, z, t)$$

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.

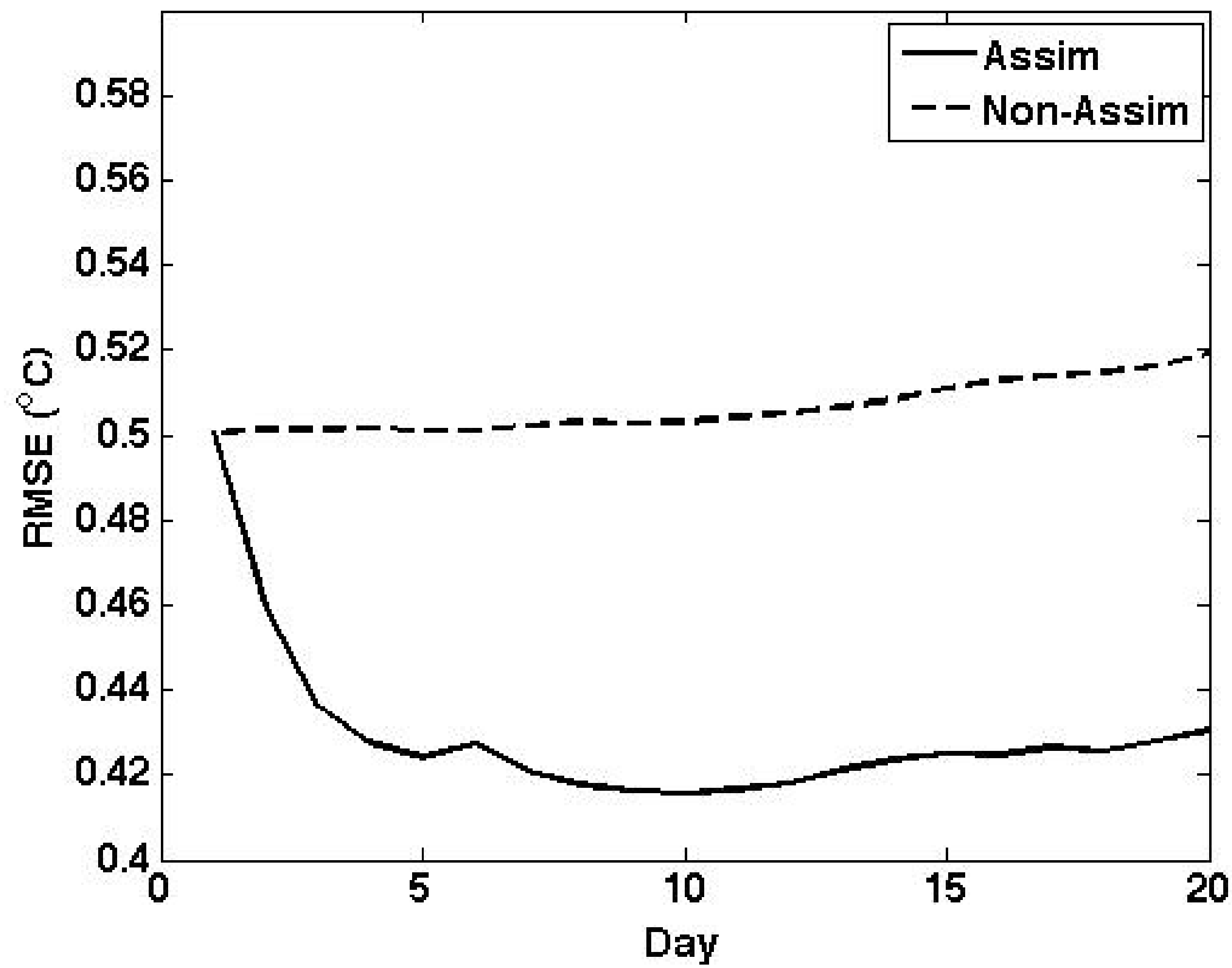


Sampling

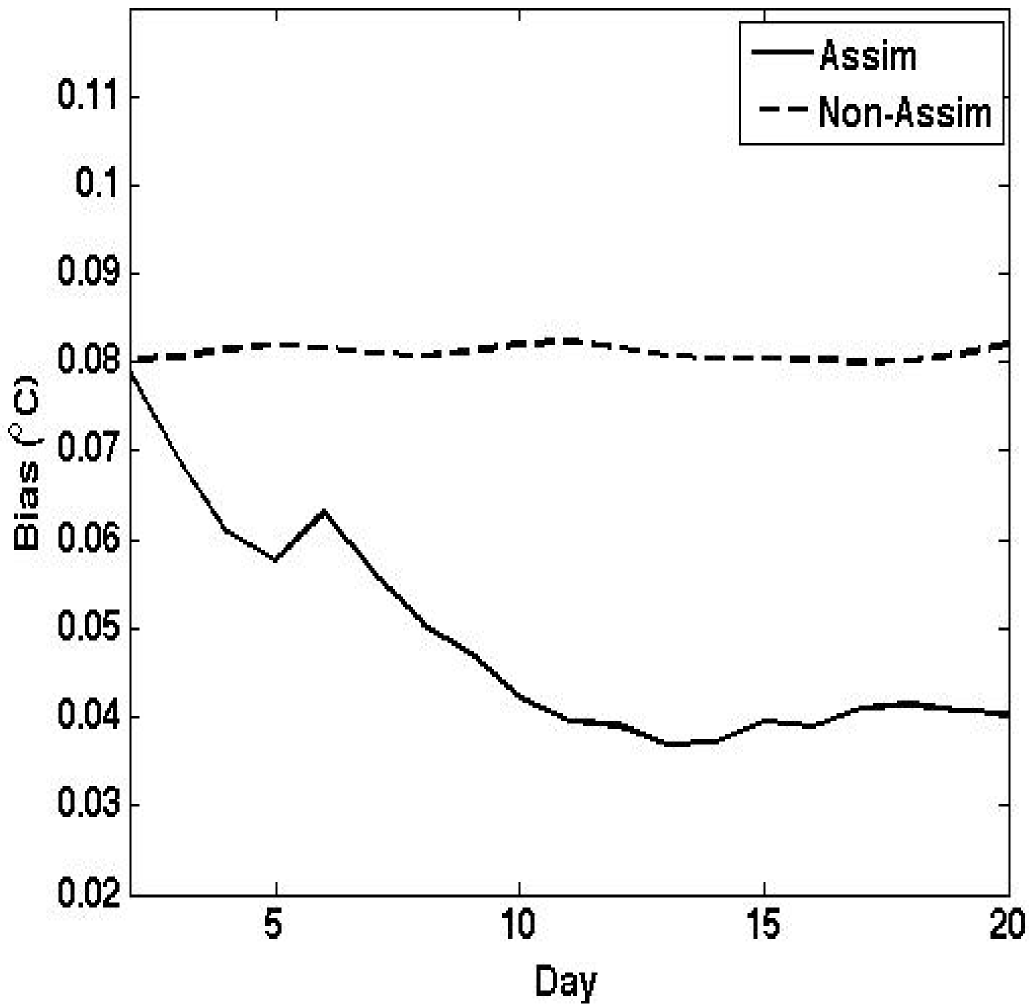
$$c_t(\mathbf{x}, z, t)$$

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

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^{o}S, and the "observational" data-poor area is south of 20^{o}S.



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

