



Objective global ocean biogeographic provinces

Matthew J. Oliver^{1,2} and Andrew J. Irwin³

Received 4 April 2008; revised 5 June 2008; accepted 25 June 2008; published 2 August 2008.

[1] Biogeographic provinces are categories used for comparing and contrasting biogeochemical processes and biodiversity between ocean regions. Provinces provide a framework for reasonable extrapolation of point or transect data to broader areas. However, their use is limited due to the non-automatic, subjective nature of province classification. Furthermore, it is unknown how province boundaries respond to seasonal and climate forcing. These issues make province related hypotheses difficult to test with static provinces. To solve this problem, we use objective classification on global remote sensing data to automatically produce time and space resolved ocean provinces. Seasonal patterns in province geography reflect well-known ocean processes. Our predictions of province boundaries are verified by in-situ ship track data and province distributions in the equatorial Pacific correlate well with ENSO indexes. This objective classification system captures spatial and temporal province dynamics and provides objective categories for cross-province biogeochemical hypotheses to be rigorously tested. **Citation:** Oliver, M. J., and A. J. Irwin (2008), Objective global ocean biogeographic provinces, *Geophys. Res. Lett.*, 35, L15601, doi:10.1029/2008GL034238.

1. Introduction

[2] Physical and biological forces interact in the sea to create a complex emergent seascape commonly characterized by conservative (temperature and salinity) and non-conservative (ocean color and nutrient concentration) properties. Despite this complexity and rich structure, oceanographers have recognized the existence of distinct biogeographic provinces in the oceans [Barber, 1988; Hooker et al., 2000; Longhurst et al., 1995; Longhurst, 1998; McGowan, 1971; Ottens, 1991; Sverdrup et al., 1942]. The province concept has provided a framework to spatially aggregate or separate data for comparisons of biogeochemical processes over broad regions of the global ocean. Province designations have been used to analyze global distributions of primary productivity, DMSP fluxes, distributions of pelagic flora and fauna, and other biogeochemically important parameters [Angel et al., 2007; Boyd and Doney, 2003; Ducklow, 2003]. Waniek et al. [2005] found that province specific limitations of primary production in the euphotic zone were good predictors of particle

flux measured by sediment traps in three northeast Atlantic provinces. Boyd and Doney [2003] and Ducklow [2003] used the Longhurstian province concept to analyze JGOFS biogeochemical fluxes and to describe regime shifts in ecological community structure. However, both of these studies suggest that understanding the temporal dynamics of province boundaries would be beneficial in estimating trends in ocean biogeochemistry. Longhurst [1998] emphasized that province boundaries he produced are not fixed in space and time and merely represent the mean location of provinces. To actually locate the boundaries between provinces in space and time requires that provinces be defined by features that can be synoptically observed from earth-orbiting remote sensors and that objective automatic methods be developed to analyze this voluminous data. Complicated spatial patterns in province boundaries are likely to be present on small as well as large scales. Automatic methods should be able to identify features missed by global analyses such as Longhurst's.

[3] Many of the predictors used to define provinces are correlated to satellite observations. Mixed layer depth, Brunt-Väisälää frequency, Rossby radius of deformation, and nutrient fields are all significantly and strongly correlated to sea surface temperature (SST) on a global scale. Water column integrated chlorophyll concentrations, photic depth, and nutrient fields are also significantly and strongly correlated to ocean color. Therefore, the global time series of satellite ocean color and sea surface temperature provide a significant amount of discrimination power for determining the locations of biogeographic provinces [Esaias et al., 2000]. In this study we adapt and extend the bioinformatic approach of Oliver et al. [2004]. We use satellite measured sea surface temperature and ocean color radiance data to objectively identify biogeographic provinces without having *a priori* knowledge of the number of biogeographic provinces present. If our approach is able to effectively detect the spatial and temporal dynamics of Longhurstian biogeographical provinces, we would expect three clear patterns to emerge from our analysis. First, we expect province patterns to emerge in our analysis that are similar to province patterns previously described by Longhurst. Second, we expect that in-situ data would verify the location of predicted boundaries between provinces. Third, we expect predicted biogeographic provinces to recognizably respond spatially and temporally to well-known climate forces such as the El Niño southern oscillation (ENSO).

