

The emergence of ocean biogeochemical provinces: A quantitative assessment and a diagnostic for model evaluation

Marcello Vichi,^{1,2} J. Icarus Allen,³ Simona Masina,^{1,2}
and Nicholas J. Hardman-Mountford³

Received 16 May 2010; revised 16 October 2010; accepted 15 February 2011; published 17 May 2011.

[1] The concept of ocean biogeochemical provinces is based on the observation that large ocean regions are characterized by coherent physical forcing and environmental conditions, which are eventually representative of macroscale ocean ecosystems. Biogeochemical models of the global ocean focus on simulating the coupling between prevalent physical conditions and the biogeochemical processes with the assumption that biological properties respond coherently to physics and therefore should produce such provinces as an emergent property. In this paper, we quantitatively assess the emergence of a reference set of predefined biogeochemical provinces in the available global data sets and propose a province-based approach to the evaluation of one of the most comprehensive models of ocean biogeochemistry. Multivariate statistical tools were applied to model and observation data, verifying the existence, distinctiveness and reliability of the predefined provinces and quantifying the correlation of model results with observations at the global scale. The analysis of similarity between provinces shows that they are statistically separable in data and model output and therefore can be used as reliable metrics. The analyses indicate that provinces can be more easily distinguished in terms of their environmental features rather than using chlorophyll concentration. The characterization of provinces by means of chlorophyll values shows a significant overlap in both the Sea-viewing Wide Field-of-view Sensor (SeaWiFS) data and the model. It is likely this is related to the choice of province boundaries based on coarse-resolution mapped data, which are not necessarily the same as those derivable from high-resolution satellite data. We also demonstrated through cluster analysis that the long-term time series data collected at Joint Global Ocean Flux Study (JGOFS) stations are representative of environmental conditions of the respective province and can thus be used to evaluate model results extracted from that province. The method shows promise for helping to overcome problems with model verification due to under sampling of most ocean biogeochemical variables but also gives indications that unsupervised clustering may be required when more spatially resolved data and models are available.

Citation: Vichi, M., J. I. Allen, S. Masina, and N. J. Hardman-Mountford (2011), The emergence of ocean biogeochemical provinces: A quantitative assessment and a diagnostic for model evaluation, *Global Biogeochem. Cycles*, 25, GB2005, doi:10.1029/2010GB003867.

1. Introduction

[2] Over the last 15 years, Ocean Biogeochemistry General Circulation Models (OBGCM) have demonstrated a capability to reproduce the major spatial features of global ocean biogeochemistry such as oligotrophic gyres and upwelling regions [e.g., *Six and Maier-Reimer*, 1996;

Aumont et al., 2003; *Gregg et al.*, 2003; *Moore et al.*, 2004; *Le Quéré et al.*, 2005; *Vichi et al.*, 2007b]. Notwithstanding the different degree of complexity of the various models, ranging from those with simple implicit representations of biology to the explicit realization of several dynamic plankton functional types (PFT), most of these models show phytoplankton biomass distributions which roughly correspond to the chlorophyll distribution observed from space. This kind of comparison (defined hereafter as face validity) is done at the level of a bulk property of the marine ecosystem and tells us very little about the intrinsic capability of the model to simulate the biogeochemical features of each single region. Further visual comparisons involving satellite-derived phytoplankton type distributions at the global

¹Centro Euro-Mediterraneo per i Cambiamenti Climatici, Bologna, Italy.

²Istituto Nazionale di Geofisica e Vulcanologia, Bologna, Italy.

³Centre for Coastal Marine Studies, Plymouth Marine Laboratory, Plymouth, UK.

scale have been attempted (e.g., *Vichi et al.* [2007b] used data from *Alvain et al.* [2005]), though a direct comparison with relevant ecosystem properties such as PFT fractionated production, rates of nutrient regeneration, etc. is still lacking [*Anderson*, 2005].

[3] Historically there has been a lack of objective, quantitative comparison between biogeochemical model results and observations. This can partly be attributed to a lack of (or ease of access to) global data and partly to a cultural acceptance that a subjective visual comparison, especially with global ocean models, is acceptable. A recent set of works on model validation and assessment (see *Lynch et al.* [2009] and papers in the “Skill Assessment for Coupled Biological/Physical Models of Marine Systems” special issue) has pointed out the necessity to move beyond face validity, and analyze whether and where these models have skill, from which we may infer reliability when these models are used to make projections of the future. *Vichi and Masina* [2009] presented an assessment of the PELAGOS model (Pelagic biogeochemistry for Global Ocean Simulations [*Vichi et al.*, 2007a]) using some of the proposed objective measures [*Friedrichs et al.*, 2009; *Stow et al.*, 2009; *Doney et al.*, 2009]. The assessment used existing data in the public domain, focusing on large-scale data sets such as satellite-derived chlorophyll concentration, and long-term time series of variables that allow a description of the transformation of organic matter (mainly primary and bacterial production).

[4] This is, however, a limited assessment for models like PELAGOS, which incorporate many functional parameterizations derived from specific laboratory experiments and require comparisons with in situ observations to be considered generally valid. The question is, therefore, how do we overcome the undersampling of ocean biogeochemical properties and use the available data as efficiently as possible for the assessment of biological parameterizations? Measuring biological data at the global scale is a challenge and it is not expected in the near future to have more than the following data types available: (1) long-term station time series at single point locations in the ocean, (2) single realizations (casts) at different times and locations (usually organized in transects), and (3) long-term time series of global coverage, satellite-derived products.

[5] Such data can be used as is or in combination, by means of gridding and merging techniques, to provide global, objectively analyzed maps. This latter method allows direct comparison with model results through visual comparison of maps (or difference maps) and via computed means, taken as representative of an entire region of interest. The questions arise, therefore: (1) How much information can be extracted from this evaluation approach? (2) How can we assess model validity at the level of ecosystem functioning when only limited data is available from most ocean regions?

[6] This issue is related to the determination of correlation length scales for large-scale biological data, which cannot currently be resolved by means of extensive surveys of the global ocean, so needs to be approached by different methodologies that combine sparse observations with more qualitative information. There have been several attempts at providing a conceptual spatial classification of the marine environment, based mainly on either the distribution of distinct pelagic or benthic taxonomic groups (e.g., the

Ocean Biogeographic Information System [*Costello and Vanden Berghe*, 2006; *Arvanitidis et al.*, 2009]) or the spatial variability of physical properties, such as temperature, salinity, mixing state and empirically derived chlorophyll concentration and primary production estimates [*Longhurst*, 1995; *Longhurst et al.*, 1995; *Sathyendranath et al.*, 1995]. *Longhurst's* [2007] partition of the oceans into four major biomes (polar, westerlies, trade winds, tropical), realized as ~50 ocean provinces (regional expressions of the different biomes), remains the most comprehensive and widely accepted classification of the pelagic ocean.

[7] The concept of biogeochemical provinces is based on the observation that large ocean regions are characterized by coherent physical forcing and biological conditions at the seasonal scale, which are representative of macroscale ocean ecosystems [*Longhurst*, 1995, 2007; *Hardman-Mountford et al.*, 2008]. The boundaries between provinces are generally persistent but are also spatially and temporally variable, because they are linked to physical properties (e.g., fronts) which are known to change position seasonally and inter-annually. The boundaries of *Longhurst's* provinces were selected subjectively and intuitively on the basis of climatological data (monthly data of mixed layer depth, solar irradiance penetration and chlorophyll concentrations from the Coastal Zone Color Scanner) and common knowledge on the biological properties extracted from scattered data in the existing literature. To overcome this subjective limitation, *Longhurst* [2007] suggested to further refine province definitions by (1) comparing the distribution of individual biota between provinces, (2) testing statistically different conditions in adjacent provinces, and (3) using analytical techniques to partition a relevant global data set.

[8] Examples of the third suggestion are now available in the recent literature. *Devred et al.* [2007] used ocean color radiometry data to partition the Northwest Atlantic according to statistically coherent provinces and used these dynamical boundaries as new regions over which to extrapolate the data collected at point stations. *Hardman-Mountford et al.* [2008] have applied a range of multivariate statistics to a satellite-derived chlorophyll concentration climatology to provide an objective classification of the ocean into several macroscale biomes associated with trophic status (eutrophic, mesotrophic, oligotrophic). By means of multivariate analysis they demonstrated that the majority of the spatiotemporal variance in satellite-derived chlorophyll data was explained by an overwhelmingly dominant spatial structure with almost no seasonal variability. The existence of persistent spatial structures was used as the basis for a supervised hierarchical classification of these biomes, in order to provide provinces which have properties consistent with ecological systems. The results of this approach are, to a large extent, visually compatible with *Longhurst's* [2007] subjective partitioning. *Devred et al.* [2007] report similar results for bulk observations and physical variables. However, in the case of primary production data, they found a significant difference between the mean values computed over their ecological provinces and the static ones defined by *Longhurst*.

[9] Similarly, *Gregg and Bodtker* [2007] have applied adaptive classification algorithms to selected physical variables from a general circulation model output in order to partition the northern Pacific into significant distinct regions. They identified contiguous oceanic regions which could be

related to known water masses, also finding significant correspondence with the spatial distribution of satellite-observed chlorophyll concentration.

[10] The approach proposed in this work stems from the widely held premise that most of the subjectively defined biogeochemical provinces have a physical coherence and correspond to real ecological structure. *Hardman-Mountford et al.* [2008] have tested this hypothesis to some degree but only using satellite-derived chlorophyll concentration. Our aim in this study is to assess whether biogeochemical provinces emerge as system properties from the available global data sets with the prospect that in future they may be used as operational units for model evaluation. The method involves the application of multivariate statistical analyses [e.g., *Allen et al.*, 2007] to global ocean data and it is based on a supervised classification of ocean regions taking *Longhurst's* [2007] province layout as an example. The following specific questions are investigated:

[11] 1. Can we statistically separate Longhurst provinces in the model and in gridded objectively analyzed observations?

[12] 2. Are the relationships between provinces (in terms of similarity/dissimilarity) the same in the model and in the data?

[13] 3. Are station field data (selected as typical of a specific province) representative of the entire province at the seasonal scale?

[14] This will be done by objectively verifying whether the physical and biological properties extracted from each predefined province represent an identifiable entity both in the model and in the observations. If these regions are physically coherent and statistically separable in both data and model output, it becomes possible to undertake a province-based assessment of model behavior and to compare model results with observations at the province level. This implies a reduction of the amount of data to be explored for information extraction and an increase in the utility value of biologically relevant information that can only be collected at point sources.

