



# Biogeographic classification of the Caspian Sea

F. Fendereski<sup>1,2</sup>, M. Vogt<sup>2</sup>, M. R. Payne<sup>2,3</sup>, Z. Lachkar<sup>2,4</sup>, N. Gruber<sup>2</sup>, A. Salmanmahiny<sup>1</sup>, and S. A. Hosseini<sup>1</sup>

<sup>1</sup>Gorgan University of Agricultural Sciences and Natural Resources, Faculty of Fisheries and Environmental Sciences, Gorgan, Iran

<sup>2</sup>Environmental Physics, Institute of Biogeochemistry and Pollutant Dynamics, ETH Zurich, Zurich, Switzerland

<sup>3</sup>Centre for Ocean Life, National Institute of Aquatic Resources (DTU-Aqua), Technical University of Denmark, Charlottenlund, Denmark

<sup>4</sup>Center for Prototype Climate Modeling, New York University, Abu Dhabi, United Arab Emirates

Correspondence to: F. Fendereski (fendereski\_f@yahoo.com)

Received: 29 January 2014 – Published in Biogeosciences Discuss.: 20 March 2014

Revised: 18 September 2014 – Accepted: 7 October 2014 – Published: 27 November 2014

**Abstract.** Like other inland seas, the Caspian Sea (CS) has been influenced by climate change and anthropogenic disturbance during recent decades, yet the scientific understanding of this water body remains poor. In this study, an eco-geographical classification of the CS based on physical information derived from space and in situ data is developed and tested against a set of biological observations. We used a two-step classification procedure, consisting of (i) a data reduction with self-organizing maps (SOMs) and (ii) a synthesis of the most relevant features into a reduced number of marine ecoregions using the hierarchical agglomerative clustering (HAC) method. From an initial set of 12 potential physical variables, 6 independent variables were selected for the classification algorithm, i.e., sea surface temperature (SST), bathymetry, sea ice, seasonal variation of sea surface salinity (DSSS), total suspended matter (TSM) and its seasonal variation (DTSM). The classification results reveal a robust separation between the northern and the middle/southern basins as well as a separation of the shallow nearshore waters from those offshore. The observed patterns in ecoregions can be attributed to differences in climate and geochemical factors such as distance from river, water depth and currents. A comparison of the annual and monthly mean Chl *a* concentrations between the different ecoregions shows significant differences (one-way ANOVA,  $P < 0.05$ ). In particular, we found differences in phytoplankton phenology, with differences in the date of bloom initiation, its duration and amplitude between ecoregions. A first qualitative evaluation of differences in community composition based on recorded presence–absence patterns of 25 different species of plankton,

fish and benthic invertebrate also confirms the relevance of the ecoregions as proxies for habitats with common biological characteristics.

## 1 Introduction

The Caspian Sea (CS) is an enclosed water body that plays an important geopolitical role in the Central Asia region (Kosarev and Kostianoy, 2005). During the last few decades, the joint action of natural and anthropogenic factors has been aggravating the environmental state in the CS (Kopelevich et al., 2004; Barale, 2008). Increasing human activities such as the oil and gas industries, especially in the northern part of the CS, fisheries, agriculture and tourism (Kopelevich et al., 2004), together with decades of environmental mismanagement (Barale, 2008), have led to the severe degradation of water quality. The unintentional introduction of the ctenophore jellyfish *Mnemiopsis leidyi* in late 1999 (Shiganova et al., 2001) has added to the environmental problems, affecting a whole trophic level, since this organism feeds voraciously on zooplankton (Kideys et al., 2008). *M. leidyi* is only one of many introduced species: since the early 20th century, more than 60 non-native invasive species ranging from phytoplankton to fishes and their parasites have become established in the CS (Shiganova et al., 2010, 2011). Thus, the CS ecosystem has recently been shown to change drastically over timescales of several decades (Shiganova et al., 2004, 2010, 2011; Karpinsky et al., 2005), with massive impacts on plankton and fish biomass, chlorophyll *a*

levels, primary production and nutrient status (Shiganova et al., 2011). For example, the change in the trophic status of the CS from oligotrophic to eutrophic (Leonov and Stygar, 2001) is attributed partly to the effects of *M. leidy* on the CS ecosystem (Saravi Nasrollahzadeh et al., 2008). Due to the severity of these changes, ecosystem management and monitoring efforts to restore the health of the CS, or at least to mitigate some of the most severe changes, are required.

In this context, the inefficiency of management efforts focusing on single species only or based on limited regions in marine environments is well known (Zacharias and Roff, 2000). For an ecosystem-based management approach, in which all the interactions between the CS ecosystem and its surrounding biotic and abiotic environment are considered (EPAP, 1996), managers need to deal with the challenge of mapping the physical and biological components of the whole ecosystem (GREG and Bodtker, 2007). However, there is still no comprehensive archive of oceanographic information on the whole CS. The existing data on oceanographic observations in the CS are heterogeneously distributed in space and time and most of the observations stem from the northern and middle regions and are confined to the coastal parts of the CS. Moreover, most of the data date back to the Soviet era, not reflecting the substantial biological and environmental changes that have occurred since then (Kosarev and Kostianoy, 2005).

In recent years, marine ecosystem mapping has been greatly facilitated through the increased availability of remote sensing data (Richardson and LeDrew, 2006; McDermid et al., 2005). Detailed broad-scale spatial and temporal variability of several environmental variables, such as sea surface temperature and chlorophyll *a* concentration, can be captured by satellites (Robinson, 2004). Satellite data can be used for a biogeophysical classification of the CS, in a fashion similar to studies done in the European seas (Hoepffner and DoWell, 2005), the North Atlantic (Roff et al., 2003; Devred et al., 2007), the North Pacific (GREG and Bodtker, 2007; Kavanaugh et al., 2013) or the World Ocean (Longhurst, 1998; Vichi et al., 2011; Spalding et al., 2012). In these studies, regions with similar environmental patterns are grouped into ecological units or so-called “ecoregions”. Each ecoregion may respond in a different way to environmental changes and management policies (Bailey, 1996).

The close association between the oceanographic and biological characteristics in pelagic systems (Day and Roff, 2000) has been described both for primary producers (e.g., Platt et al., 2005; Reynolds, 2006) as well as for the higher trophic levels of the food chain (e.g., Thrush and Dayton, 2010). The link between ocean physics and biology is of great significance, especially for the areas where the biological data are scarce. Different applications, such as the computation of primary production (Platt and Sathyendranath, 1999) or species habitat modeling (e.g., Valavanis, 2008; Irwin et al., 2012) have benefited from the association between physical/environmental factors and the biological responses.

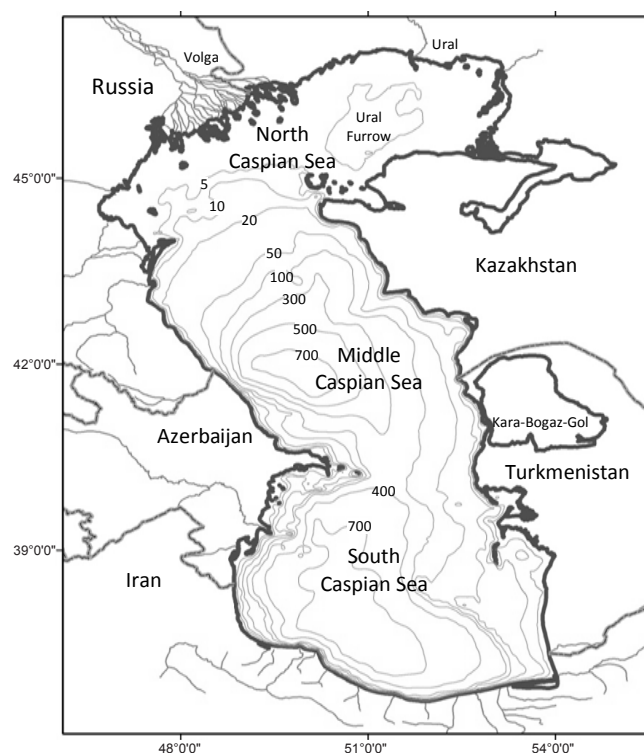
This association provided the basis for a number of studies in which biogeographic maps of the oceans have been developed based on abiotic non-taxonomic oceanographic classifications (e.g., Dietrich et al., 1963; Longhurst, 1998; Day and Roff, 2000). Ecoregions reflect the diversity in physical habitat types and hence, indirectly, the range of conditions that influence species distribution and community composition in marine ecosystems (Bredin et al., 2001). Bredin et al. (2001) linked trawl surveys and observational data on a variety of rare and endangered fish and Cetacean species of the Bay of Fundy to the physically defined ecoregions for this area. They noticed the significance of some ecoregions for the distribution of specific species. GREG and Bodtker (2007) assessed the biological relevance of their physical-based epipelagic ecoregions of the North Pacific using remote sensing measurements of chlorophyll *a* concentrations (Chl *a*). The results of GREG and Bodtker (2007) showed a significant difference in Chl *a* concentrations between ecoregions and suggested the possibility of describing regions with different biological attributes using only physical variables. The importance of environmental conditions for the distribution of *M. leidy* and several other species in the CS has been highlighted in different reports (e.g., Karpinsky et al., 2005; Shiganova et al., 2011). While episodic evidence of species occurrence and species richness exists for different locations of the CS and for a subset of dominant species on all trophic levels that inhabits the CS (e.g., Shiganova et al., 2011), a geo-referenced presence–absence or biomass data set containing information about all major dominant species across multiple trophic levels has, to our knowledge, not yet been published. A previous data synthesis and extrapolation effort was made by the Caspian Environment Program (CEP) for a set of 36 species in 2002, but the recently observed rapid changes in marine ecosystem structure and functioning indicate that habitat maps will require regular updating. As detailed above, scientific evidence suggests that species distribution and community structure is, in part, controlled by environmental conditions. If a link between environmental variables accessible through remote sensing and the biological structure of the CS ecosystem could be established, remote sensing would be of crucial importance for high-resolution ecosystem monitoring in this area.

