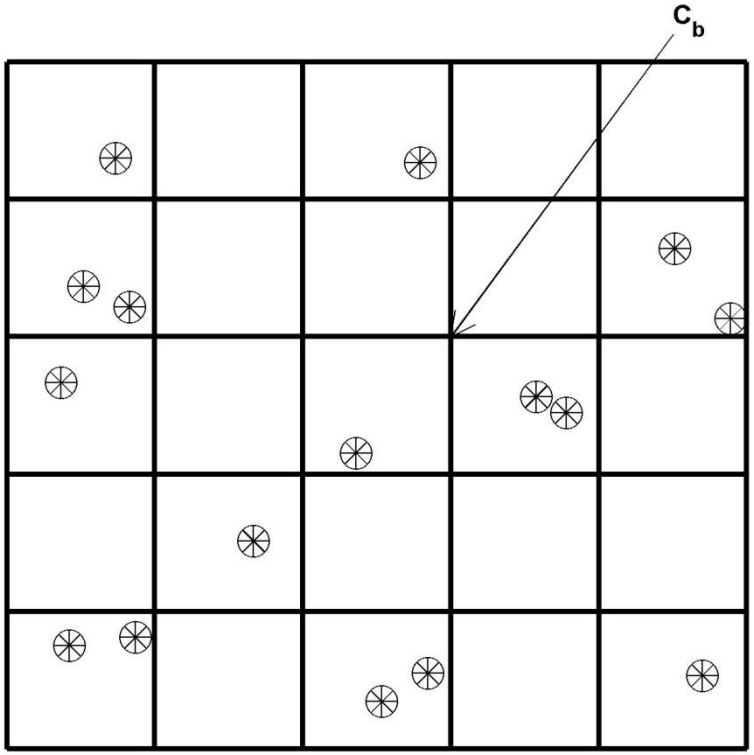
Optimal Spectral Decomposition (OSD) Data Assimilation

Atmospheric and Oceanic Science, College Park, MD, University of Maryland, 22 June 2017

(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

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 and Observational Error Covariance Matrices B and R

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

$$E^2 = \left\langle \boldsymbol{\varepsilon}_a^T \boldsymbol{\varepsilon}_a \right\rangle \rightarrow \min$$

$$\partial E^2 / \partial w_{nm} = 0$$

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

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

Gaussian Covariance Function for **B**-Matrix (Bretherton et al. 1976)

Assumption → statistically stationary and homogeneous

$\mathbf{B} = \left[b_{ij} \right]_{N \times N}$ → between two grid points $\mathbf{r}_i$ and $\mathbf{r}_j$

$$b_{ij} = \left(1 - \frac{r_{ij}^2}{r_b^2} \right) \exp\left(1 - \frac{r_{ij}^2}{r_a^2} \right), \quad r_{ij}^2 = \left| \mathbf{r}_i - \mathbf{r}_j \right|^2$$

Two user-input parameters → r_a, r_b

Observational Error Variance Matrix **R**

$$\mathbf{R} = E_o^2 \mathbf{I}, \qquad E_o^2 \equiv \left\langle \boldsymbol{\varepsilon}_o^T \boldsymbol{\varepsilon}_o \right\rangle$$

$\mathbf{I}$ → identity matrix

The third user-input parameter → E_o

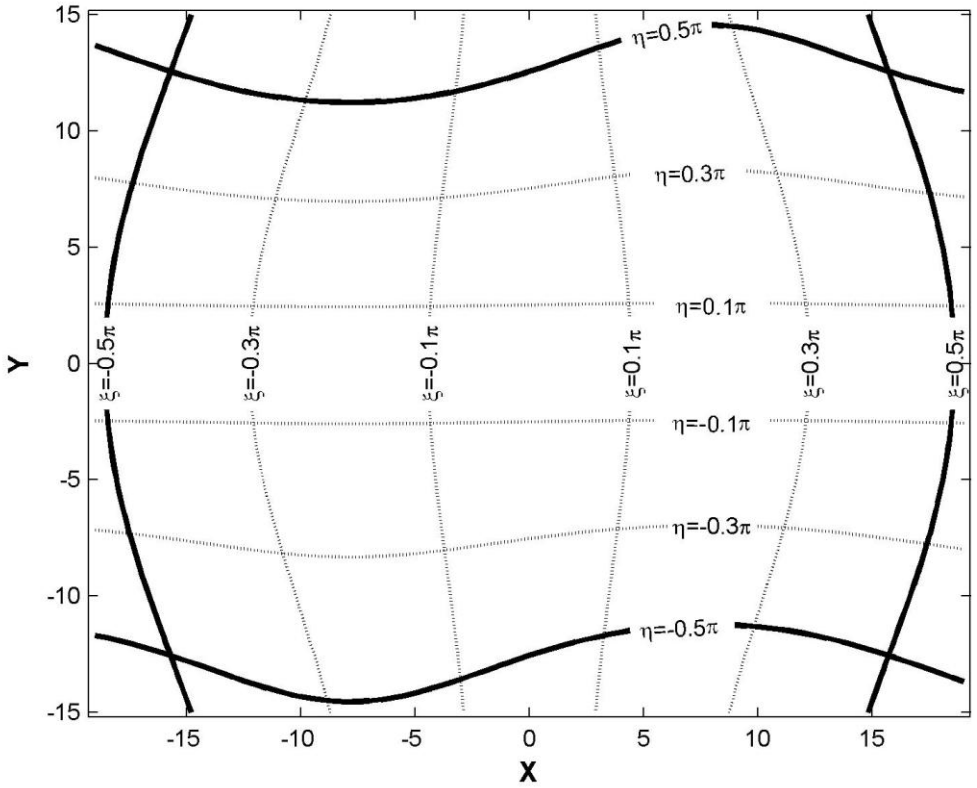
→ Instrument accuracy

OI → three user-input parameters (r_a, r_b , E_o)

(3) Test Case – A Curved Region

$$\frac{x}{10} - 0.3\cos\left(\frac{y}{8}\right)\sin\left(\frac{x}{10}\right) = \xi = \begin{cases} -\pi/2 & \text{(west)} \\ \pi/2 & \text{(east)} \end{cases}$$

$$\frac{y}{8} - 0.2\sin\left(\frac{x}{5}\right)\left[1 - \cos\left(\frac{y}{8}\right)\right] = \eta = \begin{cases} -\pi/2 & \text{(south)} \\ \pi/2 & \text{(north)} \end{cases}$$

