JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 112, C03008, doi:10.1029/2006JC003660, 2007

Multivariate reconstruction of missing data in sea surface temperature, chlorophyll, and wind satellite fields

A. Alvera-Azcárate,[1] A. Barth,[1] J.-M. Beckers,[2] and R. H. Weisberg[1]

Received 20 April 2006; revised 27 September 2006; accepted 24 October 2006; published 15 March 2007.

[1] An empirical orthogonal function–based technique called Data Interpolating Empirical Orthogonal Functions (DINEOF) is used in a multivariate approach to reconstruct missing data. Sea surface temperature (SST), chlorophyll *a* concentration, and QuikSCAT winds are used to assess the benefit of a multivariate reconstruction. In particular, the combination of SST plus chlorophyll, SST plus lagged SST plus chlorophyll, and SST plus lagged winds have been studied. To assess the quality of the reconstructions, the reconstructed SST and winds have been compared to in situ data. The combination of SST plus chlorophyll, as well as SST plus lagged SST plus chlorophyll, significantly improves the results obtained by the reconstruction of SST alone. All the experiments correctly represent the SST, and an upwelling/downwelling event in the West Florida Shelf reproduced by the reconstructed data is studied.

Citation: Alvera-Azcárate, A., A. Barth, J.-M. Beckers, and R. H. Weisberg (2007), Multivariate reconstruction of missing data in sea surface temperature, chlorophyll, and wind satellite fields, *J. Geophys. Res.*, *112*, C03008, doi:10.1029/2006JC003660.

1. Introduction

[2] Missing data reconstruction is usually performed by a univariate approach, i.e., taking into account only the information of a specific variable to infer values at missing data locations [e.g., *Reynolds and Smith*, 1994; *Beckers and Rixen*, 2003; *He et al.*, 2003; *Alvera-Azcárate et al.*, 2005; *Kondrashov et al.*, 2005]. However, oceanographic variables are often interrelated by ocean dynamics, making it potentially useful to consider these relations in the reconstruction of missing data.

[3] Multivariate optimal interpolation (OI) has been used to analyze a given variable with the help of related measures. *Gomis et al.* [2001] and *Paris et al.* [2002] worked with conductivity-temperature-depth (CTD) and acoustic Doppler current profiler (ADCP) data to study the three-dimensional geostrophic and ageostrophic circulation features of the Alboran Sea and Barbados, respectively. *Watts et al.* [2001] used OI to analyze bottom pressure and currents measurements in the Gulf Stream. *Grodsky and Carton* [2001] combined sea level and mixed layer velocities to examine the Tropical Pacific Ocean near-surface currents. These works agree in the benefit of using multiple variables in the analysis as a way to obtain more realistic fields and a better understanding of their interrelation. However, it is difficult to objectively estimate the covariance between two variables and the error statistics necessary to realize an OI analysis [*Bennett*, 2002]. Therefore assumptions are often made about the variables covariance and correlation length (e.g., isotropic and homogeneous covariances).

[4] Multivariate empirical orthogonal functions (EOFs) have also been used to improve the analysis of a given variable by the addition of related variables. Since the correlation between variables is taken into account by the EOFs, the relation between them may be studied. *Bretherton et al.* [1992] and *Wallace et al.* [1992] studied the relation between sea surface height anomalies and 500-mbar height anomalies over the Pacific using various covariance-based techniques. *Korres et al.* [2000] studied the relation between surface heat fluxes and sea surface temperature in the Mediterranean Sea using extended EOFs. *Collins et al.* [2004] studied the predictability of the sea surface temperature in the Indian Ocean, and found that a combination through EOFs of sea level pressure and sea surface height anomaly gave the best results.

[5] In this work, an EOF-based method is used to reconstruct missing data on different but related variables. Satellite data are often gappy due to the presence of clouds, lack of satellite coverage, and observations rejected by quality control. Certain satellite data, as those produced by some radiometers, do not present a cloud coverage problem, but their coverage may be subject to other limitations, such as a coastal mask or rain contamination. *Beckers and Rixen* [2003] and *Alvera-Azcárate et al.* [2005] described a technique for filling missing data using an EOF-based algorithm called Data Interpolating Empirical Orthogonal Functions (DINEOF), and they applied it to sea surface temperature (SST) data. *Kondrashov and Ghil* [2006] and *Kondrashov et al.* [2005] used this technique in a singular spectrum analysis version, and applied it to an

[1] College of Marine Science, University of South Florida, Saint Petersburg, Florida, USA.
[2] GeoHydrodynamics and Environmental Research – Astrophysique, Géophysique et Océanographie, University of Liège, Liège, Belgium.

0148-0227/07/2006JC003660$09.00

extended matrix containing time-lagged information of the variable studied. By application to a synthetic data set, as well as to several real examples, the results were shown to be improved by the addition of the time lag information. *Bergant et al.* [2005] applied this technique to tidal gauge data. Here we present a multivariate application of the method developed by *Beckers and Rixen* [2003], where the extended matrix is formed by more than one physical variable at one or at several times. The advantage of using DINEOF compared to OI application, as classically applied, to reconstruct multivariate data is that the first technique calculates the relation between variables internally, based on available data, so there is no subjective parameter estimation.

[6] We applied the multivariate DINEOF technique to a domain covering the southeast U.S. coast and the eastern Gulf of Mexico (GoM). SST, chlorophyll and winds are combined in different ways and compared to in situ data to establish the most beneficial combination. Any set of variables could have been chosen, given that they are related.

[7] A brief explanation of DINEOF, as well as its multivariate application are included in section 2. The data used in this work and the experiments carried out are explained in sections 3 and 4, respectively, followed by experimental results in section 5. As an application, section 6 then studies a series of upwelling/downwelling events on the West Florida Shelf (WFS) detected in the reconstructed data. Conclusions are offered in section 7.

2. DINEOF

2.1. Monovariate DINEOF

[8] DINEOF [*Beckers and Rixen*, 2003; *Alvera-Azcárate et al.*, 2005] is a parameter-free (i.e., the necessary parameters are derived internally from existing data), EOF-based method for the reconstruction of missing data. *Alvera-Azcárate et al.* [2005] applied DINEOF to a series of satellite SST images and compared the results to a classical OI reconstruction. The number of operations required for a global OI is $m^3\ n^3$, with m, n the spatial and temporal dimensions of the matrix being reconstructed, respectively. *Alvera-Azcárate et al.* [2005] empirically established the total cost for DINEOF as proportional to $m^{1.25}n^{1.17}$. Our experience has also shown that DINEOF is faster than local OI codes. For example, in the work by *Alvera-Azcárate et al.* [2005], DINEOF was 30 times faster than the widely used local OI package provided by Harvard Ocean Prediction System (HOPS [*Carter and Robinson*, 1987]), and both techniques performed similarly when compared to in situ data. DINEOF has proven to be an accurate technique, and it produces reliable results.

[9] DINEOF works as follows: data are stored in a matrix $\mathbf{X}$ with the temporal and spatial average subtracted a priori. Missing data are initialized to zero, to guarantee that they are unbiased with respect to $\mathbf{X}$. With this first guess, a first singular value decomposition (SVD) is realized with one ($k = 1$) EOF. The missing data are then replaced using the obtained EOF:

$$\mathbf{X}_{i,j} = \sum_{p=1}^{k} \rho_p (\mathbf{u}_p)_i \left(\mathbf{v}_p^T\right)_j \tag{1}$$

where i, j are the spatial and temporal indexes of the missing data in matrix $\mathbf{X}$; $\mathbf{u}_p$ and $\mathbf{v}_p$ are the pth column of the spatial and temporal EOF $\mathbf{U}$ and $\mathbf{V}$, respectively; and ρ_p is the corresponding singular value, with $p = 1 \ldots k$. With the new values for the missing data, the SVD decomposition is performed again. The two last steps are repeated until convergence is obtained for the missing values. Then, the whole iterative procedure is performed for $k = 2, 3, \ldots, k_{\max}$ EOFs, where $k_{\max}$ is a predefined number that should vary according to the initial matrix characteristics. For each k, an estimation for the missing values is obtained. The optimal number of EOFs retained for the reconstruction is determined by cross validation: a number of points (typically 1% of the initial data) are set aside and considered as missing. At each EOF estimation, the error between those initial points and their reconstruction is calculated, so the optimal number of EOFs minimizing this error can be determined. Along with the reconstructed data, local error fields reflecting the accuracy of the reconstruction can be also generated [*Beckers et al.*, 2006]. DINEOF calculates the EOFs using a Lanczos method [*Toumazou and Cretaux*, 2001], which reduces the calculation time. For an extended description of DINEOF please refer to *Beckers and Rixen* [2003] and *Alvera-Azcárate et al.* [2005].

