Remote Sensing Letters
Vol. 3, No. 8, 20 December 2012, 697–706

Taylor & Francis
Taylor & Francis Group

Cloud-free sea surface temperature and colour reconstruction for the Gulf of Mexico: 2003–2009

YAO ZHAO and RUOYING HE*
Department of Marine, Earth, and Atmospheric Sciences, North Carolina State University, Raleigh, NC 27695-8208, USA

(*Received 28 October 2011; in final form 9 February 2012*)

The data interpolating empirical orthogonal function (DINEOF) method is applied to concurrent MODIS sea surface temperature (SST) and chlorophyll-*a* (chl-*a*) data to produce daily, 4 km, cloud-free SST and chl-*a* analyses for the Gulf of Mexico (GOM) from 2003 to 2009. Comparisons between SST analysis and *in situ* buoy temperature measurements indicate that the DINEOF method can accurately resolve temperature variability and solve the cloud-cover problem, which is a typical issue of remote-sensing observations. Based on significant correlations between cloud-free chl-*a*, SST and sea surface height (SSH) data in the GOM, a simple chl-*a* statistical prediction model is further developed. Favourable comparisons between model solutions and independent satellite chl-*a* observations indicate that the statistical model provides a feasible means to predict GOM chl-*a* field based on existing SST and SSH information, allowing for observational gap fillings when concurrent chl-*a* data were not available.

1. Introduction

The Gulf of Mexico (GOM) is a semi-enclosed marginal sea surrounded by the North American continent and the island of Cuba. Ocean waters in the GOM are connected with the Atlantic Ocean through the Florida Straits and with the Caribbean Sea through the Yucatan Channel. The size of the Gulf basin is approximately 1.6 million km^2, with the deepest depth being 4384 m in the central gulf (Oey *et al*. 2005, Schmitz *et al*. 2005). The GOM has a wide array of valuable resources, such as natural gas, petroleum and fisheries. Material properties in the gulf are transported largely by the anticyclonic Loop Current and Loop Current rings in the deep water and by coastal circulation on the shallow continental shelves that together take up about half the size of the Gulf basin (Adams *et al*. 2004). Improved understanding on the Loop Current and shelf circulation dynamics is therefore crucial for protecting the well-being of regional marine ecosystems, exploration and management of natural resources and prevention and mitigation of both natural and man-made marine environmental problems, such as harmful algal blooms and oil spills (e.g. Walsh *et al*. 2003, North *et al*. 2011).

Satellite remote sensing is arguably the most important routine observational method in monitoring ocean circulation features. Compared with shipboard or fixed mooring samplings, satellite observations, such as sea surface temperature (SST),

*Corresponding author. Email: rhe@ncsu.edu

Remote Sensing Letters
ISSN 2150-704X print/ISSN 2150-7058 online © 2012 Taylor & Francis
http://www.tandf.co.uk/journals
http://dx.doi.org/10.1080/01431161.2012.666638

ocean colour and sea surface height (SSH), provide two-dimensional mappings of global ocean in a matter of days to several weeks, offering the only observational approach suitable for large-scale routine sampling of physical and biological properties of the ocean (McClain *et al*. 2004, Signorini and McClain 2011). However, this technology is not without limitations when considering the integrity of its data sets. For example, radiometer-based satellite data often suffer from cloud-cover problem, resulting in no observations at all or poor data quality. These problems have imposed a significant restriction on using satellite SST and chlorophyll-*a* (chl-*a*) data to study ocean physical and biological processes in the global ocean, including the GOM. For instance, we currently do not have concurrent, long-term cloud-free SST and chl-*a* time series that can be used to study important regional oceanographic processes, such as Loop Current Eddy separations and shelf-deep ocean exchanges. The objective of this study in attempting to solve this problem is thus twofold: (1) to apply a new data reconstruction method on satellite observations to generate a cloud-free SST and chl-*a* data set for the GOM and (2) to develop a statistical chl-*a* model based on existing satellite SST and SSH observations, so that the model can be used to predict chl-*a* values in the GOM when concurrent chl-*a* data are not available.

