Analysis of Remotely Sensed Ocean Data by the Optimal Spectral Decomposition (OSD) Method

Peter Chu
National Ocean Analysis and Prediction Laboratory
Department of Oceanography
Naval Postgraduate School
Monterey, CA 93943 USA

***Abstract*— A new data analysis/assimilation scheme, optimal spectral decomposition (OSD), has been developed to reanalyze fields from noisy and sparse data in a domain with open boundary conditions using two scalar representations for a three-dimensional incompressible flow. The reanalysis procedure is divided into two steps: (a) specification of basis functions in the spectral decomposition from knowledge of boundary geometry and velocity and (b) determination of coefficients in the spectral decomposition for the circulation solving linear or nonlinear regression equations. The basis functions are the eigenfunctions of the Laplacian operator with mixed boundary conditions. The optimization process is used to obtain unique and stable solutions on the base of an iteration procedure with special regularization (the filtration. The capability is demonstrated using various examples.**

***Key Words*— O**ptimal spectral decomposition, Argo drifter, Lagrangian data, satellite data, rotation method

I. INRODUCTION

Sparse and noisy ocean data need to be reanalyzed before being assimilated into numerical models. Any field (temperature, salinity, or velocity) can be decomposed into generalized Fourier series using the Optimal Spectral Decomposition (OSD) method. The three dimensional field is then represented by linear combination of the products of basis functions (or called modes) and corresponding Fourier coefficients. If a rectangular closed ocean basin is considered, the basis functions are sinusoidal functions. If a realistic ocean basin is considered, the basis functions are the eigen-values of the three-dimensional Laplace operator with real topography. The Fourier coefficients are determined from observational data through solving a set of linear algebraic equations. Major benefit of using the OSD method is that the boundary conditions for the ocean variables (temperature, salinity, velocity) are always satisfied.

II. GENERALIZED FOURIER-SERIES EXPANSION

Let (**x**, *z*) be horizontal and vertical coordinates and *t* be time. A physical variable $c(\mathbf{x}, z, t)$ at depth z_k is decomposed using the generalized Fourier series [1-6]

$$c(\mathbf{x}, z_k, t) = A_0(z_k, t) + \sum_{m=1}^{M} A_m(z_k, t)\ \Psi_m(\mathbf{x}, z_k), \quad \mathbf{x} \in R(z_k) \tag{1}$$

where *M* is the truncated mode number, $\Psi_m(\mathbf{x}, z_k)$ and $A_m(z_k, t)$ are the orthogonal basis functions (or called modes) and the spectral coefficients, respectively; $R(z_k)$ is the area bounded by the lateral boundary $\Gamma(z_k)$ at depth z_k. The basis functions $\{\Psi_m(\mathbf{x}, z_k)\}$ are eigen-functions of the horizontal Laplace operator with the basin geometry and certain physical boundary conditions. For temperature and salinity, the homogeneous Neumann boundary condition is taken at the solid boundary $\Gamma(z)$ (i.e., no heat and salt fluxes),

$$\nabla_h^2 \Psi_m = -\lambda_m \Psi_m, \quad \mathbf{n} \bullet \nabla_h \Psi_m |_\Gamma = 0, \quad m = 1, 2, ..., M, \tag{2}$$

where $\nabla_h^2 \equiv \partial^2/\partial x^2 + \partial^2/\partial y^2$, and n is the unit vector normal to $\Gamma(z)$. The basis functions $\{\Psi_m\}$ are independent of the data and therefore available prior to the data analysis. The OSD method has two important procedures: optimal mode truncation and determination of spectral coefficients $\{A_m\}$. After the two procedures, the generalized Fourier spectrum (3) is used to provide data at regular grids in space and time.