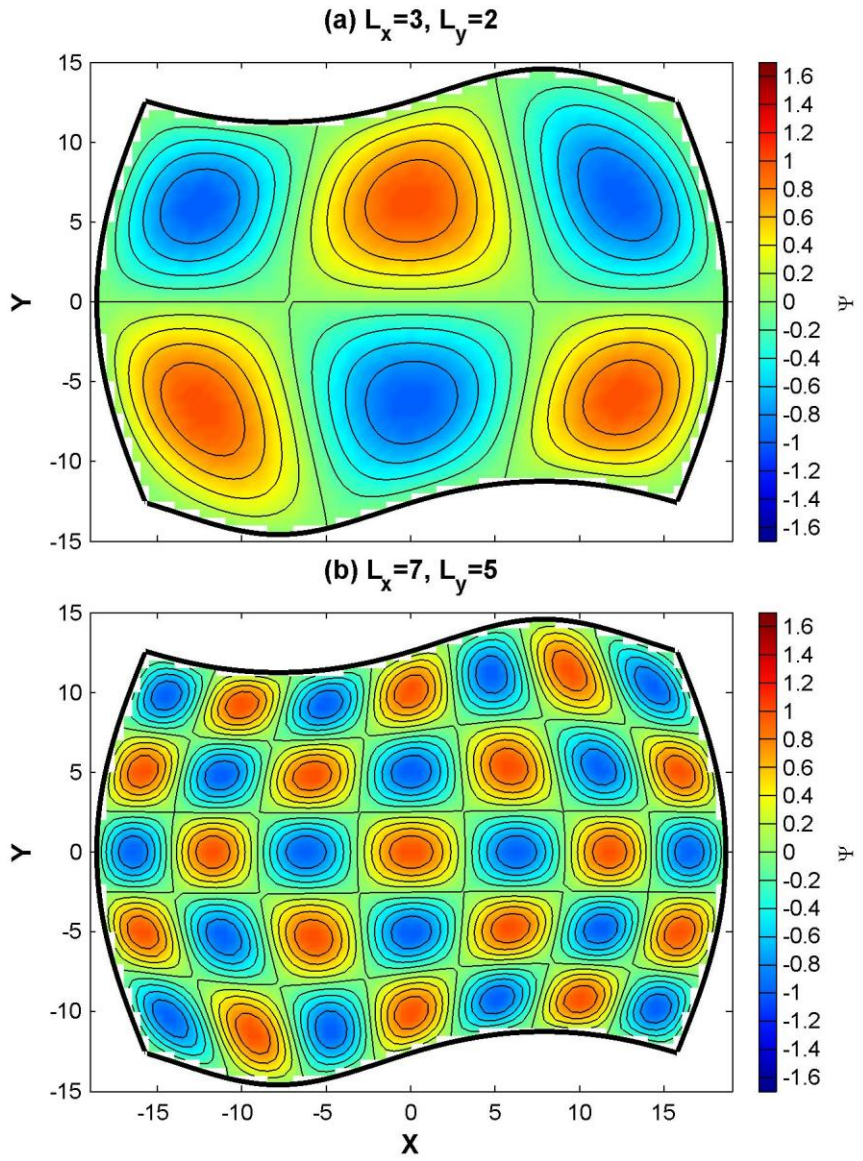


“Truth” – Analytical Streamfunction

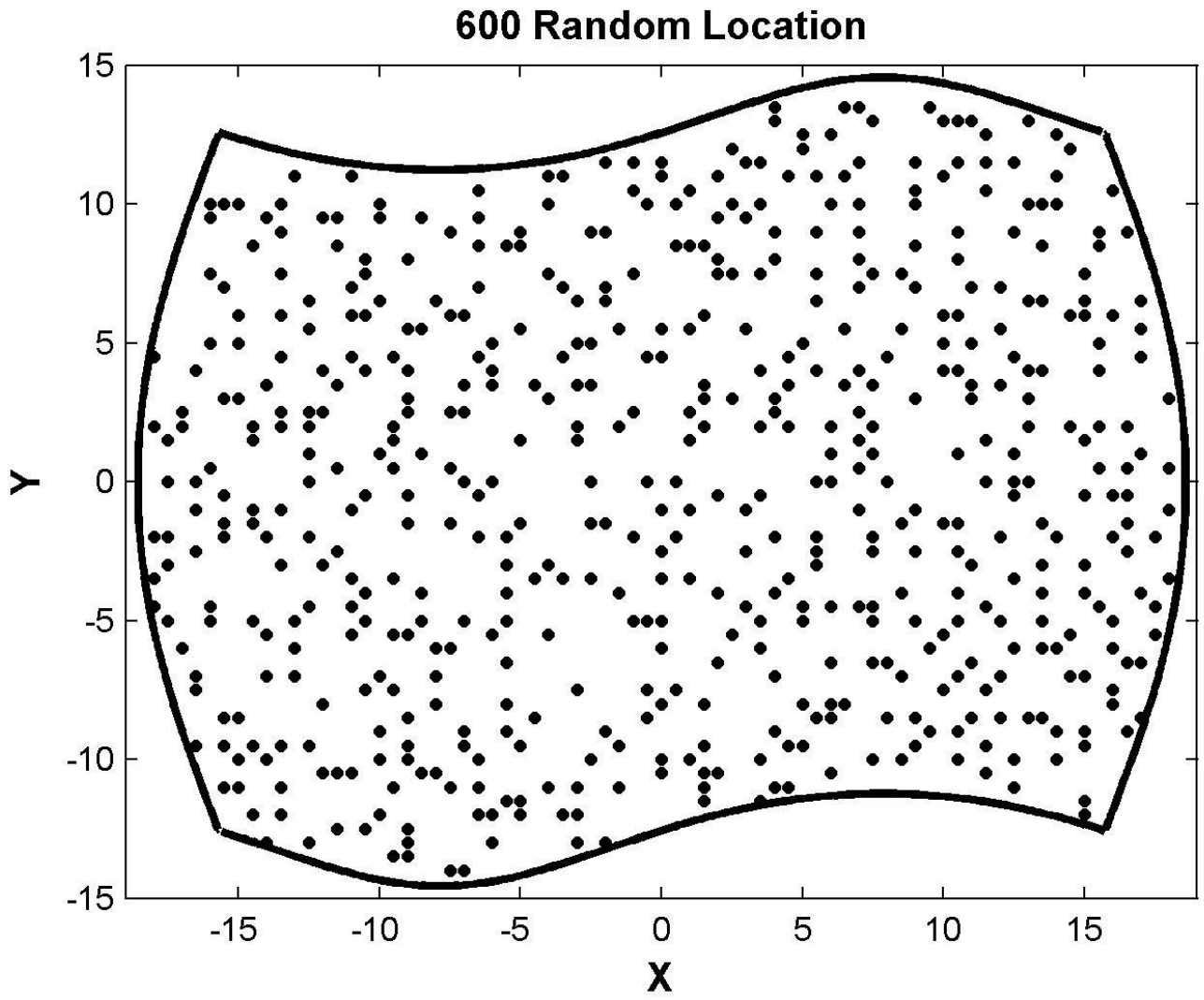
Large-Scale Eddies →

$$\begin{cases} \psi_t(x,y) = \cos\left[L_x\xi(x,y)\right]\cos\left[L_y\eta(x,y)+\beta\right] \\ \xi = \dfrac{x}{10} - 0.3\cos\left(\dfrac{y}{8}\right)\sin\left(\dfrac{x}{10}\right), \\ \eta = \dfrac{y}{8} - 0.2\sin\left(\dfrac{x}{5}\right)\left[1-\cos\left(\dfrac{y}{8}\right)\right] \end{cases}$$