2. Materials and Methods

2.1. Satellite Data Sources

[4] The satellite data we used to classify global biogeographical provinces were normalized water leaving radiance at 443 nm (nLW₄₄₃), nLW₅₅₁, and 11 μ daytime SST derived from NASA's MODIS/Aqua instrument. Monthly and annual

¹Institute of Marine and Coastal Sciences, Rutgers University, New Brunswick, New Jersey, USA.

²Now at College of Marine and Earth Studies, University of Delaware, Lewes, Delaware, USA.

³Department of Mathematics and Computer Science, Mount Allison University, Sackville, New Brunswick, Canada.

means from 2003–2006 were used as provided by the ocean color processing team (G. C. Feldman and C. R. McClain, Ocean Color Web, MODIS/Aqua Reprocessing 1.1, 2007, available at <http://oceancolor.gsfc.nasa.gov/REPROCESSING/Aqua/1.1/>). We used level-3 binned data processed with SeaDAS (<http://oceancolor.gsfc.nasa.gov/seadas/>) to a cylindrical grid of 36 km resolution at the equator. There are a host of satellite derived products that could be used as predictors of biogeographical provinces. We focused on the predictors that describe the majority of the variation in the satellite signal. A principal component analysis of MODIS radiance channels and SST revealed that three components account for 95% of the variance in the MODIS data set, and that nLW_{443} , nLW_{551} , and 11μ daytime SST were the major contributors to those three components.

2.2. Classification Algorithm

[5] Fundamental to specifying discrete biogeographic provinces is the task of transforming continuous observations (SST and ocean color) into discrete classes (provinces). Many methods require either subjective expert decisions or extensive training data with *a priori* known classifications [Devred *et al.*, 2007; Longhurst, 1998]; however, whatever theoretical progress is gained with *a priori* expert skill is lost pragmatically because expert skill does not lend itself to automation. We use an objective criterion to determine the number of classes; when our statistic reaches a threshold, we stop creating more province classes.

[6] In this study, we divide a global ocean, four-year time series of MODIS/Aqua data with monthly resolution into provinces automatically using an objective statistic. To do this, we simplify and automate a classification technique used earlier in a regional study [Oliver *et al.*, 2004]. This classification technique is modeled after bioinformatic clustering algorithms that are used to decide how the continuum of genetic sequences ought to be broken up into distinct classes (species) without *a priori* knowledge of the number of classes [Yeung *et al.*, 2001]. The essence of the algorithm is to standardize the SST and ocean color data to zero mean and unit variance (Table S1¹), calculate standardized Euclidean distance matrices in predictor space (SST, nLW_{443} , and nLW_{551}), then use clustering algorithms to divide the data into classes. As classes are specified, a Figure of Merit (which is similar in form to an RMS error) quantifies how well each class centroid predicts all other members of the class. Classification is halted when adding additional classes to the data set does not improve the ability of the class centroids to predict all other members of their classes [Oliver *et al.*, 2004].

[7] The full method is described by Oliver *et al.* [2004] with the following modifications. We reduced the number of clustering algorithms to two (Ward's linkage agglomerative clustering [Ward, 1963] and K-means divisive clustering [Hartigan and Wong, 1979]). These algorithms represent two major approaches to cluster analysis, which can isolate distinct features in data and are complementary in this application. We used the stats package in the computer program R to implement these clustering algorithms [R Development Core Team, 2006]. Clustering

algorithms use distance matrices in their calculations which require storage proportional to the square of the number of data points to be analyzed. A single MODIS global image at 36 km resolution contains more than 10^5 points, making clustering of all data infeasible. Our solution is to cluster a small subset of the data (12,000 points), sampled uniformly in the temporal domain, but skewed more heavily toward the coast in the spatial domain (Figure S1) since we expect a large amount of the variance in the satellite data set to be concentrated in shallow water. Every data point in the full-resolution images was subsequently classified according to the cluster label of the nearest point (in predictor, not geographic, space) in the sample. A single classification of the sampled data into province types is used to ensure consistent labels over the entire time-series. The resulting classes are then mapped. The same class is often mapped in many non-contiguous locations. The two clustering algorithms produce separate maps of province distributions that are overlaid into a single province map.