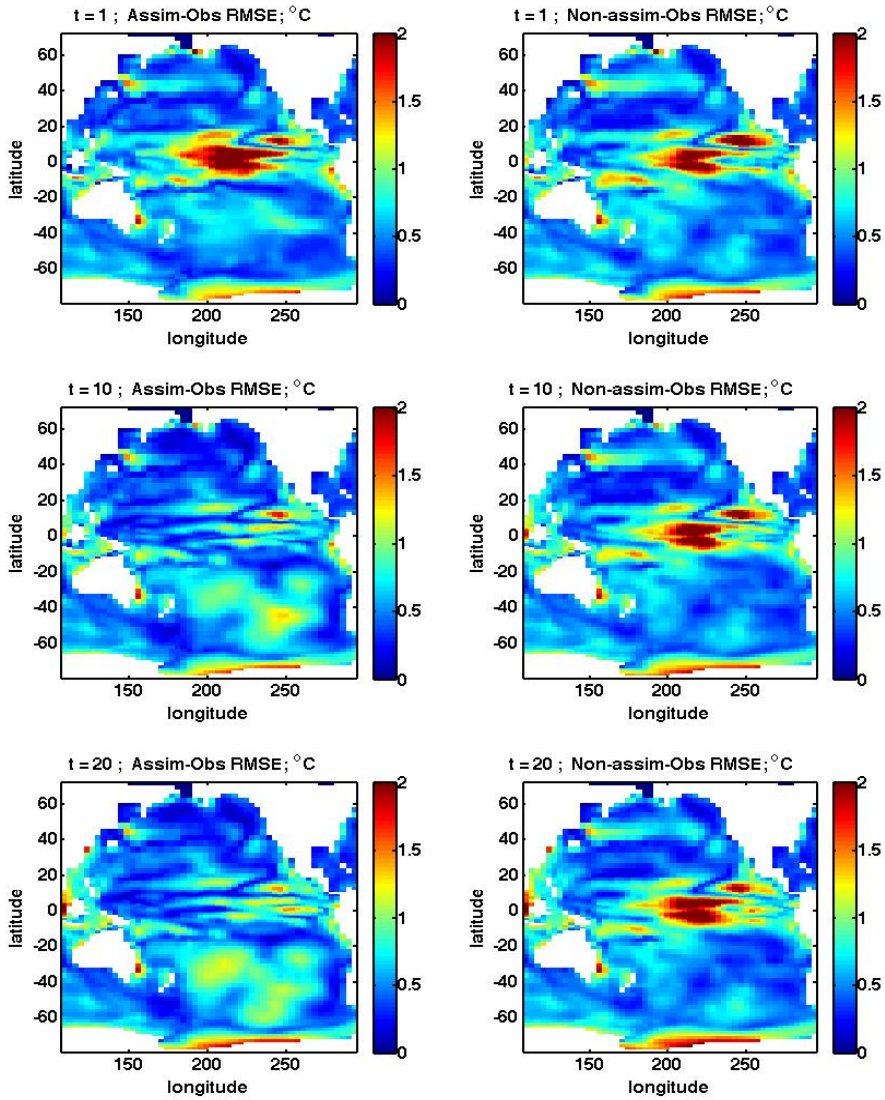


Comparison of temporally varying basin-wide biases for the assimilation and non-assimilation runs.

Day-1

Day-10

Day-20



Left → Assimilation

Right → Non-assimilation

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

Conclusions

- OSD is a useful method especially for ocean data assimilation. No background error covariance matrix B is needed.
- 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).