III. OPTIMAL MODE TRUNCATION

The optimal mode truncation number (M_{opt}) is defined as the critical mode number with the set of spectral coefficients $\{A_m\}$ least sensitive to observational data sampling and noise. For sample size of *P* and mode truncation of *M*, the spectral coefficients $\{A_m\}$ are estimated by the least square difference between observed and calculated values

0-933957-38-1 ©2009 MTS

$$J_{emp} = J(\tilde{A}_1,...,\tilde{A}_M, P, M)$$
$$= \frac{1}{P}\sum_{j=1}^{P}\left(c^{(j)} - \sum_{m=1}^{M}\tilde{A}_m(z,t)\Psi_m^{(j)}(\mathbf{x},z)\right)^2 \to \min \quad (3)$$

where the symbol "~" represents the estimated value at (**x**, t). For homogeneously sampled data with low noise and without systematic error, the empirical cost function J_{emp} should tend to 0 monotonically as M increases to infinity. The set of the spectral coefficients $\{A_m\}$ depends on the mode truncation M. Optimal estimation of $\{A_m\}$ is equivalent to the determination of M_{opt} [3,4]. A modified cost function [7],

$$J \le \frac{J_{emp}}{1-\sqrt{\left[M\left(\ln\frac{P}{M}+1\right)-\ln(1-\tau)\right]/P}}, \quad (4)$$

is used to determine the optimal mode truncation M_{opt}. Here, τ is the probability of

$$\left|J - J_{emp}\right| \to 0 \quad \text{as } M \text{ increases.}$$

Usually, for a regional sea such as the Black Sea, M_{opt} is 30–50 for the basin-scale (~300 km) variability and 150 for the mesoscale (~20 km) variability [5]. For sparse and noisy data, it is difficult to get reliable and stable estimates of all the necessary spectral coefficients, but the first few spectral coefficients $\tilde{A}_0(z_k,t), \tilde{A}_1(z_k,t)$, …, $\tilde{A}_m(z_k,t)$ are reliable and stable.

IV. ROTATION MATRIX METHOD FOR REGULARIZATION

Determination of the spectral coefficients is achieved by solving a set of linear algebraic equations of $\{\tilde{A}_m(z,t)\}$ that are obtained from the optimization procedure (1) and (4),

$$\mathbf{A}\hat{\mathbf{a}} = \mathbf{QY}. \quad (5)$$

where $\hat{\mathbf{a}}$ is the estimated state vector (L-dimensional) for the exact state vector **a**; **A** is a $P \times L$ coefficient matrix; **Q** is a $P \times P$ square matrix ($P > L$); **Y** is a P-dimensional observation vector, consisting of a signal $\overline{\mathbf{Y}}$ and a noise **Y'**,

$$\mathbf{Y} = \overline{\mathbf{Y}} + \mathbf{Y}'.$$

Due to the high level of noise contained in the observations, the set of algebraic equations is ill-posed and needs to be solved by a regularization method that requires: (a) stability (robustness) even for data with high noise, and (b) the ability to filter out errors with a-priori unknown statistics. The two known matrices **A** and **Q** are determined by the physical process or field. Let $\|\ldots\|$ be the Euclidean norm and

$$\eta_1 = \frac{\|\mathbf{QY}'\|}{\|\mathbf{Q}\overline{\mathbf{Y}}\|}, \quad \eta_2 = \frac{L}{P},$$
$$\eta_3 = \frac{\max(\text{sigular values of } \mathbf{A})}{\min(\text{sigular values of } \mathbf{A})}, \quad (6)$$

be the noise-to-signal ratio, dimension ratio, and condition number of the matrix A. For a particular system, η_2 is given. Usually, η_1 and η_3 are large (called "imperfect"),

$$\eta_1 \ge 1, \qquad \eta_3 >> 1,$$

which makes (5) difficult to solve. A new rotation method for $\eta_2 < 1$ is developed to change (5) into a new system with possibly minimum coefficient matrix and noise-to-signal ratio without *a-priori* knowledge of noise statistics. Nonsingular orthogonal transformation is conducted through multiplication of (5) by a plane rotation matrix **S** from the left,

$$\mathbf{SA}\hat{\mathbf{a}} = \mathbf{SQY}, \quad (7)$$

which changes the coefficient matrix and the source term from (**A**, **QY**) to (**SA**, **SQY**) and provides the opportunity to minimize the imperfection of the new system (7),

$$\tilde{\eta}_3^2\,(1 + \tilde{\eta}_1^2) \to \min, \quad (8)$$

Where

$$\tilde{\eta}_1 \equiv \|\mathbf{SQY}'\| / \|\mathbf{SQ}\overline{\mathbf{Y}}\|, \quad \tilde{\eta}_3 \equiv \|\mathbf{SQ}\overline{\mathbf{Y}}\| / \|\mathbf{a}\|. \quad (9)$$