2. Methods

2.1. Biogeochemical Model

[15] The global ocean biogeochemical model used in this study is PELAGOS (Pelagic biogeochemistry for Global Ocean Simulations [*Vichi et al.*, 2007a, 2007b]), which is a coupling between the OPA (Océan Parallélisé) general circulation model [*Madec et al.*, 1999] and the global ocean version of the Biogeochemical Flux Model (BFM, <http://bfm.cmcc.it>). The model grid is the irregular ORCA2 configuration [*Madec and Imbard*, 1996] with a grid mesh varying from 0.5 to 2 degrees. The biogeochemical model is a multiple nutrient; multiple plankton functional group model described in terms of biomass of carbon and the major macro and micronutrients. The model is fully detailed by *Vichi et al.* [2007a]. The climatological features of the model have been analyzed by *Vichi et al.* [2007b], and the interannual simulation (1958–2001) used in this work has been analyzed by *Vichi and Masina* [2009]. In this paper we focus on the period of the simulation when satellite ocean color data from the Sea-viewing Wide Field-of-view Sensor (SeaWiFS) are also available (1998–2001).

2.2. Observational Data

[16] Observational data used in this work consist of a selection of gridded products and station data publicly available to the scientific community.

[17] Temperature and salinity climatologies for the world ocean have been taken from the Levitus data set [*Levitus et al.*, 1998]. Global maps of nutrient data were obtained from the World Ocean Atlas 2001 data set [*Conkright et al.*, 2002]. The data consist of optimally interpolated monthly climatological fields of nitrate, phosphate and silicate at selected depths (see *Conkright et al.* [2002] for details on the methodology). The 1×1 degree maps of both physical and chemical properties have been further interpolated on the ORCA2 model grid with a nearest-neighbor procedure.

[18] Mixed layer depth data have been obtained from the global ocean analysis performed by *de Boyer Montégut et al.* [2004] according to a 0.2°C temperature criterion. The interpolated data on the ORCA2 model grid have been directly obtained from de Boyer Montégut's Web site (<http://www.locean-ipsl.upmc.fr/~cdblod/mld.html>).

[19] Monthly composites of global satellite-derived chlorophyll concentration (level 3 products), obtained from visible spectral radiometer (ocean color) data from SeaWiFS on board the OrbView-2 satellite, were obtained from NASA. The composites were spatially interpolated and averaged over the time period January 1998 to December 2001 to produce climatological means for each month that match the model grid and the temporal resolution.

[20] As an example of station data we used the Joint Global Ocean Flux Study (JGOFS) data set, which was already employed by *Ducklow* [2003] to overview the questions of biogeochemical provinces and by *Vichi et al.* [2007b] and *Vichi and Masina* [2009] as model validation sites. The JGOFS data set comprises three long-term (at least 10 years) station time series (Bermuda Atlantic Time series Study, BATS; Hawaii Ocean Time series, HOT; Station PAPA, STNP), one medium-term station on the Kerguelen Plateau (KERFIX) and several process studies from the North Atlantic Bloom Experiment (NABE), equatorial Pacific (EQPAC), Arabian Sea (ARAB), the Antarctic Polar Front Zone (APFZ) and the Ross Sea (ROSS).

2.3. Multivariate Statistical Analyses

[21] Differences between samples extracted from province pairs were tested for statistical significance using analysis of similarities (ANOSIM [*Clarke and Green*, 1988]). The ANOSIM statistic (R) compares the similarity of ranked variables within a province with the average rank of different provinces. Due to the kind of the data sets used, which may contain both physical and biological information, the similarity/dissimilarity between multivariate combinations of data (known as the resemblance matrix) is expressed as normalized Euclidean distances.

[22] The ANOSIM test is scaled to vary between -1 and $+1$. A value of $+1$ indicates that the similarity between all samples within one province is higher than similarities between provinces. The suggested arbitrary thresholds are: >0.75 total separation; $0.75-0.5$ weak overlap; $0.5-0.25$ overlap but some separation; <0.25 no separation. A level of significance is also computed by performing 1000 random permutation

Table 1. Description of the 38 Longhurst Provinces Defined in the Model Domain^a

Number	Name	Description	Biome
1	BPLR	Boreal Polar Province	Polar
2	ARCT	Atlantic Arctic Province	Polar
3	SARC	Atlantic Subarctic Province	Polar
4	NADR	North Atlantic Drift Province	Westerlies
5	GFST	Gulf Stream Province	Westerlies
6	NASW	N Atlantic Subtropical Gyral Province (W)	Westerlies
7	NATR	N Atlantic Tropical Gyral Province	Trades
8	WTRA	Western Tropical Atlantic Province	Trades
9	ETRA	Eastern Tropical Atlantic Province	Trades
10	SATL	South Atlantic Gyral Province	Trades
11	NECS	NE Atlantic Shelves Province	Coastal
12	CNRY	Canary Coastal Province	Coastal
15	NWCS	NW Atlantic Shelves Province	Coastal
16	MEDI	Mediterranean Sea, Black Sea Province	Westerlies
17	CARB	Caribbean Province	Trades
18	NASE	North Atlantic Subtropical Gyral Province (E)	Westerlies
21	FKLD	SW Atlantic Shelves Province	Coastal
30	MONS	Indian Monsoon Gyres Province	Trades
31	ISSG	Indian South Subtropical Gyre Province	Trades
32	EAFR	East Africa Coastal Province	Coastal
33	REDS	Red Sea, Persian Gulf Province	Coastal
34	ARAB	NW Arabian Upwelling Province	Coastal
50	BERS	North Pacific Epicontinental Province	Polar
51	PSAE	Pacific Subarctic Gyres Province (East)	Westerlies
53	KURO	Kuroshio Current Province	Westerlies
54	NPPF	North Pacific Polar Front Province	Westerlies
56	NPSW	North Pacific Subtropical Gyre Province (West)	Westerlies
59	SPSG	South Pacific Subtropical Gyre Province	Westerlies
60	NPTG	North Pacific Tropical Gyre Province	Trades
61	PNEC	North Pacific Equatorial Countercurrent Province	Trades
62	PEQD	Pacific Equatorial Divergence Province	Trades
63	WARM	West Pacific Warm Pool Province	Trades
64	ARCH	Archipelagic Deep Basins Province	Trades
70	SUND	Sunda	Coastal
80	SSTC	South Subtropical Convergence Province	Westerlies
81	SANT	Subantarctic Province	Westerlies
82	ANTA	Antarctic Province	Polar
83	APLR	Austral Polar Province	Polar

^aThe numbers (the same as by Longhurst [2007]) are used to identify provinces and biomes in Figures 1–9.

and estimating the p value (P) that the R value is obtained by chance.

[23] Nonmetric multidimensional scaling ordination (MDS) was used to visualize in a two-dimensional space the proximities between provinces according to the extracted multivariate set of variables. MDS is a powerful statistical tool that creates a spatial ordination of the data which carefully maintains the similarities or dissimilarities between the objects [Borg and Groenen, 2005]. In our case, the objects are the provinces, characterized by a multivariate combination of station data or means and variances of the major biogeochemical properties (see section 2.5). The realized 2-D representation is obtained through the minimization of a stress index [Borg and Groenen, 2005], which compares the relative distances of objects in the original and downscaled data set.

[24] Groups of similar provinces were also identified by hierarchical group-average clustering. This method starts from a similarity matrix and creates a dendrogram that shows the levels (on the y axis) at which samples or group of samples are considered separated. Groups within the dendrogram were separated with two different methods: (1) by taking slices at arbitrary distances in the dendrogram and (2) by applying the SIMPROF analysis [Clarke and Gorley, 2006]. The latter tests the significance of each split in the dendrogram by means of a statistic applied to a sample of the possible permutations of the variable values involved in that split. The null hypothesis is that there is no structure in the cluster configuration. The results of both groupings can be visualized in the MDS space to show the difference in the resulting classifications.

[25] The Spearman rank correlation coefficient (ρ) has been used to quantify the difference between model and data by using the respective resemblance matrices (a Mantel test, implemented in the PRIMER software as the RELATE test statistic [Sommerfield et al., 2002; Clarke and Gorley, 2006]). The investigated relationship is therefore considered in terms of similarity, e.g., whether province 1 and 5 are closest both in the model and in the data. The value of ρ ranges between -1 and $+1$, where 0 means there is no match between the similarity matrices. The test is built by means of random permutation of the (symmetric) data resemblance matrix and the probability of obtaining a correlation larger than the actual ρ is computed. One thousand permutations are used as default.

[26] Numerical computations and visualizations were performed with PRIMER v. 6 [Clarke and Warwick, 2001; Clarke and Gorley, 2006] and MATLAB. The ANOSIM MATLAB implementation is provided by D. L. Jones (Users manual for FATHOM: A MATLAB toolbox for multivariate ecological and oceanographic data analysis, 2002, available at <http://www.rsmas.miami.edu/personal/djones/>).

2.4. Longhurst Provinces in the Model Domain

[27] The name and definition of provinces listed in Table 1 are derived from Longhurst [2007]. Longhurst's provinces have been mapped on the model grid as shown in Figure 1 by assigning a unique province number to each ocean grid point. The resolution of the model has limited skill in capturing the physical dynamics of coastal biomes [see Vichi et al., 2007b; Vichi and Masina, 2009]; therefore, most of the coastal provinces have been discarded, only considering all of the open ocean provinces, for a total of 38 regions out of the ~ 50 originally proposed by Longhurst. Some coastal provinces have been retained (e.g., the Mauritanian-Moroccan upwelling, CNRY, and the NE Atlantic Shelves) as examples to test the performance of the model in these areas. The position of major physical boundaries which are known to be related to horizontal resolution issues, such as the western boundary currents (Gulf Stream and Kuroshio) and the extent of the Pacific Equatorial Divergence (PEQD) into the Western Pacific Warm Pool (WARM), was determined subjectively from the mean annual position of the physical features in the model results.