In the present study, we applied an objective classification method to cluster the Caspian Sea into ecoregions based on a set of geophysical data. The main objective of this study is to provide a map of the CS divided into different areas with homogeneous geophysical attributes, i.e., so-called ecoregions to assist the effective management of the CS. We then tested for the biological significance of the obtained ecoregions using long-term satellite-derived Chl *a* distribution and available species composition data. Throughout the manuscript, we use the terms ecoregion to imply a homogenous area of the CS with similar biogeophysical/environmental characteristics.

**Table 1.** List of the potential variables used, their units, resolution and sources.

Variable	Sensor/provider	Units	Spatial resolution	Temporal extent	Source
Sea surface temperature (SST) and its seasonal variation (DSST)	AVHRR	°C	0.08789°	1985–2003	<a href="http://apdrc.soest.hawaii.edu">http://apdrc.soest.hawaii.edu</a>
Total suspended matter (TSM) and its seasonal variation (DTSM)	MERIS	g m <sup>-3</sup>	0.04166°	2003–2010	<a href="http://www.globcolour.info/">http://www.globcolour.info/</a>
Photosynthetic active radiation (PAR) and its seasonal variation (DPAR)	MODIS Aqua	Einstein m <sup>-2</sup> day <sup>-1</sup>	0.04°	2003–2011	<a href="http://oceancolor.gsfc.nasa.gov">http://oceancolor.gsfc.nasa.gov</a>
Wind speed <sup>a</sup> (turbulence) (WSP) and its seasonal variation (DWSP)	QuikSCAT	m s <sup>-1</sup>	0.25°	2000–2007	<a href="http://apdrc.soest.hawaii.edu">http://apdrc.soest.hawaii.edu</a>
Sea surface salinity (SSS) and its seasonal variation (DSSS)	CASPCOM <sup>b</sup>	ppt	–	1995	<a href="http://www.caspcom.com">http://www.caspcom.com</a>
Bathymetry (depth)	DNO of the RF MD	m	–	–	<a href="http://www.caspianenvironment.org">http://www.caspianenvironment.org</a>
Sea ice (ICE)	AARI	%	–	2004–2012	<a href="http://wdc.aari.ru/datasets/">http://wdc.aari.ru/datasets/</a>

<sup>a</sup> QuikSCAT data are produced by Remote Sensing Systems and sponsored by the NASA Ocean Vector Winds Science Team. Data are available at [www.remss.com](http://www.remss.com). <sup>b</sup> Committee on Hydrometeorology and Pollution Monitoring of the Caspian Sea.

**Figure 1.** Map of the CS showing the three sub-regions, the major rivers (solid lines) and bathymetry (depth in meters).

## 2 Methods

### 2.1 Study area

The CS (Fig. 1) is a landlocked water body with no major outlet to the ocean. Its surface area is about 371 000 km<sup>2</sup>, with a length of about 1210 km and average width of 320 km from 37 to 47° N and 47 to 55° E (Barale, 2008). The CS

is commonly divided into three basins, hereafter denoted as the northern Caspian Basin (NCB), the middle Caspian Basin (MCB) and the southern Caspian Basin (SCB), occupying about 29 %, 36 and 35 % of the surface area of the CS, respectively. Progressing southward, the depth of the CS increases considerably, averaging only about 5 m in the NCB, 190 m in the MCB and 330 m in the SCB (UNEP, 2006). Around 60 % of the CS, mainly the area along the northern and eastern coasts, consists of shelf areas with less than 100 m depth (Fig. 1; Ibrayev et al., 2010). The Kara-Bogaz-Gol, a hypersaline lagoon at the eastern slope of the CS is not considered in this study (Kopelevich et al., 2008).

## 2.2 Ecoregion classification

### 2.2.1 Data description

We identified 12 geophysical variables as potential candidates for the ecoregion classification (Table 1). For 5 observables, we considered both the annual mean and their seasonal variations, forming 10 independent variables, while sea ice and bathymetry constituted the remaining ones.

The satellite-based observables consisted of sea surface temperature (SST) from the Advanced Very High Resolution Radiometer (AVHRR), total suspended matter (TSM) from Medium Resolution Imaging Spectrometer (MERIS; level-3 binned products), photosynthetic active radiation (PAR) from the Moderate Resolution Imaging Spectroradiometer (MODIS) Aqua level 3 products, and wind speed (WSP) from Quick Scatterometer (QuikSCAT). Following Bakun and Parrish (1991), we computed the wind-generated turbulent mixing index as the cube of surface wind speed (WSP<sup>3</sup>). This represents the rate at which wind turbulent kinetic energy is injected into the ocean and becomes available for mixing the upper thermocline (Niiler, 1975). Bathymetry data based on interpolated echo sounding measurements were

extracted from the digitalized official CS navigation map<sup>1</sup>. Because of the non-normal distribution of bathymetry data throughout the CS, we used log-transformed bathymetry data. We digitized monthly surface isohaline data of the CS to create sea surface salinity (SSS) maps (Kosarev and Tuzhilkin, 1995). Sea ice data, provided by the Centre for Ice Hydrometeorological Information at the Arctic and Antarctic Research Institute in St. Petersburg (AARI), were used to create maps of percentage of sea ice coverage throughout the year (2004 to 2012). All data were interpolated onto a  $0.1^\circ$  longitude–latitude resolution grid, i.e., re-gridded onto a  $120 \times 98$  grid with 11 760 pixels.

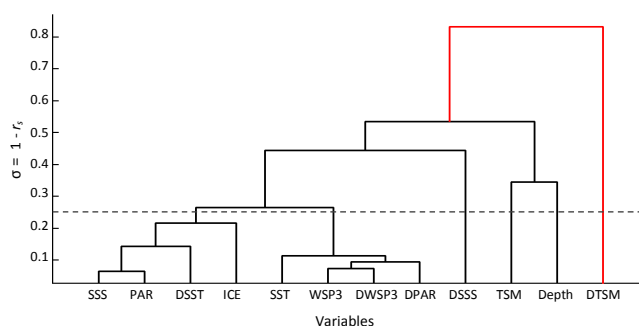
Decadal-scale interannual variability of physical parameters used are, in general, much smaller than the spatial variability and the latitudinal gradients of these parameters throughout the CS, i.e., the patterns that were used for ecoregions boundary delineation. For example, for a period from 1982 to 2000, a positive SST trend of about  $0.05$  to  $0.1^\circ\text{C}$  per year was observed in the CS (Ginzburg et al., 2005). This increase is very small in comparison with the spatial variability of climatological SST throughout the CS for the same period of time ( $4.24$  to  $24.9^\circ\text{C}$ ; see Fig. 4). In other words, despite the interannual to decadal-scale variability in physical features, like other water bodies, a generally persistent spatial pattern in the environmental variables used for this classification has been described for physical properties of the CS on decadal scales. Thus, while the available climatologies did not necessarily have the same time span, they can still be used to characterize the general spatial patterns in the CS (see Table 2). The seasonal amplitudes of the physical variables, denoted here by DSST, DSSS, DTSM, DWSP, DWSP3 and DPAR (abbreviations defined in Table 1), were computed by subtracting average values of the summer months (June, July, August) from those of the winter months (December, January, February).

## 2.2.2 Selection of classification variables

A robust classification requires the input variables to be as independent as possible, which is often not the case with environmental data sets. To eliminate redundant (strongly correlated) variables (Raftery and Dean, 2006; May et al., 2011), the collinearity between the 12 potential input variables was investigated using Spearman's rank correlation coefficient ( $r_s$ ; Supplement, Sect. S1; Wheeler and Cook, 2005). Variables were sorted and represented in a dendrogram, where the vertical axis represents the degree of independence ( $\sigma$ ) between variables based on their correlation coefficient ( $\sigma = 1 - r_s$ ; Fig. 2; see Supplement, Sect. S1 for details).

We decided to cut the dendrogram at  $\sigma = 0.25$  ( $r_s = 0.75$ ), i.e., to group/cluster variables with a correlation of  $r_s > =$

<sup>1</sup>The Navigation Map, Department of Navigation and Oceanography of the Russian Federation Ministry of Defense (DNO of the RF MD), scale: 1 : 1 000 000; 1987.



**Figure 2.** Dendrogram of all the potential variables correlation matrix (annual means). The dashed line shows the sensitivity at which the dendrogram was cut. The abbreviations are defined in Table 1. DSST, DWSP3, DPAR, DSSS and DTSM represent the seasonal amplitude of the corresponding parameter (SST, WSP3, PAR, SSS and TSM, respectively).