Minimization (8) leads to

$$J_1 = \|\mathbf{A}\|^2 - \tilde{\eta}_3^2\left[1 + \frac{2\,(\mathbf{SQ}\overline{\mathbf{Y}} * \mathbf{SQY}')}{\|\mathbf{SQ}\overline{\mathbf{Y}}\|^2} + \tilde{\eta}_1^2\right]$$
$$\to \max, \quad (10)$$

which is the procedure to obtain minimum values of $\tilde{\eta}_1$ and $\tilde{\eta}_3$ without $\|\mathbf{a}\|^2$. Here, the symbol '*' indicates the scalar product in the Euclidean space. The new transformed system (7) can be solved by usual algebraic methods such as the Gauss method.

V. ARGO DATA ANALYSIS

Between November 2003 and January 2005, over 56000 float days (cumulative) of data were collected in the North Atlantic (10^oN-60^oN) in general at three

parking depths: 1000 m, 1500 m and 2000 m. The floats parking at 2000 m, depths shallower than 1000 m, and unknown depths are excluded from the analysis. Temperature at 950 m and trajectories at 1000 m and 1500 m are extracted from all the existing Argo floats. The data from 1000 m and 1500 m were grouped together to represent the mid-depth.

The measurement cycle of an Argo profiling float includes four stages: ascending, surface drifting, diving and deep drifting. The Argo float can only get its position fixings while it ascends to the sea surface. The vector between two consecutive surface positions during the deep drifting divided by the time interval is taken as the mid-depth velocity vector. When the Argo float is diving, ascending and drifting below the sea surface, no data can be transmitted to the ground stations in real time. Velocity field after the first step analysis shows noisy circulation patterns (Figs. 1 a, b) with large spatial gaps (from 230 km to 800 km).

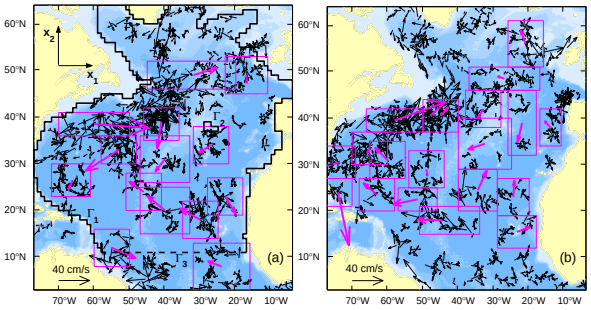

Fig. 1. Circulation velocities (tiny arrows) estimated from the original ARGO float tracks at 1000 m for (a) Dec 2003–Mar 2004 and (b) Aug 2004–Nov 2004. The figure scale is given for tiny arrows. Red arrows are circulation velocities obtained by averaged over appropriate bins [8].

Uncertainty in the Argo float data causes errors in the velocity field. First, the data extracted from the floats parking at two different levels: 1000 m and 1500 m and grouped together to represent the mid-depth (1000 m). This neglects the vertical shear. Second, the vertical shear causes increase or decrease of the distance between the points of ascending from and diving to the parking depth. Third, the sequence of float trajectory segments only approximates the real Lagrangian paths. Fourth, preliminary computations (not included here to be published in a separate paper) show that high resolution elements of circulation in the western North Atlantic, such as the northern re-circulation gyre and the Deep Western Boundary Current (DWBC) are also revealed by the Argo floats. For example Figs. 1 a, b clearly show the existence of DWBC. However, such a resolution is not available for the whole North Atlantic.

The high-energetic mesoscale eddies and narrow boundary currents are classified as "noise" and removed from the analysis. Fifth, there are large spatial gaps in Argo float trajectories. The temperature observation has higher quality than the velocity observation since the former has (a) higher resolution and (b) less spatial gaps. Fig. 2 shows typical monthly coverage of observations. Interested readers can find the detailed discussion on the navigation errors and measurement errors caused by temperature sensors in http://argo.jcommops.org.

