Methods for the Reconstruction of Vertical Profiles from Surface Data: Multivariate Analyses, Residual GEM, and Variable Temporal Signals in the North Pacific Ocean

BRUNO BUONGIORNO NARDELLI AND ROSALIA SANTOLERI

Istituto di Scienze dell’Atmosfera e del Clima, CNR, Rome, Italy

(Manuscript received 17 June 2004, in final form 10 February 2005)

ABSTRACT

Different methods for the extrapolation of vertical profiles from sea surface measurements have been tested on 14 yr of conductivity–temperature–depth (CTD) data collected within the Hawaii Ocean Time-series (HOT) program at A Long-Term Oligotrophic Habitat Assessment (ALOHA) station in the North Pacific Ocean. A new technique, called multivariate EOF reconstruction (mEOF-R), has been proposed. The mEOF-R technique is similar to the previously developed coupled pattern reconstruction (CPR) technique and relies on the availability of surface measurements and historical profiles of salinity, temperature, and steric heights. The method is based on the multivariate EOF analysis of the vertical profiles of the three parameters and on the assumption that only a few modes are needed to explain most of the variance/covariance of the fields. The performances of CPR, single EOF reconstruction (sEOF-R), and mEOF-R have been compared with the results of residual GEM techniques and with ad hoc climatologies, stressing the potential of each method in relation to the length of the time series used to train the models and to the accuracy expected from planned satellite missions for the measurement of surface salinity, sea level, and temperature. The mEOF-R method generally produces the most reliable estimates (in the worst cases comparable to the climatologies) and seems to be slightly less susceptible to errors in the surface input. Multivariate EOF analysis of HOT data also gave by itself interesting results, being able to discriminate the three major signals driving the temporal variability in the area.

1. Introduction

In recent years, different studies have investigated the coupling between surface measurements, integral quantities [such as geopotential thickness, i.e., steric height (SH), or vertical acoustic travel time (τ)] and the vertical structure of the ocean. The objective of these studies is the reconstruction of vertical profiles of hydrological parameters from surface and integrated measurements that can be obtained from spaceborne remote sensing instruments and/or from moored instrumentation, such as echo inverted sounders (IESs). This extrapolated information can be used to investigate the internal ocean dynamics on a global scale with a higher temporal coverage than that obtainable with traditional techniques and to better initialize numerical predictive models adopted for operational uses [for a wide review on data assimilation, see De Mey (1997)].

Passive infrared/microwave sensors provide measurements of surface parameters [e.g., sea surface temperature (SST)], while active microwave sensors (more exactly radar altimeters) give us a sort of integrated measure of the density along the water column. In fact, the sea surface height (SSH) variations are given by a combination of volume and total mass variations at a given location. Similarly, IESs provide a measure of the round-trip acoustic travel time, which is clearly related to the distribution of density along the water column (τ is tightly related to the surface steric height, being defined as twice the integral along the water column of the specific volume divided by the product of gravitational acceleration and sound speed). Consequently, a strong effort has been made in the last years to propagate on the vertical the information from altimetric SSH data, from combined SST and SSH measurements, and from IES data, often starting from an analysis of the correlation between in situ temperature and steric height measurements.

© 2005 American Meteorological Society
Dynamical (e.g., Hurlburt 1984), variational (e.g., Thacker and Long 1988), statistical (e.g., Hurlburt et al. 1990; Carnes et al. 1990, 1994; Gavart and De Mey 1997; Pascual and Gomis 2003; Buongiorno Nardelli and Santoleri 2004), and empirical analyses (Meinen and Watts 2000; Watts et al. 2001; Mitchell et al. 2004) were at the base of the different methods.

Most of the statistical approaches begin with the single (univariate) empirical orthogonal function (sEOF) decomposition of observed hydrological vertical profiles, and then search a correlation/regression between the amplitudes of the sEOF modes and some surface measurements (e.g., Carnes et al. 1990, 1994; Pascual and Gomis 2003).

Carnes et al. (1990, 1994), in particular, developed several different models based on the computation of the sEOF of temperature, salinity and steric height profiles and on the least squares regression of the amplitudes of the most significant modes to both linear and nonlinear functions of the surface steric height and temperature. Including possible dependencies from nonlinear terms as powers and cross terms of surface SH and SST considerably improved the results with respect to the simpler linear correlation models (Carnes et al. 1994), even if a much higher number of degrees of freedom (DOF) are absorbed by each nonlinear regression (8 DOFs per mode; see section 3 for details). This higher number of DOFs required by the method can lead to not trivial differences, in terms of significance levels, when estimating the model parameters from a limited number of data, especially since the observations could be significantly autocorrelated.

In recent years, an increasing number of investigations have been conducted on an empirical technique combining historical hydrography with integral measurements from IES. This method, called the gravest empirical mode (GEM) technique, consists of the projection on surface geopotential height space (or, better, an approximation of the geopotential height space, such as two-way acoustic travel time or steric height space) of hydrographic data (Meinen and Watts 2000). A two-dimensional empirical mode (or GEM) is built by applying to the hydrographic profiles as a function of \( \tau \) (or, equivalently, of the surface SH) a cubic smoothing spline. The smoothing is performed at one pressure level at a time and basically leads to the construction of a lookup table based on the integral measurements available.

Some refinements of the standard GEM technique have also been proposed, such as the residual GEM (Mitchell et al. 2004). Residual profiles are obtained by removing a suitable climatology from each historical hydrographic profile. The variability associated with these new profiles is clearly expected to be less than the variability observed in the original dataset. Accordingly, the error associated to a GEM field defined from these data should also be reduced. In practice, the residual GEM fields are the correctors of a climatological field taken as predictor. A further parameterization includes SST measurements to define bins on which different GEM fields can be estimated (Mitchell et al. 2004).