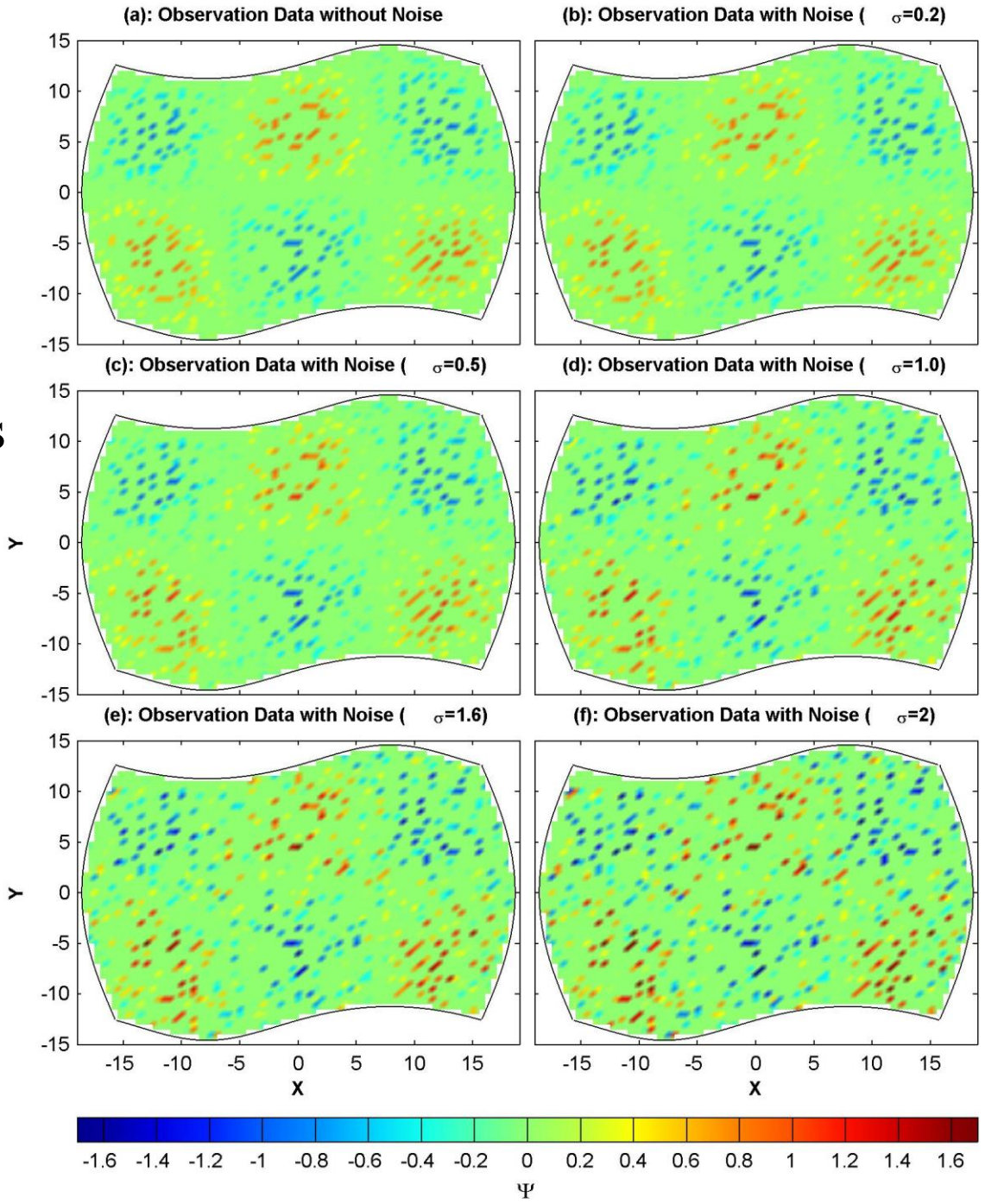
Small-Scale Eddies →



Randomly selected locations (total: 600) inside domain as “observational” points



"Observational" data (ψ_o) from "truth" (large-scale) plus white Gaussian noises of zero mean and various standard deviations: (a) 0 (i.e., no noise) (b) 0.2, (c) 0.5, (d) 1.0, (e) 1.6, and (f) 2.0.



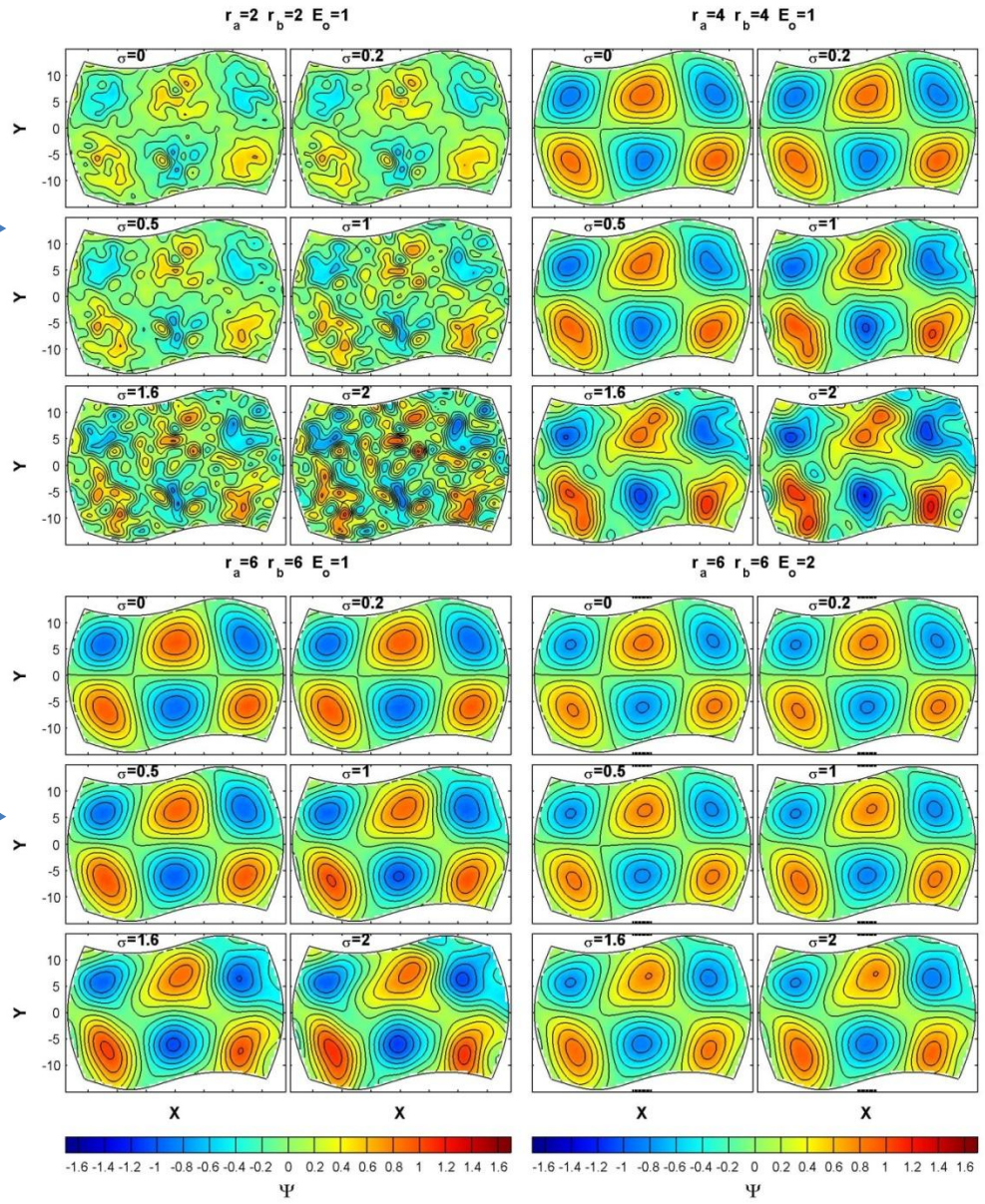
Analysis field ψ_a obtained by the OI – large eddies