2.2. Multivariate DINEOF

[10] Extended EOFs (ExEOFs) are a generalized form of the classical EOFs in which more than one data set are used simultaneously to perform an EOF analysis. Most works [e.g., *Weare and Nasstrom*, 1982; *Fraedrich et al.*, 1997; *Tourre and White*, 1997; *Smith and Reynolds*, 2003] use a lagged version of the analyzed matrix to construct the extended matrix $\mathbf{X_e}$. Here we present the extended matrix in a more general case, where different variables with lagged versions of themselves are combined to form an extended, multivariate matrix $\mathbf{X_e}$:

$$\mathbf{X_e} = \begin{pmatrix} X_1 & X_2 & .. & X_{N-2l} \\ X_{1+l} & X_{2+l} & .. & X_{N-l} \\ X_{1+2l} & X_{2+2l} & .. & X_N \\ .. & & & \\ Y_1 & Y_2 & .. & Y_{N-2l} \\ .. & & & \\ Z_1 & Z_2 & .. & Z_{N-2l} \end{pmatrix} \tag{2}$$

where X_t, X_{t+l}, X_{t+2l} are column vectors that contain all spatial points of variable $\mathbf{X}$ at times t, $t + l$ and $t + 2l$, respectively (l and $2l$ are the time lags). $\mathbf{Y} \ldots \mathbf{Z}$ are other variables. Each matrix can have a different number of spatial points (size $\mathbf{X} = M \times N$, size $\mathbf{Y} = P \times N$, size $\mathbf{Z} = T \times N$, and so on, with M, P, and T the spatial dimension and N the temporal dimension). The use of such an extended matrix to compute ExEOFs has many advantages over the use of classic EOFs in the frame of missing data reconstruction. First, they can resolve moving patterns more accurately because of the presence of future and/or past information, if those points are not missing in the lagged matrix [*von Storch and Zwiers*, 1999; *Ghil et al.*, 2002; *Jolliffe*, 2002; *Kim and Wu*, 2000]. Also, the correlation between different physically related variables can help to reconstruct missing data [e.g., *Gomis et al.*, 2001].

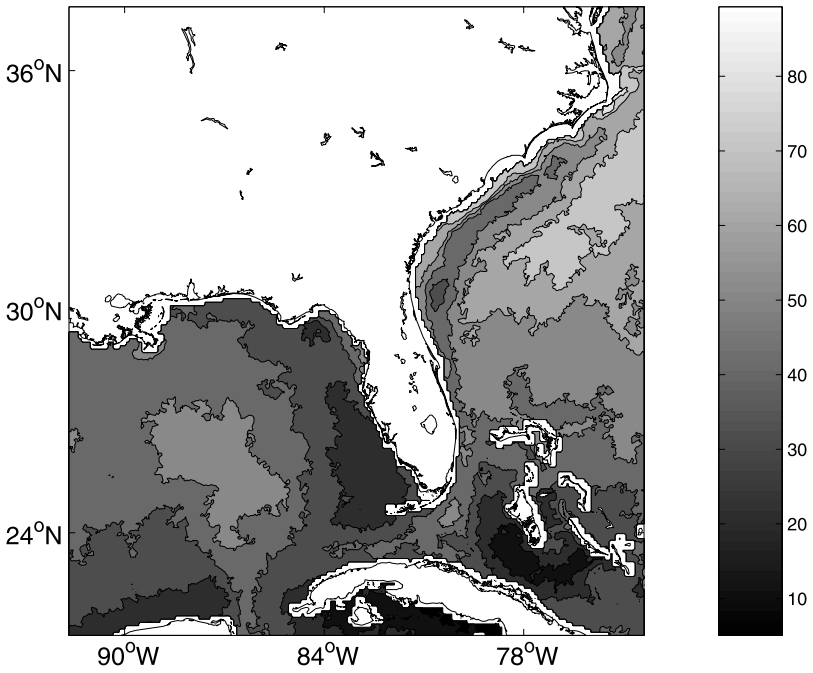

Figure 1. Spatial distribution of the percentage of cloud coverage in the original AVHRR SST data set.

[11] When the different variables included in matrix $\mathbf{X_e}$ have different units, the normalization of each variable before constructing $\mathbf{X_e}$ is necessary. In this multivariate version of the ExEOFs, a given missing point of one variable can benefit from the presence of other variable at the same time. The correlation between both variables is taken into account by the ExEOFs to reconstruct missing data.

3. Data

[12] Several satellite data sets covering the GoM and the east U.S. coast are used for the period from 1 October 2004 to 31 March 2005. SST is derived from infrared observations made by the advanced very high resolution radiometer (AVHRR) sensors on board the NOAA Polar Orbiting Environmental Satellite (POES) series. The SST is calculated using the multichannel sea surface temperature (MCSST) algorithm developed by *McClain et al.* [1985]. Chlorophyll *a* pigment concentration data are obtained from the Aqua Moderate Resolution Imaging Spectroradiometer (MODIS) satellite. These fields have a resolution of 1 km.

[13] The AVHRR SST data set has 45.3% missing data, and a slightly increasing trend on the percentage of missing data is observed in the six months of the experiment. Figure 1 shows the mean spatial percentage of missing SST during the period studied: the highest cloud coverage is concentrated along the U.S. east coast. The open Atlantic Ocean also has a high percentage of missing data. In the GoM, the zone near the Loop Current (LC) has the highest cloud concentration during this period, along with the Big Bend at the Florida coast. The chlorophyll fields have a higher mean percentage of missing data (65%) and its spatial distribution (not showed) is similar to the one described for the SST.

[14] QuikSCAT Level 3 wind fields are obtained through the Physical Oceanography Distributed Active Archive Center Web site http://podaac-www.jpl.nasa.gov/. They consist of 25 km resolution gridded fields and they cover the whole globe (up to 75° latitude) every 2 days. The winds are referenced to 10 m height. Only points where no rain is detected are retained for the study. There are typically two passes each day over the GoM. When there is more than one pass during a day, a composite with the various passes is calculated to obtain one wind field per day. Both the north-south and east-west components of the wind field are used simultaneously with the SST to form the extended matrix $\mathbf{X_e}$ for their reconstruction.

[15] QuikSCAT winds have a lower percentage of missing data (22.6%) than the SST data set, because it employs a radiometer sensor that sees through clouds. There are two main sources of missing data: the lack of satellite passes over the study zone and rain-contaminated areas. Spatially, the GoM has a higher percentage of missing data (up to ∼35% in average) than the Atlantic Ocean (∼20%). Coastal regions have higher percentage of missing data than the open ocean regions, mainly due to the coarse resolution of the product and the land-sea mask.

[16] In situ data are obtained from the Coastal Ocean Monitoring and Prediction System (COMPS http://comps.marine.usf.edu/) at the University of South Florida and the National Data Buoy Center (NDBC, http://www.ndbc.noaa.gov/). A total of seven stations measuring temperature at 1 m depth (Figure 2) are used for the validation of the results obtained with DINEOF. We also used seven stations measuring winds from NDBC (Figure 2) to validate the reconstruction of the QuikSCAT fields. The in situ winds are adjusted to a height of 10 m.

