Proceedings on 15th Symposium on Integrated Observing and Assimilation Systems for the Atmosphere, Oceans and Land Surface (IOAS-AOLS), American Meteorological Society, 23-27 January 2011, Seattle

ESTABLISHMENT OF NEAR-REAL TIME MONTHLY GRIDDED DATASET FROM GLOBAL TEMPERATURE AND SALINITY PROFILE PROGRAM (GTSPP) USING THE OPTIMAL SPECTRAL DECOMPOSITION (OSD)

Peter C. Chu [(1)], Charles Sun [(2)], Chenwu Fan [(1)]

[(1)] Department of Oceanography, Naval Postgraduate School, Monterey, CA 93943, USA, Email:pcchu@nps.edu
[(2)]NOAA/NODC, 1315 East-West Highway, Silver Spring, MD 20910, USA, Email: Charles.Sun@noaa.gov

ABSTRACT

A new data analysis/assimilation scheme, optimal spectral decomposition (OSD), has been developed to analyze fields from noisy and sparse raw data using two scalar representations for a three-dimensional incompressible flow. The analysis procedure is divided into three steps: (a) determining a set of basis functions (i.e., the eigen functions of the Laplacian operator) from the knowledge of boundary geometry and conditions, (b) optimizing the mode truncation, and (c) calculating the spectral coefficients using the observed Global Temperature and Salinity Profile Program (GTSPP) data from solving a large ill-posed algebraic equations with filtration procedure with special regularization. The capability is demonstrated using various examples.

***Key Words*— Optimal spectral decomposition, GTSPP, Argo drifter, Lagrangian data, satellite data, rotation method**

1. INTRODUCTION

Sparse and noisy ocean data need to be reanalyzed into grid points before being used for interpreting ocean conditions or being assimilated into numerical models for prediction or climate studies. 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 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. Neal-real time monthly gridded dataset was established from the Global Temperature and Salinity Profile Program (GTSPP) using the OSD method.

2. 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 (1) is used to provide data at regular grids in space and time.

3. OPTIMAL MODE TRUNCATION Font:

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

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

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} ([2]-[3]). 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 [4]. 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.

4. 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 algebraic equation (5) 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 [8],

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

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

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

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.

5. ERROR ESTIMATION

Using the OSD method, the velocity $\overline{\mathbf{u}}(\mathbf{x})$ and temperature $\overline{T}(\mathbf{x})$ are given by

$$\overline{\mathbf{u}}(\mathbf{x},t) = \overbrace{C\Psi_0 + \sum_{n=1}^{24} A_n \nabla \times \mathbf{k}\Psi_n(\mathbf{x})}^{\hat{\mathbf{u}}} + \tilde{\mathbf{u}}(\mathbf{x}) + \mathbf{u}'(\mathbf{x}), \quad (11)$$

$$\overline{T}(\mathbf{x}) = \overbrace{T_0 + \sum_{l=1}^{48} D_l \Phi_l(\mathbf{x})}^{\hat{T}} + T'(\mathbf{x}). \quad (12)$$

Here, $(\mathbf{u}', T')$ represent observational error approximated as white noise. The component $\tilde{\mathbf{u}}$ is spatial correlated field (red noise-like signal). The quality of reconstructed data can be estimated through non-dimension noise ratio between $(\boldsymbol{\alpha})$ and $(\boldsymbol{\beta})$ defined as

$$\eta(\boldsymbol{\alpha},\boldsymbol{\beta}) = \frac{\|\boldsymbol{\alpha}\|_{(P)}}{\|\boldsymbol{\beta}\|_{(P)}}, \quad (13)$$

which leads to

$$\eta(\mathbf{u}', \hat{\mathbf{u}}) \sim 7, \ \eta(\mathbf{u} - \mathbf{u}', \hat{\mathbf{u}}) \sim 1.8, \ \eta(T', T - T') \sim 0.1. \quad (14)$$

Clearly, noise strongly dominates the velocity field, but not the temperature field. Therefore, at least reasonable results for circulation are possible only if noise has the special features.