$0.75$  (Fig. 2, dashed line on the dendrogram). From each branch, one representative variable was chosen as an independent variable. DSSS, TSM, depth and DTSM were directly selected, while we chose ICE and SST from the two groups of highly correlated variables. The selection of a cutoff correlation coefficient and of one variable from each branch was a compromise between the retention of as many environmental predictor variables as possible and the avoidance of high correlations that would have confounded the interpretation of the resulting patterns. A sensitivity test of classification output to the choice of representative variables from these two groups was conducted (see Sect. 2.2.5). All the variables were normalized to unit variance in order to avoid the dominance of those variables with significantly higher variance in the data clustering (Kohonen, 2000).

## 2.2.3 Classification method

We followed a two-step classification procedure, consisting of (i) a data reduction using self-organizing maps (SOMs) and (ii) a synthesis of most relevant features into reduced number of marine provinces using the hierarchical agglomerative clustering (HAC) method. This two-step classification procedure has been successfully used in several previous studies (e.g., Saraceno et al., 2006; Leloup et al., 2007; Lachkar and Gruber, 2012).

## Self-organizing maps (SOMs)

The SOM is an unsupervised learning technique that allows the detection and visualization of the underlying structure in large- and high-dimensional data sets. To this end, SOM implements topology-preserving mapping from the higher-dimensional observation space into a lower-dimensional (here two) lattice of prototype vectors called neurons. Each neuron (prototype) represents a local summary of similar observations. The topology preservation implies that

neighboring neurons on the map represent similar observations in the input space. This is achieved thanks to neighborhood relations that connect adjacent neurons. At the end of the training, the SOM approximates the probability density function of the input data (Kohonen, 2000). A batch-training algorithm was used to train the self-organizing maps (SOM toolbox package<sup>2</sup>). Neurons were arranged on a hexagonal grid. This ensures that distance between neighboring neurons is the same in all directions. The strength of neuron connections was determined using a Gaussian neighborhood function, thus ensuring a smoother mapping and a higher generalization capability of the trained maps (Kohonen, 2000; Lachkar and Gruber, 2012). The learning rate decreases linearly with time, and the radius of the neighborhood decreases from an initial value of  $n/4$  (where  $n$  is the size of the map, i.e., number of neurons in each direction) to 1 at the end of the training.

The output of a self organizing neural network can be sensitive to the choice of additional training parameters such as the size of the map (e.g., Gonzalez et al., 1997). Yet, no standard, general principal exists for choosing these parameters (Chon, 2011; Lachkar and Gruber, 2012). To address this issue, the sensitivity of the SOM output to the neuron map size was quantified based on the average of the quantization and topological errors (Uriarte and Martin, 2005). A saturation in the reduction of the total error was observed at high numbers of neurons. The size of the neuron map was determined based on a cutoff criterion (reduction in total error after addition of further neurons < 5 %). This led to the use of a  $20 \times 20$  neuron map (Supplement, Sect. S2, Table S2 and Fig. S1 for details).

### Hierarchical ascending clustering (HAC)

Multiple scales are in interaction within and between ecosystems. With the prior knowledge of hierarchy being an important feature of ecosystems (Vichi et al., 2011), we used the hierarchical agglomerative clustering (HAC) method (Jain and Dubes, 1988) to cluster the resulting 400 neurons (Supplement Fig. S2) into the major geophysical ecoregions of the CS in a bottom-up clustering procedure. The use of hierarchical clustering here presents several advantages over other techniques of flat clustering (e.g.,  $K$  means clustering). First and foremost, HAC provides structured clustering with valuable information on the levels of similarity and relative distance between clusters. Additionally, HAC is an unsupervised classification method that requires no a priori specification of the number of classes. Finally, HAC provides a deterministic clustering insensitive to initialization (Frades and Matthiessen, 2010). An Euclidean distance metric was used for measuring dissimilarity between pairs of objects, with Ward's method applied as linkage criterion in the MAT-

LAB environment (The MathWorks, 1998). A tree-like diagram (dendrogram) was used to display the HAC clustering (Fig. S4). Cutting the hierarchical tree at different levels of similarity resulted in a different number of ecoregions.

#### 2.2.4 Number of ecoregions

The choice of the final number of clusters highly affects the quality of the clustering (Hong et al., 2011). In this study, a 10-fold cross-validation approach was conducted to determine the optimum number of ecoregions (De'ath and Fabricius, 2000). The data were divided into two unequal parts of 90 % (the training set) and 10 % (the validation set). The training data set was reduced to 400 classes ( $20 \times 20$  neurons) using the SOM (see Sect. 2.2.3 above). These 400 prototypes were further agglomerated using the HAC algorithm. Cutting the HAC dendrogram at different levels of similarity for each cross-validation iteration, a total of 140 cross-validation experiments were performed that clustered the 400 neurons into 2 to 15 ecoregions. The ecoregion of the closest neuron on the map (using the Euclidean distance, i.e. the distance measure used for training the SOM algorithm), also called best matching unit (BMU), was attributed to each observation from the validation set.

For each cross-validation experiment, the clustering error was computed as the average distance of the validation observations to the center (average) of their respective ecoregions in the training set (Sect. S3 for details). The final cross-validation error for each experiment was calculated by averaging errors for all the 10 given iterations (Table S3). The cross-validation error was minimized with 11 ecoregions (Fig. S3), which was identified as the optimal number of ecoregions. Further analysis was performed to characterize the individual ecoregions. Since we were interested in large-scale structures in the CS, we further merged ecoregions that covered less than 3 % of the CS into the statistically most-similar neighboring ecoregion. This procedure resulted in a final classification of the CS into 10 ecoregions.

We used a non-parametric, one-way analysis of variance by ranks (Kruskal–Wallis  $H$  test) to test whether ecoregions differed in their environmental conditions (Zar, 1999). A significant result of the Kruskal–Wallis  $H$  test implies that at least one ecoregion differs from all others. We then applied a multi-step a posteriori pairwise testing procedure based on studentized range statistics (Dunn's procedure) to identify which ecoregions differ significantly from others ( $P < 0.05$ ). Dunn's test is a non-parametric multiple comparison test, analogous to for example, a Tukey–Kramer test (Wheater and Cook, 2005).

#### 2.2.5 Swapping dependent variables

To estimate the effects of our choice of environmental predictor variables from the two groups of dependent variables on the results of the classification (Fig. 2, Sect. 2.2.2), several

<sup>2</sup>SomToolbox, Helsinki University of Technology, available at [www.cis.hut.fi/projects/somtoolbox](http://www.cis.hut.fi/projects/somtoolbox)

sensitivity tests were run where we exchanged one or more of the input variables with another representative from the same group (i.e., variable with a high degree of correlation; Table 3). All other parameters were kept constant, with  $20 \times 20$  number of neurons and 11 number of classes. The output of each experiment (i.e., different maps of ecoregions) was then compared with the original classification. The agreement between the two classification outputs was quantified using the Kappa index of agreement. Kappa is a chance-adjusted measure of the relative percentage of similarity between the two classification results that considers the location and value of each two correspondent pixels (Gygr and Bodtker, 2007). Possible values of Kappa range from  $-1$  to  $1$ , with  $1$  indicating perfect agreement and  $-1$  representing worse than random agreement (Sim and Wright, 2005).

## 2.3 Ecoregion validation

### 2.3.1 Biological data for ecoregion validation

To validate the geophysically based ecoregion classification with independent biological information, we used satellite chlorophyll data and available information on species distribution for a range of marine taxa. The chlorophyll *a* concentrations (Chl *a*) were obtained from ESA's GlobColour Project (<http://www.globcolour.info/>) as monthly mean standard mapped images with a spatial resolution of 4.6 km. The GlobColour data contain the merged product from three satellite sensors, namely the Sea-Viewing Wide Field-of-View Sensor (SeaWiFS), MODIS-AQUA and MERIS (Maritorena et al., 2010). The satellite-derived Chl *a* concentrations tend to be biased towards higher values in the CS (e.g., Kopelevich et al., 2004; Fendereski et al., 2010), but is unlikely to affect the relative spatial and temporal variability which is the critical property for the ecoregion classification (Kopelevich et al., 2004; Nezhlin, 2005; Thomalla et al., 2011). We produced an annual mean climatology of Chl *a* using data for the period from January 2003 through December 2010, excluding data before 2003 in order to reflect the situation after the invasion of *M. leidy*, which altered the Chl *a* pattern (Kopelevich et al., 2008).

In addition to Chl *a* concentrations, digitalized habitat maps based on in situ species observations of 36 different marine species were downloaded from [caspienvir.org](http://caspienvir.org). The observations were based on the data from different cruises, compiled and mapped by the Caspian Environment Program (CEP) in 2002, and extrapolated to the entire CS for seven of the 36 species. While this data set is limited in terms of its time span and the number of species included (36 species distribution data for periods before peak abundances of *M. leidy*; Shiganova et al., 2011), it is the only data set currently available where individual measurements of species distribution (presence-absence and biomass) cover the entire Caspian Sea. Since the raw species data was not available, we used the digitalized published habitat maps

for all available species. We separated the species for which habitat maps were available into a pelagic group and a benthic group. In shallow seas, like most parts of the CS (e.g., Ibrayev et al., 2010; Shiganova et al., 2010), the whole water column tends to be mixed, leading to only small gradients in species distribution along the vertical axis (Nybakken, 2000). However, because of their high degree of dependency on topographic features, sediment particle size, etc., (Day and Roff, 2000), benthic species were studied separately from pelagic species. Species for which sampling data were confined to a limited area of the CS were discarded from this analysis. This led to the exclusion of two phytoplankton species (*Exuviaella cordata* and *Rhizosolenia calcar-avis*), two zooplankton species (*Acartia tonsa* and *Cercopagis pengoi*), two benthic invertebrates (*Dreissena* (*Pontodreissena*) *rostriformis* and *Gammarus* (*Chaetogammarus*) *pauxillus*), two fish species (*Stenodus leucichthys leucichthys* and *Stizostedion lucioperca*) and one mammal (*Phoca* (*Pusa*) *caspiaca*). Furthermore, one phytoplankton species (*Dactyliosolen fragilissimus*, recorded as *Rhizosolenia fragilissima* in the original data set) was eliminated, because its abundance and ubiquity has drastically changed in the CS since the introduction of another diatom, *Pseudosolenia calcar-avis* in the 1930s and the jellyfish invasion in the late 20th century (Shiganova et al., 2005; Karpinsky, 2010; Shiganova, 2011). Thus, our final validation data set consisted of 25 species, including zooplankton (2 species), 9 pelagic fish species, 11 demersal fish species, and 9 benthic invertebrates (see Table S6).