The number of DOFs absorbed by the GEM techniques is 1 (given by the tension applied to the curve) plus the number of nodes used in the smoothing. The latter is generally chosen to be the maximum between a fraction of the total number of profiles in the learning dataset and a minimum of 3. As a consequence, the DOFs absorbed by a basic GEM are generally higher than 4. If different bins are defined according to any additional surface parameter, the number of degrees of freedom needs to be multiplied by the total number of GEM fields computed. This clearly means that this further parameterization can be applied only if the GEM structures can be adequately represented in parameter space.

Recently, we presented a new extrapolation method, called coupled pattern reconstruction (CPR), that allows extrapolation of the vertical profiles of temperature and steric heights from corresponding surface data (Buongiorno Nardelli and Santoleri 2004). Unlike many of the previous techniques, this method is not based on sEOF but on a statistical technique for the identification of the coupled modes of variability, known in literature as coupled pattern analysis (CPA; see Bretherton et al. 1992). CPA consists of the singular value decomposition (SVD) of the cross-covariance matrix constructed from temperature (\( T \)) anomaly and steric height (SH) anomaly profiles. The SVD identifies pairs of orthogonal vertical patterns, each one explaining as much as possible of the mean-squared covariance of the two datasets. The amplitudes of the \( T \) and SH coupled modes can be linearly related one to the other (verifying the effective coupling of the modes), and if we limit the expansions to the first two modes, vertical profiles can be estimated from \( T \) and SH surface values, solving a simple linear system. Six DOF are absorbed by the CPR, two for each regression between \( T \) and SH amplitudes and two for the constraints imposed by the linear system.

The CPR was tested on the hydrological data collected by the Dynamique des Flux de Matière en Méditerranée (DYFAMED) program (1994–2002) at a fixed location in the Ligurian Sea (northwestern Mediterranean Sea). The analyses seem to indicate that CPR generally improves climatological estimates in the up-
per layers and is less susceptible to errors associated with steric height estimation with respect to equivalent sEOF reconstruction (sEOF-R) methods, even if no comparison has been done with GEM (Buongiorno Nardelli and Santoleri 2004). Moreover, no specific studies on the sensitivity of the methods to the length of the learning/training dataset were performed, and the conclusions of that analysis clearly cannot be generalized to any region or dataset. The capability of any of these methods in the reconstruction of the vertical structure of the sea is in fact strongly dependent, on one hand, on the signals that characterize the area considered and/or that can be resolved by available in situ data. On the other hand, it depends on the number of independent measurements needed to accurately train the model, with respect to the DOF absorbed by the model itself.

Consequently, the first objective of the present paper is the comparison of these techniques applied to a different dataset. In particular, our work is focused on the analysis and reconstruction of a long time series of CTD data collected in the framework of the Hawaii Ocean Time-series (HOT) program in the North Pacific Ocean, using the in situ data as surface input, and evaluating the errors associated with the extrapolation when using training datasets of different lengths. The sensitivity of the various techniques to the surface input accuracy has also been investigated. Data and validation strategies are presented in section 2.

The North Pacific area is characterized by three main components modulating the temporal variability of the hydrological profiles: an intra-annual signal related to the westward propagation of a baroclinic signal (Chiswell 1996); decadal/interannual oscillations linked to the Pacific decadal oscillation (PDO), to the El Niño–Southern Oscillation (ENSO), and to the variability in the associated winter rainfall (Lukas 2001); and a weaker component due to local seasonal forcings (Bingham and Lukas 1996). As we will see in section 3a, a strong salinity signal dominates the steric height variations at HOT, making it unfeasible to apply the standard methods for the reconstruction of the profiles from surface temperature and sea level alone. We have therefore hypothesized additional measurements of the sea surface salinity (SSS) to be available. From an operational point of view, these changes could come from drifting buoys, voluntary ships, and/or from the Soil Moisture and Ocean Salinity (SMOS) satellite missions, scheduled for launch by the European Space Agency (ESA) in early 2006, and Aquarius/SAC-D, which has been recently approved for launch by the National Aeronautics and Space Administration (NASA).

Accordingly, a modified CPR (based on the CPA of salinity and steric height profiles), a modified sEOF-R (using SSS instead of SST to estimate the amplitudes), and proper residual GEM techniques have been applied to the HOT time series in section 3b. A new technique, called multivariate EOF reconstruction (mEOF-R), is described in section 4; mEOF modes for selected learning periods are discussed in section 5; and a summary of the results and conclusions are presented in section 6.

2. Data and validation methods

The HOT program started in 1988 in the framework of the Joint Global Ocean Flux Study (JGOFS) and World Ocean Circulation Experiment (WOCE), aimed at the study of the interactions between physical, chemical, and biological processes that modulate the earth climate and can possibly modify the environment [for a complete description of the HOT program see Karl and Lukas (1996) and all papers in a special issue of Deep-Sea Res. (1996, vol. 43B, no. 2–3)]. The HOT experiment is based on the regular sampling of various hydrological and biochemical parameters at some stations located near the center of the North Pacific subtropical gyre. In situ surveys have been conducted nearly monthly since 1988.

In this work we will concentrate on the CTD measurements collected at A Long-term Oligotrophic Habitat Assessment (ALOHA) station, situated ~100 km north of Oahu, Hawaii (22.75°N, 158°W), between 1988 and 2001 (Fig. 1). CTD data have been preliminarily processed directly by HOT personnel. They are quality controlled and binned at 2-db intervals. [See also Santiago-Mandujano et al. (2002) and the HOT
Data Reports, all available online at http://www.soest.hawaii.edu/HOT_WOCE/data_report.html. All measurements can be freely obtained at www.soest.hawaii.edu/HOT_WOCE/ftp.html.] The CTD profiler used is an SBE911plus (Sea-Bird Electronics) that is calibrated regularly. We selected only the CTD deeper than 2000 db and computed an average profile for each cruise in order to filter out as much as possible of the higher-frequency variability (associated with internal waves and baroclinic tides). In situ density was computed using the standard equation recommended by the United Nations Educational, Scientific and Cultural Organization (UNESCO 1981). Steric heights were referenced to the 2000-db surface.