2.5. Model and Data Sampling Strategy

[28] The analysis considers two groups of variables: environmental and biological. Environmental variables (ENV

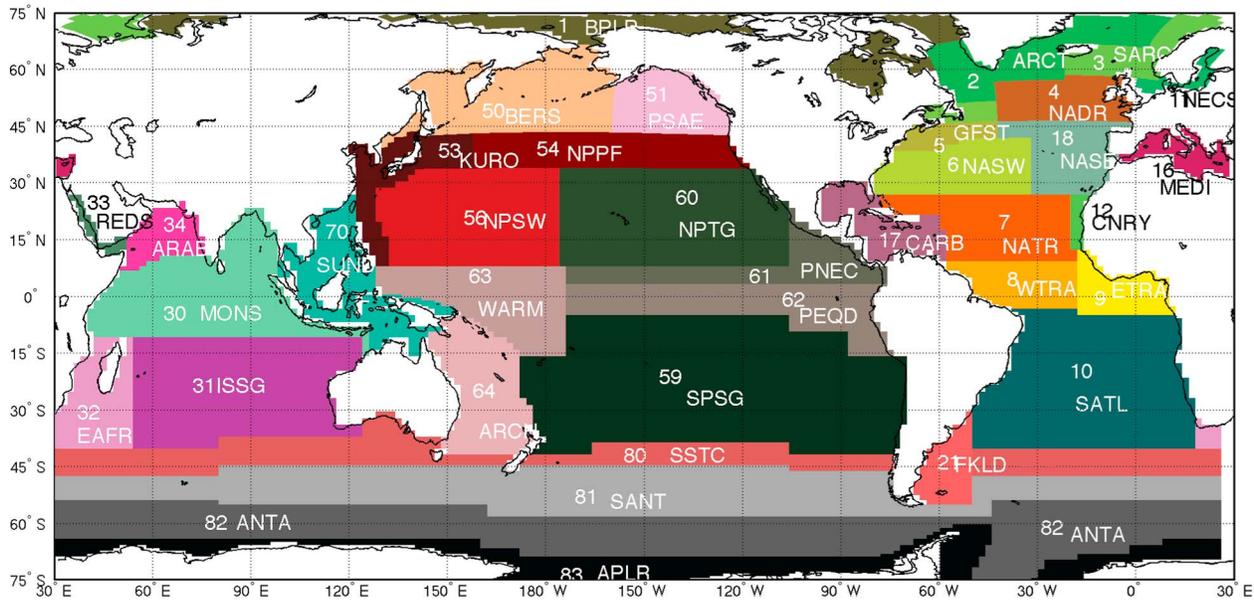


Figure 1. Location, names, and numbering of selected Longhurst's [2007] provinces on the model grid.

data set) are considered to be the physical and hydrochemical indicators of the state of the water masses. The physical variables available both in the observations and in the model are: sea surface temperature, salinity and mixed layer depth. Photosynthetically Available Radiation (PAR) data were also used by Longhurst in his classification, but it was not used here as the physical model is forced by this surface information and it was preferred to use derived variables only. Hydrochemical properties are also considered environmental variables because of their potential bottom-up control on the living functional groups. However, it must be considered that the actual concentration is the final result of both hydrodynamic processes and uptake/remineralization by biota. The hydrochemical properties available in the model and in the observations are the surface concentrations of phosphate, nitrate and silicate. The term surface indicates here the first level of the model (5 m) and

the first level of observational data sets (0 m in the World Ocean Atlas).

[29] The only observable biological variable that is available at the global scale is total chlorophyll concentration derived from satellite data (section 2.2). In the model, satellite-like total chlorophyll is derived as described by Vichi et al. [2007b] by considering the vertically integrated chlorophyll concentration and the related attenuation coefficient (CHL data set). Further biological variables are available from the model output (BIO data set): particularly we considered surface concentrations of chlorophyll and carbon content of each phytoplankton group and carbon biomass of heterotrophs. Phytoplankton chlorophyll and carbon concentrations are considered separately since PELAGOS computes variable ratios of these two variables [Vichi et al., 2007a].

[30] Model and observation data sets are distinguished by adding the suffix “-mod” or “-obs,” respectively, to the data

Table 2. Summary of Variables for Each Data Set Used in the Analyses

Name	Variable	Depth
ENV-obs and ENV-mod	Temperature	surface
	Salinity	surface
	Mixed layer depth	-
	Phosphate concentration	surface
	Nitrate concentration	surface
BIO-mod	Silicate concentration	surface
	Diatoms carbon and chlorophyll concentration	surface
	Flagellates carbon and chlorophyll concentration	surface
	Picophytoplankton carbon and chlorophyll concentration	surface
	Bacteria carbon concentration	surface
	Heterotrophic nanoflagellates carbon concentration	surface
	Microzooplankton carbon concentration	surface
	Mesozooplankton carbon concentration	surface
CHL-mod	Sum of phytoplankton chlorophyll concentration	integral over optical depth
CHL-obs	Total satellite-derived chlorophyll	-

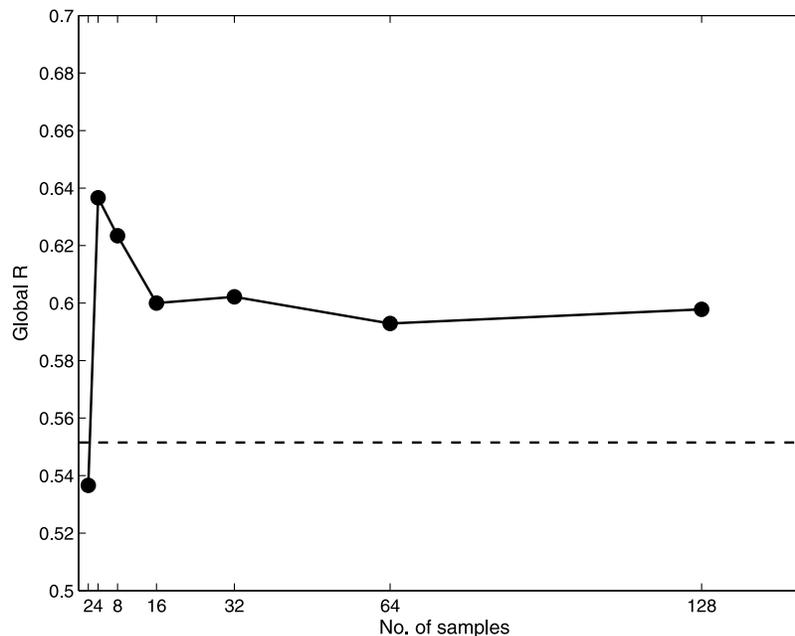


Figure 2. Dependence of the global R statistics from the number of samples randomly extracted from each region. The dashed line indicates the global R value of the area-weighted sampling.

set names. A summary of the data sets used in the analyses is given in Table 2. The target temporal window is the seasonal scale and monthly means of the relevant variables over the 4 year period 1998–2001. All global data sets are interpolated onto the model grid with a nearest neighbor interpolation.

[31] Different strategies have been used to extract model properties from the given set of provinces. (1) Fixed sampling refers to the random extraction of a constant number of stations from each province; this is done by randomly selecting a given number of grid points from the ones that have been assigned to a certain province (section 2.4). (2) Bulk sampling refers to the computation of mean values and standard deviations as measures of the variability within each province; descriptive statistics consider the spatial standard deviation of the grid point means across time for each province and the temporal standard deviation over the year of the province spatial mean. (3) Area-weighted sampling, in which the number of extracted samples is proportional to the surface area of the province, with a given minimum of samples from the smallest province. In this case, the largest province named Southern Pacific Subtropical Gyre (SPSG, province 59, see Figure 1 and Table 1), has 136 randomly distributed samples and the smallest (like the coastal Canary province 12, CNRY) have the fixed minimum of 10.

3. A Province-Based Analysis of Physical-Chemical Variables

3.1. Partitioning

[32] The statistical separation between the provinces was assessed with the ANOSIM test (section 2.3) [Clarke and Green, 1988] by comparing samples extracted from the model output and observations. Each sample consists of the annual mean and standard deviation of the monthly values.

[33] The dependency on the sampling procedure was first analyzed. The data set combining biological and environmental data from the model (BIO_ENV-mod) was subsampled with a progressively increasing number of points (fixed sampling) and the ANOSIM test was performed (Figure 2). The global R value (section 2.3) converges when the number of extracted points is above 16. The two-point sampling is indicative of the results that one may obtain by characterizing a province with one station (although at least two points are needed for the ANOSIM test to work). The global R value is lower in this case as it is intuitively more difficult to separate provinces on the basis of one single extraction. The area-weighted random sampling, which increases the number of samples in the largest provinces, gives a global R value comparable to the two-point sampling. This implies that a nonuniform sampling tends to bias the results. Therefore, the fixed sampling scheme with 16 randomly extracted points from each province was adopted in the following results.

[34] The two-dimensional MDS structure of the 16-point sampling from each province is shown in Figure 3 for the ENV-obs and ENV-mod (normalized) data sets. These ordinations give hints at similar clustering of samples for certain provinces both in the observations and in the model. For instance, province 83 (Austral Polar, Table 1) has the largest scattering and all the trades/tropical provinces are clustered together. It is, however, difficult to assert whether the provinces are statistically different from each other on the basis of the intraprovince differences, and analysis of similarity is therefore needed.

[35] The results from the ANOSIM test are listed in Table 2 where the global R statistic (section 2.3) is reported for all the data sets and combinations. Almost all of the analyzed province pairs were statistically separable ($P < 0.01$) for any considered data set. The global R values and