2. Daily, cloud-free SST, chl-*a* reconstructions

We utilized SST and chl-*a* observations from MODIS Aqua satellite to reconstruct daily, continuous, cloud-free SST and chl-*a* data sets. For the first time in satellite history, MODIS Aqua measures long-term concurrent ocean temperature and colour fields, allowing examinations of covariation of SST and chl-*a*. However, like most ocean satellite sensors, MODIS also suffers the cloud-cover problem.

Our analysis was based on MODIS SST and chl-*a* 4 km resolution, Level 3 products. Daily data of 7 years from 1 January 2003 to 31 December 2009 were collected in a 325 pixel × 455 pixel grid covering the GOM between 22–38° N and 73–82° W. The GOM has significant cloud-cover problem due to strong convection associated with warm gulf ocean temperature. For instance, the average data-missing ratio due to cloud contamination is about 70% in 2003. To ensure the quality of cloud-free reconstruction, daily data with cloud contamination ratio exceeding 95% were removed from both the SST and chl-*a* data sets. With this procedure, among an initial data set of 2557 days, 2335 daily images were retained for our reconstruction analysis.

We used data interpolating empirical orthogonal function (DINEOF) method, which is a parameter-free, empirical orthogonal function (EOF)-based method for the reconstruction of missing values in satellite data (Alvera-Azcárate *et al*. 2007). Compared with classic methods such as optimal interpolation (OI), DINEOF was shown to be able to provide similar results more efficiently (30 times faster) (Alvera-Azcárate *et al.* 2005). Another advantage of DINEOF over OI is that it does not need a user to supply any prior information, such as decorrelation scales and covariance functions, and both are computed internally in DINEOF (Beckers *et al.* 2006).

Interested readers can refer to Beckers and Rixen (2003), Alvera-Azcárate *et al*. (2005) and more recently, Miles *et al.* (2009) and Miles and He (2010) for more detailed method descriptions and applications. The basic procedures of DINEOF can be summarized as follows: first, the observational data are stored in a matrix $\mathbf{X}$ ($m \times n$) with m, n the spatial and time dimensions, respectively. All missing data in matrix $\mathbf{X}$ are replaced by 0. Second, the single value decomposition (SVD) method is used to generate the spatial and temporal EOFs $\mathbf{U}$ and $\mathbf{V}$ as well as the square matrix $\mathbf{D}$ which

contains the singular values. Third, the first optimal number (N) of EOFs is truncated to obtain the interpolated value at the missing data points through the equation:

$$\mathbf{X}_a = \mathbf{U}_N \mathbf{D}_N \mathbf{V}'_N = \sum_{k=1}^{N} \rho_k u_k v'_k. \tag{1}$$

The optimal number ($N = 7$) is determined by cross-validation. Then, with the new values for the missing data, the SVD is performed again for $\mathbf{X_a}$. These last two steps are repeated until convergence (a preset threshold of the absolute value of the difference between the SVD of the current and previous iterations) is obtained for the missing values.

As demonstrated by Alvera-Azcárate *et al.* (2007), multivariate DINEOF can significantly improve the quality of reconstruction. In such a case, extended EOFs can be calculated from an extended matrix $\mathbf{X_e}$, where different variables are combined to form an extended, multivariate matrix. With these extended EOFs, a given missing point of one variable can benefit from the presence of other variables at the same time. The correlation between different related variables can thus help to reconstruct missing data. Additionally, extended EOFs can resolve moving patterns more accurately because of the presence of future and/or past information (Alvera-Azcárate *et al*. 2007). In this study, the multivariate DINEOF used was based on the combination of concurrent SST and chl-*a* fields observed by MODIS. After reconstruction, only missing (cloud-covered) data are filled in and original valid data are retained. Figure 1, as an example, demonstrates DINEOF's ability to construct time- and space-continuous SST and chl-*a* fields.