Since abundance or biomass estimates were not available for all species, recorded observations were transformed into presence patterns and were subsequently used for the validation of our ecoregions (Table S6). To this end, the individual species distribution maps were mapped onto the ecoregions. For each ecoregion, two vectors representing the presence of pelagic (Table S6, zooplankton and pelagic fish) and benthic (Table S6, demersal fish and invertebrates) species were created. The presence (or absence) of each species in each of the 10 ecoregions was defined on the basis of the coordinate match between occurrence and ecoregion data, with “1” (“NaN”) defining the presence (absence) of a species in a given ecoregion. Since the lack of presence of species in an ecoregion does not necessarily mean its absence in that ecoregion (it can be due to, for example, the lack or imprecision of sampling in that area for the target species), ecoregions in which the given species were not observed were assigned “NaN” for that species (see also Sect. 4.3 for a thorough discussion on data availability and data quality). Those ecoregions for which species presence data were solely located near/on its borders were assigned “1” for that species.

### 2.3.2 Biological evaluation of the ecoregions

To test whether physically different ecoregions may host different type of communities (Bredin et al., 2001), we



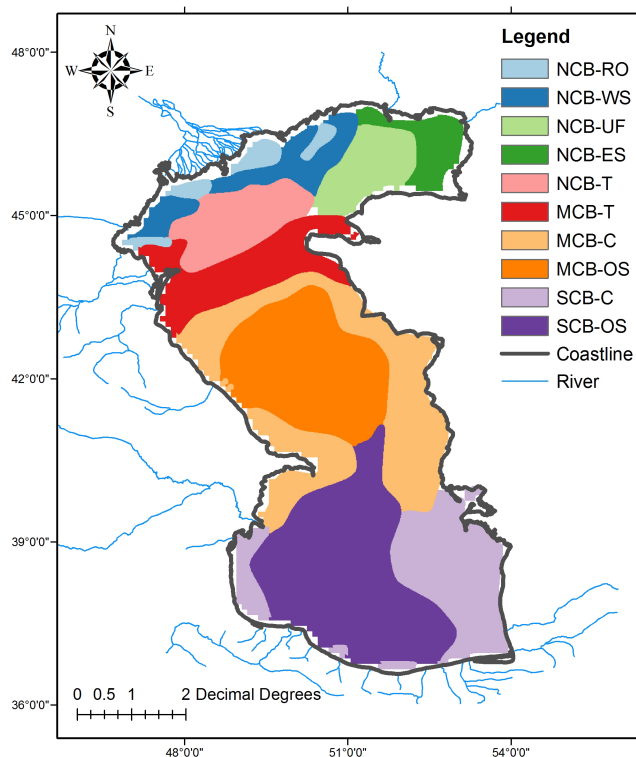
compared the chlorophyll and ecosystem composition between ecoregions. Given the different nature of the chlorophyll and species data, two different methods had to be employed. For chlorophyll, a parametric, one-way analysis of variance (ANOVA) was used to identify whether there are significant differences between the 10 ecoregions in their spatial distribution of annual mean Chl *a*. Because of the lognormal distribution of Chl *a* concentrations in the ocean (Campbell, 1995), Chl *a* data was transformed logarithmically prior to the test. When overall significant differences were observed between ecoregions, pair-wise comparisons of ecoregions were conducted using the Tukey–Kramer multiple comparison test (Kramer, 1956). Ecoregions were then analyzed in terms of the initiation, duration and amplitudes of the phytoplankton bloom. For each ecoregion, Chl *a* annual median was calculated in two steps: first, the temporal median over full-year monthly mean climatologies (2003–2011) was calculated, and then, the spatial median within the individual ecoregion was found. The initiation of the net growth phase date was defined as the period of the year when an increase in Chl *a* monthly mean climatology values (2003–2011) was observed relative to the annual median (Thomalla et al., 2011).

For the evaluation of differences in the community structure, the degree to which each pair of ecoregions were similar based on their pelagic and benthic species composition was measured using the Sørensen–Dice similarity coefficient (Sect. S4; Sørensen, 1948; Dice, 1948). This statistical measure for comparison of two samples was designed to compare binary presence–absence data. The Sørensen–Dice coefficient ranges from 0 to 1, with 0 representing no overlap and 1 showing a complete agreement between two sets of samples, A and B. The resulting Dice coefficient matrix of all the potential pair-wise combinations of the different ecoregions was projected on a dendrogram (Fig. 10; see Sect. S4 for details).

### 3 Results

#### 3.1 Ecoregions of the CS

The 10 identified geophysically based ecoregions form spatially coherent units and sub-divide the three previously defined basins (see e.g., Kopelevich et al., 2004) in a natural manner (Fig. 3 and Table 2). Using names according to their geographical characteristics, we found five of the distinguished ecoregions in the NCB (NCB River Outflows (NCB-RO), NCB Western Shelf (NCB-WS), NCB Ural Furrow (NCB-UF), NCB Easternmost Shelf (NCB-ES) and NCB Transition (NCB-T)), three in the MCB (MCB Transition (MCB-T), MCB Coastal (MCB-C) and MCB Offshore (MCB-OS)) and two in the SCB (SCB Coastal (SCB-C) and SCB Offshore (SCB-OS)). In general, ecoregions in the NCB are smaller in size than those identified in the two other



**Figure 3.** Ecoregions of the CS derived from the SOM classification algorithm with six independent clustering variables (SST, ICE, DSSS, depth, TSM and DTSM).

basins. The spatial coherence strongly supports our classification method, as the geographical location of the individual pixels was not used in any steps of the classification procedure. Spatially coherent patterns in the distribution of ecoregions reflect the gradients in physical environmental variables in the CS, which is also evident from the distribution maps of these sets of physical input variables (Fig. 4; see Sect. 3.3 for description of the spatial distribution of physical variables with regard to ecoregions).

The 10 ecoregions are clearly organized in a hierarchical manner. Indeed, a close inspection of the HAC dendrogram (Fig. S4) and the hierarchical splitting of clusters (Fig. 5) shows that ecoregions in the NCB, on the one hand, and those in the MCB and SCB, on the other hand, belong to two separate branches in the hierarchy of clusters. This indicates large contrasts in the physical and environmental conditions between the NCB and the rest of the CS, as well as a higher degree of similarity between ecoregions of the MCB and SCB. The dichotomy between the NCB on the one hand and the MCB and SCB on the other hand is further evidenced by a non-metric multidimensional scaling analysis (see Sect. S5 and Fig. S5), which highlights the underlying distances among the different ecoregions in a two-dimensional projection of the multidimensional predictor variable space.

### 3.2 Sensitivity of the classification to the choice of input physical variables

To estimate the sensitivity of our classification to the choice of representative input variables from the two groups of dependent variables (Fig. 2), several sensitivity experiments were conducted where one or more of the input variables were replaced by another representative from the same group (Table 3, see Sect. 2.2.5). The respective sets of ecoregions resulting from these sensitivity experiments were compared with the results of the original classification (Fig. 3) using the Kappa index of agreement. The Kappa score, ranging from 0.6 to 0.8 (Table 3), indicated a substantial agreement between the results of the experimental maps and the reference maps (Landis and Koch, 1977). It suggested that little variation occurred in the classification output when dependent variables from the same cluster were swapped, and that while the relative size of individual ecoregions depended on this choice, the general pattern remained robust (Sim and Wright, 2005). However, the percentage of agreement varied depending on which variables had been swapped. In most cases, the variation in the Kappa index was a direct function of degree of correlation between the exchanged dependent variables.

### 3.3 Geophysical differences between ecoregions

The maps of the environmental conditions depicted in Fig. 4 reveal how the gradients in the different physical variables contributed to the classification. This is most obvious with bathymetry, which separated the offshore MCB and SCB from the nearshore regions (Fig. 4). The mean conditions of the different variables differ substantially between the 10 regions (Fig. 6, Table S4) as revealed by a Kruskal–Wallis test ( $P < 0.05$ ). Figure 6 depicts the results of the 45 pairwise comparisons on six independent biogeophysical variables for all ecoregions (Dunn's test; see Sect. 2.2.4). For each variable, the same letters were assigned to those ecoregions that did not show significant differences in a given variable ( $P < 0.05$ ). Overall, we failed to reject 7 of the 45 pairwise comparisons for SST, 10 for DSSS, 12 for depth, 5 for TSM, 7 for DTSM and 12 for ICE ( $P > 0.05$ ). Except for a few cases, significant differences between ecoregions in the NCB and those in the MCB and SCB were detected for all six environmental variables. The physical characteristics of all ecoregions are discussed below.