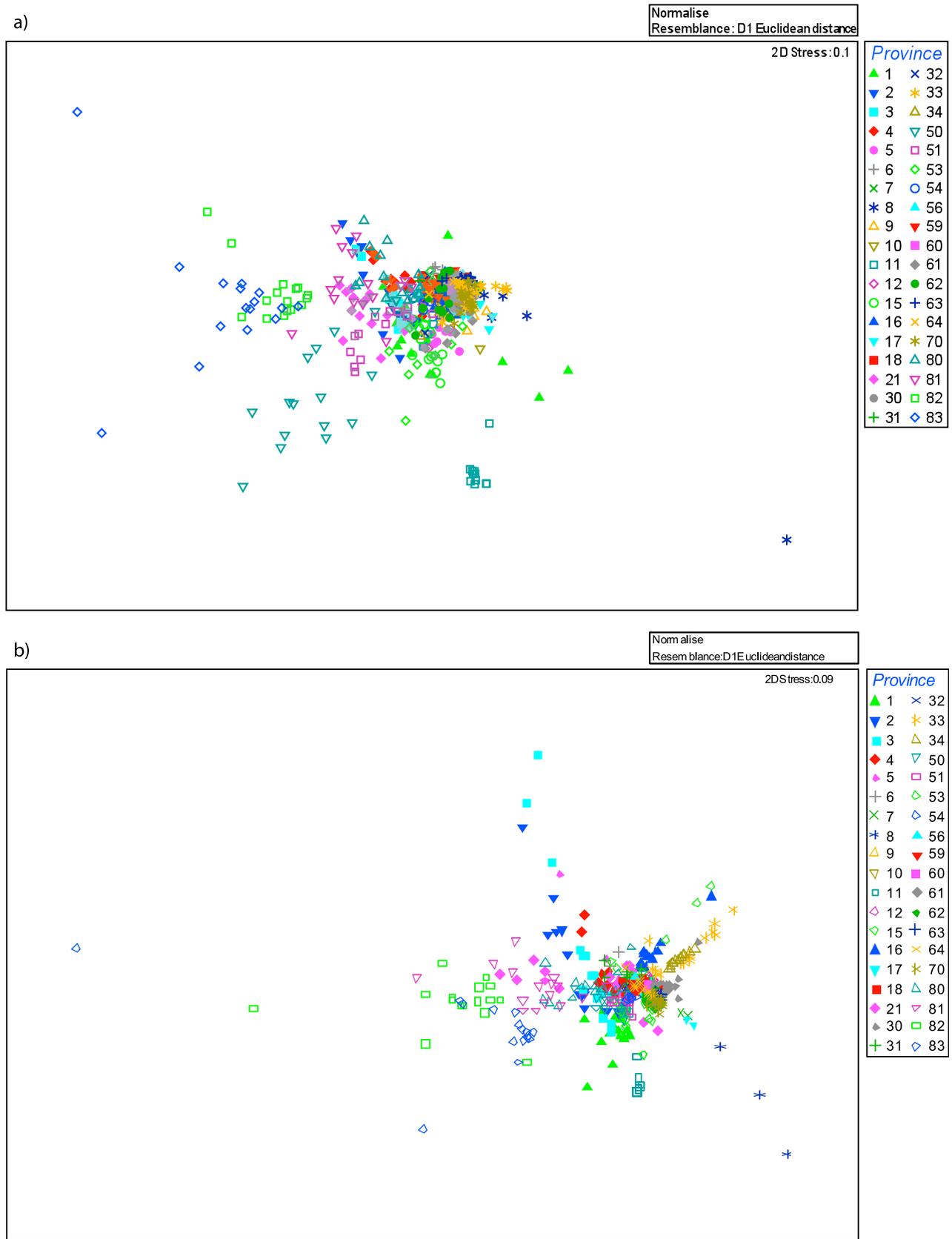


Figure 3. MDS ordination of the multivariate samples extracted from (a) observations and (b) model data with the 16-point fixed sampling. The symbols indicate the province number where the sample was extracted from.

Table 3. Results of the ANOSIM Test Performed on the Observations (ENV-obs) and ENV, BIO, and ALL Model Data by Extracting 16 Random Samples From Each Province^a

	Global R	P < 0.01	>0.75	0.50–0.75	0.25–0.50	<0.25
<i>Data Set (Global)</i>						
ENV-obs	0.59(0.003)	0.98	0.47	0.27	0.19	0.05
ENV-mod	0.60(0.003)	0.98	0.46	0.31	0.18	0.04
BIO-mod	0.52(0.004)	0.93	0.37	0.25	0.22	0.09
ENV_BIO-mod	0.57(0.003)	0.98	0.48	0.29	0.18	0.04
<i>Data Set (Tropics)</i>						
ENV-obs	0.43	0.97	0.08	0.32	0.48	0.09
ENV-mod	0.39	0.92	0.03	0.27	0.44	0.18
BIO-mod	0.42	0.92	0.26	0.20	0.35	0.12
ENV_BIO-mod	0.44	0.95	0.24	0.17	0.39	0.15

^aThe global R value is estimated as the mean and standard error (in parentheses) of 12 repeated ANOSIM computations for each data set. P < 0.01 entries report the fraction of province pairs with significant resemblance and the remaining entries are the fractions of significant pairs which fall within arbitrary separation thresholds [Clarke and Green, 1988] (>0.75 totally separate; 0.75–0.5 weak overlap; 0.5–0.25 overlap but some separation; <0.25 no separation). The tropics data set presents the results of the same analyses performed over the subset of provinces belonging to Longhurst's [2007] trades biome. Global R values are given for one single sampling.

the percentage of statistically separable provinces for the environmental data (ENV-obs) and model output (ENV-mod) are very similar; global R values were 0.59 and 0.60 for observations and model, respectively, suggesting overall separation with some overlap of the provinces. Only 5% out of the possible 703 province pairs are not distinguishable on the basis of the extracted random samples and close to 50% of the pairs are totally separate. We notice here and discuss further in section 6 that the addition of biological variables to the environmental data set (ENV_BIO-mod in Table 2) reduces the global R value to 0.57 (P < 0.001) because the predefined set of provinces is less separate by considering the BIO-mod data set only.

[36] ANOSIM was also performed on the tropical provinces 7, 8, 9, 10, 17, 30, 31, 60, 61, 62, 63, 64 classified as trades in Table 1 (Table 3). The global separation index R for the environmental data sets is lower than the one for all the provinces. There is a very small fraction of completely separated provinces both in the observations and in model results, and most of the province pairs fall in the range of 0.5–0.75 that indicates overlap. It illustrates that physical conditions are more similar at tropical-equatorial latitudes, even between provinces that belong to different oceans. It is, however, interesting to note that the R index increases when biological model data are added to the analysis, while it decreased in the global analysis that included the temperate and polar regions. This indicates that contrasting trophic structures are possible in the model, even if the physical and environmental conditions are more similar.

[37] The global analysis indicates that the chosen partition is largely composed of statistically separable provinces in both observations and model and that for the model this separation is more robust on the basis of the seasonal characteristics of the physical and hydrochemical variables rather than the biological variables.

3.2. Assessment

[38] A corollary of the previous results is that the 16 samples of environmental data are representative of each province's physical-chemical conditions because they are statistically less different from each other than the differences between provinces. Any overlaps between provinces

are likely related to the superimposition of subjective, fixed boundaries, but the high level of similarity between intraprovince station properties is common to both the observations and the model output. Nonetheless, the two cannot be compared by a visual inspection of Figure 3, so a more objective measure is needed.

[39] Data and models can be related to each other by means of a Spearman rank correlation coefficient (ρ) on the elements of the normalized resemblance matrices derived from each data set (equivalent to the RELATE test of Clarke and Gorley [2006, section 2.3]). To account for the variation due to the random extraction of the samples from each province, the computation of ρ was done on a set of 1000 different samplings of both observation and model data sets. The distribution of the coefficient perfectly fits a normal shape with mean and standard deviation 0.64 ± 0.02 . The significance of this number against the null hypothesis that there is no relationship between the two matrices is always very high (P < 0.001).

[40] However, it is important to point out that the observation data set used as reference for this test is itself defined as a random extraction of samples from each province. Therefore, there is an associated uncertainty in the reference data set, and the expected value for a perfect fit cannot be $\rho = 1$. It is possible to estimate a reference correlation value from the distribution of autocorrelation coefficients between several random extractions all done from the reference ENV-obs data set. The resulting value of 0.7 can be compared with the value obtained above with the random correlation test (Figure 4). The two means are of course statistically different (otherwise the model would be perfect) but the closeness indicates that the correlation in terms of province characterization between model and data is very good.

3.3. Significance

[41] As discussed by Hardman-Mountford *et al.* [2008], one property of systems is hierarchical organization, which in ecosystem terms implies that biomes can be further subdivided into classes with distinctive properties (i.e., provinces). We can test this hierarchical grouping by checking whether the predefined provinces are indeed

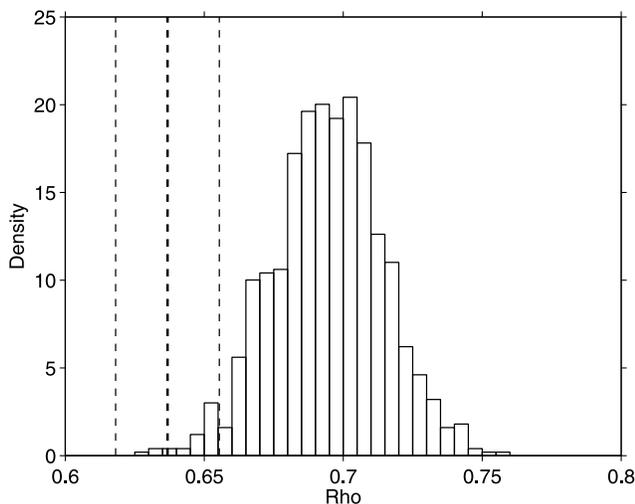


Figure 4. Empirical distribution of the rank correlation coefficient values obtained with pairs of random samples extracted from the biogeochemical provinces of the ENV-obs data set. The dashed thick line is the value of ρ for the correlation between ENV-obs and ENV-mod with the standard deviation interval (gray dashed lines).

arranged into larger biomes, either in the model or the observations.

[42] Figure 3 is a homomorphic projection of model and observation data that is, however, difficult to interpret in terms of biomes. It is possible to improve this description and gain further insights into data and model biogeography by analyzing, using a two-dimensional MDS ordination, the multivariate bulk properties of each predefined province. As explained in section 2.5, to account for the spatial and temporal variability within each province, both standard deviations are considered as additional variables of the multivariate bulk data sets.

[43] This analysis was done on the ENV data set only, because it allows the comparison between observations and model data. Figure 5a shows that there is a discrepancy in the observations between the groups identified by either the MDS or the hierarchical cluster analysis (here shown with an arbitrary choice of distance isolines) and the classification of provinces according to *Longhurst's* [2007] biomes (Table 1). In particular, at the scale of the mapped observations it is not possible to separate the coastal biome provinces because they are scattered in almost all the other groups. *Longhurst* classifies the subtropical Pacific gyres (NW province 56 and south province 59, Figure 1 and Table 1) as belonging to the westerlies biome. Our analysis based on the seasonal multivariate observations described in section 2.2 indicates that their characteristics are more similar to the trades regime provinces.

[44] The unsupervised grouping identified by the SIMPROF analysis (Figure 5b, see section 2.3) reveals in fact a different classification than *Longhurst's* [2007], indicating that there is a statistically significant separation also between provinces that intuitively belong to the same biome. For instance, the western boundary currents are well separated from the other provinces of the westerlies biome, as

also occurs for other groups such as the North Atlantic provinces, the upwelling regions and the Southern Ocean provinces, which are divided into polar front and Antarctica (provinces 21, 51 and 80–81). Some provinces (indicated with their acronyms in the legend of Figure 5b) cannot be grouped into common biomes on the basis of the available environmental data. This may imply either that they are characterized by physical-chemical properties that are so different as to become a separated biome or that their classification into larger biomes requires additional biological information (as, e.g., done subjectively by *Longhurst* [2007]). Additionally, as also pointed out by *Hardman-Mountford et al.* [2008], the trades biome is too “wide.” One single biome cannot account for the environmental seasonal variability found in tropical provinces, upwelling provinces (mostly equatorial) and especially the southern subtropical gyre provinces (SATL-SPSG, Figure 5b).