[17] The spatial correlation between the different data sets used in this work is presented in Figure 3. The advantage of using an EOF-based method is that, with a small computational cost, the spatial correlation between two variables is taken into account, and not only the correlation between two

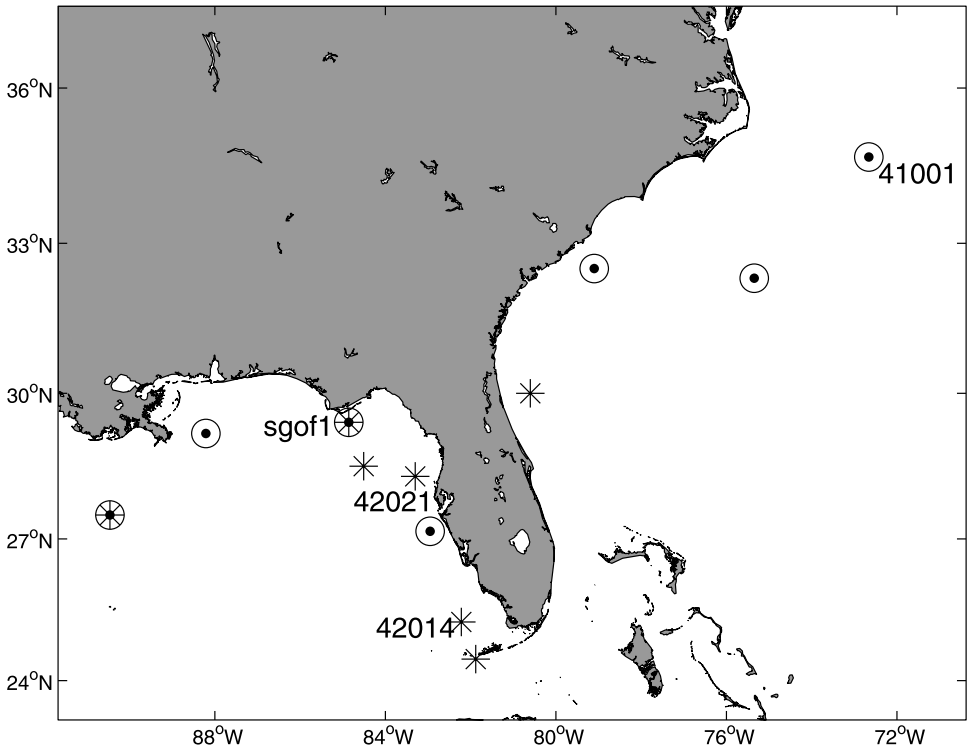

Figure 2. In situ stations used for the validation of the SST and wind reconstruction. Asterisks show SST station positions, and circles show the wind station positions. Stations 42021 and 42014 are COMPS buoys. Buoy sgof1 is a C-Man station owned by NDBC, and station 41001 is an NDBC buoy.

points at the same geographical location. For figure clarity, the correlation fields have been filtered, so only the main features are retained.

[18] The correlation between wind and temperature is more important near the coast, where the wind plays an important role in the ocean dynamics. The correlation between SST and the north-south component of the wind (Figure 3, V) has a clear positive/negative correlation indicating upwelling/downwelling along the eastern/western coast of Florida, respectively, with the highest values near the coast. The whole WFS presents this high correlation. The negative correlation on the GoM north coast also indicates upwelling/downwelling dynamics. The correlation between SST and the east-west component of wind (Figure 3, U) shows also an influence of the winds on the SST. The correlation between SST and the east-west component of wind is mainly positive in the GoM, in agreement with the upwelling/downwelling influence of the wind along the west coast of Florida. The negative correlation seen in the Atlantic Ocean in Figure 3 (U) shows a relation between the SST and the easterly winds coming from the central Atlantic Ocean.

[19] The correlation between SST and chlorophyll (Figure 3, CHL) is almost everywhere negative, meaning that cold waters have high chlorophyll concentration and warm waters have low chlorophyll concentration. The magnitude of the correlation is larger than that observed with winds, due to the direct relationship between these two variables. A multivariate reconstruction using SST and chlorophyll can be beneficial for both variables, with the disadvantage of both having a similar high cloud contamination. It is thus interesting to assess which combination leads to better results: SST with nearly complete wind fields, or SST with highly related chlorophyll, but with similar cloudiness.

4. Experiments

[20] Section 5 presents the results of the multivariate reconstruction of missing data using DINEOF, based on

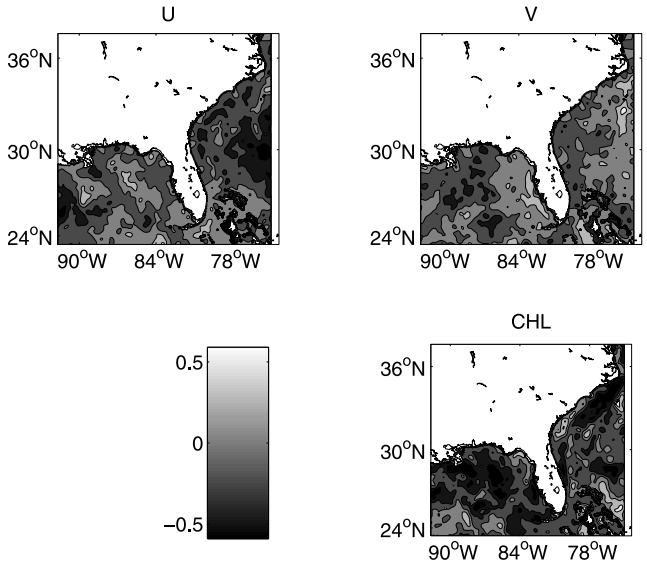

Figure 3. Correlation between the SST and both components of (top) wind and (bottom) the chlorophyll.

Table 1. Summary of the Errors Calculated Between the Reconstructed Fields and in Situ Data Over Initially Missing Points

	RMS, °C	Bias, °C	Correlation	SD,[a] °C
Only SST	0.76	−0.09	0.59	1.04
SST + 1 day lag wind	0.75	0.01	0.6	1.00
SST + CHL	0.62	0.05	0.71	0.93
SST +1 day lag SST + CHL	0.6	−0.01	0.69	0.89

[a]In situ data 0.97°C.

several combinations of satellite data, summarized here as follows: (1) only SST; this reconstruction is used as a control data set, to assess the impact of the different multivariate combinations on the reconstruction of SST; (2) SST + chlorophyll, the first of our multivariate reconstructions; matrix $\mathbf{X_e}$ is formed by SST and chlorophyll at time t; (3) SST + lagged winds; the effect of winds is not directly transferred to the ocean, so past winds are better correlated to current SST; in this case, SST at time t and both components of wind at time $t - 1$ (1-day lag) form matrix $\mathbf{X_e}$; the 1-day lag was chosen because it minimizes the error of the reconstruction when compared to in situ data; and (4) SST + chlorophyll + lagged SST; the SST can benefit from the presence of a lagged version of itself. In this experiment we use SST at time t with SST at time $t + 1$ and chlorophyll at time t.

5. Results

5.1. Temperature Reconstruction Results

[21] Table 1 presents the comparison between the reconstructed SST data and the in situ SST from the stations shown in Figure 2. The error is calculated at initially missing points. For reference, the RMS error between the original noncloudy SST data and in situ data is 0.65°C, which can be used as a reference value for the reconstructions. All multivariate reconstructions improve the results obtained by the univariate SST reconstruction, in terms of RMS, bias and correlation. The largest improvement corresponds to the combination of SST + chlorophyll + lagged SST, although all reconstructions are satisfactory quantitatively, since the reconstruction error of all experiments is of the order of the expected difference between the satellite and in situ data. We have performed a Student's t test to know if the improvement of the multivariate approach compared to the univariate approach is significant. We calculate the squared differences

$$d^i = \left(T^o - T^i\right)^2 \tag{3}$$

with T^o the observed temperature and T^i, ($i = 1\ldots4$) the reconstructed temperature (1 for SST only; 2 for SST + chlorophyll + lagged SST; 3 for SST + chlorophyll; 4 for SST + lagged winds). We then calculate the variance of the difference between the monovariate reconstruction and each of the multivariate reconstructions $x^i = d^1 - d^i$; $i = 2\ldots4$:

$$s_i^2 = \frac{1}{N-1}\sum_{n=1}^{N}\left(x_n^i - \bar{x^i}\right)^2 \tag{4}$$

with N the number of points of the time series x^i (816 in this case). The mean $\bar{x^i}$ is defined by

$$\bar{x^i} = \frac{1}{N}\sum_{n=1}^{N} x_n^i \tag{5}$$

We assume that x^i follows a Gaussian distribution, which in our case is approximately true. The t value is calculated as

$$t^i = \frac{|\bar{x^i}|}{s_i/\sqrt{N}} \tag{6}$$

With the time step being one day it is possible that a serial correlation is present in the time series x^i. This correlation can result in an overestimation of the t value. One way to account for serial correlation is to transform N into the effective sample size [see, e.g., *Wilks*, 1995] by multiplying N by the factor $(1 - \rho_1)/(1 + \rho_1)$, with ρ_1 the 1-day lag correlation of the time series x^i. The effective sample size becomes $N = 698$, the value used in the calculation of the t value. The t value is 3.4 for case 2, 4.8 for case 3 and 1.4 for case 4. Cases 2 and 3 exceed the threshold for significance of 1.96 at the 5% level. The error reduction in the SST +