$r_a = 2$
$r_b = 2$
$E_o = 1$

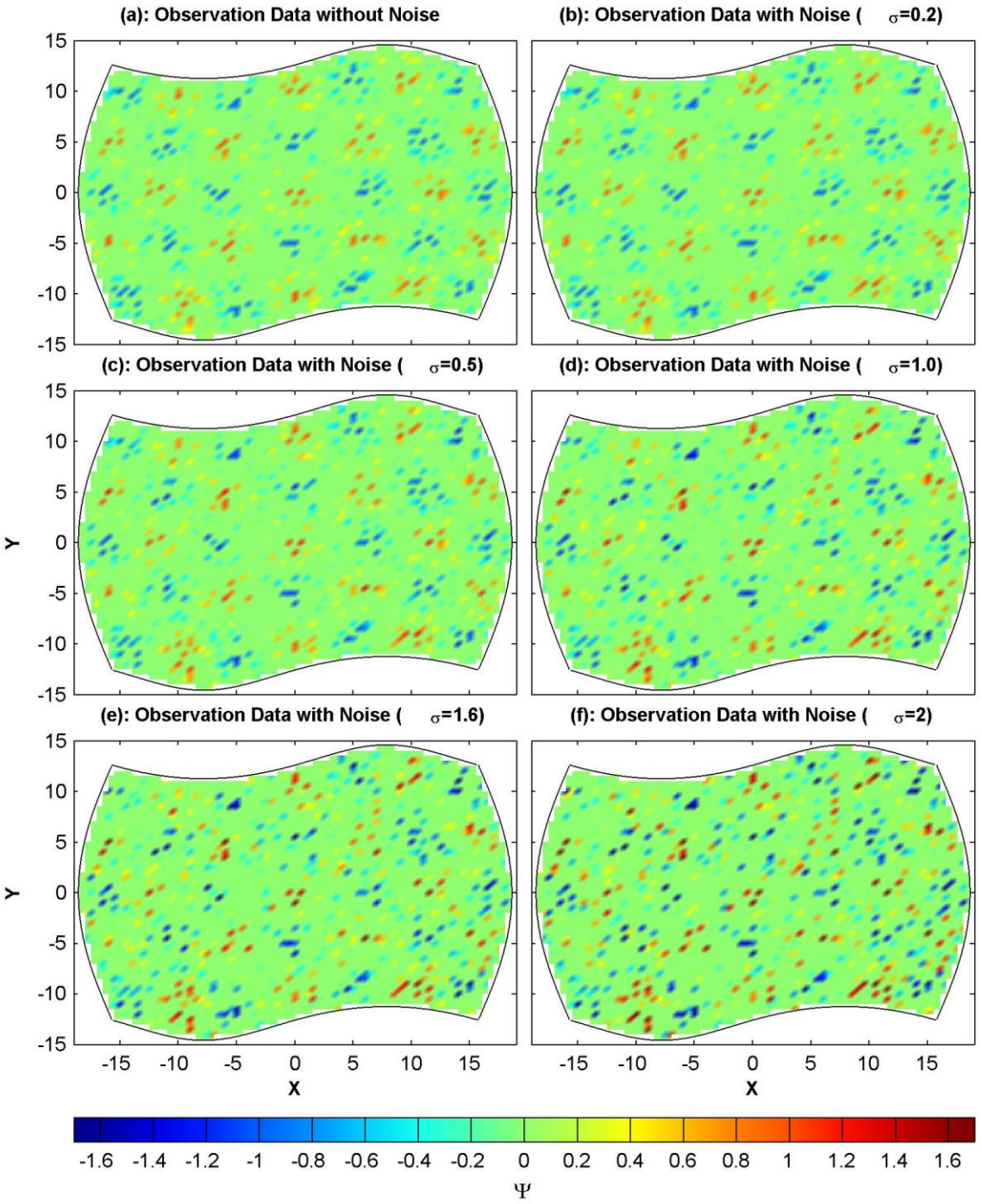
$r_a = 4$
$r_b = 4$
$E_o = 1$

$r_a = 6$
$r_b = 6$
$E_o = 1$

$r_a = 6$
$r_b = 6$
$E_o = 2$

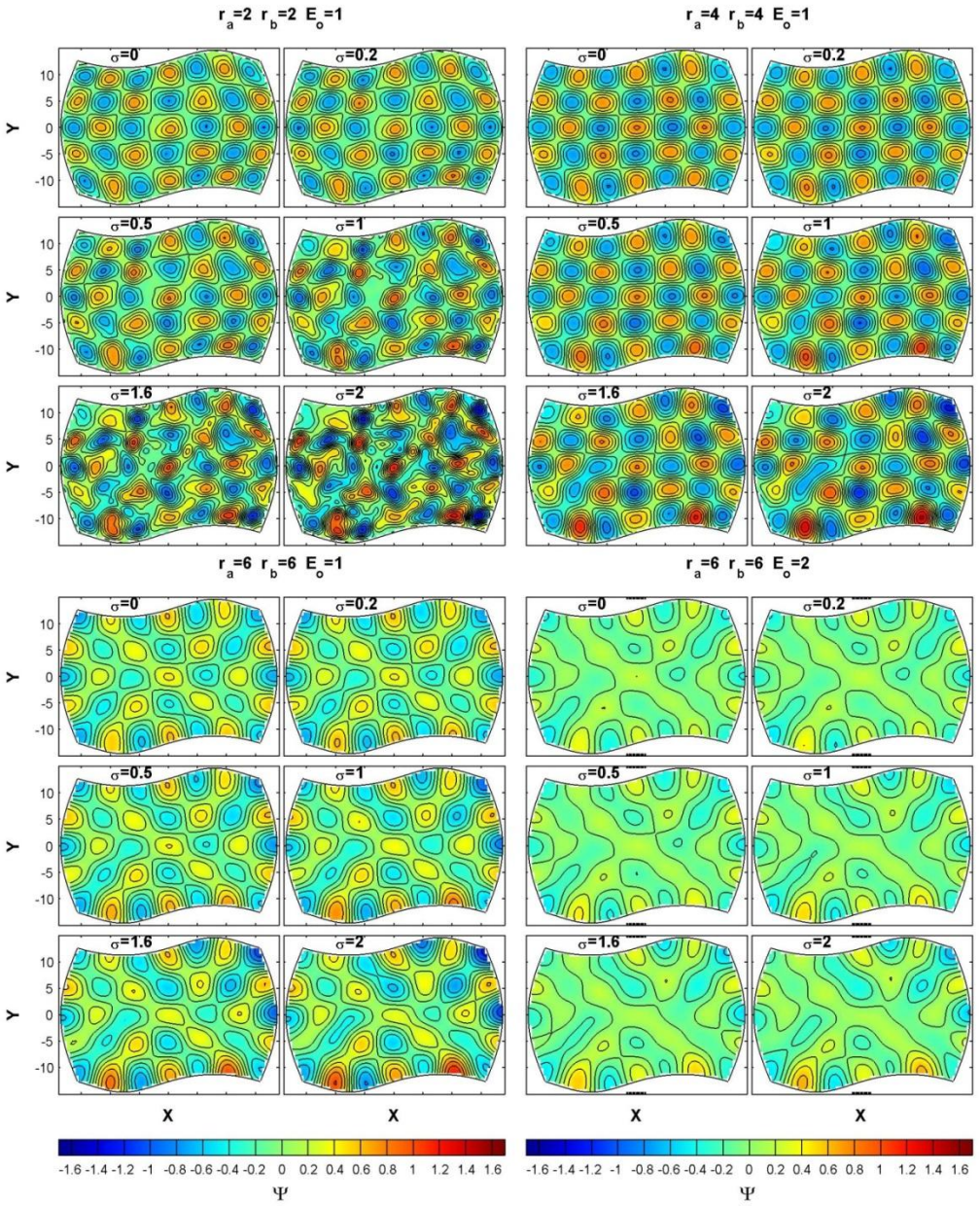


"Observational" data (ψ_o) from "truth" (small-scale) plus white Gaussian noises of zero mean and various standard deviations: (a) 0 (i.e., no noise) (b) 0.2, (c) 0.5, (d) 1.0, (e) 1.6, and (f) 2.0.