[45] The same MDS analysis applied to the model (Figure 6) shows a smaller number of groups which are more difficult to link to the *Longhurst* biomes. In contrast with the observations, the MDS and hierarchical clustering give similar classification, and the number of significant clusters is lower according to the SIMPROF test. The North Atlantic Provinces are grouped together as in the observations and the same is true for the Antarctic frontal provinces (80 and 20), although the mean properties of the sub-Antarctic province SANT (81) are more similar to the North Atlantic cluster. This evidences a model bias in province 81, which is due to an enhanced early stratification during springtime that renders this province more similar to the characteristics of the North Atlantic [*Vichi and Masina*, 2009]. The other large cluster can be further separated into equatorial and subtropical gyre regions, with more similarity between the southern hemisphere subtropical gyres (provinces 59 and 10). On the basis of the mean bulk properties and their spatial and temporal variations, the model has less skill in separating the northern Indian provinces from the Mediterranean and also from the Kur-oshio (53), which is surprisingly distant from the Gulf Stream province (5). These provinces are characterized by well-known coastal or mesoscale processes. It is thus more likely that the coarse resolution of the model hinders the development of distinct physical-chemical features.

4. Chlorophyll-Derived Provinces

[46] Chlorophyll observations were deliberately not used in the previous analyses based on the assumption that bottom-up control by physical drivers defines the biogeochemical provinces [*Longhurst*, 2007]. In section 3 we have seen that there are consistent bottom-up physical and hydrochemical features characterizing the set of predefined provinces, supporting this assumption in both observations and model. It is therefore interesting to analyze whether the chlorophyll signal is also characteristic in each province, which might also imply that there is a direct connection between bottom-up forcings and the ecosystem (although here the ecosystem is limited to phytoplankton and to chlorophyll concentration as a proxy for biomass). We have thus repeated all the province-based analyses presented above on the mapped chlorophyll data (model and SeaWiFS, section 2), after log transformation to equally account

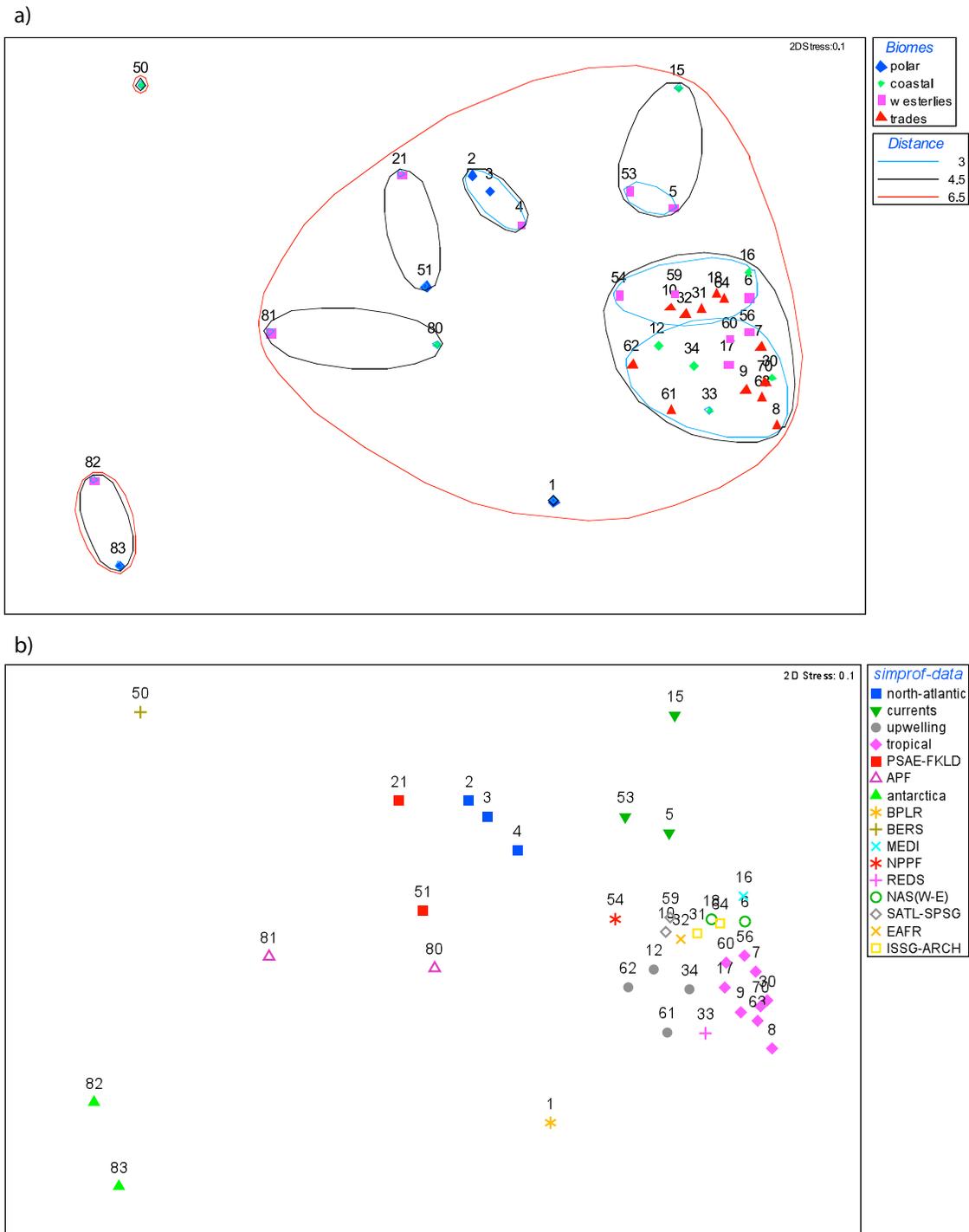


Figure 5. MDS ordinations of the province bulk samples extracted from the observation data set (ENV-obs). Labels indicate the province number. (a) Symbols represent the corresponding biome according to Longhurst's [2007] classification, and contours are the results of a hierarchical cluster analysis drawn at 3 arbitrary distances. (b) Symbols are the classification obtained from the SIMPROF analysis.

for extreme values and computation of resemblance matrices with the Euclidean distance.

[47] Particularly, we checked through fixed random sampling and the ANOSIM test (section 2.5) whether provinces are statistically separable in terms of chlorophyll values from the model and satellite data sets, independently

from the environmental data. The test results are shown in Table 4. Global R statistics are both lower than 0.5 (0.3) and less than found with environmental data (0.44, see Table 2), which implies that the chlorophyll-based provinces statistically overlap more than the environmental ones, though there is still some separation. CHL-obs is characterized by a

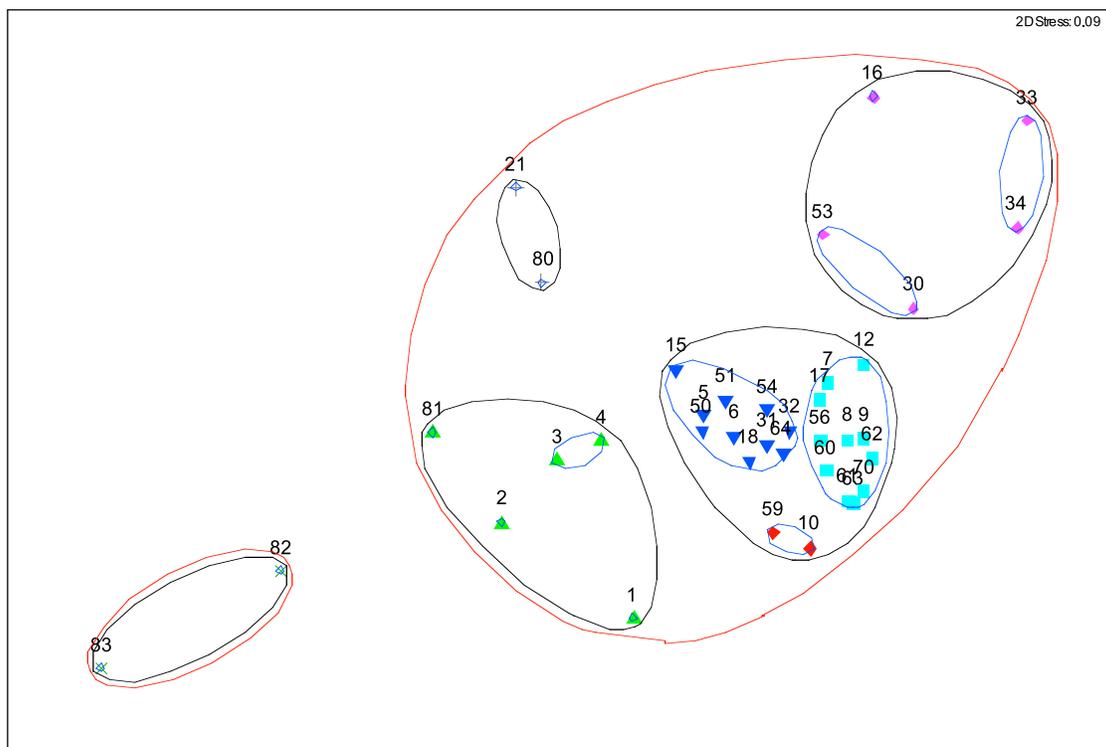


Figure 6. MDS ordinations of the province bulk samples extracted from the model environmental data set (ENV-mod). Labels indicate the province number, the symbols are the groups obtained with the SIMPROF analysis, and the contours are the results of a hierarchical cluster analysis drawn at three arbitrary distances.

very small fraction of fully separated province pairs, whereas this fraction is substantially larger in the model data (CHL-mod). If the environmental data are also analyzed together with the Chl data (ENV_CHL data sets in Table 4), the separation increases as expected in both data sets, indicating that chlorophyll concentration alone is not a sufficient criterion for province separation according to this predefined layout.

[48] The comparison between the CHL-obs and CHL-mod data sets operated by means of one single fixed sampling extraction and the computation of resemblance matrices (RELATE test, section 2.3) gives a ρ value of 0.19 ($P > 0.001$) which changes to 0.31 ($P > 0.001$) if the coastal provinces 11 and 12 with the highest values in CHL-obs are removed. This implies that model and data are weakly related through chlorophyll, yet the matching pattern is statistically significant at 99% level in the open ocean areas.