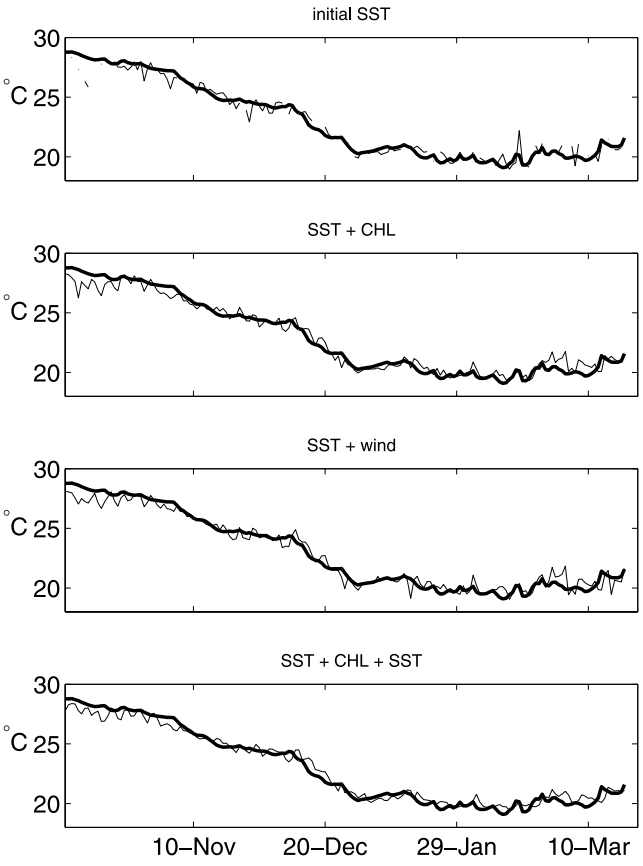

Figure 4. Comparison of the SST with buoy 42014. From top to bottom panels, initial (cloudy) SST, SST + chlorophyll, SST + lagged winds, and SST + chlorophyll + lagged SST. The thick line represents data from buoy 42014, and the thin line represents the SST interpolated at the position of this buoy.

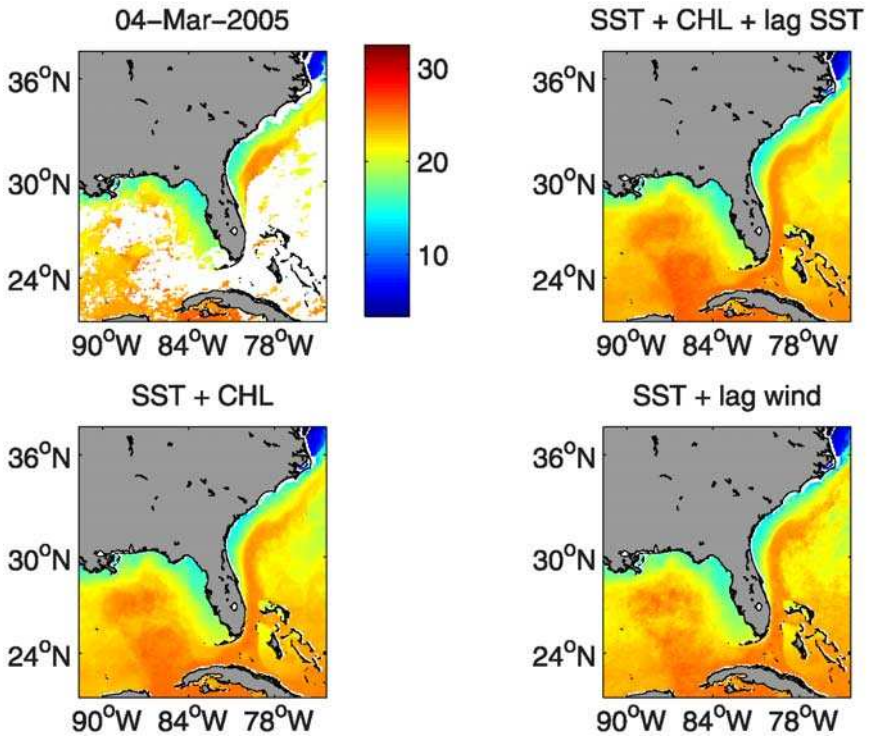

Figure 5. (top left) Original SST and (top right) its DINEOF reconstruction using SST + chlorophyll + lagged SST, (bottom left) SST + chlorophyll, and(bottom right) SST + lagged winds on 4 March 2005. Units are °C.

chlorophyll, and SST + chlorophyll + lagged SST reconstructions is then significantly smaller than the reconstruction of SST alone. The improvement achieved by the combination of SST with winds is very small, and it does not show a significant improvement with respect to the reconstruction of SST alone. However, it will be shown that the three multivariate combinations give very similar results when qualitatively compared.

[22] In Figure 4 we can see the comparison of all the presented reconstructions with buoy 42014 (Figure 2). The buoy sensor is located at 1 m depth, so we can observe that in general, the satellite data are influenced more strongly by short-term atmospheric fluctuations. This can be seen in both the initial data and the reconstructions. In addition, at the beginning of the time series there is a bias between the buoy and the satellite data, caused by the difference between the bulk temperature and the skin temperature, which is still important in October in the GoM. This bias is present in the initial SST data as well as in all reconstructions. Besides these differences between in situ and satellite data, the reconstructed SST data sets accurately follow the observations.

[23] For a qualitative comparison of the different SST reconstructions, Figure 5 shows the SST on 4 March 2005 for all data sets. Although the original SST is very cloudy in the GoM and off the U.S. east coast, the reconstruction of SST + chlorophyll + lagged SST, SST + chlorophyll, and SST + lagged winds all capture the main features of the SST distribution correctly. A large eddy is detached from the LC, which is clearly seen in all three reconstructed data sets, although it is not visible in the cloudy SST (the signature of this eddy is visible for about a month in the original SST, although with a cloudiness similar to what is observed in Figure 5). Since the LC consists of warm, oligotrophic water, characteristics easily detected in the SST and chlorophyll, it is not surprising that the two reconstructions combining SST with chlorophyll show a sharper contrast in the LC and the eddy, better defined than in the SST + lagged winds reconstruction. The benefit of using winds along with SST is clearer near the coast, as observed in the correlation between those variables (Figure 3), where the wind plays a stronger role in determining the characteristics of the surface water, through upwelling and surface mixing. The effect of wind on the SST reconstruction will be shown later with an example.