Because there is no *in situ* chl-*a* data for use, *in situ* temperature data from the National Data Buoy Center (NDBC) (http://www.ndbc.noaa.gov/) were utilized to validate cloud-free SST reconstruction. A total of six buoys (42001, 42002, 42003, 42040, 42055 and 42056 with locations shown in figure 1(*b*)) in the GOM were selected. Buoy-observed hourly time series were averaged into daily means for comparison with DINEOF-reconstructed cloud-free daily SST time series. Temporal correlations between DINEOF SST and buoy-measured SST were calculated. Additionally, to examine temperature comparison on synoptic scale, we also removed temperature seasonal cycles derived by the harmonic analysis from both *in situ* observed and DINEOF-analysed SST time series. As shown in table 1, when seasonal cycles are retained, correlation coefficients between the two are all above 0.92. When seasonal cycles are removed, correlation coefficients are all statistically significant, ranging from 0.49 to 0.69 at six stations. Some potential errors could contribute to the difference between DINEOF SST and buoy-measured SST, including spatial offsets between buoy locations and the DINEOF SST 4 km footprint and differences between satellite-derived ocean skin temperature and buoy bulk temperature measured 1–2 m below the surface. In the light of these potential errors, the decent correlations between analysis and *in situ* data suggest that DINEOF is a robust method for daily, cloud-free data reconstruction.

3. Building a statistical prediction model for chl-*a*

With long-term (9 years), daily, cloud-free and concurrent chl-*a* and SST DINEOF reconstruction fields, we can analyse their intrinsic covariations. The calculation of temporal correlation between SST and the log-transformed chl-*a* at each grid point