Analysis field ψ_a obtained by the OI – small eddies

$r_a = 2$
$r_b = 2$
$E_o = 1$



$r_a = 4$
$r_b = 4$
$E_o = 1$

$r_a = 6$
$r_b = 6$
$E_o = 1$

$r_a = 6$
$r_b = 6$
$E_o = 2$

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

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

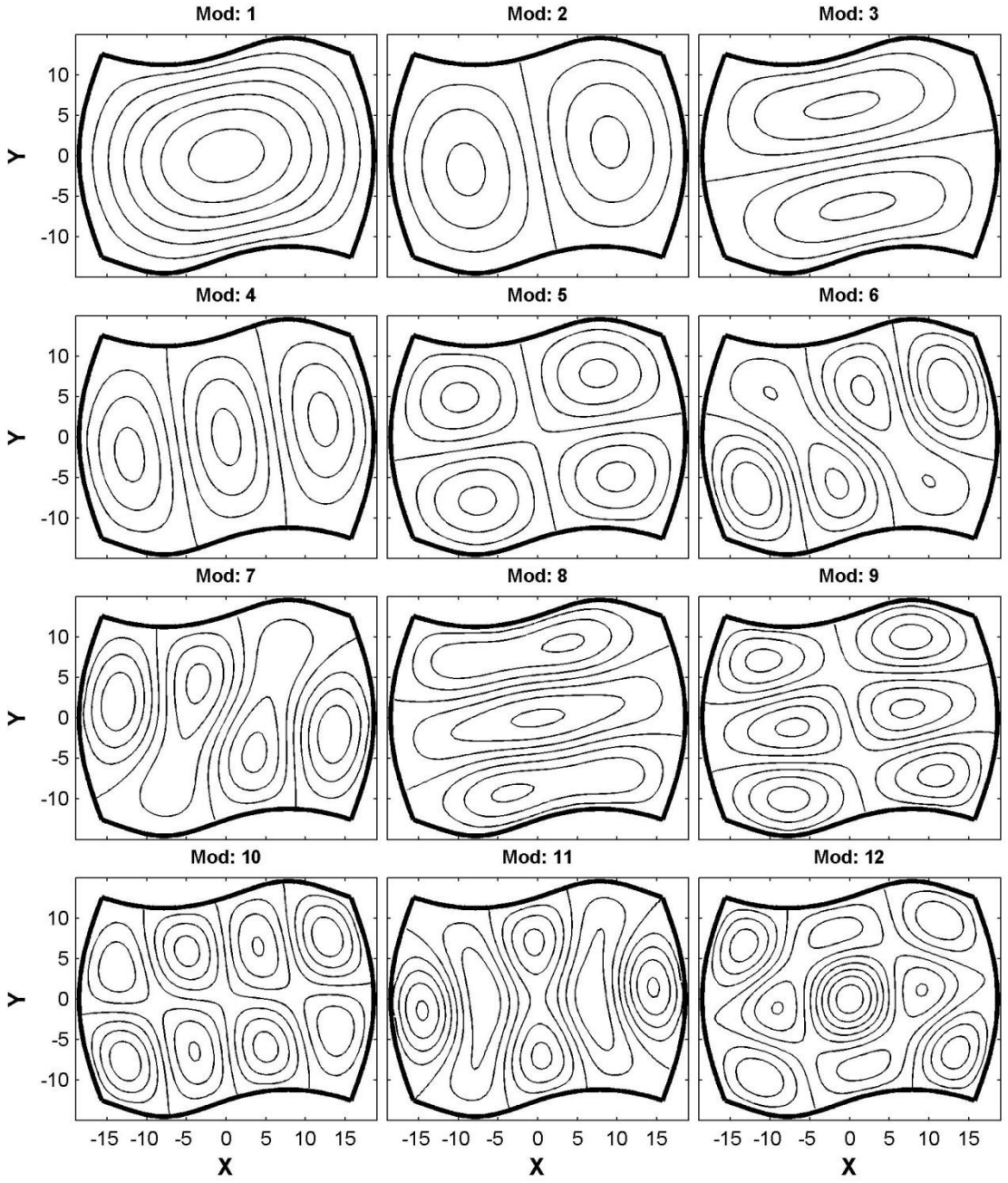
ϕ_k → The eigen functions of the 2D Laplacian Operator

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



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/OI (KF) Data Assimilation Equations