Three quality control steps are performed to identify and remove temperature profiles corrupted by large measurement errors. (1) Explicitly absurd profiles were removed after a visual inspection. (2) A portion of temperature profiles, which were outside of the prescribed accuracy of the climatic data [9], were also excluded from the further analysis. (3) Temperature snapshots computed by the Optimal Spectral Decomposition (OSD) method [3,10] should not show explicit outbreaks in temperature structure. Any temperature profile contributing to such outbreaks was a subject to removing.

A parking depth for each Argo float was extracted from a "meta" file as the variable "PARKING_PRESSURE" (http://www.Argo.ucsd.edu). To control the parking depth the variable "PRES" (a pressure measured along the float trajectory) from a file containing float trajectory data, was used. Most floats launched in the area of interest have measured this variable.

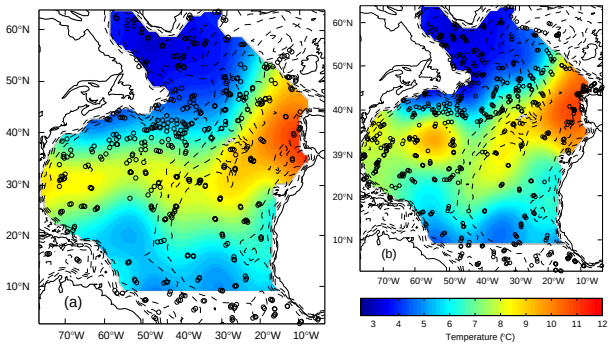

Fig. 2. Monthly temperature at 950 m for (a) Feb 2004 and (b) Dec 2004. Circles are points where temperature was measured [8].

Identification of long Rossby wave of tropical North Atlantic at mid-depth (~1000 m) from the Argo track data is taken as an example to demonstrate the usefulness of the OSD method to reanalyze sparse and noisy ocean data [11]. Argo float data (subsurface tracks and temperature profiles collected from March, 04 through May, 05) are used to detect signatures of long Rossby waves in velocity of the currents at 1000 m depth and temperature, between the ocean surface and 950 m, in the zonal band of 4°N -24°N in the Tropical North Atlantic. Different types of long Rossby waves (with the characteristic scales between 1000 km and 2500 km) are identified in the western [west of the Mid-Atlantic Ridge (MAR)] and eastern [east of the MAR] sub-basins.