5.2. QuikSCAT Winds Reconstruction Results

[24] QuikSCAT winds used along with the SST are also reconstructed. Although they present a much smaller amount of missing data, some locations, as near the coast, are particularly problematic. Figures 6 and 7 show the comparison of the reconstructed QuikSCAT winds with stations 41001 and sgof1, respectively (Figure 2) from January to March 2005. The initial gappy data are also shown. The original QuikSCAT data are already very similar to the in situ data, capturing the overall wind field as well as high-frequency variations. At station sgof1 the original QuikSCAT data are particularly gappy (Figure 7), because of its proximity to coast and also because there is a higher amount of missing data in the GoM region. Even in these conditions the wind reconstruction realized by the SST + lagged winds shows good agreement with the in situ

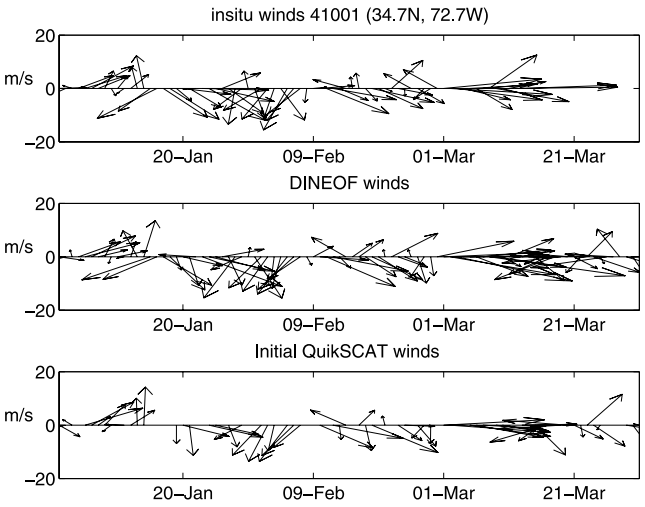

Figure 6. (top) Vector plot of buoy 41001, (middle) the reconstructed winds, and (bottom) the original QuikSCAT winds at the buoy position.

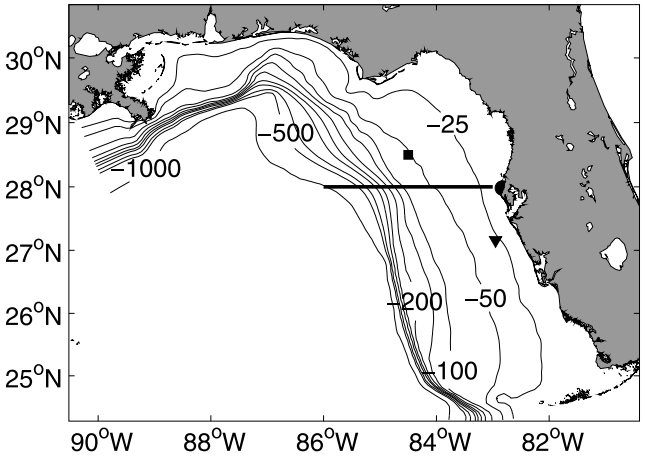

Figure 8. Transect studied in an upwelling/downwelling event in the West Florida Shelf (WFS). Contours and labels show the bathymetry. The square indicates buoy 42036 (owned by NDBC), the triangle shows buoy 42013 (owned by COMPS), and the dot at the right of the transect shows the position of the Clearwater tide gauge.

data given by the top panel of Figure 7. Considering the seven stations measuring wind shown in Figure 2, the RMS error in speed is 2.1 m/s for the initial data, 2.3 m/s for the whole reconstruction and 2.8 m/s for only the initially missing data. The mean direction difference between the initial gappy data and in situ data is 8.8°, and 14.6° for the reconstruction.

6. Study of an Upwelling/Downwelling Event in the West Florida Shelf

[25] From 30 December 2004 to 15 January 2005, we observe a series of upwelling/downwelling-favorable wind events over the WFS. NDBC buoy 42036 (Figure 8) shows a series of alternating upwelling/downwelling winds for the duration of the mentioned period (Figure 9). Starting on 4 January 2005, we can see two important downwelling-favorable wind periods.

[26] To study the impact of these events on the WFS, and the accuracy of the reconstructed data sets, we looked to a transect perpendicular to the coast, located at 28°N (Figure 8). The reconstructed QuikSCAT winds at this transect (Figures 10 and 11 for east-west and north-south components, respectively) show the upwelling/downwelling pulses observed at buoy 42036. The east-west component is predominantly negative, and the north-south component presents an alternation of positive and negative values, more evident near the coast. Given the orientation of the coast at 28°N, the north-south wind component is the most important in defining an upwelling or downwelling-favorable event.

[27] The partially clouded SST along 28°N is shown in Figure 12. The SST decreases with time near the coast, and

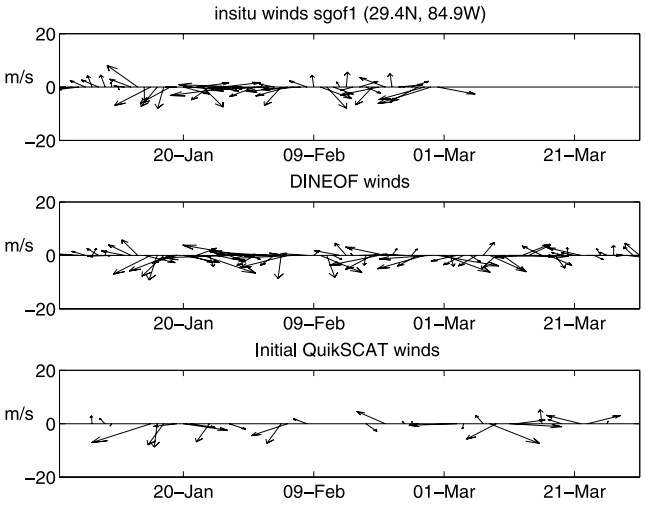

Figure 7. (top) Vector plot of buoy sgof1, (middle) the reconstructed winds, and (bottom) the original QuikSCAT winds at the buoy position.

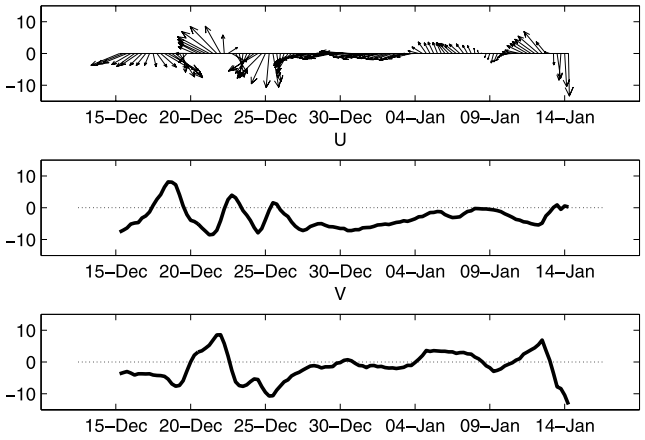

Figure 9. Wind at buoy 42036. (top) Winds vector plot; (middle) east-west component; and (bottom) north-south component. Units are m/s.

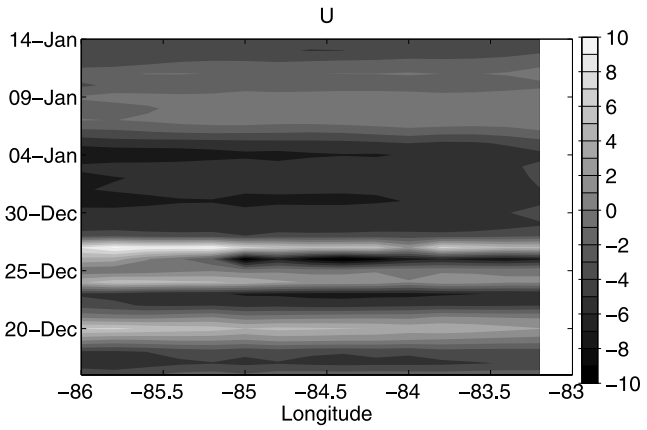

Figure 10. QuikSCAT reconstructed winds (east-west component) at 28°N on the WFS. Units are m/s.