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

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

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

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

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

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

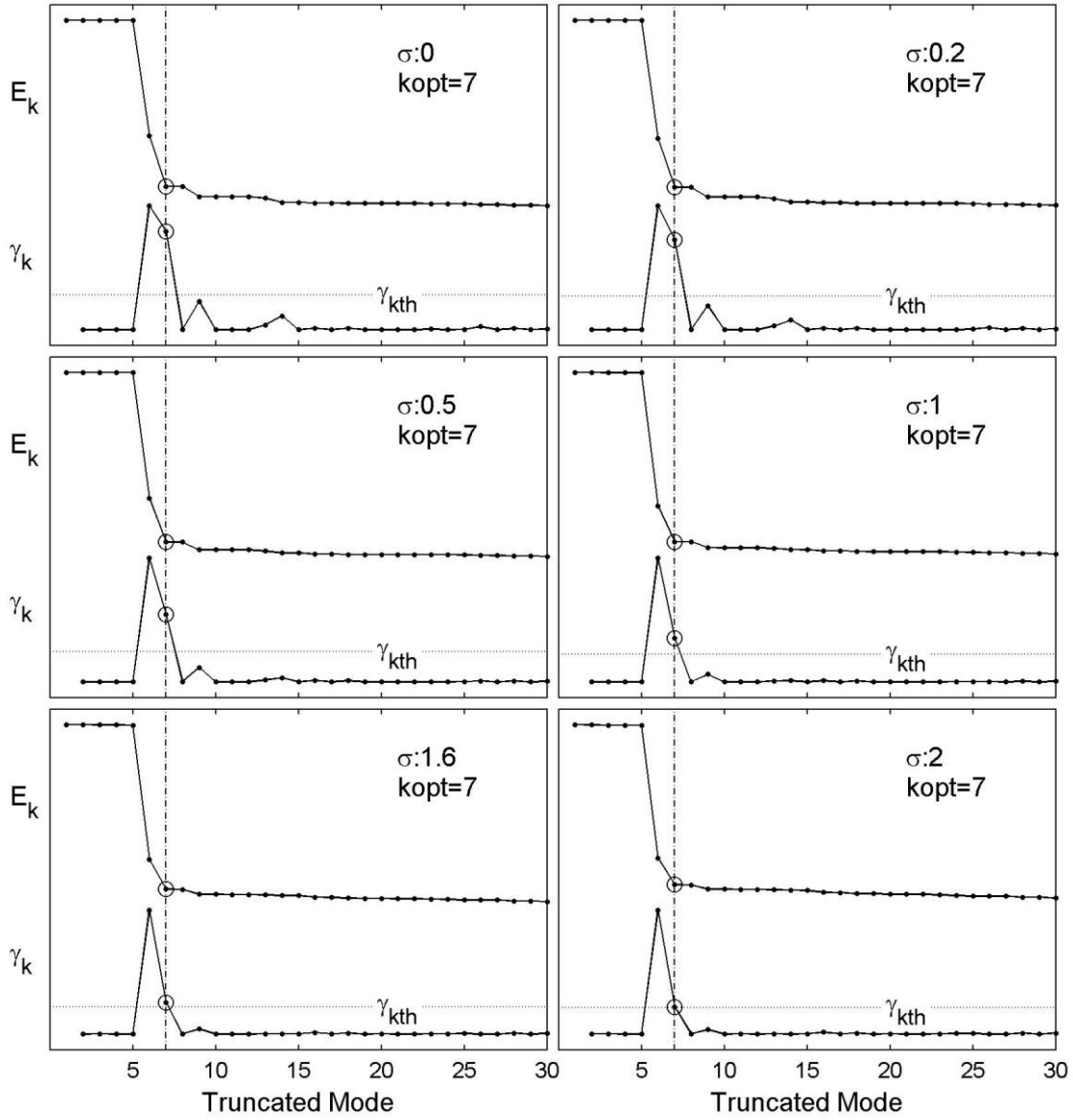
Two Issues in OSD

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

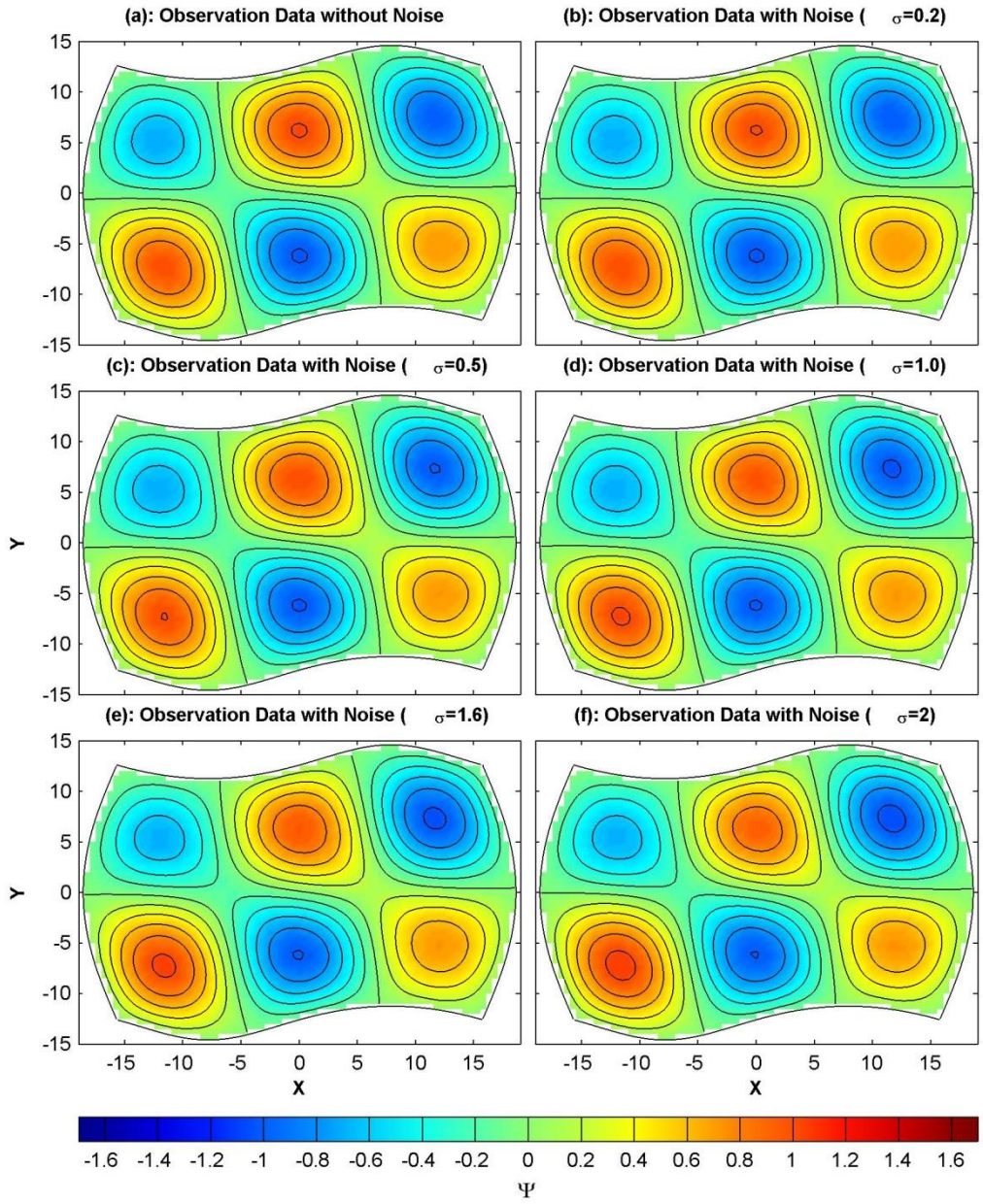
K = Kopt

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

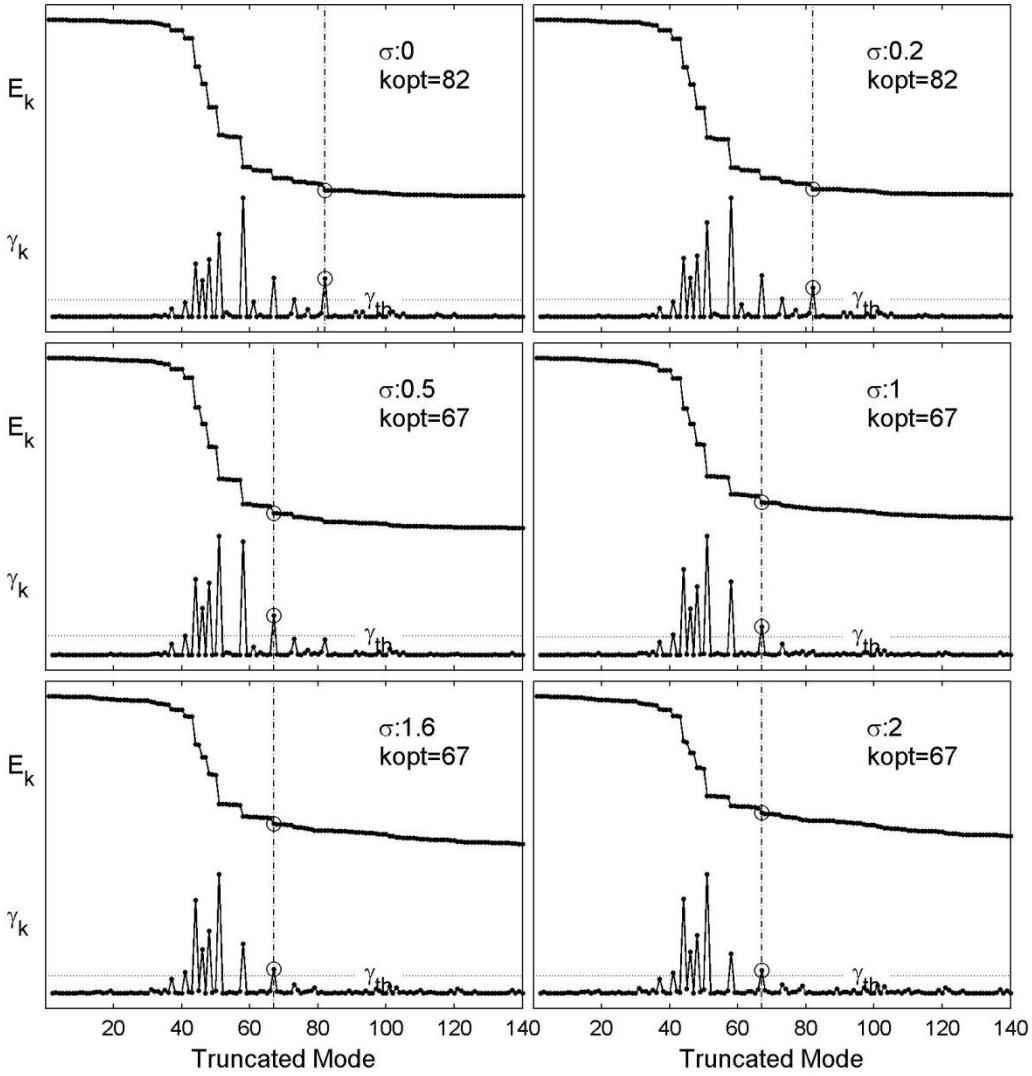
Mode truncation for large-scale “observational data”