2.3. Province Boundaries

[8] Province boundary locations show the spatial and temporal extent of a given province and geographically distinguish one province from the next. In addition to province boundary location, we estimated the relative strength of the boundaries between provinces. Gradients between adjacent provinces were computed along boundaries as the Euclidean distance between standardized SST, nLW_{443} and nLW_{551} averaged over the adjoining provinces. The greater the difference between the mean predictors in neighboring provinces, the stronger the boundary between them. The boundary strength was mapped as quartiles over a grayscale to increase contrast and aid in identification of the strongest boundaries (Figure 2).

2.4. Boundary Validation

[9] We verified province boundary locations predicted by this analysis by overlaying independent in-situ ship tracks of salinity and density on boundary maps computed from monthly means of MODIS/Aqua imagery. The three sources of ship-track data we used were from the Integrated Science Data Management branch of Canada's Federal Department of Fisheries and Oceans (www.meds-sdmm.dfo-mpo.gc.ca), the Coriolis project for operational oceanography (www.coriolis.eu.org/cdc/tsg_and_buoy_data.htm), and the National Oceanic and Atmospheric Administration, Atlantic Oceanographic and Meteorological Laboratory (www.aoml.noaa.gov/phod/tsg/index.php). Some mismatch was anticipated since the cluster boundaries were drawn from monthly average data and compared with instantaneous in situ observations.

2.5. ENSO Indexes

[10] To determine if provinces were responding to well-known climate forces, we correlated equatorial province areas with three different indexes of ENSO. We computed the area of the dominant provinces corresponding to the equatorial upwelling within the region bounded by 10°S , 10°N , 165°E , and 110°W . The three sources of ENSO indexes we used were the Southern Oscillation Index

¹Auxiliary materials are available in the HTML. doi:10.1029/2008GL034238.

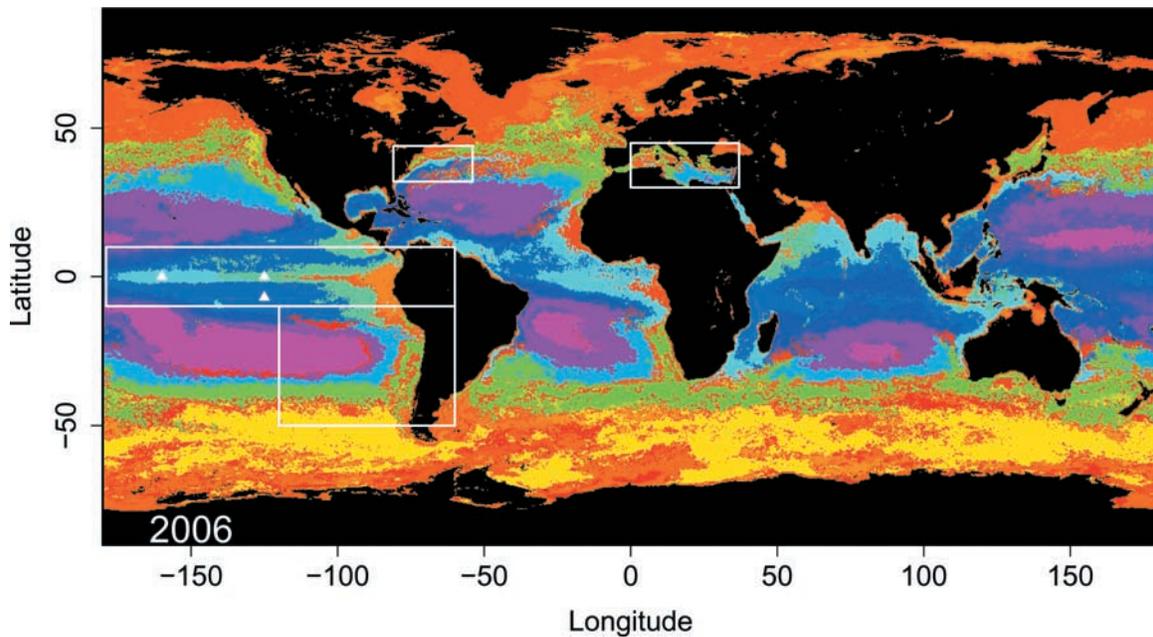


Figure 1. Objective global ocean biogeographic provinces determined from MODIS/Aqua 2006 annual average data. Colors identify distinct classifications (province types) and do not represent the value of any of the predictor variables. White boxes indicate regions used to verify province boundaries with ship-track data in Figure 2. Equatorial provinces identified with a white triangle are further analyzed in Figure 3.