Chu et al. ([2], [3], [8]) pointed out that the OSD method effectively filtrates any noise if the noise is orthogonal to the reconstructed signal. Our computations show that the angle θ between the vectors $\mathbf{u}'$ and $\hat{\mathbf{u}}$ is approximately 89^o since

$$\cos(\theta) = \frac{(\hat{\mathbf{u}} \cdot \mathbf{u}')}{\|\hat{\mathbf{u}}\|_{(P)} \ \|\mathbf{u}'\|_{(P)}} \sim 0.0175, \quad (15)$$

which shows that $\mathbf{u}'$ is a white noise but not necessary Gaussian because

$$\Lambda = \frac{1}{P}(\hat{\mathbf{u}} \cdot \mathbf{u}') \to 0, \qquad \text{if } P \to \infty. \quad (16)$$

Increase of P from 6000 to 18000 leads to reduction of Λ from 0.100 to 0.012.

Decrease of P makes the angle (θ) departure from 90^o. Therefore, the four month average time period applied to the float velocity data analysis is a compromise between maximizing P and getting high temporal resolution of the reconstructed circulation. After projection of $\mathbf{u}'$ onto a plane defined by 24 basis functions $\{\Psi_n, n = 1, \ldots, 24\}$, the noise ratio $\eta(\mathbf{u}', \hat{\mathbf{u}})$ equals 0.07 only. Therefore, the white noise is not the main source of reconstruction error. The angle between $\hat{\mathbf{u}}$ and $\tilde{\mathbf{u}}$ is 80^o; $\eta(\tilde{\mathbf{u}}, \hat{\mathbf{u}})$ is about 0.31 with projection onto the plane defined by the basis functions. Clearly, that red noise-like signal gives maximum inaccuracy of estimate. For temperature, the ratio of $\eta(T', T - T_0)$ is a priory low. Therefore, temperature patterns can be reconstructed with more accuracy and more spatial resolution than circulation ones.

6. GTSPP

The following information was obtained from the website of the International Oceanographic Commission of UNESCO (IODE) http://www.iode.org/. GTSPP is a cooperative international project. It seeks to develop and maintain a global ocean Temperature-Salinity resource with data that are both up-to-date and of the highest quality possible. Making global measurements of ocean temperature and salinity (T-S) quickly and easily accessible to users is the primary goal of the GTSPP. Both real-time data transmitted over the Global Telecommunications System (GTS), and delayed-mode data received by the NODC are acquired and incorporated into a continuously managed database. Countries contributing to the project are Australia, Canada, France, Germany, Japan, Russia, and the United States. Canada's Marine Environmental Data Service (MEDS) leads the project, and has the operational responsibility to gather and process the real-time data. MEDS accumulates real-time data from several sources via the GTS. They check the data for several types of errors, and remove duplicate copies of the same observation before passing the data on to NODC. The quality control procedures used in GTSPP were developed by MEDS, who also coordinated the publication of those procedures through the Intergovernmental Oceanographic Commission (IOC).

The GTSPP handles all temperature and salinity profile data. This includes observations collected using water samplers, continuous profiling instruments such as CTDs, thermistor chain data and observations acquired using thermosalinographs. These data will reach data processing centres of the Program through the real-time channels of the IGOSS program or in delayed mode through the IODE system. Real-time data in GTSPP are acquired from the Global Telecommunications System

in the bathythermal (BATHY) and temperature, salinity & current (TESAC) codes forms supported by the WMO. Delayed mode data are contributed directly by

member states of IOC (Sun, 2008). Fig. 1 shows increasing of observational stations especially the TESAC due to input of Argo floats (Fig. 2).

The GTSPP went through quality control procedures that make extensive use of flags to indicate data quality. To make full use of this effort, participants of the GTSPP have agreed that data access based on quality flags will be available. That is, GTSPP participants will permit the selection of data from their archives based on quality flags as well as other criteria. These flags are always included with any data transfers that take place. Because the flags are always included, and because of the policy regarding changes to data, as described later, a user can expect the participants to disseminate data at any stage of processing. Furthermore, GTSPP participants have agreed to retain copies of the data as originally received and to make these available to the user if requested (GTSPP Working Group, 2010).