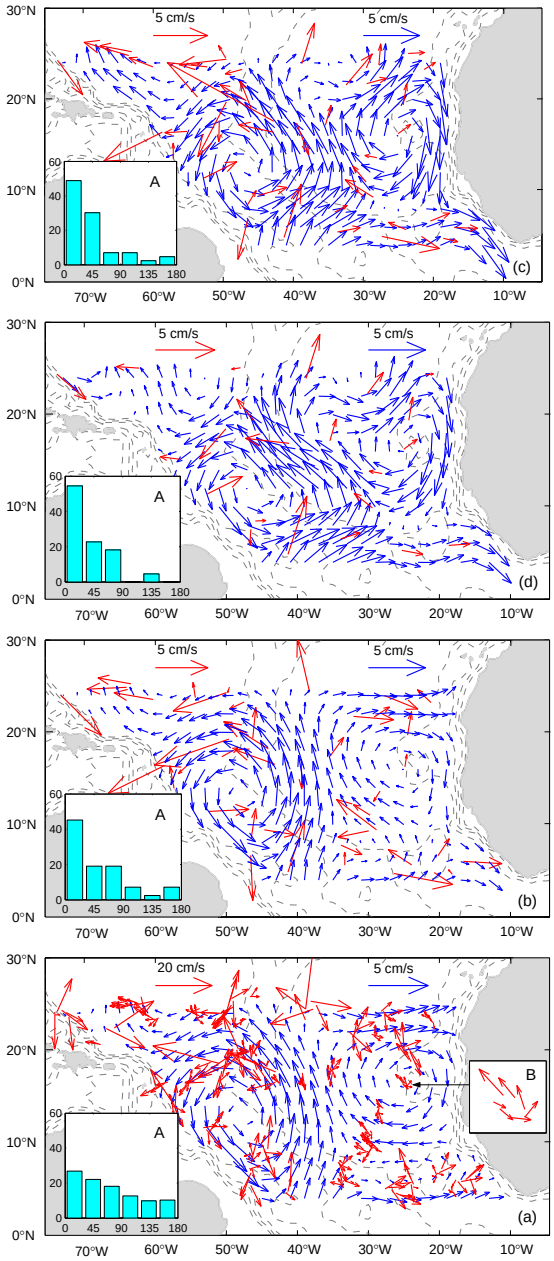

Fig. 3. Sensitivity of the reconstructed circulation patterns to filtration of the original data: OSD is applied to (a) the original data (November-December, 04); (b) the original data (November-December, 04) filtered with a 2-month window; (c) the original data (October-December, 04) filtered with a 3-month window; (d) the original data (October-December, 04) filtered by $4^o \times 4^o$ bin averaging; Blue and red arrows correspond to the reconstructed circulation and original/filtered data. For α -histograms (inserts A) the x-axes is the angle α and the y-axes is the number of comparisons (%) [10].

Current velocities computed along the original (non-smoothed) Argo tracks in November-December, 2004 are shown in Fig. 3 as red arrows. A strong contribution from intensive eddies, such as that shown in inset B, narrow jets and measurement errors is clearly identified here. Visually, the velocity pattern corresponding to the original data looks quite chaotic, and there are 500-600 km spatial gaps in observation coverage. To understand the reconstruction skill for such data we applied three criteria: (1) the formal mean square error (the reconstruction error) computed by the "laminar ensemble" technique, (2) statistics of angle (α) between the reconstructed and observed velocity at float locations, and (3) stability degree of the reconstructed snapshot on observation sampling. Rossby wave propagation is also identified from reanalyzed temperature field at 550 m depth (Fig. 4).

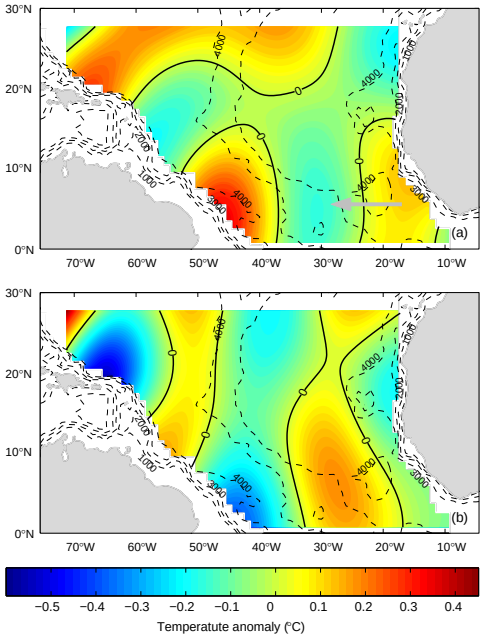

Fig. 4. Spatio-temporal structure of semi-annual temperature anomaly at 550 m depth: (a) May 15th and (b) June 29th 2004. Here, the gray arrow shows the direction of the positive temperature anomaly [10].

VI. GLOBAL SURFACE CURRENT VECTOR DATA

Near real time ocean surface current (0 – 30 m depth) analyses – real time (OSCAR) has been produced by NOAA on 5-day interval from sea surface height (TOPEX/Poseidon), surface vector wind (SSM/I) and sea surface temperature (AVHRR 1 in situ measurements) with a diagnostic model using quasi-linear and steady physics and the absolute velocity using the mean dynamic height inferred from the *World Ocean Atlas* (WOA). This satellite-derived global ocean surface current vector data, representing combined geostrophic, Ekman, and Stommel shear dynamics, has been undergone thorough quality control procedures such as error analysis using drifter, ship, and moored observations. Detailed information can be found at the website: http://www.oscar.noaa.gov/. During constructing the OSCAR data, the lateral boundary condition is never considered. This leads to the absence of major western boundary currents such as Gulf Stream, Kuroshio, Somali Current, California Current, Brazil Current, etc. (Fig. 5b). After reanalyzing the OSCAR data with the OSD method, the surface ocean current field is shown in Fig. 5a (June 15, 2007). Comparison

between Figs.5a and 3b shows the powerfulness of the OSD method.