(SOI) (www.cpc.ncep.noaa.gov/data/indices/reqsoi.for), the Multivariate ENSO Index (MEI) (www.cdc.noaa.gov/people/klaus.wolter/MEI/table.html) and the Oceanic Nino Index (ONI) (www.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ensoyears.shtml).

3. Results and Discussion

3.1. Global Biome Distribution

[11] The global scale biogeographic provinces predicted by our analysis of 2006 MODIS/Aqua average data (Figure 1) reflect many well-known large scale hydrographic features. These include the five major ocean gyres, the equatorial upwelling zone, western boundary currents such as the Gulf Stream and Kuroshio, and large river plumes from the Amazon and Congo Rivers. Our ocean province classes span ocean basins and hemispheres; which is different than the location specific province descriptions described previously. In the work presented here, same province types that are not geographically connected would be considered different provinces. It is also important to note that Figure 1 reflects only the average signal, and that many oceanographic features such as river plumes that exhibit strong seasonal and inter-annual variability are better described in the time resolved Animation S1. The general topology of Longhurst's classification is reproduced by our analysis, including many of the dominant and regional provinces, *e.g.*, the North Atlantic Gyral (NATR), Caribbean (CARB), Atlantic Arctic (ARCT), Canary Coastal (CNRY), South Atlantic Gyral (SATL), Southwest Atlantic Shelves (FKLD), Benguela Current Coastal (BENG), Red Sea, Persian Gulf (REDS), Northwestern Arabian Upwelling (ARAB), California Current (CALC), Pacific Equatorial Divergence (PEQD), and the South Pacific Subtropical Gyre (SPSG). What is

critical here is that these regions have been automatically and objectively detected without knowing the number of provinces present *a priori*. This represents a significant advance over province algorithms that require the number of provinces to be imposed on the data, rather than the data dictating the number of provinces [Devred *et al.*, 2007]. Many of our boundaries are irregular compared to Longhurst's boundaries. Our provinces commonly intrude into their neighbors with small isolated patches or cyclonic fingers. Some regions, such as the Southern Ocean and North Pacific show extensive heterogeneity in our analysis which may reflect a higher level of provincial discrimination or be a consequence of remote sensing sampling bias due to frequent and irregular cloud cover. Our automatic classification identifies 81 province types, of which 17 account for the vast majority of the global ocean. This classification is based on simple observational data (temperature and ocean radiance) as opposed to parameters such as chlorophyll that are derived from complex empirical algorithms.

3.2. In-situ Verification of Province Boundaries

[12] While our analysis is in good agreement with historically recognized ocean provinces, the key advantage of our approach is the elucidation of temporal trends in province boundaries (Animation S1). A province boundary implies a difference in hydrography between two provinces. In this study, we used ship-of-opportunity transects to verify province boundary locations. We overlaid salinity and density ship-of-opportunity transects on province boundaries derived from monthly averaged MODIS imagery. Boundaries were divided into quartiles by their strength with the darkest boundaries representing the strongest boundaries. Salinity and density are technically independent of the satellite data we used in the algorithm; however,