#### 3.3.1 Ecoregions in the NCB (NCB-UF, NCB-W, NCB-T, NCB-E and NCB-RO)

Ecoregions located in the NCB are characterized by lower SST and higher DSSS and TSM than ecoregions located in the MCB and SCB (Fig. 6). Despite the differences between ecoregions in the NCB in almost all of the environmental variables (see below), these ecoregions are more similar to each other than to ecoregions located in the MCB and SCB

(Fig. 6 and Table S4). All ecoregions located in the NCB are covered with sea ice during the winter season. The percentage of ice cover in NCB-T is lowest compared to other ecoregions in the NCB ( $P < 0.05$ ; Fig. 6f and Table S4). The other ecoregions in the NCB do not show significant differences in this variable ( $P < 0.05$ ; Fig. 6f).

The NCB does not include any open ocean ecoregions (water depth  $< 100$  m) as the entire NCB consists of continental shelf waters with an average water depth of only 5 m (Barale, 2008). In the NCB, SST is highest in the southern part (NCB-T) and lowest in the eastern part (NCB-ES; Table S4). Due to the spring floods from the Volga and to a lesser extent from the Ural River (Kosarev, 2005; Kara et al., 2010), high negative values for DSSS are observed in all ecoregions of the NCB (Fig. 6b). In NCB-WS, NCB-T and NCB-RO in the western part of the NCB, DSSS is significantly higher than those in the eastern part of this basin (NCB-UF and NCB-ES,  $P < 0.05$ ; Fig. 6 and Table S4). Seasonal variability in SSS decreases southwards in the western part of the NCB (Table S4). In general, TSM in the shallow NCB is significantly higher than in the two other basins ( $P < 0.05$ ; Fig. 6d and Table S4). Similar to SSS, distance from major river outflows is important in determining the patterns in TSM and its seasonal variation. NCB-WS, NCB-ES and NCB-RO show a higher mean TSM than other ecoregions in this basin, and NCB-WS has much smaller seasonal variations in TSM than NCB-ES and NCB-RO (Fig. 6e and Table S4).

#### 3.3.2 Ecoregions in the MCB (MCB-OS, MCB-C and MCB-T) and SCB (SCB-OS and SCB-C)

Ecoregions located in the MCB and SCB are characterized by higher SST but lower DSSS and TSM than ecoregions located in the NCB (Fig. 6 and Table S4). Ecoregions in the MCB and SCB differ significantly between each other in all of the environmental predictor variables, except for ICE ( $P < 0.05$ ; see Fig. 6). Few of the ecoregions in the SCB and MCB are covered by sea ice (Fig. 6f); the exception is MCB-T in the northern part of the MCB, which is (partly) ice covered during winter (Figs. 4f, 6f and Table S4).

The CS primarily consists of continental shelf waters (62 %; Fig. 6c; Ibrayev et al., 2010); thus, only the MCB-OS and SCB-OS regions are characterized by water depths equal or larger than 100 m, thus being significantly deeper than other ecoregions located in the continental shelf areas of the CS ( $P < 0.05$ ; Fig. 6c and Table S4). Ecoregions in the SCB feature significantly higher annual mean SST than other ecoregions in the CS (Fig. 6a and Table S4). This is because the CS has a large north–south gradient in SST during winter, while in summer there are no major differences in the SST throughout the CS (Ibrayev et al., 2010). In general, in the ecoregions of the MCB and SCB, salinity can be as high as 10–13 PSU (practical salinity units; Tuzhilkin et al., 2005), showing significantly lower seasonal variability



**Table 2.** Ecoregions and their characteristics in terms of the environmental variables.

Ecoregion	Acronyms	Characteristics
NCB River Outflows	NCB-RO	Shallow, dominated by the Volga River outflows, high in SSS seasonality, high in TSM and its seasonality, covered in ice in winter
NCB Western shelf	NCB-WS	Shallow, low in SSS and high in its seasonality, high in TSM, covered in ice in winter
NCB Ural Furrow	NCB-UF	Shallow, low in SST, low in SSS and its seasonality, covered in ice in winter
NCB Easternmost shelf	NCB-ES	Shallow, low in SST, low in DSSS, high in TSM and its seasonality, highly ice covered in winter
NCB Transition	NCB-T	Shallow, high in SSS seasonality, partly ice covered in winter
MCB Transition	MCB-T	Medium SST, medium DSSS, medium TSM and DTSM, partly ice covered in winter
MCB Coastal	MCB-C	Continental shelf area, High in SSS and low in DSSS, low in TSM and its seasonality
MCB Offshore	MCB-OS	Deep, low in SST, low in TSM and its seasonality
SCB Coastal	SCB-C	Continental shelf area, High in SST and SSS, low in TSM and its seasonality
SCB Offshore	SCB-OS	Deep, high in SST, low in TSM and its seasonality

**Table 3.** Kappa index of agreement (0–1) for comparisons of the classification outputs after swapping dependent input variables (underlined) and the final classification output used in this study as the reference.

Input variables	Kappa index of agreement
SST, ICE, DSSS, depth, TSM, DTSM	1.0000
SST, <u>SSS</u> , DSSS, depth, TSM, DTSM	0.6535
SST, <u>PAR</u> , DSSS, depth, TSM, DTSM	0.6018
SST, <u>DSST</u> , DSSS, depth, TSM, DTSM	0.6350
<u>WSP3</u> , ICE, DSSS, depth, TSM, DTSM	0.8010
<u>DPAR</u> , ICE, DSSS, depth, TSM, DTSM	0.6332
<u>WSP3</u> , <u>SSS</u> , DSSS, depth, TSM, DTSM	0.7386

than in the NCB (DSSS; Fig. 6b and Table S4). The elevated SSS values in summer in almost all ecoregions in these two basins (except for MCB-T; Fig. 6b and Table S4) can be attributed to their distance from major river outflows and a higher rate of evaporation in summer compared to winter. In the MCB and SCB, TSM in the continental shelf regions (MCB-C and SCB-C) is higher than in the deeper basins (MCB-OS and SCB-OS; Fig. 6d and Table S4). However, in all of these ecoregions, TSM is higher in winter than in summer (Fig. 6e), with the highest variability in MCB-C.

### 3.4 Biological validation of ecoregions

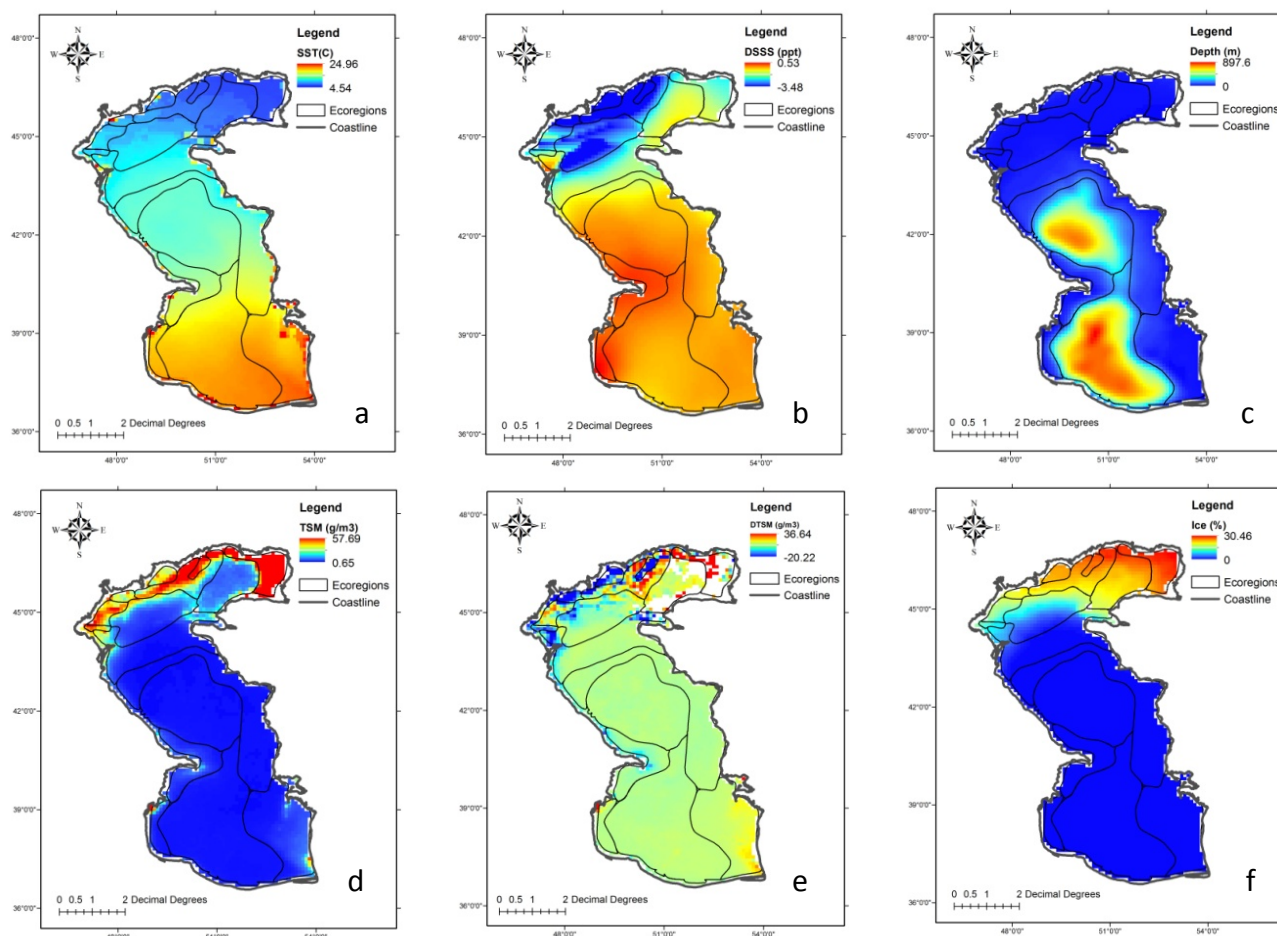
To test the biological significance of ecoregions, we compared climatological annual mean (2003–2011; Fig. 7) and monthly mean Chl *a* of the different ecoregions. We also at-

tempted a first qualitative evaluation of differences in ecosystem composition based on recorded presence–absence patterns of 25 different species of plankton, fish and invertebrate species.