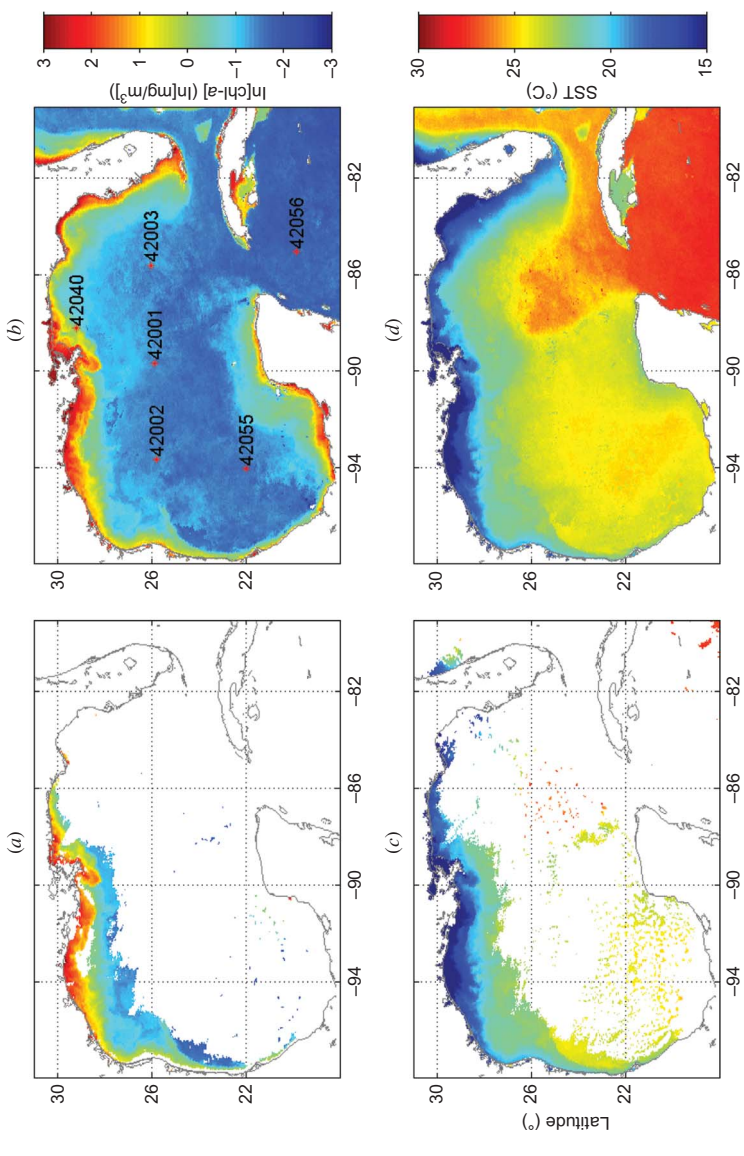

Figure 1. Comparisons of MODIS chl-*a* concentration (*a*) and (*c*) SST and their DINEOF-reconstructed cloud-free chl-*a* concentration (*b*) and SST (*d*) on 3 January 2003. Panel (*b*) also shows locations of six NDBC buoys that provide *in situ* surface temperature data for validation.
Note: SST, sea surface temperature; DINEOF, data interpolating empirical orthogonal function; NDBC, National Data Buoy Center.

Table 1. Point-by-point comparisons between buoy-observed SST and DINEOF SST.

Buoy number	Available data time series	Observed/DINEOF SST correlation coefficient (seasonal cycle retained/removed)
42001	1 January 2003–30 December 2009	0.95/0.69
42002	1 January 2003–30 December 2009	0.97/0.49
42003	1 January 2003–30 December 2009	0.92/0.57
42040	1 January 2003–30 December 2009	0.97/0.63
42055	1 January 2005–30 December 2009	0.96/0.49
42056	1 January 2005–30 December 2009	0.92/0.56

Notes: SST, sea surface temperature; DINEOF, data interpolating empirical orthogonal function; NDBC, National Data Buoy Center.
Column 1 lists the NDBC numbers of each buoy being considered, Column 2 shows the comparison period, given the data availability of each buoy, and Column 3 shows the correlation coefficient computed with SST seasonal cycle retained/removed from the time series.

reveals that a significant negative (∼–0.6) correlation exists in the GOM (figure 2(*a*)), indicating that cold (warm) SSTs correspond to high (low) pigment concentration. Such a correlation dominates the deep water and a significant portion of shelf regions as well, although it starts to break down in the coastal areas of west Florida shelf, Mississippi delta, coast of Cuba, Bay of Campeche and Campeche Bank, where variations of chl-*a* concentration are known to be dominated by terrestrial influence (i.e. riverine nutrient input) rather than open ocean physical processes.

Similar temporal correlation pattern (figure 2(*b*)) is seen in the same 7-year period between the log-transformed DINEOF chl-*a* and SSH time series. For the latter SSH data, we used 1/12 degree, daily HYCOM/NCODA model analysis (Chassignet *et al.* 2005) that assimilates satellite altimetry data (and thus gap-free and continuous). Although correlation coefficients are not as high as those with SST, negative correlations are significant (∼−0.3 to 0.4) at 95% confidence level. Some positive correlations are seen in similar coastal regions mentioned above, but overall the negative correlations between log(chl-*a*) and SST (SSH) in most of the GOM areas provide the statistical basis for constructing a chl-*a* prediction model based on SST and SSH information.

To test this idea, a multivariate regression model was proposed as follows:

$$\ln(\text{chl-}a) = a_0 + a_1 \times \text{SST} + a_2 \times \text{SSH}. \tag{2}$$

We used DINEOF-reconstructed cloud-free SST and chl-*a* fields and HYCOM/NCODA SSH fields from 2004 to 2009 as the training data sets to derive fitting functions (a_0, a_1, a_2) in the least square method. To account for temporal variations, these data sets were grouped and analysed on a monthly basis, resulting in 12 sets of monthly fitting functions (a_0, a_1, a_2) that can be used to predict the chl-*a* on any given day falling in its corresponding month.