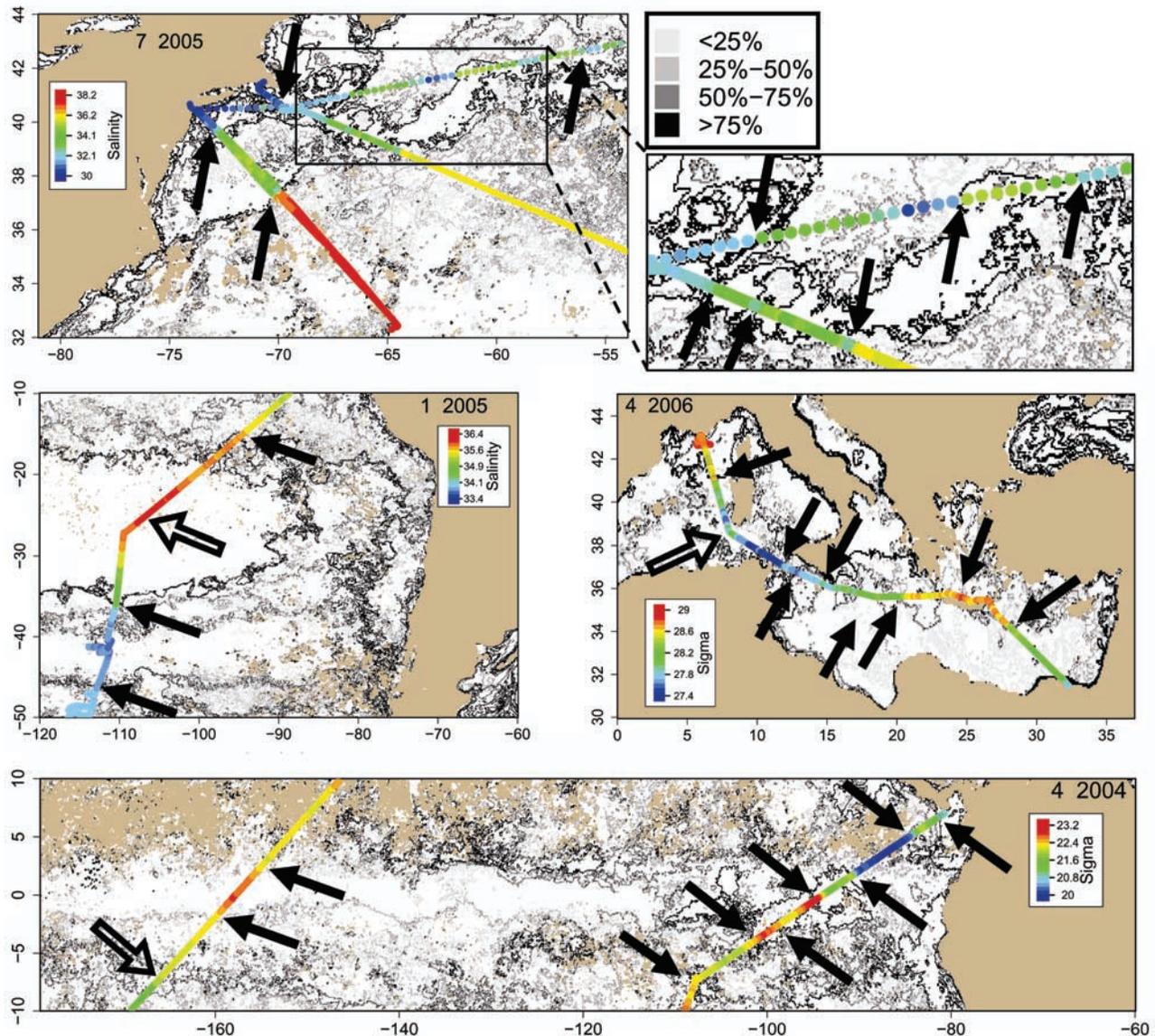


Figure 2. *In situ* salinity and density transects (colored lines) overlaid on maps of province boundaries in selected regions of the global ocean (Figure 1). Grey lines are province boundaries, with shading determined by the strength of the gradient of standardized SST and water leaving radiances and indicated by the grayscale bar legend. Darkest boundaries are the strongest boundaries. Colored lines are salinity or density indicated by color bar on each plot. Solid arrows highlight areas of boundaries consistent between *in situ* data and province analysis. Open arrows indicate a few locations where gradients in *in situ* data are not reflected in province boundaries.

satellite data contain partial salinity and density information because temperature is needed to compute both salinity and density from a CTD. Figure 2 shows examples of these hydrographic overlays from a western boundary region, a gyre, a marginal sea, and an equatorial region. Figures S2–S16 show the same regions as Figure 2, but at different times with different province boundary locations. In general, changes in hydrographic measurements are co-located with boundary locations on multiple spatial scales. Slight mismatches in boundary location are likely due to overlaying instantaneous ship track information on province boundaries derived from monthly averaged MODIS imagery. Our technique has a difficult time deriving boundaries where

hydrographic data show a very smooth transition over large space scales such as in the salinity data in the South Pacific Gyre (Figure 2). In this particular case, density measured in this same transect was nearly uniform across the gyre (Figure S16). However, since the purpose of our approach is to force continuous ocean data into classes, it is not surprising that in areas where hydrographic transitions are extremely smooth the approach does not do well.