CTD-derived temperature, salinity and geopotential thickness data were first used both to train the various models and to “simulate” the surface input used by the models when testing their performances. However, a simple residual evaluation, that is, using the same dataset to train the model and to validate it (hindcast validation), does not give a reliable indication of how well a model will do when it is asked to make predictions for data it has never seen. For this reason, we decided to judge the methods’ performance using an improved form of holdout cross validation, that is, a simple form of k-fold cross validation. The holdout method requires that the original dataset is separated into two sets, called the learning/training dataset (used to train the model) and the testing dataset (used to independently evaluate the model performance). The method adopted here consists of repeating the holdout validation k times each time the learned dataset is shifted by a few months. The advantage of this method is that it is more robust than both the residual and the simple holdout methods, even if it introduces only few complications from the computational point of view.

Different training period lengths (2, 5, 8, and 11 yr) have been chosen out of the original 14-yr series to test the techniques. CTD profiles corresponding to the training periods were also used to compute ad hoc monthly climatologies, built up by simply averaging the casts for each month. The climatological profiles were used as reference for the evaluation of the performance of the different techniques considered and for the estimation of residual GEM fields.

All the methodologies presented aim at estimating the vertical structure of the sea directly from SST and eventually SSS data measured from space, or by drifters/buoys and/or ships of opportunity measurements, coupled with altimeter estimates of SSH or with ρ measured by IES. As a consequence, it is important to emphasize that a basic validation based on virtually “error-free” input data (surface and integrated parameters measured/computed directly from CTD data) seems inadequate, as fully independent data could be contaminated by errors of various kind. In particular, SST accuracy from space is of the order of 0.5°C (Kearns et al. 2000), while satellite SSS measurements (still only planned with SMOS and Aquarius/SAC-D) plan to reach an accuracy of 0.1 psu with a spatial resolution of 200 km every 10 days (Kerr et al. 2001), even if there are still many uncertainties about the real capability of these sensors in achieving these requirements. On the other hand, the accuracy in estimating steric heights from altimeter data is affected both by instrumental errors and by the method adopted to adjust altimeter SSH to a steric level (which is quite a complicated problem, clearly beyond the objective of the present paper). Hence, we present here a simple analysis of the sensitivity of the methods to the surface SH, SST, and SSS errors. This has been done analogously to our previous work on CPR (Buongiorno Nardelli and Santoleri 2004), that is, adding a random noise to the observed surface values used as input for the vertical reconstruction. The white noise errors were generated through the Interactive Data Language (IDL; version 5.3, Research Systems, Inc.) RANDOMN function, which uses the Box-Muller method for generating normally distributed (Gaussian) random numbers. Two different noise levels have been added to the independent testing datasets, one corresponding to the accuracy expected for future satellite data and a second one considering a more realistic estimation of present capabilities. The rms errors added to surface SH, SST, and SSS in the two cases are 2 cm (3 cm), 0.4°C (0.5°C), 0.1 psu (0.2 psu).

3. CPR, sEOF-R, and residual GEM applied to ALOHA data

a. Reconstruction of vertical profiles knowing surface SH and SST

The first method that has been applied to ALOHA time series is the CPR of temperature and steric heights. The CPA of T and SH profiles (limited to the first 500 db) showed that most of the covariability (on average more than 93%) between the two parameters is explained by the first two modes, with the first mode generally explaining more than 85% of the covariance, whatever the length of the learning dataset considered. These percentages clearly satisfy the first hypothesis required by the CPR (most of the covariability explained by the first two modes). However, the CPR did not perform quite acceptably in the reconstruction of T profiles, with rms errors often comparable to or above...
those associated with the climatology, always up to more than 1°C around 80 db, and also worse (over 1.2°C) in the layers below 300 db with 2 yr of training. Better results were found in the layer between 200 and 300 db (rms ~0.5–0.7°C) and in the first 30–40 db. Effectively, the method is forced to reproduce values imposed at the surface, so it is not surprising that the climatological error was soon approached, adding a noise to the input surface SH and temperature values (Figs. 2a,d).

The second method tested here is the sEOF-R of temperature profiles from SST and surface SH as described by Carnes et al. (1994), reported in their paper as model 3 (or simplified model 6). The EOF of $T(z, t)$ identified three main modes, generally explaining ~90% of the variance in any training period. The re-
gression of the amplitude of the first three modes to surface steric height and temperature has been computed considering a nonlinear model of the form:

$$a_i(t) = a_i T(0, t) + b_i T(0, t)^2 + c_i T(0, t)^3 + d_i SH(0, t) + e_i SH(0, t)^2 + f_i SH(0, t)^3 + g_i SH(0, t) T(0, t) + \text{const},$$

where, $a_i$, $b_i$, $c_i$, $d_i$, $e_i$, $f_i$, $g_i$, and const, are the unknown coefficients for the $i$th mode. Using the different training datasets and applying the reconstruction to the independent testing data, a significant improvement with respect to the climatological means is generally found, with an rms error comparable to that of the climatology only in the lower layers and around 80 db in the case of the 2-yr training (Fig. 3a), otherwise attaining on average below 0.7°C (Figs. 3c,d), even if introducing an error in the surface input data.