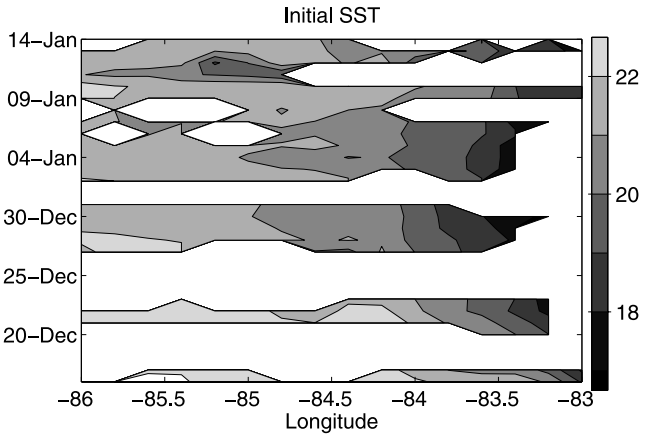

Figure 12. Cloudy SST at 28°N. Units are °C.

we notice a series of warm water/cold water pulses over the entire shelf. However, the cloudiness in this data set makes it difficult to study the duration and strength of these events. The different SST reconstructions (in Figures 13, 14, and 15 for SST + chlorophyll + lagged SST, SST + chlorophyll, and SST + lagged winds, respectively) are able to reproduce these events clearly. The pulses become stronger on 4 January 2005, when the two long downwelling wind periods, observed in Figure 9, start. The correlation between the SST and the reconstructed north-south wind component is very weak (∼0.01) from 15 December 2004 to 15 January 2005. However, if we only consider the period from 30 December to 15 January, when the warm/cold pulses are more evident in the SST, the correlation is 0.35, showing that the SST upwelling/downwelling pulses are caused by winds. This is further confirmed by the response of the sea surface height (SSH) at 28°N. Figure 16 shows the Hybrid Coordinate Ocean Model (HYCOM [*Bleck*, 2002]) SSH over the WFS transect. An oscillatory signal is observed, with high SSH corresponding to warm SST episodes, typical of a downwelling case, and vice versa. The response of the WFS under upwelling-favorable winds develops in a few hours [*Li and Weisberg*, 1999; *Weisberg et al.*, 2001a], which explains the wave-like cold water pulses alternating with the warm water, as wind changes from upwelling to downwelling-favorable direction.

[28] The three SST reconstructions are able to reproduce the series of upwelling and downwelling events, although there are some differences with the initial cloudy SST, as well as between the different reconstructions. For example, on 11 January 2005 there is a cold signal in the cloudy SST that is not observed in the reconstructions. DINEOF reconstructs missing data using a truncated EOF series, which results in the elimination of noise from the initial data set. Some small-scale features can be also smoothed by this procedure, and the cold water on 11 January 2005 is one example. However, this cold temperature can be also an artefact of the SST introduced by the vicinity of clouds, as none of the three multivariate reconstructions support this feature.

[29] A difference between the three SST reconstructions can be observed on 26 December 2004. The reconstruction of SST + chlorophyll, and SST + lagged winds show a cold event around this date, not seen in the SST + chlorophyll + lagged SST. The model SSH showed in Figure 16 presents low values over the entire WFS around this date, up to 85°W. Atmospheric total net heat flux from the National Centers for Environmental Prediction (NCEP) reanalysis model (available at http://www.cdc.noaa.gov/cdc/reanalysis/

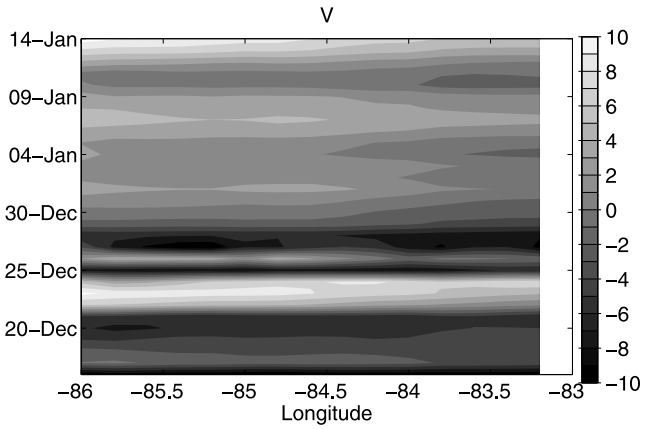

Figure 11. QuikSCAT reconstructed winds (north-south component) at 28°N on the WFS. Units are m/s.

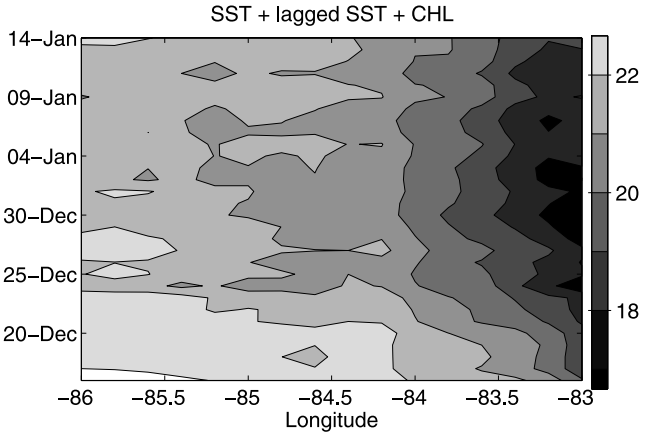

Figure 13. SST from the SST + chlorophyll + lagged SST reconstruction at 28°N. Units are °C.

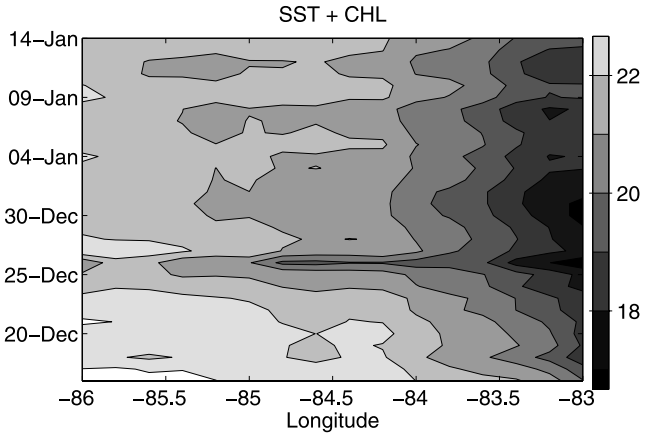

Figure 14. SST from the SST + chlorophyll reconstruction at 28°N. Units are °C.

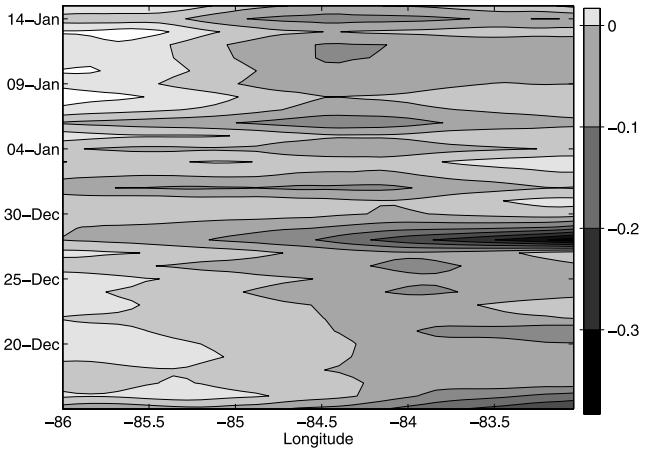

Figure 16. SSH from Hybrid Coordinate Ocean Model (HYCOM), at 28°N. Units are cm.

reanalysis.shtml, data not shown) shows a cooling event from 24 to 26 December 2004. This event is also supported by observations: buoy 42013, situated south of the transect (Figure 8), shows a temperature decrease of about 1.5°C on this date (Figure 17). An upwelling-favorable wind event can be observed at buoy 42036 (Figure 9), as well as in the reconstructed QuikSCAT winds (Figures 10 and 11). Finally, a tide gauge situated at Clearwater (Figure 8) also shows a sharp decrease in the water level on this date (Figure 18), typical of an upwelling event. The temperature decrease is represented in the SST + chlorophyll, and SST + lagged winds data sets, and the various data sources mentioned support this event. However, the upwelling happens during a gap on the initial data set, so we cannot be certain that this event is real though all indirect indications suggest so. We have observed that this kind of cold events happen in fall/winter on the WFS. Figure 19 shows a sequence of cloudy SST on November 2005. This time the clouds do not cover the entire region, so we are able to see a rapid decrease in temperature on 2 November 2005, a decrease that lasts only one day. The HYCOM SSH, the NCEP atmospheric forcings and the temperature at buoy 42013 all support this event in the same way as for the one reconstructed by DINEOF on December 2004. The absence of the 26 December 2004 event in the SST + chlorophyll + lagged SST reconstruction can be explained by the high cloud coverage observed over the whole WFS. The SST reinforced with 1-day lagged SST does not add information because of the cloudiness during those dates, resulting in a smoother SST reconstruction. As seen in Table 1, the reconstruction of SST + chlorophyll + lagged SST presents a lower standard deviation than the other reconstructions. A low-variability reconstruction can result in a better RMS error, but it is also likely to miss localized and extreme events such as this strong upwelling event.