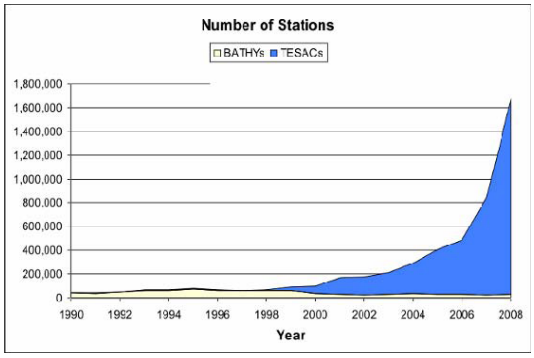

Fig. 1. The number of stations reported as BATHYs and TESACs (from Sun, 2008).

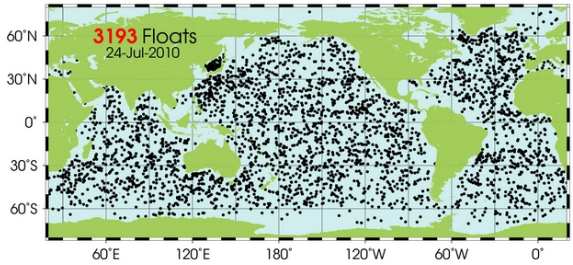

Fig. 2. World-wide distribution of Argo floats (from the website: http://www.argo.ucsd.edu/).

7. NEAL-REAL TIME MONTHLY TEMPERATURE, SALINITY, AND VELOCITY (STSV) DATA

Ocean temperature, salinity, and currents available at the present time do not have sufficient resolution to describe the variability in Meridional Overturning Circulation (MOC) and other climate change related processes. With our unique data access, strong experience in analysis of sparse and noisy ocean data (the OSD method), and our data system already in place at NOAA/NODC and the Naval Postgraduate School (NPS), we are producing monthly varying three dimensional temperature, salinity, and velocity fields using the Global Temperature-Salinity Profile Program (GTSPP) and Argo profile and track data together with the Navy's Master Observational Oceanographic Data Set (MOODS) since 1990, at spatial resolution comparable to or higher than the standard product ($1^o \times 1^o$) at the standard level as the NODC World Ocean Atlas (WOA) [11]. For illustration, we only show monthly (January, July) temperature (Fig. 3) and salinity (Fig.4) anomalies (relative to the climatological monthly mean) of the Pacific Ocean for 1990, 1996, 2002, and 2009. Interannual variability of temperature and salinity can be easily identified from the synoptic (monthly time interval) (T, S) data ([12-[16]).

Similarly, monthly varying velocity vector field at mid-depth (1000 m depth) can be constructed using the OSD method from the Argo trajectory data. Synoptic three dimensional velocity field can be easily obtained from the velocity at 1000 m depth (as the reference velocity) and the (T, S) data using the geostrophic calculations ([17]-[18]).

With the reanalyzed three dimensional ocean fields for two decades (1990-2010), we can indentify temporal and spatial variability of MOC and access the current status of MOC by computing the meridional heat transport. Since rapid changes in MOC could have implications for regional changes of climate, correlation analysis between our reanalyzed datasets (3D ocean fields) and the surface atmospheric data (such as NCEP reanalyzed wind stress, air-ocean heat and moisture fluxes) will improve understanding of the physical mechanisms behind fluctuations in the MOC. A recently identified phenomenon (mid-depth long baroclinic Rossby wave propagation) from the Argo track data [6] will be further investigated, especially its effect on the subtropical gyre and in turns on MOC. Correlation analysis between our reanalyzed data and the large-scale atmospheric variability (characterized by the NAO index) will further characterizing the potential impacts of the MOC fluctuations on regional and global climate. With the new knowledge added to the seasonal and interannual variability of the MOC and to the linkages between the tropical and mid-latitude oceans, we may identify physical mechanisms for transferring climatic signals from seasonal to inter-annual time scale and in turn to predict rapid climatic change. During this phase of work, we will focus on the role of propagation of long baroclinic Rossby waves on the MOC variability and in turn on the rapid climate change. At the same

time, the reanalyzed datasets will provide background ocean information to support the initial planning and development of large national program on MOC.