3.3. Province Response to Climate Forcing

[13] The time series of monthly province geographic distributions (Jan 2003 to December 2006, Animation S1) documents the dramatic temporal variation in location and size of biogeographic provinces. Large-scale seasonal oscil-

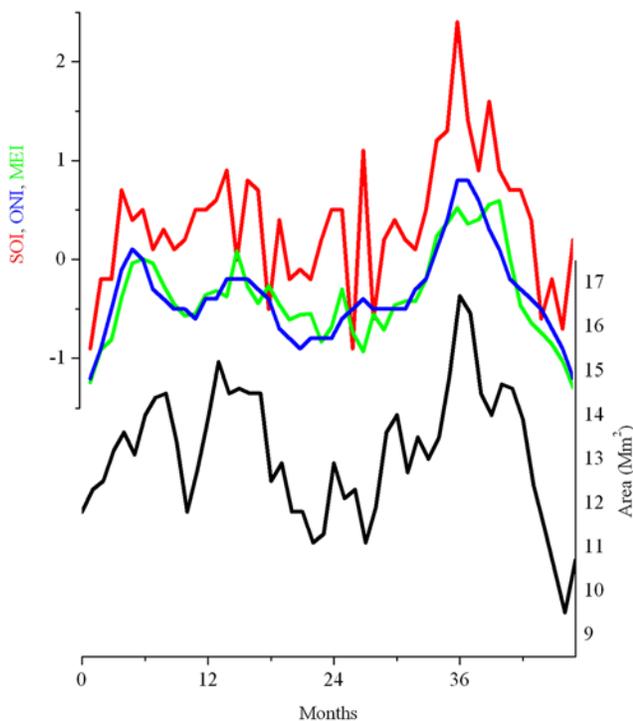


Figure 3. Temporal variability in the area of Pacific equatorial biogeographic provinces from 2003 – 2006 with monthly resolution and one month time lag. The area in 10^{12} m^2 of three provinces identified with a triangle on Figure 1 (black line) correlates well with El Niño indexes (colored lines). Missing data (cloud cover) amounts to $<3\%$ of the area in each month and is not correlated with province area or El Niño indexes.

lations in province location, movements of eddies, and current meanders are identified easily in this dynamic presentation. One of the most striking visual patterns is the inter-annual oscillation pattern of provinces in the equatorial Pacific. The areas of the three provinces coincident with the El Niño phenomenon (identified with triangles in Figure 1) oscillate together with three indexes of ENSO (Figure 3). A one-month time lag of province area shows the strongest correlation with these three indexes ($r^2 = 0.45, 0.58, 0.58$ for SOI, MEI, and ONI, respectively). The strong correlations indicate that provinces are responding spatially and temporally to multi-annual climate forces. The one month time lag in maximum correlation also suggests biogeographic provinces respond relatively quickly to ENSO forces. This confirms that methods we are using in this analysis are not masking well-known climate forces.

4. Conclusion

[14] We described a simple objective automatic method based on remote sensed data for discriminating global ocean biogeographic provinces that are supported by three major forms of evidence. First, our annual average province distributions generally conform to province distributions previously described. This shows that much of the information needed to discriminate ocean provinces is contained in the satellite data streams. Second, province boundary

locations are supported by independent in-situ observations of hydrography from a wide range of oceanic environments. This verification supports our method and predicted province locations, and demonstrates that we can dynamically and quantitatively identify provinces. Third, temporal changes in province areas appear to be tightly coupled to major climate forces in the global ocean as the area of key equatorial Pacific provinces track ENSO indexes with a lag of one month. This shows provinces to be dynamic reporters of climate conditions and allows them to be used as diagnostic factors for climate change.

[15] One of the main advantages for specifying ocean provinces is that it enhances comparative analysis of ocean processes and attributes such as carbon flux and chlorophyll concentrations. Hypotheses about the differential effects of climate change on biogeochemical fluxes or biodiversity could be simply tested across different province types with comparative methods. However, the validity of comparative statistical tests is compromised if the class identities are uncertain. If provinces in a region of interest are highly spatially and temporally dynamic, the results of cross-province comparative methods could be due to simple mis-classification of province type, thus not reflecting any real difference between provinces. Misclassification of province type would lead to both type 1 and type 2 errors. To guard against these errors, cross province comparisons have been generally restricted to global scale climatological studies near province centers to ensure that uncertainties about province edges did not affect the results of the comparison [Ducklow, 2003]. Our approach solves this issue by producing objective and time-resolved provinces, allowing for oceanographic experiments of both global and local scales and experiments in regions where provinces are highly dynamic to use simple cross-province comparative methods. These results are a step toward understanding how ocean provinces and the species that are contained within them are altered in space and time. Having an objective and automatic approach to province detection provides a standard metric for planning and analyzing oceanographic studies in an ever-changing seascape.