[30] We observe (Figure 17) that the water column on the WFS evolves from well-mixed during the first half of the study period to a stratified situation after the cold event on 26 December 2004. Under stratified conditions, the response to upwelling-favorable winds is found to exceed the response to downwelling-favorable winds [*Weisberg et al.*, 2001b], a result that is also seen in this data set: whereas the downwelling winds for events starting on 4 January 2005 and 10 January 2005 are longer and more persistent than the intervening upwelling winds (Figure 9), the SST warm pulses are similar in strength to the cold pulses

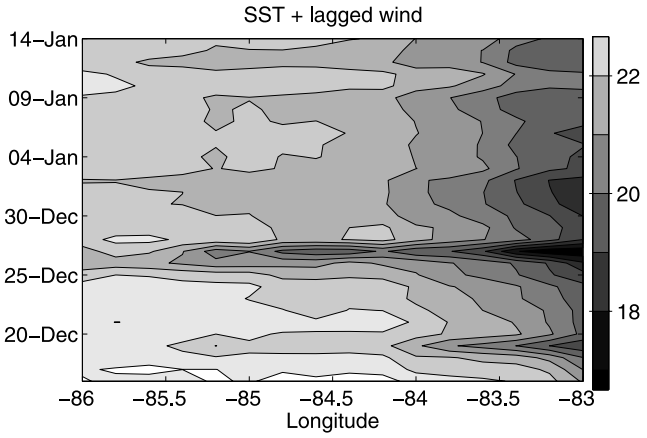

Figure 15. SST from the SST + lagged winds reconstruction at 28°N. Units are °C.

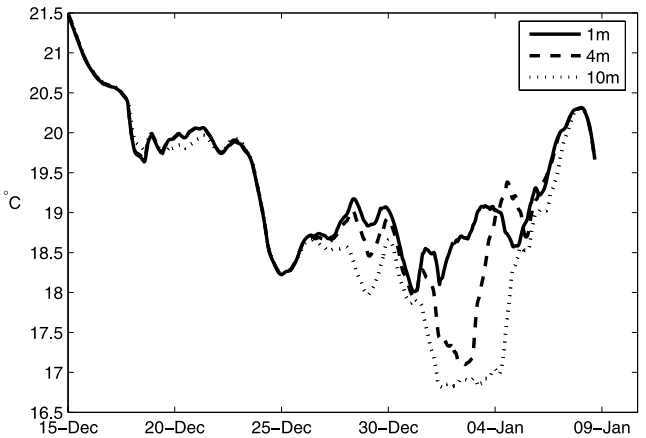

Figure 17. Temperature measured at buoy 42013.

produced by much weaker winds (as seen, for example, in Figure 14).

7. Conclusions

[31] We successfully applied a multivariate EOF-based technique, DINEOF, to reconstruct the missing data of a series of satellite products. To our knowledge, this is the first time that a multivariate EOF-based technique is used to reconstruct missing data, and we showed that the results improve upon a univariate approach. Sea Surface Temperature (SST), winds and chlorophyll were used in three combinations: SST +1-day lagged winds, SST + chlorophyll, and SST +1-day lagged SST + chlorophyll. These reconstructions were compared to in situ data over the Gulf of Mexico (GoM) and the U.S. east coast. The SST + chlorophyll + lagged SST reconstruction gave the best results in terms of RMS, bias and correlation. All three experiments improved the reconstruction of SST alone by the same technique. The SST + chlorophyll + lagged SST case, however, presents smoother results. The three reconstruction experiments gave qualitatively similar results.

[32] We have observed that the length of the data set does not influence the outcome of the SST + chlorophyll + lagged SST, and SST + chlorophyll cases. We analyzed 3-month and a 6-month experiments, the latter one presented here, and the results were comparable qualitatively and quantitatively, which shows the robustness of the method. In the case of the SST + lagged winds reconstruction, however, the results improved with the 6-month experiment. This finding indicates that the correlation between the SST and winds, weaker than between SST and chlorophyll, is not fully explained in short time periods, and thus DINEOF needs more information to converge to the solution.

[33] When using lagged versions of the same variable, like in the SST + chlorophyll + lagged SST, one should establish which one of the two reconstructions of the lagged variable gives better results, as there is more than one estimate. We checked this for the SST + chlorophyll + lagged SST case, but the skill measured was very similar for both SST versions.

[34] For the results obtained, we conclude that the combination of SST with related variables gives better results than when using the SST alone. The combination of SST + chlorophyll + lagged SST misses some localized events due to a lower variability of the reconstructed SST. The SST + chlorophyll reconstruction presents a higher variability. The choice of SST with winds is qualitatively very similar to the other two experiments, but the reduction of error with respect to the reconstruction of SST alone is not significant. However, this combination gave good results near the coast, where the wind plays a more important role in the SST evolution.

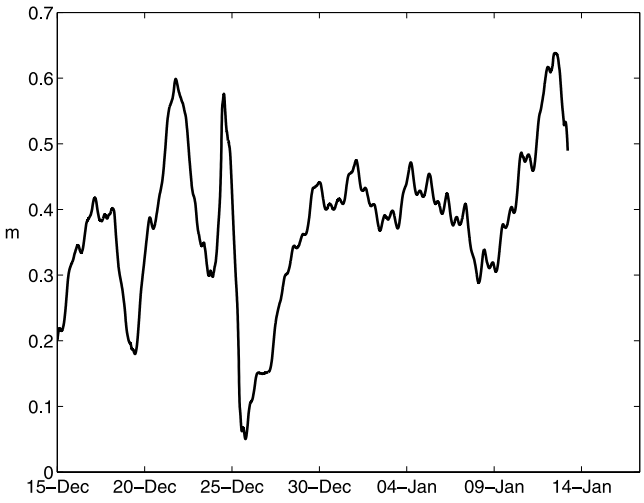

Figure 18. Water level at Clearwater.

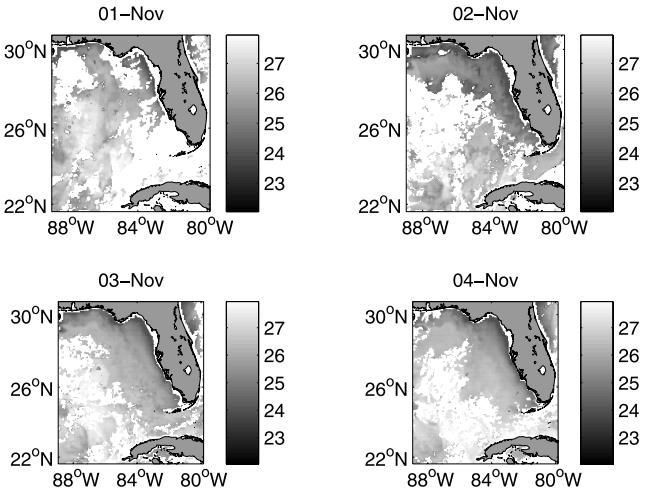

Figure 19. Cloudy SST from 1 to 4 November 2005. A rapid decrease in temperature is observed on 2 November.

[35] A series of upwelling/downwelling events were detected in the West Florida Shelf (WFS). These events, of high ecological importance for the WFS, are supported by observations, and all multivariate SST reconstructions correctly represent them. The reconstructed QuikSCAT winds are also in accordance with observations.