The sEOF-R of salinity profiles using SST and surface SH as surface input was tested next (Figs. 3e,h). The first three EOF modes of $S(z, t)$ explain on average ~90% of the variance, with a ~60% in the first mode, and ~20% and ~10% in the second and third modes, respectively.

Much worse results were obtained in this case, as surface temperature is very low correlated to the salinity variations, especially in the upper 100 db (Figs. 10 and 11a). The reason for this is the presence of a strong interannual signal in the salinity data (well visible in Fig. 10a). Actually, the error is comparable to that associated with the climatological profiles in the first 100 db and only slightly lower between 200 and 300 db. However, climatology clearly cannot be considered significant at HOT, as it is absolutely not able to reproduce the main signals observed in the data (e.g., see the 11-yr climatology in Fig. 10c). It can also be noticed that it is probably the effect of these salinity variations on density, not correlated to the temperature signal, that led to the low performance of the CPR of $T$.

The next step consisted of building GEM models fitted to the training temperature and salinity profiles. Both standard GEM and residual GEM techniques were applied, based on the procedures described by Meinen and Watts (2000) and Mitchell et al. (2004). Actually, we observed that these methods, applied to HOT data, were much more sensible to the choice of the number of nodes than to the degree of smoothing (spline tension). As a consequence, the tension was taken constant at all levels (0.1), chosen to minimize the error, and the number of nodes used in the smoothing was set as the maximum between 3 and $n/20$ (where $n$ represents the number of observations in the training dataset). An additional model that includes also SST measurements to define different GEM fields on different SST bins has also been tested. In that case, three SST bins were selected mainly on the basis of the SST distribution, so that the GEM structures could always be well represented in parameter space. From sensitivity tests, it came out that this criterion minimizes the error, while the only evaluation of the hindcast rms error never yielded decisive results, as similar performances could be obtained even with very different choices of the bins.

The standard and residual GEM gave quite similar results, which do not differ much from the climatology (see Fig. 4), while the use of the surface temperature as an additional parameter led to even worse results (not shown). This fact is probably to be ascribed not only to the strong salinity signal (poorly correlated to the temperature) that dominates the variability of the geopotential thickness, but mainly to the fact that the temperature and salinity signals on the vertical are characterized by processes that occur at quite different time scales and with different vertical structures (e.g., baroclinic Rossby waves, interannual signals, etc.) (Chiswell 1996; Lukas 2001). The same surface steric heights are thus likely to be associated with very different profiles, even after filtering the seasonal signal. In these cases, a simple “smoothing” could be a poor substitute for fitting data to a more complex model. The inferior performance of the GEM when introducing the SST parameterization is then a mere consequence of the reduced number of observations available to estimate the GEM fields in each bin. GEM accuracy was not affected significantly by the introduction of a noise on the surface inputs, as the correction introduced by the technique to the climatology is always very little. However, we must observe that as soon as a learning period shorter than 11 yr is chosen, an increasing number of profiles is given by the sole climatology, as the GEM technique does not allow prediction of a correction related to a particular steric height value, if that value lies outside of the limits observed during the learning period. This percentage is around 1%, 5%, and 33% with a training of 8, 5, and 2 yr, respectively.

b. Reconstruction of vertical profiles from surface SH, SST, and SSS

The successive step has thus been the reconstruction of vertical profiles, making the hypothesis that surface measurements of salinity were also available. The techniques applied at first are obviously a modified CPR, which is based on the coupled pattern analysis of salinity and steric height profiles; and the sEOF-R, which evaluates the amplitudes of salinity modes from a re-
Fig. 3. Rms error of sEOF-R temperature profiles extrapolated from SST and surface SH and associated with climatology, computed from (a) 2-, (b) 5-, (c) 8-, and (d) 11-yr learning datasets; rms of sEOF-R salinity profiles extrapolated from SST and surface SH and associated with climatology, computed from (e) 2-, (f) 5-, (g) 8-, and (h) 11-yr learning datasets; rms error of sEOF-R salinity profiles extrapolated from SSS and surface SH and associated with climatology, computed from (i) 2-, (j) 5-, (k) 8-, and (l) 11-yr learning datasets.
gression with surface salinity and steric heights and residual GEM, including different bins defined according to the additional SSS parameter. The results are presented in Figs. 2e,h; 3i,l; and 4e,h, respectively. The CPA of $S$ identified two modes that include most of the covariability, with the first one explaining on average more than 80% of the covariance and the second one attaining around 10%. The CPR of $S$ always gives much better results than climatology in the first 100–150 db (Figs. 2e,h), while below 200 db it generally performs worse, with errors up to 0.25 psu around 250 db that raise to more than 0.4 psu when a noise is added to the surface input.

The sEOF-R method significantly improves salinity estimates along the whole water column when the learning dataset is at least 8 yr long (Figs. 3k,l), per-
forming slightly better than the climatology also with very noisy surface input. On the other hand, it fits very poorly the data with a shorter training period. This fact is clearly related to the difficulties in estimating correctly the 24 coefficients required by the sEOF-R from datasets of comparable dimensions. This is crucial if one considers that learning datasets could also contain data that are cross-correlated (not fully independent).

The residual GEM with surface salinity as an additional input parameter has been applied similarly to the SST case described in section 3a. Anyway, even in this case it does not improve much the climatology with 11-yr training (with an rms decrease only for salinity of about 0.05 psu in the first 80 db), while it leads to poorer results when 8-, 5-, and 2-yr training periods are used or when surface noise is introduced. The lower performance of the GEM when introducing the SSS parameterization cannot be explained only as a consequence of the inferior number of observations available to estimate the GEM fields in each bin but more likely to the nature of the GEM technique itself.