[16] **Acknowledgments.** This work was supported by NASA ROSES NNG06GH75G, program element A.4, Terrestrial Ecology and Biodiversity, and made possible by the ongoing data collection and analysis of the NASA and the ocean color research team (<http://oceancolor.gsfc.nasa.gov/>). AJI was also supported by NSERC Canada. Helpful discussions from Oscar Schofield, Paul Falkowski, Zoe Finkel, Josh Kohut, and Mark Moline are gratefully acknowledged. Both authors contributed equally to this paper.

References

- Angel, M. V., et al. (2007), Changes in the composition of planktonic ostracod populations across a range of latitudes in the north-east Atlantic, *Prog. Oceanogr.*, 73, 60–78.
- Barber, R. T. (1988), Ocean basin ecosystems, in *Concepts of Ecosystem Ecology: A Comparative View*, *Ecol. Stud.*, vol. 67, edited by L. R. Pomeroy and J. J. Alberts pp. 171–193, Springer, Berlin.
- Boyd, P. W., and S. C. Doney (2003), Impact of climate change and feedback processes on the ocean carbon cycle, in *Ocean Biogeochemistry*, edited by M. Fasham, pp. 157–193, Springer, New York.
- Devred, E., S. Sathyendranath, and T. Platt (2007), Delineation of ecological provinces using ocean colour radiometry, *Mar. Ecol. Prog. Ser.*, 346, 1–13.
- Ducklow, H. (2003), Biogeochemical provinces: Toward a JGOFS synthesis, in *Ocean Biogeochemistry*, edited by M. Fasham pp. 3–17, Springer, New York.
- Esaias, W., et al. (2000), Ocean province classification using ocean colour data: Observing biological signatures of variations in physical dynamics, *Global Change Biol.*, 6, 39–55.

- Hartigan, J. A., and M. A. Wong (1979), Algorithm AS 136: A *K*-means clustering algorithm, *Appl. Stat.*, *28*, 100–108.
- Hooker, S. B., et al. (2000), An objective methodology for identifying oceanic provinces, *Prog. Oceanogr.*, *45*, 313–338.
- Longhurst, A. R. (1998), *Ecological Geography of the Sea*, 398 pp. Academic, San Diego, Calif.
- Longhurst, A., et al. (1995), An estimate of global primary production in the ocean from satellite radiometer data, *J. Plankton Res.*, *17*, 1245–1271.
- McGowan, J. A. (1971), Oceanic biogeography of the Pacific, in *The Micropalaeontology of Oceans*, edited by S. M. Funnell and W. R. Riedel, pp. 3–74, Cambridge Univ. Press, Cambridge, U. K.
- Oliver, M. J., S. Glenn, J. T. Kohut, A. J. Irwin, O. M. Schofield, M. A. Moline, and W. P. Bissett (2004), Bioinformatic approaches for objective detection of water masses on continental shelves, *J. Geophys. Res.*, *109*, C07S04, doi:10.1029/2003JC002072.
- Ottens, J. J. (1991), Planktonic foraminifera as North Atlantic water mass indicators, *Oceanol. Acta*, *14*, 123–140.
- R Development Core Team (2006), *R: A Language and Environment for Statistical Computing*, R Found. for Stat. Comput., Vienna.
- Sverdrup, H. U., M. W. Johnson, and R. H. Fleming (1942), *The Oceans, Their Physics, Chemistry, and General Biology*, Prentice-Hall, Englewood Cliffs, N. J.
- Waniek, J. J., et al. (2005), Long time series of deep water particle flux in three biogeochemical provinces of the northeast Atlantic, *J. Mar. Syst.*, *56*, 391–415.
- Ward, J. H. (1963), Hierarchical grouping to optimize an objective function, *J. Am. Stat. Assoc.*, *41*, 236–244.
- Yeung, K. Y., et al. (2001), Validating clustering for gene expression data, *Bioinformatics*, *17*, 309–318.
-
- A. J. Irwin, Department of Mathematics and Computer Science, Mount Allison University, Sackville, NB E4L 1E4, Canada.
- M. J. Oliver, College of Marine and Earth Studies, University of Delaware, 700 Pilottown Road, Lewes, DE 19958, USA. (moliver@udel.edu)