To evaluate the utility of this statistical prediction model, we then compared DINEOF chl-*a* data in 2003 (that we purposely withhold as the validation set) against those chl-*a* values predicted by the model based on SST and SSH data in 2003. Model skills were evaluated using snapshot comparisons between model-predicted chl-*a* and DINEOF-reconstructed chl-*a* and time series of their spatial correlation

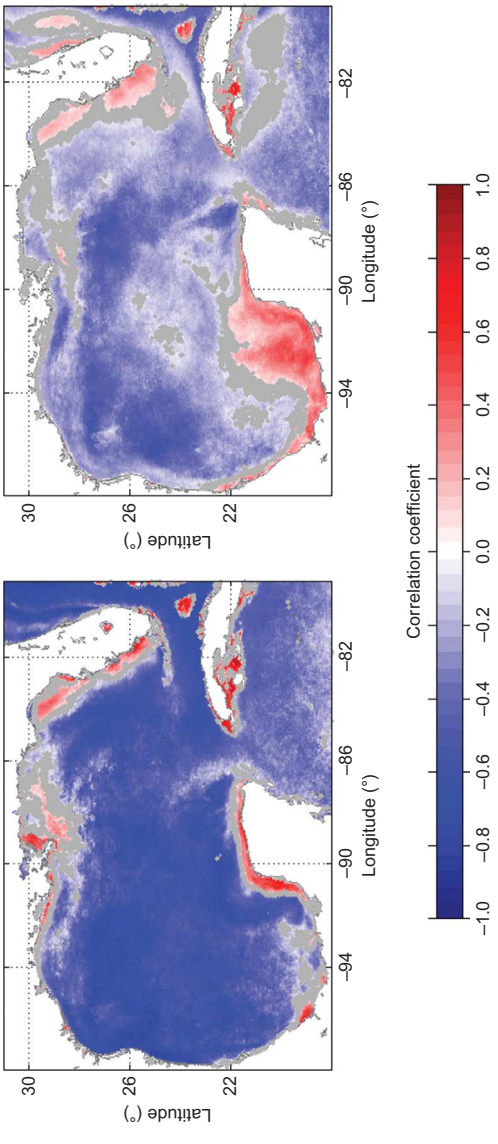

Figure 2. Temporal correlations of (*a*) chl-*a* and SST and (*b*) chl-*a* and SSH, obtained by computing the correlations between ln(chl-*a*) and SST and between ln(chl-*a*) and SSH at each pixel over the entire GOM region. The calculations use daily time series of ln(chl-*a*), SST and SSH over the period of 2003–2009. Regions with insignificant correlation (at 95% confidence level) are highlighted in grey.

Note: SST, sea surface temperature; SSH, sea surface height; GOM, Gulf of Mexico.