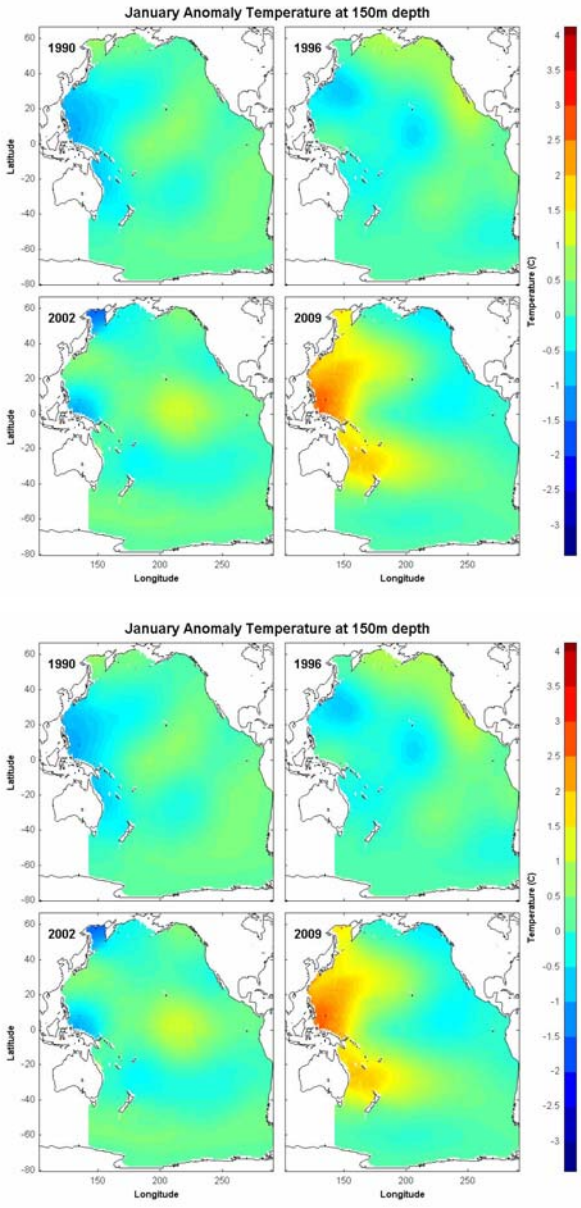

Fig. 3. Monthly (January, July) temperature anomalies (relative to the climatological monthly mean) of the Pacific Ocean for 1990, 1996, 2002, and 2009.