It is also interesting to compute the reference ρ value of autocorrelation from an empirical distribution of 1000 successive fixed random samplings of the CHL-obs data set (as done in section 3.2 for the environmental data and shown in Figure 4). The autocorrelation coefficient with the chlorophyll data is lower (0.44 ± 0.03 , confirming the large heterogeneity within provinces evidenced by the ANOSIM) and the average correlation between the data sets is 0.27 ± 0.03 , still removing the coastal provinces as above.

[49] The clustering of provinces into biomes was also analyzed as in section 3.3 (Figures 5 and 6) using the bulk sampling (province means, monthly standard deviation and spatial annual variability, section 2.5). Interestingly, there is very little grouping into larger biomes either in the data and in the model (Figure 7), as it might also be argued by the marked overlap found in the analysis of similarity (Table 4). The layout of provinces in satellite data is similar to that

Table 4. Results of the ANOSIM Test Performed on the Chlorophyll Data Sets and on the Environmental Variables and Chlorophyll Data Sets^a

Data Set	Global R	$P < 0.01$	>0.75	0.50–0.75	0.25–0.50	<0.25
CHL-obs	0.31	0.77	0.08	0.21	0.27	0.20
CHL-mod	0.44	0.83	0.29	0.21	0.23	0.11
ENV_CHL-obs	0.54	0.98	0.42	0.27	0.21	0.07
ENV_CHL-mod	0.54	0.97	0.41	0.32	0.19	0.05

^aSee Table 3. Chlorophyll data sets alone are model and SeaWiFS observations and the environmental variables and chlorophyll data sets are 16 random samples from each province are considered.

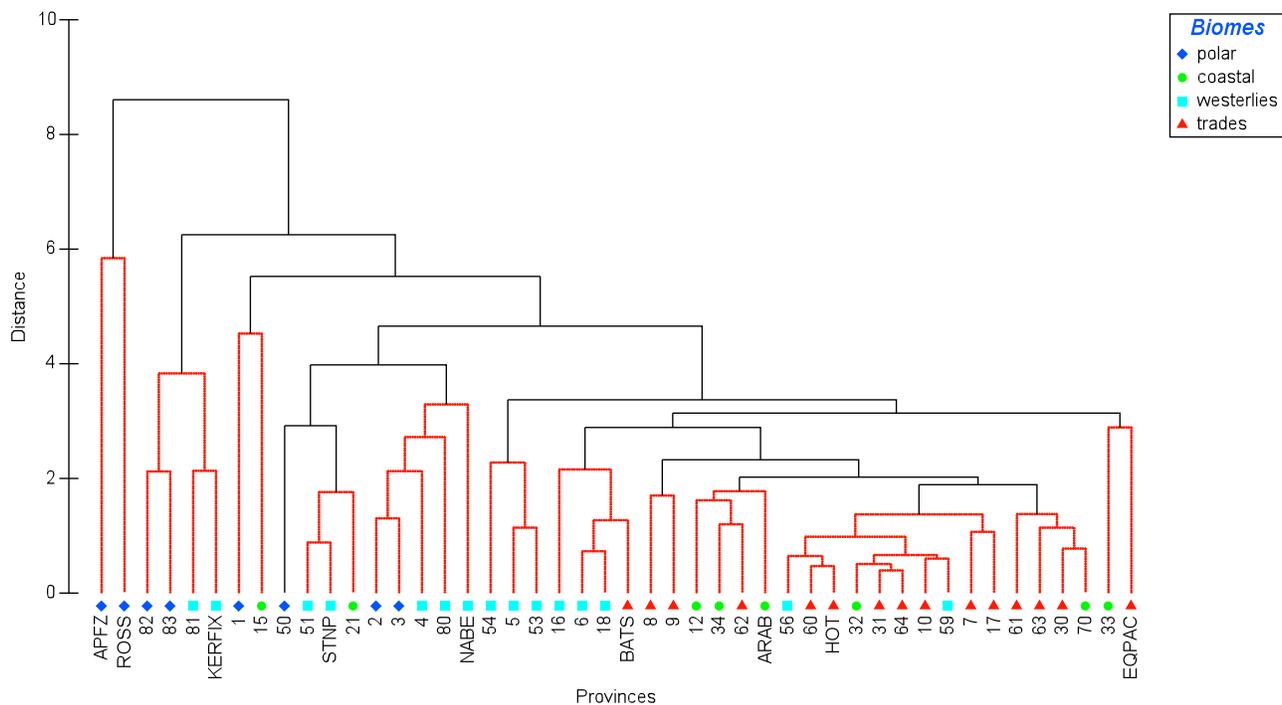


Figure 8. Cluster analysis of the seasonal variability in environmental variables of provinces and JGOFS stations. The Longhurst biomes are also given as reference.

the lower reliability of satellite data at high latitudes, as also discussed by *Hardman-Mountford et al.* [2008].

5. The Usage of Single-Point Time Series Data for Province Characterization

[50] In sections 2–4 we have compared gridded data with model results. In this section, we ask the question of how useful are time series data for the validation of global ocean biogeochemical models verifying whether the observed data can be considered representative of the entire province. The underlying assumption is that gridded data sets are the best available spatially resolved description of ocean properties.

[51] We assume, as in section 3.1, that physical features are the basic expression of the province, and therefore, we limit the analysis to the environmental variables. Note, however, that this is not the same analysis done in section 3.2; here we have applied bulk sampling considering the annual mean and the temporal standard deviation only (section 2.5). Spatial variability is neglected as we compare mean province properties with a single station. The JGOFS sites have also been added to the data set, considering both the time series and process study (compare section 2.2). The MDS and hierarchical cluster analyses as described above have been performed and, since they are comparable, we discuss here the results of the cluster analysis only (Figure 8). The permutation method SIMPROF was also used to derive the significance of a branching in the dendrogram and the clusters identified by red lines are significant at 95% level.

[52] The analysis suggests that no provinces can be represented by the APFZ and ROSS process study data, as they are distant from all the other data. This implies that these JGOFS studies are local and should be used for

colocated comparisons, avoiding any large-scale extrapolation. This is likely to occur also for KERFIX, which although has a time series a couple of years longer, it is located in a frontal zone with high temporal and spatial variability difficult to be described by coarse resolution data mapping. KERFIX is in our analysis more related to the polar biogeochemical provinces defined by Longhurst, particularly with the sub-Antarctic province 81 that is instead classified as westerlies by *Longhurst* [2007]. The analysis at this spatial scale indicates that the separation between the other southern Antarctic provinces is less significant.

[53] Moving to the tropics, the other process study EQPAC is also not associated with the other provinces from the trades biome. The grouping with the Red Sea province is incidental and probably caused by the common absence of seasonality that characterize these regions. On the other hand, Arabian Sea ARAB data span a period of about 2 years covering an entire seasonal cycle, and therefore, they are more linked to the characteristics of the other provinces of tropical upwelling (12, 34 and 62).

[54] BATS is equivalent, in terms of seasonal variability of the environmental parameters to the subtropical regions of the North Atlantic (6 and 18). BATS geographically belongs to the trades biome, even if, given the large seasonal variability is probably better classified as transitional region between the trades and westerlies provinces [*Brix et al.*, 2006]. Model results averaged over province 6 (NASW) can thus be compared with data from BATS, though it might be argued that the boundary with province 16 should be moved eastward to enhance the difference between the eastern and western sides of the Atlantic.

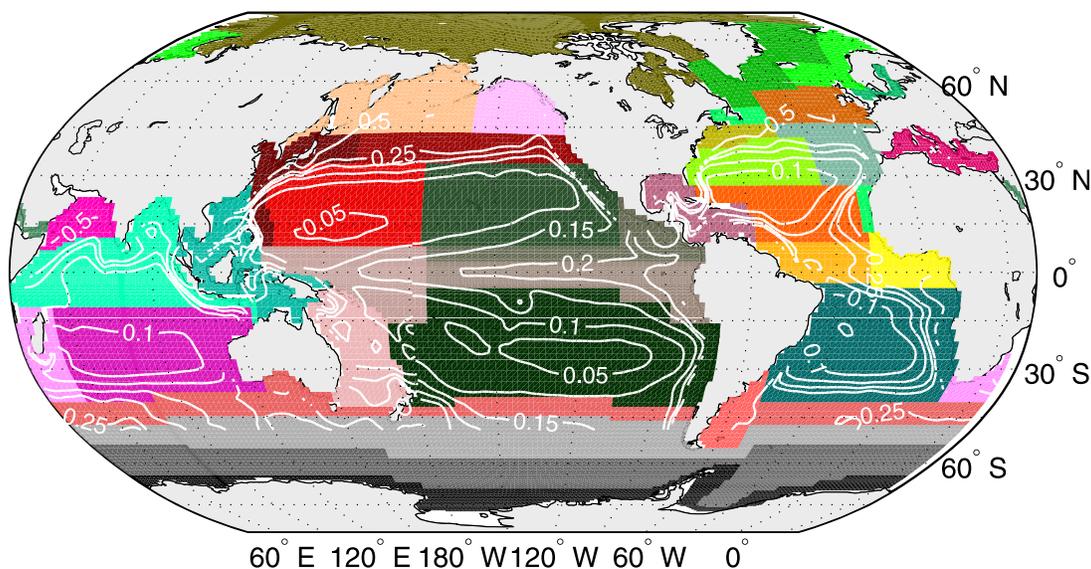


Figure 9. Province layout from Figure 1 with superimposed mean annual chlorophyll concentration from SeaWiFS.

[55] The Pacific has two long-term monitoring stations, STNP and HOT. The environmental variables at STNP in the northwestern Pacific have a similar seasonal cycle as the mean of province 51 that contains it. Data collected at STNP can thus be compared at a 95% confidence level with averages extracted from that province. The same considerations are valid for station Aloha, the Hawaii Ocean Time series (HOT). HOT is part of the large cluster of equatorial and intertropical provinces identified in the MDS analysis of Figure 5a. This implies that there are several provinces with similar seasonal cycles in the physical and nutrient variables. In this case, the method indicates that the province where HOT is located geographically (province 60) is the more similar, although this similarity does not reach the 95% level of significance.