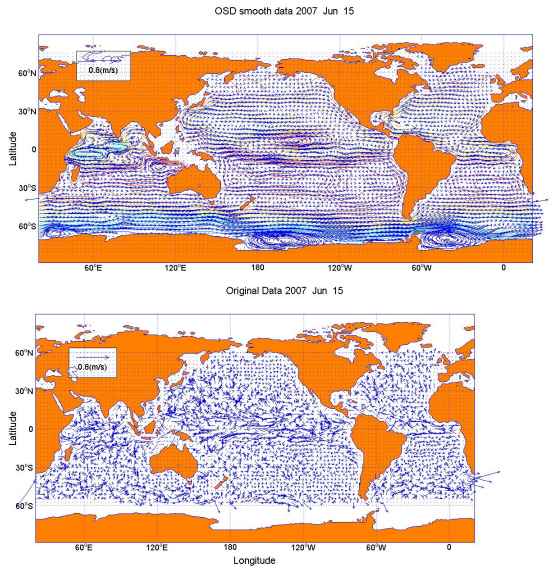

Fig. 5. Comparison of global surface ocean currents on June 15, 2007 from OSCAR data (a) with OSD and (b) without OSD.

VII. LANGRANGIAN DRIFTER DATA

The OSD method was also to process the current velocity data collected from the Texas-Louisiana Shelf Physical Oceanography Program (LATEX-A) and first segment of the Surface Current and Lagrangian Drift Program (SCULP-1) into temporally-spatially gridded data. With this data set, the TLCS current structure and reversal may be identified. In conjunction with the local surface wind data, the wind effect on the TLCS synoptic current reversal may also be identified.

Data for the TLCS fall/winter current reconstruction were collected from 31 near-surface current meter moorings during LATEX-A at 10-15 m depths (Cho et al. 1998) from April 1992 to December 1994 (Fig. 6) and from drifting buoys deployed during SCULP-1 from October 1993 to July 1994. The current meter moorings sample the current velocity (speed and direction), temperature, and salinity every 5-min to 2-hour (mostly 30 min). The drifting buoys record the location at various times with inhomogeneous area coverage. The data were interpolated into a uniform temporal grid with time step of 3 hours.

Meteorological data are from 7 buoys (represented by squares in Fig. 6) from the National Data Buoy Center (C-MAN) in the LATEX-A area. Wang et al. (1998) have estimated the zero-crossing spatial scales of winds (~350 km) and other meteorological parameters over the TLCS shelf after analyzing extensively the meteorological data collected during the LATEX-A Program observation period. Since the scale of the atmospheric systems is larger than the scale of the continental shelf, the horizontal mean wind speed and direction from 7 buoys are used for analysis.

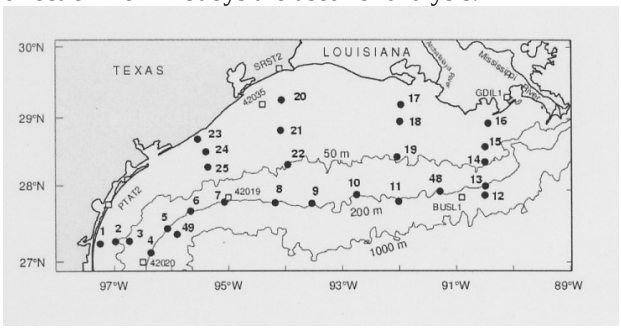

Fig. 6. Geography and topography of Texas-Louisiana continental shelf, LATEX-A current meter stations (represented by the symbol ' ● '), and meteorological buoys (represented by the symbol ' □ ') [5].

Trajectories of 77 SCULP-I drifting buoys show current reversals during January 2-7, 1994 (Fig. 7). The OSD reconstructed horizontal velocity vector field shows dynamical characteristics of the synoptic current. The flow pattern on 30 December 1993 is characterized as a cyclonic gyre with a strong westward TLCS flow as the northern flank (Fig. 8a). This current completely reverses to the east on 3 January 1994 (total reversal, Fig. 8b), and the cyclone re-occurs (6 January 1994) with the westward current at the north-central shelf and the eastward currents at the western shelf and shelf break (partial reversal, Fig. 8c). The total reversal has a width up to 200-m isobath (depth for shelf break), and the partial reversal has a width smaller than 200-m isobath. During that period (30 December 1993 to 6 January 1994), southwesterly winds prevail, and no evident offshore eddy appears at the shelf edge.