The first three methods discussed in section 3a are based on the availability of surface measurements of sea temperature and salinity and can be effectively applied also for operational nowcasts of the ocean structure using existing remotely sensed data: infrared/microwave passive sensors for SST and active microwave altimeters for sea level (see, e.g., Fox et al. 2002). On the other hand, the last three methods described here imply a measure of salinity, which is presently only planned from space with the SMOS and Aquarius/SAC-D missions, but already available, even if with much lower coverage, through drifting/moored buoys and thermosalinographs mounted on voluntary ships. However, once we make the hypothesis that SSS is known, there is no reason why we should not use all available information on the surface together. Including both SST and SSS in the linear and nonlinear terms of the regressions used in the sEOF-R could be a way, but a very high number of DOF would then be absorbed by the method. We have then developed and tested a new method for the simultaneous extrapolation of temperature, salinity, and steric height profiles from corresponding surface data that is fully discussed in the following.

4. Multivariate EOF reconstruction (mEOF-R)

The new method described here, is called multivariate EOF reconstruction, is based on the analysis of the “multicoupled” variability of three parameters (temperature, salinity, and geopotential thickness). The novelty of this method mainly resides in the choice of the state vector used to characterize the covariance of the system. In fact, by including the whole steric heights profile in the state vector (i.e., a proxy of the geopotential streamfunction at each pressure level) a closer link between the conservative water properties variability and the dynamical modes can be expected.

In practice, a multivariate matrix X is constructed, putting the three sets of data (T, S, and SH anomaly profiles) in a single matrix. Data are preliminarily normalized, dividing each parameter by its standard deviation (computed for the whole profile). The dimensions of each original matrix of data is m×n, with n measurements (stations) and m vertical levels. The new matrix has dimensions equal to 3m×n, and each column consists of the three normalized profiles of T, S, and SH taken at the same location:

$$
X = \begin{bmatrix}
T(0,t_1) & T(0,t_2) & \ldots & T(0,t_n) \\
T(z_{m,t_1}) & T(z_{m,t_2}) & \ldots & T(z_{m,t_n}) \\
S(0,t_1) & S(0,t_2) & \ldots & S(0,t_n) \\
S(z_{m,t_1}) & S(z_{m,t_2}) & \ldots & S(z_{m,t_n}) \\
SH(0,t_1) & SH(0,t_2) & \ldots & SH(0,t_n) \\
SH(z_{m,t_1}) & SH(z_{m,t_2}) & \ldots & SH(z_{m,t_n})
\end{bmatrix}
$$

The further step is the computation of the EOF or, equivalently, the SVD of the “multivariance” matrix $$\Gamma_X$$ calculated from this new dataset. More details on the multivariate EOF can be found in Bretherton et al. (1992). In substance, the multivariate matrix consists of variance matrixes of the single variables along the principal diagonal and submatrixes that represent the cross covariances between all the couples of data and their transposes, above and under the diagonal. The eigenvalue problem is $$\Gamma_X U = U A$$, where $$U = (u_k)$$ represent the multicoupled modes, each containing the three patterns corresponding to the parameters considered, $$u_k = (L_k, M_k, N_k)$$, and $$A = (a_l)$$ is the diagonal matrix of the eigenvalues.

Similarly to CPA $$T(z,t), S(z,t),$$ and SH($$z,t)$$ can thus be expanded in terms of these three series of patterns, but with the same coefficients $$a_l$$:

$$
T(z,t) = \sum_{k=1}^{n} a_k(t)L_k(z); \\
S(z,t) = \sum_{k=1}^{n} a_k(t)M_k(z); \\
SH(z,t) = \sum_{k=1}^{n} a_k(t)N_k(z). 
$$
Now, exactly as for the CPR and sEOF-R, suppose that these expansions can be reduced to the sum of a very limited number of modes. In particular, if we limit to the first three modes, the vertical profiles can be estimated from the surface values \((z = 0)\) of the three parameters, solving the system (4) for \(a_1, a_2, \) and \(a_3:\)

\[
\begin{align*}
a_1(t)L_1(0) + a_2(t)L_2(0) + a_3(t)L_3(0) &= T(0,t) \\
a_1(t)M_1(0) + a_2(t)M_2(0) + a_3(t)M_3(0) &= S(0,t) \\
a_1(t)N_1(0) + a_2(t)N_2(0) + a_3(t)N_3(0) &= SH(0,t)
\end{align*}
\]

The mEOF-R thus poses a very low number of constraints, as only 3 DOF are absorbed by the linear system (4).

5. mEOF-R application to ALOHA time series

The first three mEOF modes computed from ALOHA time series explain on average slightly less than 90% of the multivariate, whatever the length of the training periods considered, so the condition required by the mEOF-R is reasonably realized. The mEOF-R has thus been applied to HOT data, and the rms error with respect to observed profiles for all independent testing periods has been evaluated. In the present paper, we concentrate on the temperature and salinity profiles reconstruction alone (not on steric height profiles reconstruction). This choice is related to the fact that \(T\) and \(S\) profiles are more commonly assimilated in numerical/operational models. In this sense, synthetic profiles estimated along altimeter tracks could improve our forecasting capabilities. On the contrary, there is no particular interest in the extrapolation of the time series of steric heights profiles at a single location, as transports can be estimated only from horizontal gradients of the steric level (geostrophic component).