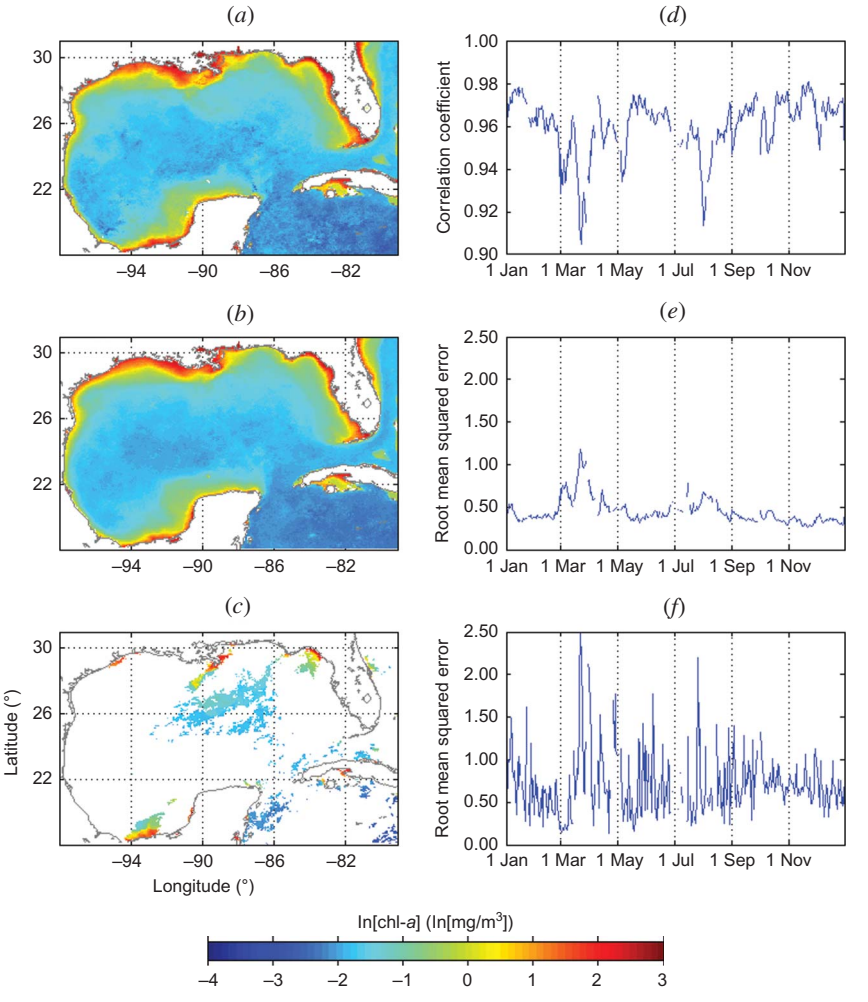

Figure 3. Comparison among (*a*) DINEOF-reconstructed cloud-free chl-*a* concentration, (*b*) statistical model-predicted chl-*a* concentration and (*c*) raw MODIS chl-*a* concentration on 16 December 2003. To highlight the spatial distribution, all these three chl-*a* fields are transformed into the log scale. Also shown on the right are (*d*) time series of spatial correlation coefficients between daily DINEOF chl-*a* concentration and statistical model predicted chl-*a* concentration in 2003; (*e*) time series of RMSE between DINEOF chl-*a* concentration and statistical model-predicted chl-*a* concentration in 2003; and (*f*) time series of RMSE between statistical model-predicted chl-*a* concentration and raw MODIS chl-*a* concentration in 2003.

Note: RMSE, root mean square error; DINEOF, data interpolating empirical orthogonal function.

in 2003. As a very representative example, a direct snapshot comparison between MODIS raw chl-*a*, DINEOF-reconstructed cloud-free chl-*a* and model-predicted chl-*a* on 16 December 2003 is given in figure 3(*a*) and 3(*b*). It is seen that the model-predicted and DINEOF-reconstructed chl-*a* show high resemblance, and they both provide much more complete spatial information than the cloud-covered MODIS raw data (figure 3(*c*)). Throughout the year of 2003, the spatial correlations between

model-predicted and DINEOF-reconstructed chl-*a* are all above 0.94 (figure 3(*d*)) and the corresponding root mean square errors (RMSEs) between model-predicted and DINEOF-reconstructed chl-*a* and between model-predicted and MODIS-observed chl-*a* are all smaller than 2.5 mg/m^3, indicating that the simple chl-*a* prediction model developed above is capable of reproducing the temporal and spatial variability of chl-*a* in the GOM.

4. Summary and conclusions

Daily, cloud-free SST and chl-*a* data sets are reconstructed for the GOM by applying a recently developed DINEOF method on MODIS satellite observations. The complete reconstructed SST and chl-*a* time series from this study are available for download by interested readers at http://omgrhe.meas.ncsu.edu/RSL. Favourable comparisons between SST reconstructions and *in situ* buoy temperature observations confirm the utility of DINEOF method in solving the cloud-cover problem. Significant correlations are identified between DINEOF chl-*a* and DINEOF SST, as well as HYCOM/NCODA-based SSH analysis for a major portion of the GOM, providing a statistical basis for the construction of chl-*a* prediction model based on SST and SSH information. The model-predicted chl-*a* fields compare well with the validation data set in 2003, suggesting that the statistical model developed here provides a feasible means to predict GOM chl-*a* fields based on existing SST and SSH information and to fill in the data gaps when concurrent chl-*a* observations were not available in the GOM.