#### 3.4.1 Annual mean Chl *a* distribution in each ecoregion

Based on a one-way ANOVA test, the 10 ecoregions are significantly different from one another in terms of their annual mean Chl *a* distributions ( $P < 0.05$ ). This was supported by a Tukey–Kramer multiple comparison test (depicted in Fig. 8).

In general, only 4 of the 45 pair-wise Chl *a* comparisons do not show significant differences in annual mean Chl *a* concentration ( $P > 0.05$ ; Fig. 8). Results show significant differences in the annual mean Chl *a* concentration between ecoregions of the NCB and those of the MCB and SCB (Fig. 8). The highest concentrations of Chl *a* are observed in the NCB, specifically in the two ecoregions at the mouth of the Volga River (NCB-WS and NCB-RO; Fig. 8). This river supplies a large amount of nutrients via 80 % of the total river runoff to the CS (Tuzhilkin et al., 2005). On the west side of the NCB, surface Chl *a* concentrations gradually decrease southward. A pronounced west–east and north–south decrease in Chl *a* concentrations is observed with increasing distance from the Volga River (Table S5). In the MCB and SCB, ecoregions located on the continental shelf (i.e., MCB-C and SCB-C) do not show significant differences in their annual mean Chl *a* distributions ( $P < 0.05$ ; Fig. 8). Ecoregions comprising the deeper basins of the MCB and SCB (i.e., MCB-OS and SCB-OS) are statistically different from each other ( $P < 0.05$ ; Fig. 8) and MCB-OS has lower annual mean Chl *a* concentrations than SCB-OS (Table S5).



**Figure 4.** Independent input variables for classification the CS: (a) annual mean climatology SST ( $^{\circ}\text{C}$ ); (b) SSS seasonal amplitudes (ppt); (c) depth (m); (d) annual mean climatology TSM ( $\text{g m}^{-3}$ ); (e) TSM seasonal amplitudes ( $\text{g m}^{-3}$ ); (f) ice cover percentage (%). Black contour lines superimposed on data represent the borders of the ecoregions identified in this study (see Sect. 3.1).

### 3.4.2 Seasonal variations in Chl *a* concentration in ecoregions

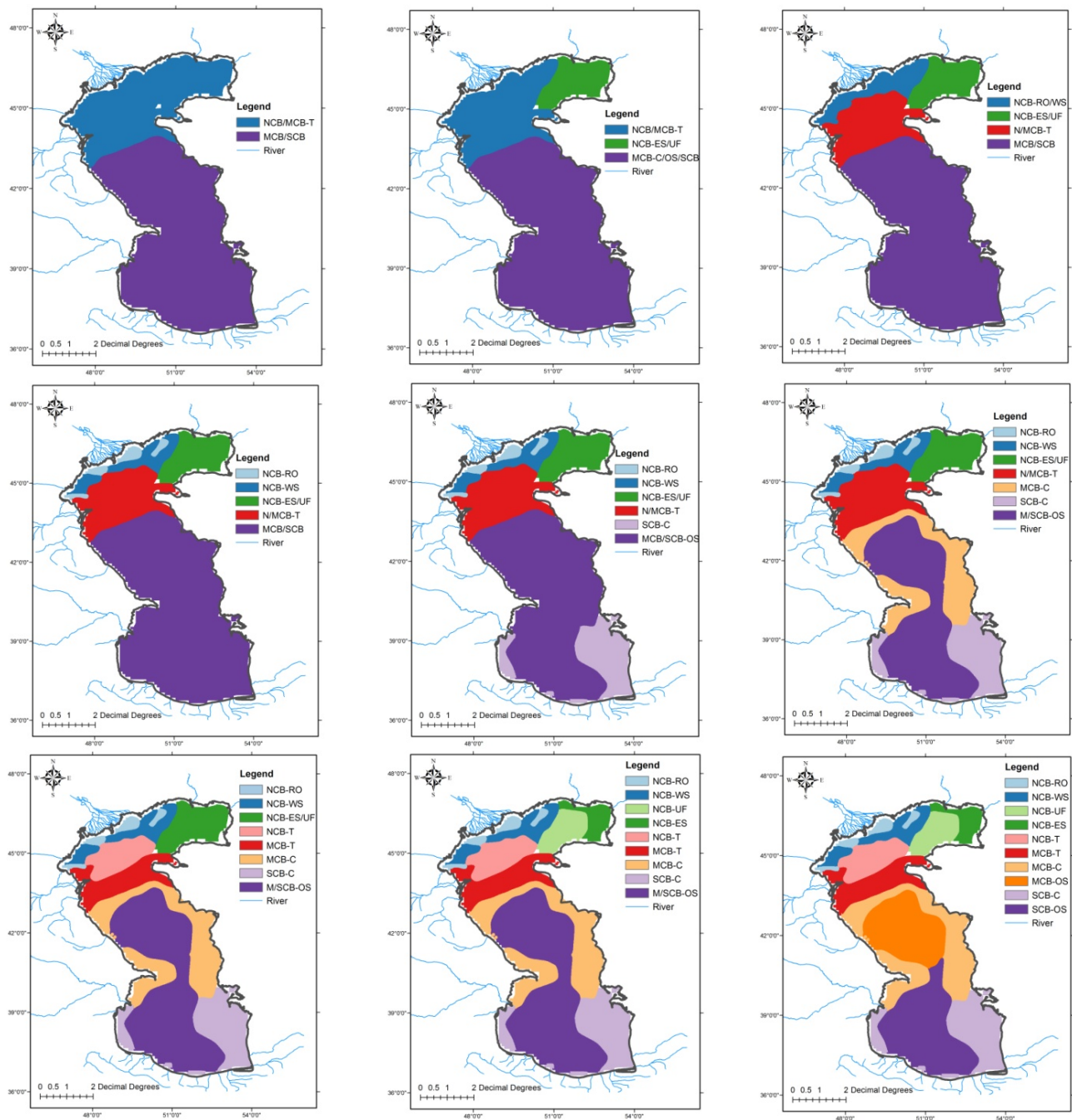
Ecoregions were then analyzed in terms of phytoplankton phenology. Two distinct periods are observed throughout the CS: a period of the net growth (the “bloom phase”) in summer and fall and a period with low Chl *a* concentrations during winter and spring (Fig. 9).

Considerable differences in the date of bloom initiation, bloom duration and amplitude are found between ecoregions (Fig. 9). In the NCB, the bloom starts between May (in NCB-WS; Fig. 9b) and July (NCB-RO and NCB-UF; Fig. 9a and c, respectively). In all of the ecoregions in the NCB, Chl *a* reaches peak concentrations in July, the exception being NCB-ES, where phytoplankton biomass reaches its maximum concentration in August (Fig. 9d and Table S5). The duration of the phytoplankton bloom in the NCB differs between ecoregions, lasting from 4 months in NCB-WS (Fig. 9b) and NCB-ES (Fig. 9d) to 6 months in NCB-UF (Fig. 9c, see Table S5). In the continental shelf and open

waters of the SCB (SCB-C and SCB-OS; Fig. 9i and j, respectively) and MCB-T (Fig. 9f), the bloom initiates in July, while in the MCB (MCB-C and MCB-OS; Fig. 9g and h, respectively), the bloom period starts with a delay of 1 month, in August. Blooms in the MCB and SCB last from 5 months in MCB-OS (Fig. 9h) to 8 months in MCB-C (Fig. 9g). Within all ecoregions in the MCB and SCB, Chl *a* shows its maximum amplitude in October, the exception being SCB-C, where the peak is observed in August.