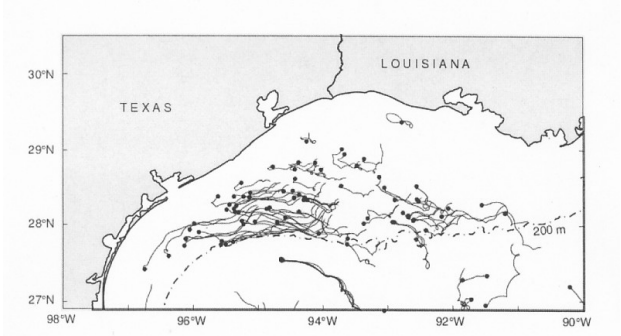

Fig. 7. TLCS current reversals detected from SCULP-1 buoy trajectories during January 2-7, 1994. The black dots show the starting positions of buoys [5].

Recurrence of the TLCS current reversals is estimated using the OSD reconstructed velocity fields (230 days of duration). A period T (from 5 to 40 days) is selected as the time window. The number of total (or partial) reversals is counted in each T day window. Let (n_0, n_1, n_2, …) be the numbers of 0-current reversal, 1-current reversal, 2-current reversals, and m be the all realizations for a given T day window. The probabilities for 0-, 1-, and 2-current reversals are calculated by

$$P_0(T) = \frac{n_0}{m},\ P_1(T) = \frac{n_1}{m},\ P_2(T) = \frac{n_2}{m}, \qquad (11)$$

where P_0(T), P_1(T), and P_2(T) depend on the period T (Fig. 5). With T larger than 20 days, the probability for zero current reversal is less than 0.2. With T of 15 days, the probability for one reversal reaches 0.5. The probability $P_k(T)$ is fitted to the Poisson distribution (Fig. 9)

$$P_k(T) = \frac{1}{k!}(\mu T)^k \exp(-\mu T),\ k = 0, 1, 2; \qquad (12)$$

where μ = 0.08 day^{-1} is the mean rate of current reversal.

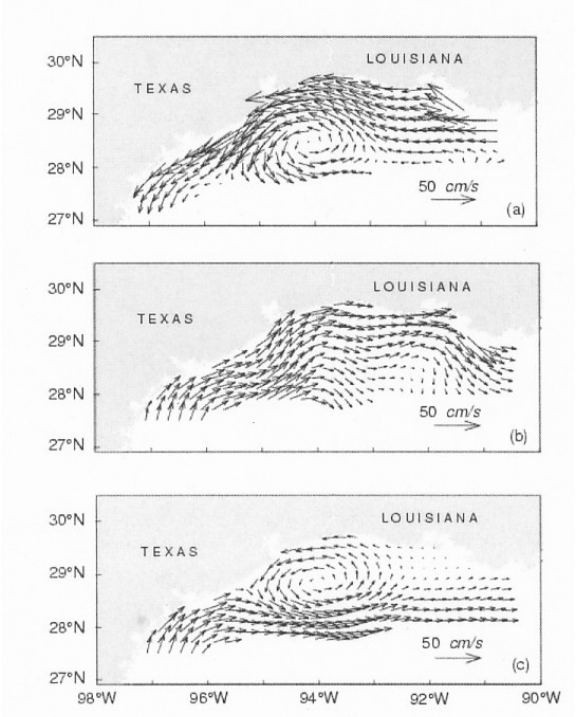

Fig. 8. OSD reconstructed circulation: (a) cyclonic gyre on December 30, 1993, (b) total reversal on January 3, 1994 , and (c) partial reversal on January 6, 1994 [5].