The results are shown in Fig. 5. Both the error on the extrapolation of the temperature and that of salinity are significantly lower than the error associated with the climatological estimates when the learning periods are at least 5 yr long. In those cases, the rms error remains on average below 0.7°C and 0.1 psu, respectively, even when introducing the highest levels of noise at surface. On the opposite, with 2-yr training, mEOF-R improves with respect to the climatology only in the upper 100 db ut is found to have significantly higher errors between 200 and 300 db (up to 2.5°C and 0.5 psu). The mEOF-R performance is always comparable (almost identical) to that of the sEOF-R of \(T\) from SST and surface SH, and slightly better than that of the sEOF-R of \(S\) from SSS and surface SH for the 8- and 11-yr training. However, as already shown in section 3, in the case of salinity reconstruction from SSS and surface SH, sEOF-R performance dramatically drops when the time period is reduced to 5 or 2 yr, while mEOF-R gives more reliable estimates. Higher errors are found also extrapolating salinity with the sEOF-R from SST and surface SH, except for the 2-yr case.

6. Multivariate EOF modes at ALOHA

To get more insight into the dynamical processes (eventually) resolved by the mEOF, a more complete analysis of the mEOF modes that have been identified at HOT has been carried out. Nevertheless, a complete analysis would be highly redundant and difficult to present, so only the modes determined from the learning datasets starting at day 1 of the time series (January 1988) are discussed. The impact of the time series length on the principal modes identification has also been evaluated for that particular case. The standard deviations used for the data normalization required by the mEOF are presented in Table 1. A higher variability is observed in the three parameters as the length of the period considered is increased. Conversely, computed modes and amplitudes display fundamentally no differences as soon as a 5-yr learning dataset is used (Figs. 6 and 7). As a consequence, we will concentrate on and describe in detail first those obtained from the longer time series (11-yr learning) and then evidence what signals were “left out” by the analysis of the shorter datasets.

As already underlined in the introduction, several studies have analyzed the variability observed in the hydrological profiles at HOT, evidencing three main signals. It is noteworthy that the mEOF identifies and discriminates these phenomena, quantifying the impact of each in terms of contribution to the total variability of the system. Almost 90% of the variability is associated with these first three modes, clearly indicating that the signals related to other physical mechanisms are not fully detectable with a monthly sampling or have a low influence on the vertical structure of the sea at ALOHA.

The most important signal consists of an intra-annual oscillation with a ~100 day period that has been related to the westward propagation of Rossby waves. Actually, the theoretical prediction for the period of the baroclinic Rossby waves in Hawaii is higher than the observed one. Chiswell (1996) and Mitchum (1996), however, explain this frequency shift as a consequence of the Doppler effect caused by a weak (~2 cm s⁻¹) background zonal current [a wider review and analysis of the effect of slowly varying mean flows and bottom topography on wave propagation processes can be...
found in Killworth and Blundell (2003). The first mEOF mode, accounting for 65.0% of the multivariate in the 11-yr case, catches the signal associated to these baroclinic waves, displaying the same temporal behavior of the first sEOF computed by Chiswell (1996) for the years 1991–93, even if the first 100 db were excluded by his analysis (Fig. 6d, top). Analo-

TABLE 1. Temperature, salinity, and steric height standard deviations used for the data normalization required by the mEOF.

<table>
<thead>
<tr>
<th></th>
<th>2 yr</th>
<th>5 yr</th>
<th>8 yr</th>
<th>11 yr</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_T$ (°C)</td>
<td>0.83</td>
<td>0.85</td>
<td>0.90</td>
<td>0.93</td>
</tr>
<tr>
<td>$\sigma_S$ (psu)</td>
<td>0.09</td>
<td>0.10</td>
<td>0.11</td>
<td>0.11</td>
</tr>
<tr>
<td>$\sigma_{SH}$ (cm)</td>
<td>3.46</td>
<td>3.67</td>
<td>4.28</td>
<td>4.36</td>
</tr>
</tbody>
</table>
gously to those results, the vertical patterns of $T$ and $S$ corresponding to the first mEOF present a minimum around 240 db (Figs. 7d,h). The salinity profile changes sign above $\sim 120$ db, remaining positive up to the surface, while it gradually vanishes below the minimum. The temperature pattern is always negative, decreasing in absolute value toward the surface and below the minimum, down to 500 db. The corresponding vertical pattern of SH (Fig. 7n) clearly indicates that the major part of SH variability is explained by this first mode.

**Fig. 6.** Temporal coefficients associated with the first three mEOF modes computed from the (a) 2-, (b) 5-, (c) 8-, and (d) 11-yr learning datasets, starting in 1988. A vertical line separates the training and testing periods.
both at depth and at the surface, with a monotone increase from a surface minimum, which never vanishes down to the maximum depth considered.

A much lower percentage of the multivariance is explained by the second mode (15% with 11 yr of training), which is characterized by a more superficial signal (Figs. 7d,h,l) and by clear interannual trends (Fig. 6d, middle). The salinity pattern presents the highest absolute values, attaining around −1.0 and −0.5 between the surface and ~100 db, presenting a slight relative minimum (between 250 and 300 db), and finally vanishing around 400 db (Fig. 7d). Conversely, the tem-
temperature pattern displays positive (~0.3) values down to ~60 db and changes sign around 180 db, with a very weak minimum between 250 and 300 db (Fig. 7h). Actually, looking at the SH pattern, it comes out that $T$ and $S$ variations at depth are compensated in density (Fig. 7l). Given the strong interannual variation characterizing this mode, we searched a relation with the signals dominating the climate variability in the Pacific
Ocean. In particular, we computed the lagged correlation between the Southern Oscillation index (SOI), the Pacific decadal oscillation index (PDOI), and the temporal coefficient associated to the mode.