[36] To conclude, the multivariate DINEOF reconstruction gives accurate results, and it allows an easy combination of different variables. Other multivariate techniques, like multivariate optimal interpolation in its generally used form, require a priori knowledge of the statistics of each variable, as well as of the correlation between variables. DINEOF is more objective because it calculates these statistics internally and based on available data.

[37] **Acknowledgments.** This research has been funded by Office of Naval Research grants N00014-02-1-0972 and N00014-05-1-0483. The authors wish to thank Vincent Toumazou for providing the code for the EOF decomposition, iterative EOF, which allowed reduction of the computational cost of DINEOF. The National Fund for Scientific Research (FNRS) is also acknowledged. J.-M. Beckers is Honorary Research Associate at the FNRS, Belgium. AVHRR SST and chlorophyll data are collected, processed, and distributed by the University of South Florida (http://imars.usf.edu). QuikSCAT wind data were obtained from the Physical Oceanography Distributed Active Archive Center (PODAAC) at the NASA Jet Propulsion Laboratory, Pasadena, California (http://podaac.jpl.nasa.gov). In situ winds and temperature data were obtained from the Coastal Ocean Monitoring and Prediction System (COMPS http://comps.marine.usf.edu/) at the University of South Florida and from the National Data Buoy Center (NDBC, http://www.ndbc.noaa.gov/). The HYCOM group is acknowledged for the NAT HYCOM SSH data used in this work. NCEP Reanalysis data were provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, from their Web site at http://www.cdc.noaa.gov/. MARE publication MARE049.

References

Alvera-Azcárate, A., A. Barth, M. Rixen, and J. M. Beckers (2005), Reconstruction of incomplete oceanographic data sets using empirical orthogonal functions: Application to the Adriatic Sea, *Ocean Modell.*, *9*, 325–346.

Beckers, J.-M., and M. Rixen (2003), EOF calculations and data filling from incomplete oceanographic data sets, *J. Atmos. Oceanic Technol.*, *20*(12), 1839–1856.

Beckers, J.-M., A. Barth, and A. Alvera-Azcárate (2006), DINEOF reconstruction of clouded images including error maps: Application to the sea surface temperature around Corsican Island, *Ocean Sci.*, *3*(4), 735–776.

Bennett, A. F. (2002), *Inverse Modeling of the Ocean and the Atmosphere*, 234 pp., Cambridge Univ. Press, New York.

Bergant, K., M. Susnik, I. Strojan, and A. Shaw (2005), Sea level variability at adriatic coast and its relationship to atmospheric forcing, *Ann. Geophys.*, *23*, 1997–2010.

Bleck, R. (2002), An oceanic circulation model framed in hybrid isopycnic-Cartesian coordinates, *Ocean Modell.*, *1*, 55–88.

Bretherton, C., C. Smith, and J. Wallace (1992), An intercomparison of methods for finding coupled patterns in climate data, *J. Clim.*, *5*, 541–560.

Carter, E. F., and A. R. Robinson (1987), Analysis models for the estimation of oceanic fields, *J. Atmos. Oceanic Technol.*, *4*(1), 49–74.

Collins, D. C., C. J. C. Reason, and F. Tangang (2004), Predictability of Indian Ocean sea surface temperature using canonical correlation analysis, *Clim. Dyn.*, *22*, 481–497.

Fraedrich, K., J. L. McBride, W. M. Frank, and R. Wang (1997), Extended EOF analysis of tropical disturbances: TOGA COARE, *J. Atmos. Sci.s*, *54*, 2363–2372.

Ghil, M., et al. (2002), Advanced spectral methods for climatic time series, *Rev. Geophys.*, *40*(1), 1003, doi:10.1029/2000RG000092.

Gomis, D., S. Ruiz, and M. A. Pedder (2001), Diagnostic analysis of the 3D ageostrophic circulation from a multivariate spatial interpolation of CTD and ADCP data, *Deep Sea Res., Part I*, *48*, 269–295.

Grodsky, S. A., and J. A. Carton (2001), Intense surface currents in the tropical Pacific during 1996–1998, *J. Geophys. Res.*, *106*(C8), 16,673–16,684.

He, R., R. H. Weisberg, H. Zhang, F. E. Muller-Karger, and R. W. Helber (2003), A cloud-free, satellite-derived, sea surface temperature analysis for the West Florida Shelf, *Geophys. Res. Lett.*, *30*(15), 1811, doi:10.1029/2003GL017673.

Jolliffe, I. T. (2002), *Principal Component Analysis*, 487 pp., Springer, New York.

Kim, K. Y., and Q. Wu (2000), Optimal Detection using cyclostationary EOFs, *J. Clim.*, *13*(5), 938–950.

Kondrashov, D., and M. Ghil (2006), Spatio-temporal filling of missing data in geophysical data sets, *Nonlinear Processes Geophys.*, *13*, 151–159.

Kondrashov, D., Y. Feliks, and M. Ghil (2005), Oscillatory modes of extended Nile River records (A. D. 622–1922), *Geophys. Res. Lett.*, *32*, L10702, doi:10.1029/2004GL022156.

Korres, G., N. Pinardi, and A. Lascaratos (2000), The ocean response to low-frequency interannual atmospheric variability in the Mediterranean Sea. part II: Empirical orthogonal functions analysis, *J. Clim.*, *13*, 732–745.

Li, Z., and R. Weisberg (1999), West Florida continental shelf response to upwelling favorable wind forcings: 2. Dynamics, *J. Geophys. Res.*, *104*(C10), 23,427–23,442.

McClain, E., W. Pichel, and C. Walton (1985), Comparative performance of AVHRR-based multichannel sea surface temperatures, *J. Geophys. Res.*, *90*(C14), 11,587–11,601.

Paris, C. B., R. K. Cowen, K. M. M. Lwiza, D.-P. Wang, and D. B. Olson (2002), Multivariate objective analysis of the coastal circulation of Barbados, West Indies: Implication for larval transport, *Deep Sea Res., Part I*, *49*, 1363–1386.

Reynolds, R. W., and T. Smith (1994), Improved global sea surface temperature analyses using optimum interpolation, *J. Clim.*, *7*, 929–948.

Smith, T. M., and R. W. Reynolds (2003), Extended reconstruction of global sea surface temperatures based on COADS data (1854–1997), *J. Clim.*, *16*, 1495–1510.

Toumazou, V., and J. F. Cretaux (2001), Using a Lanczos eigensolver in the computation of empirical orthogonal functions, *Mon. Weather Rev.*, *129*(5), 1243–1250.

Tourre, Y. M., and W. B. White (1997), Evolution of the ENSO signal over the Indo-Pacific domain, *J. Phys. Oceanogr.*, *27*, 683–696.

von Storch, H., and W. Zwiers (1999), *Statistical Analysis in Climate Research*, 484 pp., Cambridge Univ. Press, New York.

Wallace, J., C. Smith, and C. Bretherton (1992), Singular value decomposition of wintertime sea surface temperature and 500-mb height anomalies, *J. Clim.*, *5*, 561–576.

Watts, R., X. Qian, and K. Tracey (2001), Mapping abyssal current and pressure fields under the meandering Gulf Stream, *J. Atmos. Oceanic Technol.*, *18*, 1052–1067.

Weare, B. C., and J. S. Nasstrom (1982), Examples of extended empirical orthogonal function analysis, *Mon. Weather Rev.*, *110*, 481–485.

Weisberg, R. H., B. Black, and Z. Li (2001a), An upwelling case study on Florida's west coast, *J. Geophys. Res.*, *105*(C5), 11,459–11,469.

Weisberg, R. H., Z. Li, and F. Muller-Karger (2001b), West Florida shelf response to local wind forcing: April 1998, *J. Geophys. Res.*, *106*(C12), 31,239–31,262.

Wilks, D. S. (1995), *Statistical Methods in the Atmospheric Sciences*, 467 pp., Elsevier, New York.

A. Alvera-Azcárate, A. Barth, and R. H. Weisberg, College of Marine Science, University of South Florida, 140 7th Avenue South, Saint Petersburg, FL 33701, USA. (aalvera@marine.usf.edu)

J.-M. Beckers, GHER-AGO, University of Liège, Allée du 6 Août 17, B5, Sart Tilman, B-4000 Liège, Belgium.