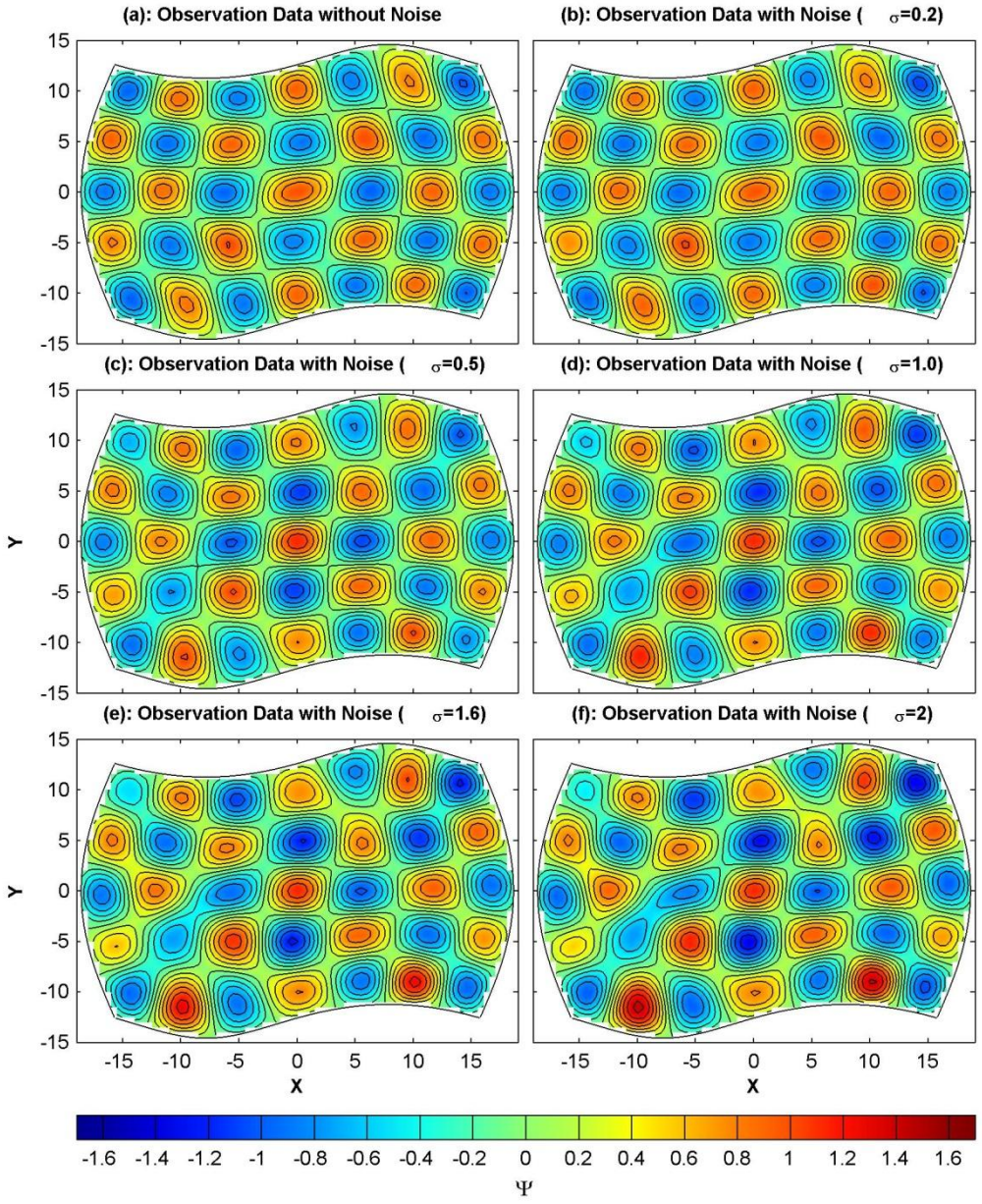
Analysis field ψ_a obtained by the OSD Method



Mode truncation for small-scale "observational data"



Analysis field ψ_a obtained by the OSD



Error Comparison Between OI and OSD

$$R^{OSD}(\sigma) = \sqrt{\frac{1}{N}\sum_{n=1}^{N}\left[\psi_a^{OSD}(\mathbf{r}_n,\sigma) - \psi_t(\mathbf{r}_n)\right]^2}$$

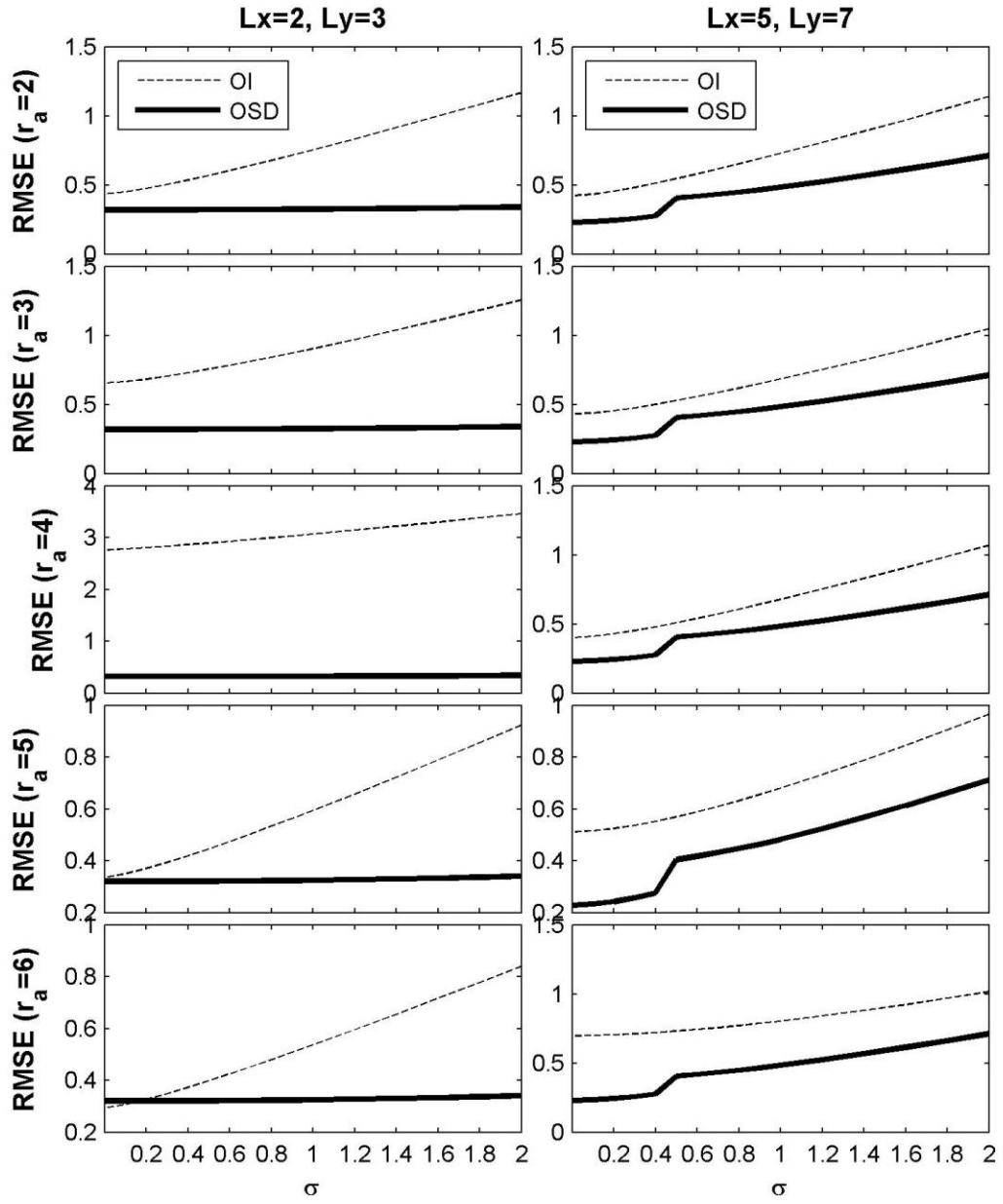
$$R^{OI}(\sigma, r_a) = \sqrt{\frac{1}{NP_bP_E}\sum_{r_b}\sum_{E_o}\sum_{n=1}^{N}\left[\psi_a^{OI}(\mathbf{r}_n,\sigma,r_a,r_b,E_o) - \psi_t(\mathbf{r}_n)\right]^2}$$

$$R^{OI}(\sigma, r_b) = \sqrt{\frac{1}{NP_aP_E}\sum_{r_a}\sum_{E_o}\sum_{n=1}^{N}\left[\psi_a^{OI}(\mathbf{r}_n,\sigma,r_a,r_b,E_o) - \psi_t(\mathbf{r}_n)\right]^2}$$