VIII. CONCLUSIONS

Without knowing the background field and decorrelation scale, the OSD method can process sparse and noisy data. This method has three components: (1) determination of the basis functions, (2) optimal mode truncation, and (3) determination of the Fourier coefficients. Determination of basis functions is to solve the eigen-value problem. Chu et al. [3,4] also developed a theory to obtain the basis functions with open boundaries. The basis functions are only dependent on the geometry of the ocean basin, not dependent on the oceanic variables. This is to say, no matter which variable (temperature, salinity, or velocity) is concerned, the basis functions are the same, and can be pre-determined before the data analysis. For data without error, the more the modes, the more the accuracy of the processed field. For data with error, this rule of the thumb is no longer true. Inclusion of high-order modes leads to increasing error. The Vapnik variational principal is used to determine the optimal mode truncation. After the mode truncation, optimal field estimation is to solve a set of a linear algebraic equation of the Fourier coefficients. This algebraic equation is usually ill-posed. The rotation method is employed to change the matrix of the algebraic equation from ill-posed to well-posed such that a realistic set of the Fourier coefficients are obtained. The OSD method is a power tool to process temperature, salinity, and velocity data from Lagrangian and Eulerian, remotely sensed and in situ observed ocean data.

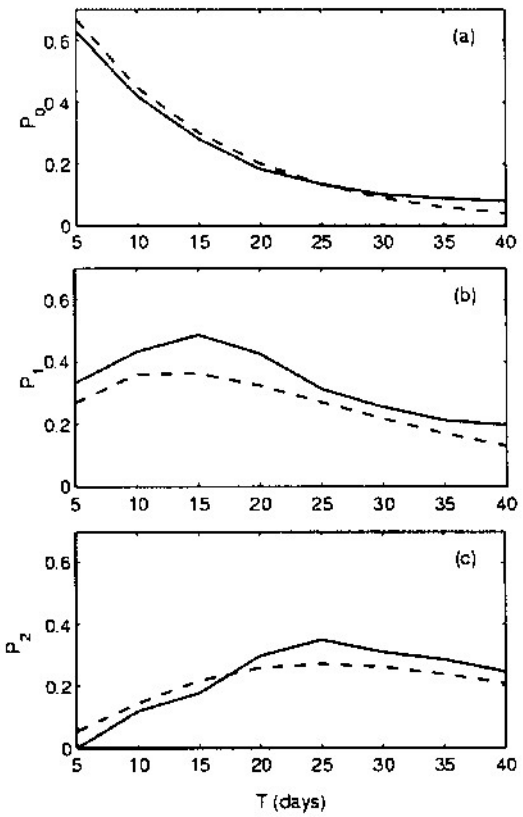

Fig. 9. Empirically (dashed curve) and theoretically (Poisson distribution, solid curve) estimated recurrence probabilities: (a) $P_0(T)$, (b) $P_1(T)$, and (c) $P_2(T)$ as functions of duration T [5].

ACKNOWLEDGMENT

This work was sponsored by NOAA/NODC.

REFERENCES

[1] Chu, P.C., Fundamental circulation functions for the determination of open boundary conditions. Third Conference on Coastal Oceanic and

Atmospheric Prediction, American Meteorological Society, 389-394, 1999.

[2] Ivanov, L. M., and P.C. Chu, 2008: On stochastic stability of regional ocean models to finite-amplitude perturbations of initial conditions. Dynamics of Atmosphere and Oceans, in press

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

[4] Chu, P.C., L.M. Ivanov, T.M. Margolina, T.P. Korzhova, and O.V. Melnichenko, 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, 492-512, 2003.

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

[6] Chu, P.C., L.M. Ivanov, and T.M. Margolina, 2005: Seasonal variability of the Black Sea Chlorophyll-*a* concentration. Journal of Marine Systems, 56, 243-261.

[7] Vanpnik, V.N., 1982: Estimation of Dependencies Based on Empirical Data. Springer-Verlag, New York, 379 pp.

[8] Chu, P. C., L. M. Ivanov, O. V. Melnichenko, and R.-F. Li, 2008: Argo floats revealing bimodality of large-scale mid-depth circulation in the North Atlantic. *Acta Oceanologica Sinica*, **27** (2), 1-10

[9] Levitus, S., T.P. Boyer, M.E. Conkright, T. O' Brien, J. Antonov, C. Stephens, L. Stathoplos, D. Johnson, R. Gelfeld, NOAA Atlas NESDIS 18, World Ocean Database 1998: VOLUME 1: INTRODUCTION, U.S. Gov. Printing Office, Wash., D.C., 346 pp, 1998.

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

[11] Chu, P.C., L.M. Ivanov, O.V. 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.