### 3.4.3 Differences in ecosystem composition between ecoregions

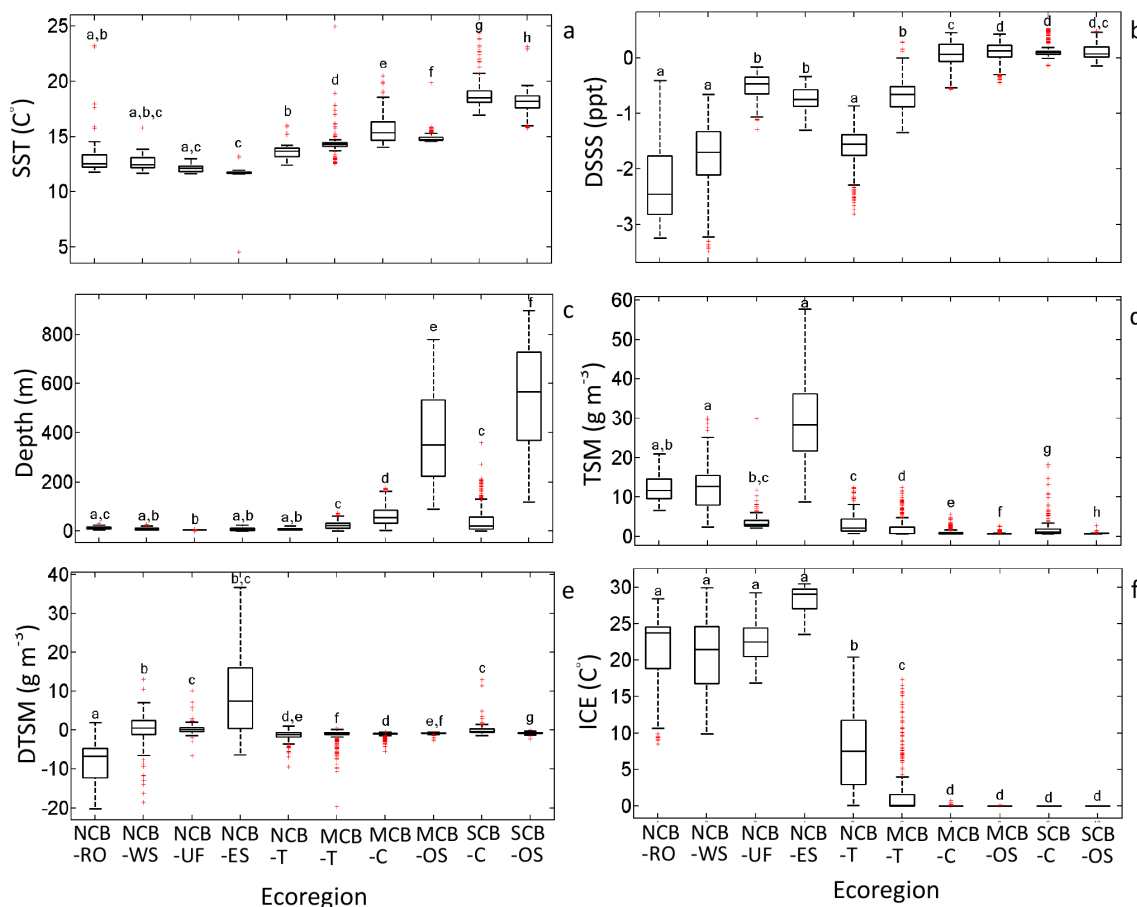
Based on the species data discussed in Sect. 2.3.1, we investigated differences in species composition between ecoregions. Almost all of the species under study are found in MCB-T, and most of the species are observed in NCB-T. These ecoregions act as a transition zone between the NCB and the MCB (Araujo, 2002).



**Figure 5.** Hierarchical sequence (from upper left, levels 1 to 9) of ecoregions formation (2 to 10 ecoregions).

The agreement between each pair of ecoregions based on their pelagic and benthic species composition was measured using the Dice similarity coefficient (see Methods, Sect. 2.3.2; Sørensen, 1948; Dice, 1948). Figure 10 shows the dendrograms used to visualize the degree of similarity/dissimilarity between ecoregions based on their pelagic and benthic species composition data.

For pelagic species, with the exception of NCB-RO, ecoregions can be grouped into two major groups, primarily composed by the NCB ecoregions and the rest of the CS ecoregions (Fig. 10a). This clustering of ecoregions according to their similarity in species composition reflects the results of our physical classification, where ecoregions from the MCB and SCB were more similar to each other in their physical attributes than to those from the NCB (Fig. S5). The planktonic



**Figure 6.** Box plots of ecoregions (x axis) against the physical variables (y axis) (a: SST, b: DSSS, c: depth, d: TSM, e: DTSM and f: ICE). The line in the middle of each box represents the ecoregion median. The top and bottom limit of each box are the 25th and 75th percentiles, respectively. The lines extending above and below each box, i.e., whiskers, represent the full range of non-outlier observations for each variable beyond the quartile range. Red pluses (+) are observations that have been classified as outliers. The results of Dunn's multiple comparison test between different ecoregions (1–10) for each physical variable are shown in figures (a) through (f). Identical letters depict ecoregions without statistically significant differences between the climatological input variables at the 95 % level ( $P < 0.05$ ), whereas different letters depict significant differences between ecoregions.

species included in this study, i.e., *E. grimmi* and *M. leidyi*, and the brackish water fish subspecies, *Salmo trutta caspius* (Caspian brown trout) and *Clupeonella engrauliformis*, were not observed in any of ecoregions in the NCB, except for NCB-T (Table S6). The presence of these species is known to be restricted by low salinity (CEP, 2002), which is a prominent feature of the NCB (Kara et al., 2010). In ecoregions in the MCB, SCB and NCB-T, a higher number of species were recorded than in the other ecoregions of the CS. All planktonic species were observed in these ecoregions, but the pelagic fish species included in this study were found simultaneously only in NCB-T, MCB-T, and MCB-C (Table S6).

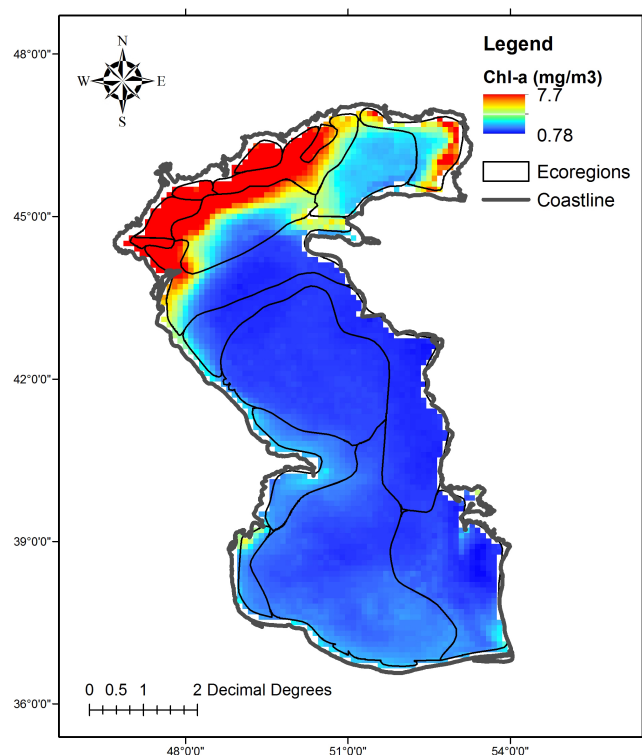
Based on the distribution of benthic species, ecoregions could be grouped into three major groups: NCB-ES and NCB-RO, MCB-OS and SCB-OS, and ecoregions from all the three basins (i.e., NCB-UF, NCB-WS, NCB-T, MCB-T, MCB-C and SCB-C; Fig. 10b). These results only partially

reflect the physical classification, where a strong separation was observed between ecoregions in the NCB and those in the MCB and SCB (Fig. 6). This suggests that controls on the distribution of benthic species might depend on properties not included in our analysis, such as sediments size and type (Day and Roff, 2000). Depth seems to be one of the key environmental factors determining benthic species composition (CEP, 2002), thus limiting their distribution to the shallower areas of the MCB and SCB.

#### 4 Discussion

Ecoregions are geographic regions with specific biotic and abiotic attributes (Day and Roff, 2000). In this study, the CS was partitioned into 10 distinct ecoregions using annual mean climatologies and seasonal amplitudes of physical



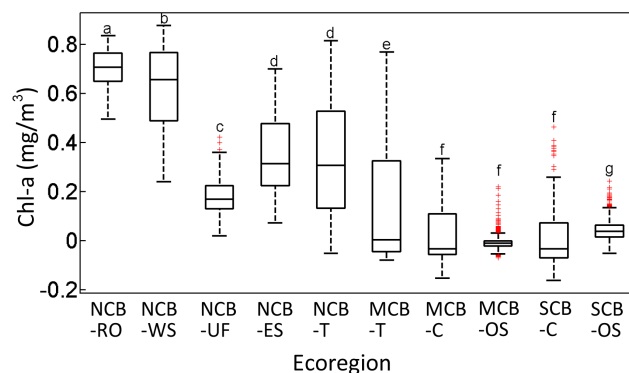


**Figure 7.** Distribution of Chl *a* concentration ( $\text{mg m}^{-3}$ ) in the CS. Black contour lines superimposed on data represent the borders of the ecoregions identified in this study (see Sect. 3.1).

oceanographic variables together with bathymetry data. Although the geographical coordinates of individual grid cells have not been included in the classification, the partitioning resulted in coherent, spatially uniform ecoregions. Ecoregions vary in size, with smaller ecoregions located in the NCB, which shows a higher spatial variability in the input variables. The MCB and SCB are relatively uniform, containing three and two large ecoregions, respectively. In agreement with our findings, a greater level of biodiversity has also been reported for the northern part of the CS in comparison to the two other basins (UNEP, 2006).