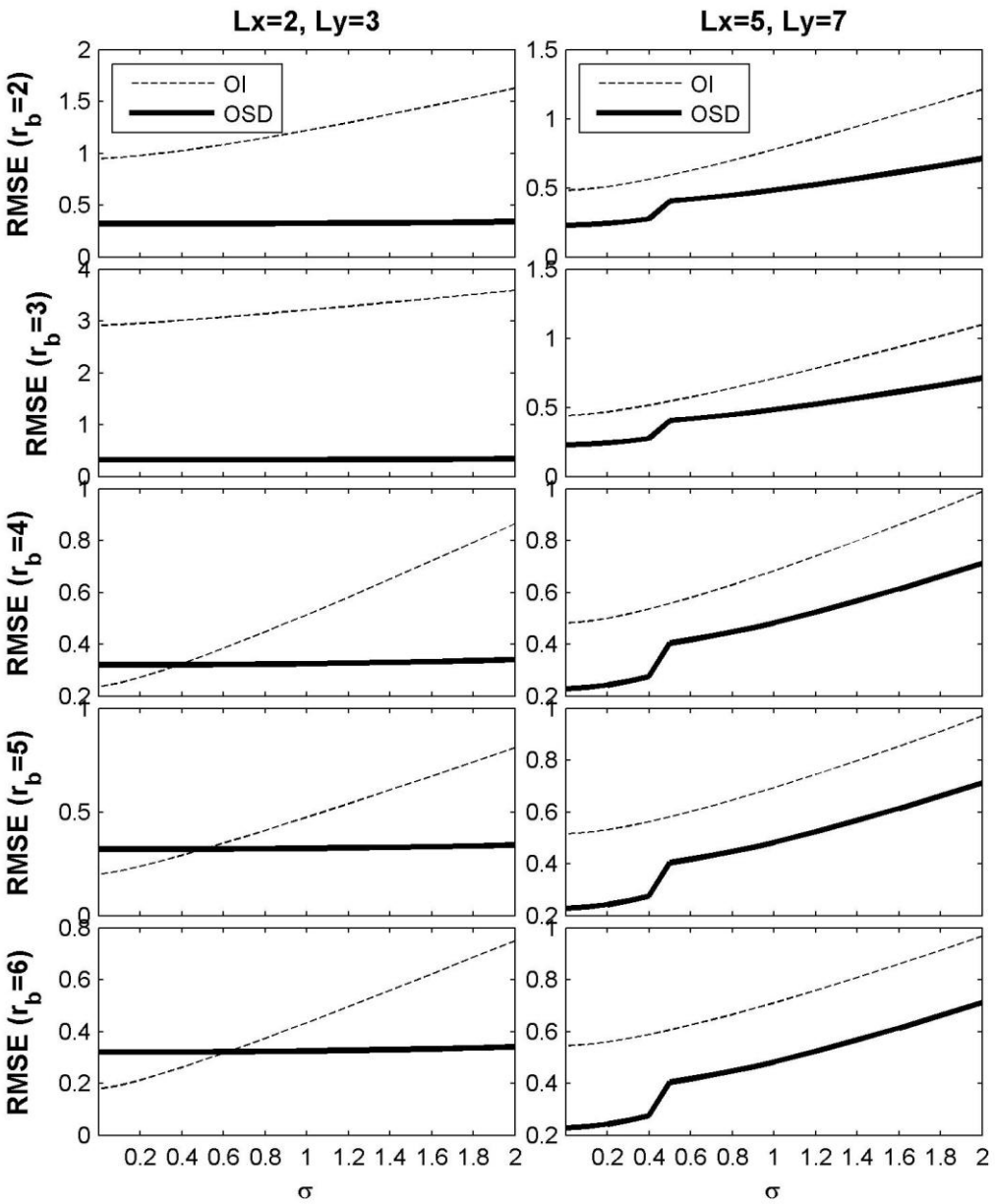
$$R^{OI}(\sigma, E_o) = \sqrt{\frac{1}{NP_aP_b}\sum_{r_b}\sum_{r_b}\sum_{n=1}^{N}\left[\psi_a^{OI}(\mathbf{r}_n,\sigma,r_a,r_b,E_o) - \psi_t(\mathbf{r}_n)\right]^2}$$

$(P_a, P_b, P_E) \rightarrow$ total numbers of (r_a, r_b, E_o)

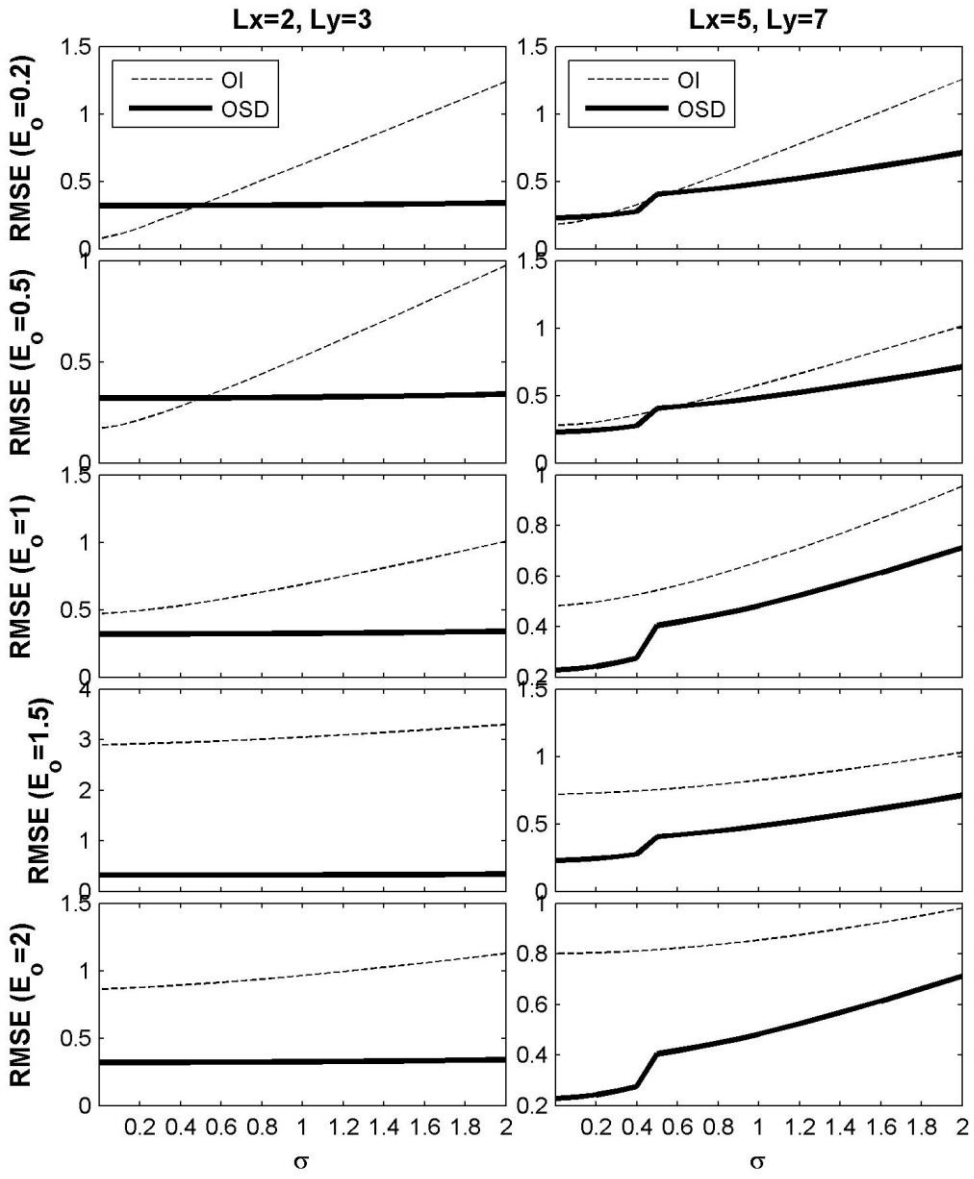
Comparison between $R^{OI}(\sigma, r_a)$ and $R^{OSD}(\sigma)$



Comparison between $R^{OI}(\sigma, r_b)$ and $R^{OSD}(\sigma)$



Comparison between $R^{OI}(\sigma, E_o)$ and $R^{OSD}(\sigma)$



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).

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

DBDB5

Dirichlet boundary condition at the southern boundary (Antarctic),

Newmann boundary condition elsewhere