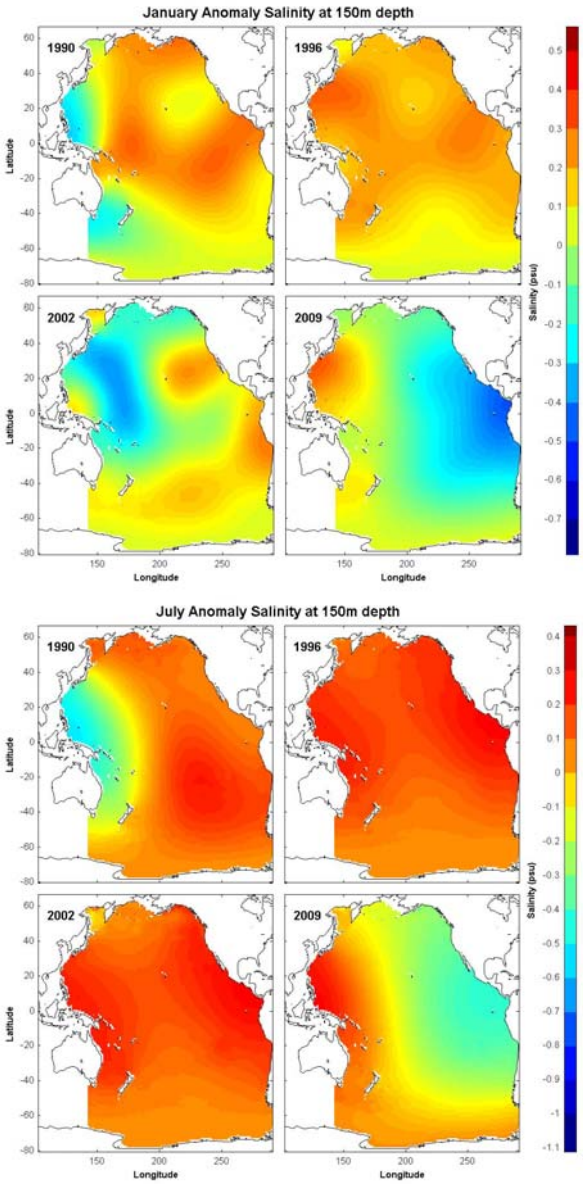

Fig. 4. Monthly (January, July) salinity anomalies (relative to the climatological monthly mean) of the Pacific Ocean for 1990, 1996, 2002, and 2009.

ACKNOWLEDGEMENTS**.** We acknowledge the sponsorship by NOAA/NODC and Central Weather Bureau, Taiwan, ROC.

8. REFERENCES

[1] Chu, P.C. (1999). Fundamental circulation functions for the determination of open boundary conditions. Third Conference on Coastal Oceanic and Atmospheric Prediction, American Meteorological Society, 389-394.

[2] Chu, P.C., Ivanov, L.M., Margolina, T.M., Korzhova, T.P., and Melnichenko, O.V. (2003a). Analysis of sparse

and noisy ocean current data using flow decomposition. Part 1. Theory. Journal of Atmospheric and Oceanic Technology, 20, 478 – 491.

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

[4] Chu, P.C., Ivanov, L.M., and Melnichenko, O.M. (2005a). Fall-winter current reversals on the Texas-Louisiana continental shelf. Journal of Physical Oceanography, 35, 902-91.

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

[6] Ivanov, L. M., and Chu, P.C. (2007) On stochastic stability of regional ocean models to finite-amplitude perturbations of initial conditions. Dynamics of Atmosphere and Oceans, 43 (3-4), 199-225.

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

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

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

[10] Chu, P.C., Sun, C., and Melnichenko, O.V. (2007). Optimal spectral decomposition (OSD) for Argo data analysis. Twenty Sixth International Union of Geodesy and Geophysics, Perugia, Italy, 2-13 July 2007.

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

[12] Chu, P.C., Fan, C.W., and Sun, C. (2008). Variability in Atlantic meridional overturning circulation and heat transport detected remotely from Argo floats. IEEE International Geoscience & Remote Sensing Symposium, Boston, Massachusetts, 6-11 July 2008.

[13] Chu, P.C., and Sun, C. (2008). Optimal spectral decomposition (OSD) method for analyzing sparse and noisy ocean data. Annual GTSPP Workshop, Honolulu, Hawaii, 27 October 2008 (invited).

[14] Chu, P.C., Sun, C., and Fan, C.-W. (2009). Detection of thermohaline structure and meridional overturning circulation above and below the ocean surface. Twenty First Conference on Climate Variability and Change, American Meteorological Society, Phoenix, Arizona, 11-15 January 2009.

[15] Chu, P.C., Sun, C., and Fan, C.W. (2009). Observations of meridional overturning circulation above and below the ocean surface. American Geophysical Union Fall Meeting, San Francisco, California, 14-18 December 2009.

[16] Chu, P.C., Fan, C.W., and Sun, C. (2010). Meridional Overturning Circulation (MOC) Detected from In-Situ and Argo Measurements. 42nd International Liege Colloquium on Ocean Dynamics, Liege, Belgium, April 26-30, 2010.

[17] Chu, P.C., Fan, C.W., and Sun, C. (2010). Changes of heat and freshwater contents in the world oceans (1990 – 2009) identified from in situ and Argo measurements. 42nd International Liege Colloquium on Ocean Dynamics, Liege, Belgium, April 26-30, 2010.

[18] Chu, P.C., and Sun, C. (2010). Quality control on ocean temperature and salinity data. Global Temperature and Salinity Profile Program Workshop, IODE Program Office, Oostende, Belgium, 5-7 May 2010.