6. Discussion

[56] For the purpose of this study, Longhurst provinces are taken as useful reference structures and they are not considered as permanent biogeographic units of the global ocean. We recognize that the predefined set of provinces is arbitrary, as are the boundaries between them, and more importantly, they are not homogeneous oceanographic units. This is why we derive them by aggregating mapped data and model results of higher horizontal resolution than the province areal extension. In their original conception, biogeochemical provinces are regions of the ocean where the physical processes have similar seasonal dynamics and give rise to similar biogeochemical processes, and not homogeneous units that can be described by zero-dimensional or vertical one-dimensional biogeochemical models.

[57] This is exactly what the ANOSIM test demonstrated; that the separation between Longhurst's [2007] provinces, based on the chosen set of variables, is statistically significant, implying that provinces are present in both the global environmental data sets and the model results. Moreover,

the PELAGOS model has skill in simulating the relationships between the different seasonal variability observed in each province, as quantified with the rank correlation analysis of the RELATE test. Therefore, the supervised partitioning of the ocean described by Longhurst has some relationship with the properties emerging from a comprehensive physical-biogeochemical model of the global ocean.

[58] However, the same analysis of similarity conducted on the satellite chlorophyll data set resulted in a significantly smaller separation between provinces, which may suggest that the province's founding concept of alignment between coherent physical properties and biological processes (in this case exemplified by chlorophyll concentration, the only globally available data) is not substantiated. This apparent mismatch may be explained by the different origin of the data sets used for the analysis. Nutrient and physical data derive from gridding procedures that imply the usage of analytical correlation functions that merge the sparse information, while chlorophyll concentration fields are interpolated from high-resolution sensor data. Therefore, in most of the provinces, chlorophyll gradients are usually stronger than physical gradients. In addition, as for instance shown in Figure 9, the same chlorophyll gradients are observed in provinces that are known to have significantly different physical properties (as for instance in the eastern and western part of the northern Pacific subtropical gyre or in the northern Atlantic, Figure 1).

[59] The separation between the subjectively determined Longhurst provinces is fixed a priori and does not necessarily coincide with the provinces that can be directly derived from the SeaWiFS data [e.g., Hardman-Mountford et al., 2008]. Hardman-Mountford et al. [2008], in fact, found a significant dissimilarity between SeaWiFS data samples extracted from different putative biomes in the Pacific Ocean. However, they used samples of relatively small areal dimension ($2^\circ \times 2^\circ$), and our results suggest that the averaging procedure within provinces reduces the sep-

aration. If the provinces are defined on the basis of the average chlorophyll value and standard deviations, the overlap between provinces is usually larger (results not shown). The usage of a fixed sampling and the comparison of resemblance matrices partly avoid the averaging bias that tends to give more weight to the choice of the boundaries.

[60] The analysis of model results only, illustrates an additional feature of the usage of predefined provinces. As shown in Table 3, provinces are more significantly separated in the environmental model data than in the biology, also when considering other variables besides chlorophyll (such as bacterial and heterotroph carbon biomass, section 2.5). We may speculate that, also in real systems, provinces obtained from considerations on the physical drivers may not be the same as macroecological province definitions. Provinces are defined partly by environmental characteristics, but also by species composition or by a combination of several ecological factors as for instance trophic interactions [Longhurst, 2007]. This is a possible reason we did not find a complete hierarchical organization in biomes as laid out by Longhurst, and some provinces have been singled out as having peculiar characteristics that do not fit into larger biomes (Figure 5). It may well be that statistically different seasonal variability in the physical drivers may lead to similar macroecological assemblages that are qualitatively classified as belonging to the same biome, because the specific ecological and physiological interactions realize a similar functional structure in the biomass distribution. The converse may also be true: provinces that are substantially similar in environmental features may host contrasting biological structures as a result of trophic and ecological interactions at smaller spatial scales. This is actually what happens if the ANOSIM is performed on the tropical provinces only (section 3.1). The dissimilarity between provinces increases when biological information is added. This can also partly explain the mismatch between our classification and Longhurst's, who actually incorporated ecological considerations in his analysis. Another implication is that biogeochemical models with more detailed trophic structures may be required in tropical regions in order to capture the observed biogeographic layout.

[61] The above discussion underlines some limitations of supervised classification methods using predefined provinces. They have been proven to be a useful concept for studying the global ocean biological processes [e.g., Ducklow, 2003] and, as proposed in this paper, are valuable for model validation issues. However, supervised classification suffers from the specific methods used for the spatial mapping of observations (gridding, averaging) and cannot completely account for the spatial variability of observed chlorophyll concentration.

[62] A further step in this approach is therefore to derive the provinces directly from the model results and then comparing with the classifications obtained using mapped data sets or satellite observations [e.g., Hardman-Mountford et al., 2008; Devred et al., 2007; D'Ortenzio and Ribera d'Alcalá, 2009]. An initial example was done by Sarmiento et al. [2004] defining the regions by means of an empirically derived, linear combination of physical properties such as SST, mixed layer depth, etc. Gregr and Bodtke [2007] refined this method by substituting the interpolation with a multivariate adaptive classification algorithm. A likely alternative

would be to use unsupervised neural networks such as Self-Organizing Maps [Kohonen, 2001] in a similar way to that of Allen et al. [2007]. Grid points with similar bulk properties are likely to be grouped together, defining the boundaries in a more coherent way which would thus be dependent on the model dynamics.

7. Conclusions and Recommendations

[63] The approach presented in this paper applies the concept of biogeochemical provinces as a diagnostic tool for the analysis and validation of global marine biogeochemistry models. It is proposed as a method of overcoming the limitations to model verification imposed by data scarcity and the general undersampling of relevant ocean biogeochemical properties. The analysis has shown that intuitive provinces derived by a priori considerations, are coherently distinguished in the environmental data (physical and hydrochemical variables) to a significant degree and that the same relationships between provinces are found in the results of the PELAGOS biogeochemistry model. A comparison carried out at the level of biogeochemical provinces demonstrates that the correlation between model and observations is significant and quite high if provinces are defined on the basis of a limited number of randomly extracted stations containing environmental data. The correlation is still significant against satellite chlorophyll data but much lower.

[64] It is important to note that the spatial and temporal variability of the real and the simulated systems have to be considered in the comparison exercise. Comparing model and observations on the basis of means over a certain province is less reliable, especially for comparisons against high-resolution satellite-derived chlorophyll data. This problem is especially relevant at high latitudes, due to the lack of information on seasonal variability from satellite sensors, which partly hinders the interpretation of model skills. On the other hand, long-term monitoring stations (BATS, HOT and STNP) appear to be representative of their respective provinces and thus, in the limit of the currently available environmental data, can be used as reference data sets for understanding the provinces themselves.

[65] The application of multivariate techniques (such as MDS and the related statistical tests for the analysis of similarity and rank correlation), provides a powerful tool for the interpretation of model results. We recommend their use in the validation process of OBGCM and especially in objective comparisons with data. This implies the use of multivariate data sets with global ocean coverage, which are only partly available. The first-order analysis has to be forcibly based on bulk properties, because these are currently the only data with sufficient spatial and temporal coverage. The new available global products of PFT distributions [e.g., Sathyendranath et al., 2004; Alvain et al., 2005, 2008; Uitz et al., 2006; Aiken et al., 2007; Hirata et al., 2008; Brewin et al., 2010] will provide a substantial aid for the objective validation of the PFT model results. Additionally, the strength of our approach is that once any possible province is verified as statistically separate from the others and characterized by consistent and coherent properties, other more ecologically relevant information (size-fractionated growth rates, biomass flows through PFTs, etc.)

can be extracted from the model results and compared with more local data sets [e.g., *Bouman et al.*, 2006; *Olguin et al.*, 2006; *Hirata et al.*, 2008] and long-term station data, which are usually more reliable than sparse observations. This might help to overcome the mismatch between station data and large-scale model results without the need to extrapolate sparse, low-frequency measurements.

[66] **Acknowledgments.** This work was initiated thanks to a short-term visiting grant given to M.V. by the EUR-OCEANS network of excellence funded by the EU FP6 program. M.V. and S.M. were partly funded by the EU FP7 GreenSeas Project (265294). J.I.A. and N.H.M. were partly funded by theme 9 of the UK Natural Environment Research Councils (NERC) Oceans2025 program, the NERC MARQUEST program, the NERC Centre for Observation of Air-Sea Interactions and fluxes (CASIX), and the UK National Centre for Earth Observation (NCEO). SeaWiFS data used in this publication were produced by the SeaWiFS project at Goddard Space Flight Center. The data were obtained from the Goddard Earth Sciences Distributed Active Archive Center under the auspices of NASA. Use of these data is in accord with the SeaWiFS Research Data Use Terms and Agreements. M.V. wishes to thank D. L. Jones for making publicly available his toolbox for multivariate analyses. We thank two anonymous reviewers for their comments on an early version of the manuscript.