The SOI is defined as the mean monthly difference between the atmospheric pressure at Darwin (Australia) and that at Tahiti, and it is strictly correlated to the cyclic heating and cooling observed in the central-eastern Pacific, known as El Niño and La Niña, respectively (see Fig. 8a). No correlation was substantially found between the SOI and the second mEOF amplitude at HOT (correlation never exceeded 0.35). Consequently, the same analysis was repeated with PDOI (Fig. 8b), defined as the temporal coefficient associated with the first principal component of the monthly mean SST in the northern Pacific (north of 20°N). The PDO is a climatic phenomenon quite similar to ENSO in terms of spatial structure, but it occurs on much wider time scales (20–30 yr compared to 6–18 months). Unlike ENSO, it influences more the northern Pacific and has minor effects on the equatorial/tropical zone (Mantua et al. 1997; Zhang et al. 1997). It is not surprising then that the second mEOF mode at HOT is correlated to the PDOI, with a time lag of 8–9 months (the correlation coefficient displays a clear maximum, even if values are not particularly high; see Fig. 9). In addition to this, we can underline that the results of this analysis are coherent with those reported by Lukas (2001), who related the decadal variations observed during HOT to the variability in the precipitation field over the northern Pacific. Lukas showed that the associated salinity anomalies appear with a certain delay in the lower layers, probably as a consequence of the thermocline ventilation, while they are almost exactly compensated in density by temperature.

The third mEOF mode is clearly related to a seasonal signal (Fig. 6d, bottom). It is characterized by a very low percentage of the multivariate (≈8%, in the case of the 11-yr learning) and is dominated by temperature changes. In fact, $T$ vertical pattern displays very high values in the first 50 db (Fig. 7h) and is perfectly co-

![Fig. 8. (a) Southern Oscillation index and (b) Pacific decadal oscillation index for the period 1988–2001.](image_url)

![Fig. 9. Lagged correlation between the second mode temporal coefficient and the PDOI, in the 1988–99 learning case.](image_url)
Fig. 10. Time series of (a) observed, (b) reconstructed from the 11-yr learning period starting in 1988, and (c) corresponding climatological salinity (psu) profiles at the HOT ALOHA site. The climatological time series has been obtained by taking the climatological profile instead of the observed one, at the same instant. A bold vertical line separates the training and test periods.
Fig. 11. Time series of (a) observed, (b) reconstructed from the 11-yr learning period starting in 1988, and (c) corresponding climatological temperature (°C) profiles at the HOT ALOHA site. The climatological time series has been obtained by taking the climatological profile instead of the observed one, at the same instant. A bold vertical line separates the training and test periods.
herent with the annual component obtained by Bingham and Lukas (1996) through a fit of the CTD data to a seasonal model that includes annual and semiannual harmonics. Both T and S profiles present a maximum (not very pronounced) around 220 bar, while the major variations in the SH profile are found in the layer between the surface and 50 db.

No relevant differences are found either in the temporal coefficients or in the vertical patterns obtained from the 5- and 8-yr learning datasets with respect to the 11-yr period (Figs. 6 and 7). On the contrary, when limiting the training period to the first two years, some significant differences are found. In particular, while the amplitude of the first mode remains almost the same, the second and third display a different behavior (Fig. 6a). In fact, we could imagine the decadal signal (second mode) not to be resolved accurately with a 2-yr series, but also, unexpectedly, the seasonal signal observed in the third mode was not well identified. Correspondingly, the vertical patterns differ significantly, too. The most evident difference is that all patterns associated with the third mode present an abrupt inversion around 120 db (Figs. 7a,e,i). These differences obviously explain why the mEOF-R applied to 2-yr learning periods was not able to correctly reproduce the vertical structure at HOT.

An example of the pressure–time transect of observed temperature and salinity profiles, and climatological and extrapolated data for the 1988–99 learning case, are presented in Figs. 10 and 11. They give us a more clear idea of the high potential of mEOF-R in reconstructing the temporal evolution of the two parameters below the surface, even considering the high variability of the observed fields, especially in the first 100 db.

7. Summary and conclusions

In this work we have tested different methods for the extrapolation of vertical profiles from surface data on 14 yr of CTD measurements collected in the North Pacific Ocean within the HOT program. The aim of the work is to identify the advantages and limits of each methodology, with a particular attention to the sensitivity of the various techniques to the length of the time series used to train the model and to the errors on the surface measurements used as input. The methods considered are the sEOF-R proposed by Carnes et al. (1994), the recently proposed CPR technique (Buongiorno Nardelli and Santoleri 2004), the GEM techniques described by Meinen and Watts (2000), Watts et al. (2001), and Mitchell et al. (2004), and a new technique called multivariate EOF reconstruction (mEOF-R).

The majority of these methods is based on the analysis of the variability of a series of profiles and on the identification of some “principal” modes driving it, while GEM and residual GEM techniques search for an empirical relation between the vertical tracer field and an integrated parameter.

GEM consists of the historical hydrographic data projection on surface geopotential thickness space, through which lookup tables are obtained by a smoothing procedure. As a matter of fact, when the data are smoothed in order to fit to an arbitrary cubic spline, the observations are simply parameterized, and the method is not learning anything more fundamental on the processes driving their variability. In the HOT case, GEM and residual GEM temperature and salinity fields were not able to improve the climatological estimates, even with the inclusion of a parameterization that takes into account an additional surface parameter, such as SST or SSS.

On the contrary, the other methods discussed here are based on analyses that do not necessarily provide information a priori on the dynamics of the system, due to their statistical nature. Statistical modes, however, can identify physically meaningful processes, especially if these are characterized by very different temporal scales and vertical structures, and if the observation sampling is adequate. Anyway, some differences exist between the various statistical models compared here.