The covariation between chl-*a* and SST and SSH has been documented by other investigations in other regions of the global ocean. For example, McGillicuddy *et al.* (2001) showed that higher (lower) pigment biomass occurs in mesoscale features consisting of cold (warm) temperature anomalies, indicating an inverse relationship between SST and chl-*a* in the Sargasso Sea. He *et al.* (2010) studied the concurrent SST and colour observations obtained by MODIS for the Mid-Atlantic Bight shelf-break and similarly detected the negative correlation between SST and surface pigment variation. Like in the GOM, the negative correlation between chl-*a* and SSH has been detected elsewhere on a variety of scales ranging, for instance, from upwelling areas with higher productivity and to the mesoscale cyclonic eddies with an increased productivity observed in the thermocline doming regime (e.g. Wilson and Adamec 2002). Some positive correlations between chl-*a* and physical variables were also detected by several studies (e.g. Coles *et al*. 2004, Uz and Yoder 2004) and were attributed mostly to the complex nutrient conditions due to coastal river runoff, dust deposition and a variety of grazing pressures.

In developing the chl-*a* statistical prediction model, we also tried to group training data sets either yearly or seasonally in deriving the least square fitting functions (a_0, a_1, a_2). Results from their respective models show that the monthly fitting approach adopted in this study provides the best prediction of the temporal and spatial variations of chl-*a* in 2003. We note that significant correlations between SST, SSH and chl-*a* break down in some shallow shelf regions, where terrestrial input (e.g. fresh water, nutrient input) become increasingly important. The nutrient content is likely another important factor to be considered in a future refinement of the statistical model.

Long-term, cloud-free chl-*a* data generated either by DINEOF reconstruction of raw satellite observations or by the statistical prediction model when raw satellite data

were not available, in conjunction with long-term SSH and cloud-free SST time series, can be extremely valuable to better understand and quantify many important physical (such as defining the Loop Current and Loop Current Eddy boundary and to quantifying its shedding frequency) and biological (e.g. harmful algal bloom) processes in the GOM. These time series can also be used to detect long-term trends, thereby addressing climate change questions (Signorini and McClain 2011).

Acknowledgements

We are grateful to the funding support provided by NSF through grant OCE-1044573 and by DOE/RPSEA through GOMEX_PPP project. Result analyses presented here are part of Y. Zhao's master thesis research at North Carolina State University.

References

Adams, C.M., Hernandez, E. and Cato, J.C., 2004, The economic significance of the Gulf of Mexico related to population, income, employment, minerals, fisheries and shipping. *Ocean and Coastal Management*, **47**, pp. 565–580.

Alvera-Azcárate, A., Barth, A., Beckers, J.M. and Weisberg, R.H., 2007, Multivariate reconstruction of missing data in sea surface temperature, chlorophyll, and wind satellite fields. *Journal of Geophysical Research-Oceans*, **112**, C05099, doi: 10.1029/2007JC004243.

Alvera-Azcárate, A., Barth, A., Rixen, M. and Beckers, J.M., 2005, Reconstruction of incomplete oceanographic data sets using empirical orthogonal functions: application to the adriatic sea surface temperature. *Ocean Modelling*, **9**, pp. 325–346.

Beckers, J.M., Barth, A. and Alvera-Azcárate, A., 2006, DINEOF reconstruction of clouded images including error maps – application to the sea-surface temperature around Corsican island. *Ocean Science*, **2**, pp. 183–199.

Beckers, J.M. and Rixen, M., 2003, EOF calculations and data filling from incomplete oceanographic datasets. *Journal of Atmospheric and Oceanic Technology*, **20**, pp. 1839–1856.

Chassignet, E.P., Hurlburt, H.E., Smedstad, O., Barron, C.N., Ko, D.S., Rhodes, R.C., Shriver, J.F., Wallcraft, A.J. and Arnone, R.A., 2005, Assessment of data assimilative ocean models in the Gulf of Mexico using ocean color. In *Circulation in the Gulf of Mexico: Observations and Models*, W. Sturges and A. Lugo-Fernadez (Eds.), AGU Monograph Series, vol. 161, pp. 87–100 (Washington, DC: American Geophysical Union).