References

- Aiken, J., J. R. Fishwick, S. Lavender, R. Barlow, G. F. Moore, H. Sessions, S. Bernard, J. Ras, and N. J. Hardman-Mountford (2007), Validation of MERIS reflectance and chlorophyll during the BENCAL cruise October 2002: Preliminary validation of new demonstration products for phytoplankton functional types and photosynthetic parameters, *Int. J. Remote Sens.*, *28*, 497–516, doi:10.1080/01431160600821036.
- Allen, J. I., P. J. Somerfield, and F. J. Gilbert (2007), Quantifying uncertainty in high-resolution coupled hydrodynamic-ecosystem models, *J. Mar. Syst.*, *64*, 3–14, doi:10.1016/j.jmarsys.2006.02.010.
- Alvain, S., C. Moulin, Y. Dandonneau, and F. M. Breon (2005), Remote sensing of phytoplankton groups in case 1 waters from global SeaWiFS imagery, *Deep Sea Res., Part I*, *52*, 1989–2004, doi:10.1016/j.dsr.2005.06.015.
- Alvain, S., C. Moulin, Y. Dandonneau, and H. Loisel (2008), Seasonal distribution and succession of dominant phytoplankton groups in the global ocean: A satellite view, *Global Biogeochem. Cycles*, *22*, GB3001, doi:10.1029/2007GB003154.
- Anderson, T. R. (2005), Plankton functional type modelling: Running before we can walk?, *J. Plankton Res.*, *27*, 1073–1081.
- Arvanitidis, C., et al. (2009), Biological geography of the European seas: Results from the MacroBen database, *Mar. Ecol. Prog. Ser.*, *382*, 265–278, doi:10.3354/meps07955.
- Aumont, O., E. Maier-Reimer, P. Monfray, and S. Blain (2003), An ecosystem model of the global ocean including Fe, Si, P co-limitations, *Global Biogeochem. Cycles*, *17*(2), 1060, doi:10.1029/2001GB001745.
- Borg, I., and P. Groenen (2005), *Modern Multidimensional Scaling: Theory and applications*, 2nd ed., Springer, New York.
- Bouman, H. A., et al. (2006), Oceanographic basis of the global surface distribution of *Prochlorococcus* ecotypes, *Science*, *312*, 918–921, doi:10.1126/science.1122692.
- Brewin, R. J. W., S. Sathyendranath, T. Hirata, S. J. Lavender, R. Barciela, and N. J. Hardman-Mountford (2010), A three-component model of phytoplankton size class for the Atlantic Ocean, *Ecol. Modell.*, doi:10.1016/j.ecolmodel.2010.02.014.
- Brix, H., N. Gruber, D. M. Karl, and N. R. Bates (2006), On the relationships between primary, net community, and export production in subtropical gyres, *Deep Sea Res., Part II*, *53*(5–7), 698–717, doi:10.1016/j.dsr2.2006.01.024.
- Clarke, K. R., and R. N. Gorley (2006), PRIMER v6: User manual/tutorial, PRIMER-E, Plymouth, U. K.
- Clarke, K. R., and R. H. Green (1988), Statistical design and analysis for a biological effects study, *Mar. Ecol. Prog. Ser.*, *46*, 213–226, doi:10.3354/meps046213.
- Clarke, K. R., and R. M. Warwick (2001), A further biodiversity index applicable to species lists: Variation in taxonomic distinctness, *Mar. Ecol. Prog. Ser.*, *216*, 265–278, doi:10.3354/meps216265.
- Conkright, M., H. Garcia, T. O'Brien, R. Locamini, T. Boyer, C. Stephens, and J. Antonov (2002), *World Ocean Atlas 2001*, vol. 4, *Nutrients*, NOAA Atlas NESDIS, vol. 52, U.S. Govt. Print. Off., Washington, D. C.
- Costello, M. J., and E. Vanden Berghe (2006), 'Ocean biodiversity informatics': A new era in marine biology research and management, *Mar. Ecol. Prog. Ser.*, *316*, 203–214, doi:10.3354/meps316203.
- de Boyer Montégut, C., G. Madec, A. S. Fischer, A. Lazar, and D. Iudicone (2004), Mixed layer depth over the global ocean: An examination of profile data and a profile-based climatology, *J. Geophys. Res.*, *109*, C12003, doi:10.1029/2004JC002378.
- Devred, E., S. Sathyendranath, and T. Platt (2007), Delineation of ecological provinces using ocean colour radiometry, *Mar. Ecol. Prog. Ser.*, *346*, 1–13, doi:10.3354/meps07149.
- Doney, S. C., I. Lima, J. K. Moore, K. Lindsay, M. J. Behrenfeld, T. K. Westberry, N. Mahowald, D. M. Glover, and T. Takahashi (2009), Skill metrics for confronting global upper ocean ecosystem-biogeochemistry models against field and remote sensing data, *J. Mar. Syst.*, *76*(1–2), 95–112, doi:10.1016/j.jmarsys.2008.05.015.
- D'Ortenzio, F., and M. Ribera d'Alcalá (2009), On the trophic regimes of the Mediterranean Sea: A satellite analysis, *Biogeosciences*, *6*(2), 139–148, doi:10.5194/bg-6-139-2009.
- Ducklow, H. W. (2003), Biogeochemical provinces: Towards a JGOFS synthesis, in *Ocean Biogeochemistry*, edited by M. J. R. Fasham, pp. 3–17, Springer, New York.
- Friedrichs, M. A., et al. (2009), Assessing the uncertainties of model estimates of primary productivity in the tropical Pacific Ocean, *J. Mar. Syst.*, *76*, 113–133, doi:10.1016/j.jmarsys.2008.05.010.
- Gregg, W. W., P. Ginoux, P. S. Schopf, and N. W. Casey (2003), Phytoplankton and iron: Validation of a global three-dimensional ocean biogeochemical model, *Deep Sea Res., Part II*, *50*, 3143–3169, doi:10.1016/j.dsr2.2003.07.013.
- Gregg, E. J., and K. M. Bodtker (2007), Adaptive classification of marine ecosystems: Identifying biologically meaningful regions in the marine environment, *Deep Sea Res., Part I*, *54*, 385–402, doi:10.1016/j.dsr.2006.11.004.
- Hardman-Mountford, N., T. Hirata, K. Richardson, and J. Aiken (2008), An objective methodology for the classification of ecological pattern into biomes and provinces for the pelagic ocean, *Remote Sens. Environ.*, *112*, 3341–3352, doi:10.1016/j.rse.2008.02.016.
- Hirata, T., J. Aiken, N. Hardman-Mountford, T. J. Smyth, and R. G. Barlow (2008), An absorption model to determine phytoplankton size classes from satellite ocean colour, *Remote Sens. Environ.*, *112*(6), 3153–3159, doi:10.1016/j.rse.2008.03.011.
- Kohonen, T. (2001), *Self-Organizing Maps*, Springer Ser. Inf. Sci., vol. 30, 3rd ed., Springer, Berlin.
- Le Quéré, C., et al. (2005), Ecosystem dynamics based on plankton functional types for global ocean biogeochemistry models, *Global Change Biol.*, *11*, 2016–2040.
- Levitus, S., T. Boyer, M. Conkright, T. O'Brien, J. Antonov, C. Stephens, L. Stathoplos, D. Johnson, and R. Gelfeld (1998), Introduction, in *NOAA Atlas NESDIS 18*, vol. 1, 346 pp., U.S. Gov. Print. Off., Washington, D. C.
- Longhurst, A. R. (1995), Seasonal cycles of pelagic production and consumption, *Prog. Oceanogr.*, *36*, 77–167, doi:10.1016/0079-6611(95)00015-1.
- Longhurst, A. R. (2007), *Ecological Geography of the Sea*, 2nd ed., Academic, Burlington, Vt.
- Longhurst, A., S. Sathyendranath, T. Platt, and C. Caverhill (1995), An estimate of global primary production in the ocean from satellite radiometer data, *J. Plankton Res.*, *17*, 1245–1271, doi:10.1093/plankt/17.6.1245.
- Lynch, D. R., D. J. McGillicuddy Jr., and F. Werner (2009), Skill assessment for coupled biological/physical models of marine systems: Preface, *J. Mar. Syst.*, *76*, 1–3, doi:10.1016/j.jmarsys.2008.05.002.
- Madec, G., and M. Imbard (1996), A global ocean mesh to overcome the North Pole singularity, *Clim. Dyn.*, *12*, 381–388, doi:10.1007/BF00211684.
- Madec, G., P. Delecluse, M. Imbard, and C. Levy (1999), OPA8.1 ocean general circulation model reference manual. Notes du Pole de Modelisation, Lab. d'Océanogr. Dyn. et de Climatol., Paris. (Available at <http://www.lodyc.jussieu.fr/opa>.)
- Moore, J. K., S. C. Doney, and K. Lindsay (2004), Upper ocean ecosystem dynamics and iron cycling in a global three-dimensional model, *Global Biogeochem. Cycles*, *18*, GB4028, doi:10.1029/2004GB002220.
- Olguin, H. F., D. Boltovskoy, C. B. Lange, and F. Brandini (2006), Distribution of spring phytoplankton (mainly diatoms) in the upper 50 m of the southwestern Atlantic Ocean (30–61°S), *J. Plankton Res.*, *28*, 1107–1128, doi:10.1093/plankt/fbl045.
- Sarmiento, J. L., et al. (2004), Response of ocean ecosystems to climate warming, *Global Biogeochem. Cycles*, *18*, GB3003, doi:10.1029/2003GB002134.

- Sathyendranath, S., A. Longhurst, C. M. Caverhill, and T. Platt (1995), Regionally and seasonally differentiated primary production in the North Atlantic, *Deep Sea Res., Part I*, 42, 1773–1802, doi:10.1016/0967-0637(95)00059-F.
- Sathyendranath, S., L. Watts, E. Devred, T. Platt, C. Caverhill, and H. Maass (2004), Discrimination of diatoms from other phytoplankton using ocean-colour data, *Mar. Ecol. Prog. Ser.*, 272, 59–68, doi:10.3354/meps272059.
- Six, K. D., and E. Maier-Reimer (1996), Effects of plankton dynamics on seasonal carbon fluxes in an ocean general circulation model, *Global Biogeochem. Cycles*, 10, 559–583, doi:10.1029/96GB02561.
- Somerfield, P. J., K. R. Clarke, and F. Olgard (2002), A comparison of the power of categorical and correlational tests applied to community ecology data from gradient studies, *J. Anim. Ecol.*, 71(4), 581–593, doi:10.1046/j.1365-2656.2002.00624.x.
- Stow, C. A., J. Jolliff, D. J. McGillicuddy Jr., S. C. Doney, J. I. Allen, M. A. M. Friedrichs, K. A. Rose, and P. Wallhead (2009), Skill assessment for coupled biological/physical models of marine systems, *J. Mar. Syst.*, 76, 4–15, doi:10.1016/j.jmarsys.2008.03.011.
- Uitz, J., H. Claustre, A. Morel, and S. B. Hooker (2006), Vertical distribution of phytoplankton communities in open ocean: An assessment based on surface chlorophyll, *J. Geophys. Res.*, 111, C08005, doi:10.1029/2005JC003207.
- Vichi, M., and S. Masina (2009), Skill assessment of the PELAGOS global ocean biogeochemistry model over the period 1980–2000, *Biogeosciences*, 6, 2333–2353, doi:10.5194/bg-6-2333-2009.
- Vichi, M., N. Pinardi, and S. Masina (2007a), A generalized model of pelagic biogeochemistry for the global ocean ecosystem. Part I: Theory, *J. Mar. Syst.*, 64, 89–109, doi:10.1016/j.jmarsys.2006.03.006.
- Vichi, M., S. Masina, and A. Navarra (2007b), A generalized model of pelagic biogeochemistry for the global ocean ecosystem. Part II: Numerical simulations, *J. J. Mar. Syst.*, 64, 110–134, doi:10.1016/j.jmarsys.2006.03.014.
-
- J. I. Allen and N. J. Hardman-Mountford, Centre for Coastal Marine Studies, Plymouth Marine Laboratory, Prospect Place, West Hoe, Plymouth PL1 3DH, UK.
- S. Masina and M. Vichi, Centro Euro-Mediterraneo per i Cambiamenti Climatici, Istituto Nazionale di Geofisica e Vulcanologia, Viale Aldo Moro 44, I-40127 Bologna, Italy. (marcello.vichi@bo.ingv.it)