The sEOF-R technique analyzes separately the variability of each parameter along the water column and hypothesizes a nonlinear relation between the amplitude of the modes and a combination of surface parameters. In substance, it correlates surface parameters and subsurface values through a nonlinear least squares regression. The sEOF-R methods absorb a high number of DOFs (8 for each mode). On the other hand, both CPR and mEOF-R directly search coupled/multi-coupled” modes. In practice, the correlation between different parameters is evaluated on the whole water column, and the amplitude of the modes is estimated solving an algebraic system whose known terms are the surface parameters. Clearly, the relation considered between the different profiles in the multivariate methods used here is supposed to be linear, but the strength and novelty of these methods resides fundamentally in the choice of the state vector used to characterize the covariance of the system. In fact, by including a proxy of the geopotential streamfunction at each pressure level in the state vector, a closer relation between the conservative water properties variability and the dynamical modes can be expected. Moreover, a reduced number of constraints need to be taken into account to extrapolate synthetic profiles (6 and 3 DOFs are absorbed by
CPR and mEOF-R, respectively). This is a crucial point when trying to infer the salinity structure from a limited learning time series.

Nonetheless, with regard to the extrapolation of vertical profiles of temperature, our tests did not evidence a “best” statistical method. In particular, the application to HOT time series of CPR and sEOF-R for the extrapolation of temperature profiles indicated a better performance of the sEOF-R. On the other hand, mEOF-R accuracy in extrapolating temperature profiles was shown to be comparable to that of the sEOF-R, even if it requires the SSS measure as an additional input. This fact is clearly explained by the fundamental role of salinity variations in determining the steric height profiles (see section 3). CPR was forced to find a correlation between temperature and steric heights, neglecting necessarily salinity variations, which were proven not to be correlated with temperature, and thus leading to worse performances with respect to sEOF-R. Conversely, mEOF-R was able to take into account the effects of both temperature and salinity variations on the steric height signal. Moreover, mEOF-R of T and sEOF-R of T did not produce much worse estimates, adding a noise on the surface data.

Actually, a strong influence of the salinity on sea level variations had already been stressed in some works by Maes (1998, 1999), who has also, in a more recent work, proposed the use of altimeter data and temperature profiles to improve the estimates of salinity variations (Maes and Behringer 2000). Our analysis indicated that the sEOF-R of S from SST and surface SH is not able to improve the climatological estimates of the vertical salinity structure at HOT. For this purpose, a surface measurement of salinity is necessary. Consequently, the four methods that have been considered in the successive step all imply a measure of surface salinity. The methods are a residual GEM parameterized with surface SSS; a modified sEOF-R, which estimates the EOF amplitudes including SSS in the least squares regression; a modified CPR (applied to S and SH profiles); and the mEOF-R, which searches the modes that maximize a sort of “multivariance” between SH, T, and S profiles (section 4).

The performances of mEOF-R and sEOF-R for the extrapolation of salinity profiles significantly improved with respect to climatology when using error-free surface data, with 5-, 8-, 11-yr, and 8-, 11-yr learning periods, respectively. However, the higher number of DOFs required by the sEOF-R method leads to strong differences when estimating the model parameters from a limited number of data and makes the application of sEOF-R much more risky. As a matter of fact, sEOF-R of S at HOT can lead to rms errors of the order of 1.5 psu or more (especially when noise-free input surface data are not available), as soon as time series shorter than (or equal to) 5 yr are used to train the method. Moreover, mEOF-R is slightly less susceptible than sEOF-R to errors in the surface input.

In addition to this, the mEOF analysis leads to interesting results by itself, being able to discriminate the main dynamical signals observed in the North Pacific Ocean (at the HOT station ALOHA) and to quantify the relative importance of each to the total variability. In particular, analyzing the mEOF computed from the 1988–99 learning set, we found that the first mode accounts for 65.0% of the multivariance and is associated with the westward propagation of baroclinic Rossby waves, characterized by an intra-annual oscillation of ~100 day period. The second mEOF mode explains ~15% of the variance, and results correlated with a delay of 8–9 months to the Pacific decadal oscillation. The last mEOF mode considered is clearly related to a seasonal signal and accounts for a lower percentage of the multivariance (8%).

Even if all these analyses cannot be considered exhaustive, given the limited number of data and the particular dynamical characteristics of the area considered, some interesting conclusions can be drawn. As a matter of fact, the in situ data sampling, in relation to the local dynamics, clearly affect the methods’ performance, so it is not obvious that a “best” method can be found independently of the dataset considered, simply based on the hypothesis required by each technique. However, GEM methods were never able to improve HOT climatology, while it seems that the sEOF-R methods are more likely to produce completely wrong predictions if they are trained with datasets that are not optimal. On the other hand, mEOF-R generally produced reliable estimates (in the worst cases comparable to the climatologies), even if it requires SSS measurements also to infer temperature profiles.

To generalize the technique and apply it to wider areas, specific studies should be performed to evaluate which is the optimal method for the extraction of the “historical/statistical” information from available datasets of hydrological profiles. In our tests, we concentrated on the comparison between different methodologies applied on time series of data at a fixed position, while future work must somehow include the spatial variability over given/limited time periods, also evaluating the possibility of more complex analyses, such as studying the temporal variability associated with a particular surface water mass. A careful analysis of the barotropic signals that appear in altimetric sea level measurements that do not appear in dynamic height is also fundamental and will be the subject of a
paper presently in preparation (Buongiorno Nardelli et al. 2005, manuscript submitted to J. Geophys. Res.). The next steps will consequently be the application of the various extrapolation methodologies to wider datasets and the effective use of remotely sensed data as surface input.

Acknowledgments. We wish to thank both the anonymous reviewers for their valuable and useful comments. We are particularly grateful to reviewer number 2, whom we did not have a chance to acknowledge for his helpful suggestions when reviewing a previous version of this paper, originally submitted to the Journal of Geophysical Research. This publication is based upon Hawaii Ocean Time-series observations supported by the U.S. National Foundation under Grant OCE-0117919. The present work was funded by the Italian Space Agency under Contract I/R/179/02.

REFERENCES