Coles, V., Wilson, C. and Hood, R., 2004, Remote sensing of new production fuelled by nitrogen fixation. *Geophysical Research Letters*, **31**, L06301, doi: 10.1029/2003GL019018.

He, R., Chen, K., Moore, T. and Li, M., 2010, Mesoscale variations of sea surface temperature and ocean color patterns at the Mid-Atlantic Bight shelfbreak. *Geophysical Research Letters*, **37**, L09607, doi: 10.1029/2010GL042658.

McClain, C.R., Signorini, S.R. and Christian, J.R., 2004, Subtropical gyre variability observed by ocean-color satellites. *Deep-Sea Research (Part II, Topical Studies in Oceanography)*, **51**, pp. 281–301.

McGillicuddy, D.J., Kosnyrev, V.K., Ryan, J.P. and Yoder, J.A., 2001, Covariation of mesoscale ocean color and sea-surface temperature patterns in the Sargasso Sea. *Deep-Sea Research (Part II, Topical Studies in Oceanography)*, **48**, pp. 1823–1836.

Miles, T. and He, R., 2010, Seasonal surface ocean temporal and spatial variability of the South Atlantic Bight: revisiting with MODIS SST and chl-*a* imagery. *Continental Shelf Research*, **30**, pp. 1951–1962.

Miles, T.N., He, R. and Li, M., 2009, Characterizing the South Atlantic Bight seasonal variability and cold-water event in 2003 using a daily cloud-free SST and chlorophyll analysis. *Geophysical Research Letters*, **36**, L02604, doi: 10.1029/2008GL036396.

North, E.W., Adams, E.E., Schlag, Z., Sherwood, C.R., He, R. and Socolofsky, S.A., 2011, Simulating oil droplet dispersal from the deepwater horizon spill with a Lagrangian approach. In *Monitoring and Modeling the Deepwater Horizon Oil Spill: A Record-Breaking Enterprise*, Y. Liu and A. McFadden (Eds.), AGU Monographs, pp. 217–226 (Washington, DC: American Geophysical Union).

Oey, L., Ezer, T. and Lee, H., 2005, Loop current, rings and related circulation in the Gulf of Mexico: a review of numerical models and future challenges. *Geophysical Monograph Series*, **161**, pp. 31–56.

Schmitz, W.J., Biggs, D.C., Lugo-Fernandez, A., Oey, L.-Y. and Sturges, W., 2005, A synopsis of the circulation in the Gulf of Mexico and on its continental margins. In *Circulation in the Gulf of Mexico: Observations and Models*, W. Sturges and A. Lugo-Fernadez (Eds.), AGU Monograph Series, vol. 161, pp. 11–30 (Washington, DC: American Geophysical Union).

Signorini, S.R. and McClain, C.R., 2011, Subtropical gyre variability as seen from satellite. *Remote Sensing Letters*, **3**, pp. 471–479.

Uz, B.M. and Yoder, J.A., 2004, High frequency and mesoscale variability in SeaWiFS chlorophyll imagery and its relation to other remotely sensed oceanographic variables. *Deep Sea Research (Part II, Topical Studies in Oceanography)*, **51**, pp. 1001–1017.

Walsh, J.J., Robert, H., Dieterle, D.A., He, R., Darrow, B.P., Jolliff, J.K., Lester, K.M., Vargo, G.A., Kirkpatrick, G.J., Fanning, K.A., Sutton, T.T., Jochens, A.E., Biggs, D.C., Nababan, B., Hu, C. and Muller-Karger, F.E., 2003, The phytoplankton response to intrusions of slope water on the West Florida shelf: models and observations. *Journal of Geophysical Research*, **108**, 3190, doi: 10.1029/2002JC001406.

Wilson, C. and Adamec, D., 2002, A global view of bio-physical coupling from SeaWiFS and TOPEX satellite data, 1997–2001. *Geophysical Research Letters*, **29**, 1257, doi: 10.1029/2001GL014063.