#### 4.1 Comparison between marine and adjacent terrestrial ecosystems

The spatial coherence of our ecoregions is due to natural gradients in the physical variables used for our classification, and the identified patterns are also reflected in the climate regions of the surrounding terrestrial biosphere. Our marine ecoregions are geographically consistent with their surrounding terrestrial ecoregions (Bailey, 1996). This can be explained by the importance of climate as a major driving force for both ecosystems. Moderately warm and arid climates surround the western and eastern parts of the MCB, respectively (Kosarev, 2005). Ecoregions in the MCB (MCB-T, MCB-C and MCB-OS) are adjacent to tropical/subtropical

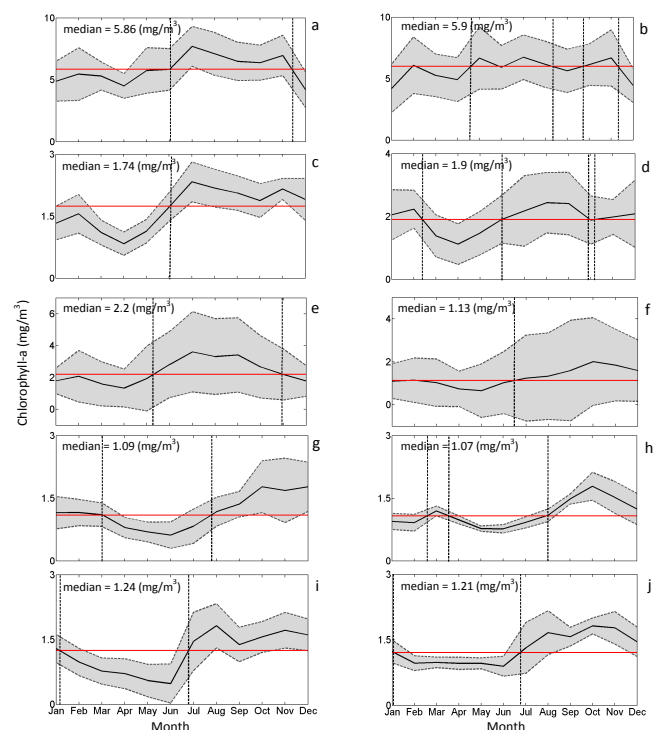


**Figure 8.** Box plot of ecoregions (x axis) against Chl *a* concentrations (y axis). The ecoregion median (line in the middle of each box) and the 25th and 75th percentiles (tops and bottoms of each box, respectively) are shown as in Fig. 6. The two lines extending vertically from the box (whiskers) depict the variability of non-outlier observations beyond the quartiles. Outliers are depicted by red pluses (+). The results of Tukey–Kramer multiple comparison test between different ecoregions (1–10) for Chl *a* concentrations are shown by differences in letters as in Fig. 6, with different letters above the boxes indicating significant differences between ecoregions at  $P < 0.05$ .

deserts and prairie regime mountains in the west and temperate deserts in the east (Bailey, 1996). The SCB (SCB-C and SCB-OS) is surrounded by Bailey's subtropical regime mountains to the south, and this area is dominated by a subtropical climate zone, characterized by warm and humid conditions (Kosarev, 2005). SCB-C is bordered by desert climate (Kosarev, 2005) and tropical/subtropical steppe to the east (Bailey, 1996).

However, there are also a few areas where the location of terrestrial ecoregions does not match that of the adjacent marine ecoregions. Differences in extent and location between marine and terrestrial ecosystems can be explained by geochemical and regional factors other than the climatic ones that affect the marine environment, such as the distance from river inflows, water depth and currents. For example, terrestrial regions are fairly uniform around the NCB (Bailey's temperate deserts), where we found the greatest environmental diversity in the marine ecosystem. The distance from the Volga and Ural rivers (with 80 and 15 % of total riverine runoff to the CS, respectively; Kosarev, 2005) is important in shaping the NCB's ecoregions. In the MCB and SCB, water depth separates deep ecoregions (MCB-OS and SCB-OS) from shallow ecoregions in continental shelf waters of these basins (MCB-T, MCB-C, and SCB-C). Cyclonic gyres in the MCB and SCB and a southward current along the western coast (Ibrayev et al., 2010) are important factors in further structuring MCB-C and SCB-C, while differences in climate between the eastern and western parts of both MCB-C and SCB-C were not captured in the current classification.

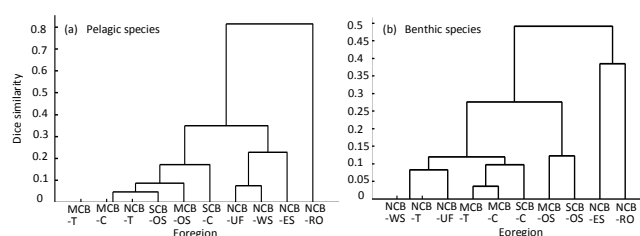




**Figure 9.** Seasonal cycle of Chl *a* concentrations from January 2003 to December 2010 in the 10 ecoregions. Panels (a) NCB-RO, (b) NCB-WS, (c) NCB-UF, (d) NCB-ES, (e) NCB-T, (f) MCB-T, (g) MCB-C, (h) MCB-OS, (i) SCB-C, and (j) SCB-OS. The red line represents the median of annual mean climatological Chl *a* in each ecoregion. The period when Chl *a* values are above the median was defined as the bloom phase. The dashed lines depict the duration of bloom in each ecoregion. The dashed blue line represents the spatial standard deviation of Chl *a* in each ecoregion. Grey filling denotes the spatial variability within an ecoregion.

## 4.2 Biological validation of CS ecoregions

Chl *a* is one of the few biological parameters that can currently be detected by satellite observations. The overall agreement between spatial distribution of Chl *a* in the CS and patterns captured in our classification indicates a biological significance of our ecoregions in terms of phytoplankton biomass and productivity (GREGG and BODTKER, 2007). We further detected significant differences in the seasonal variability in Chl *a* between ecoregions (date of bloom initiation, bloom duration and amplitude). Such differences in Chl *a* seasonality have previously been documented for the CS on the basin scale (e.g., KOPELEVICH et al., 2004; NEZLIN, 2005). The spring phytoplankton bloom in the NCB can be explained by the increase in temperature and light in April and May, in combination with the inflow of nutrient-rich waters after the ice melt, and spring floods of the Volga River between May and June (Kosarev, 2005). In the MCB and SCB, the increase in phytoplankton biomass during summer, i.e., when a deep thermocline limits phytoplankton growth in



**Figure 10.** Dendrogram of the similarity between ecoregions based on the available (a) pelagic and (b) benthic species composition data. Labels along the *x* axis represent the ecoregions.

these regions (Tuzhilkin and Kosarev, 2005) has previously been explained by the increase in the biomass of the invasive jellyfish, *M. leidyi* with increasing SST (Kideys et al., 2008; see Shiganova et al., 2004 for a spatiotemporal distribution map of *M. leidyi* in the CS). This jellyfish causes phytoplankton biomass to increase during summer through a trophic cascade effect, with *M. leidyi* releasing phytoplankton from the grazing pressure of zooplankton (Kideys et al., 2008). Conversely, the decrease in jellyfish biomass during winter relieves zooplankton from its intensive grazing pressure, resulting in low phytoplankton biomass.

Since the invasion of this jellyfish, first detected in the CS in 1999, anomalous algal blooms (AABs) have been a common feature in satellite-based Chl *a* images in the late summer and the early fall in the SCB (Kopelevich et al., 2004, 2008; Nezlin, 2005). Although we have not used any data of the jellyfish distribution in our classification, the area where AABs are frequently reported has been separated in our classification (SCB-OS). This may indicate that the physical forcing used for defining our ecoregions also controls jellyfish distribution patterns.

Hypothesizing that ecologically different regions support different species assemblages, we used a very simple data set to show that individual ecoregions exhibit differences in species composition. Despite the inevitable deficiency in our species data (see Sects. 2.3.1 and 4.3 below), our study gives a first indication that environmental (bottom-up) factors may influence species distribution and community structure patterns in the CS. Other studies in different ocean regions, such as those presented by Bredin et al. (2001) in the Bay of Fundy and Verfaillie et al. (2009) in the Belgian part of the North Sea, have demonstrated similar results. The observed differences in community structure provide some first evidence for the value of our physical classification approach for application in ecosystem management and conservation, especially in the areas where biological data are scarce.

## 4.3 Caveats of our approach

In analogy to persistent large-scale ecological patterns in terrestrial ecosystems (Bailey, 1996), we hypothesized that there are persistent ecological patterns in the oceans that can

be summarized in a classification based on annual mean and seasonal climatologies. Our static classification method, similarly to the approach presented by Longhurst (1998) for his classification of the pelagic realm of the World Ocean, has been criticized for its inability to reflect the highly dynamic nature of the oceans (Hardman-Mountford et al., 2008). Dynamic methods for delineation of ecological units in the oceans have been developed to address this issue (e.g., Devred et al., 2007; Gregr and Bodtker, 2007; Kavanaugh et al., 2013). The classification approach presented in this paper provides a description of the large-scale patterns of the CS ecosystem, which may be useful for many administrative purposes, such as longer-term monitoring and policy development (Spalding et al., 2012), management reporting and socioeconomic statistics (Hoepffner and Dowell, 2005). However, for shorter-term and local applications, seasonal or even realtime classification methods, like those presented by Devred et al. (2007) and Gregr and Bodtker (2007), may result in a more detailed characterization.

We developed an objective classification approach in which physical input variables were selected based on statistical analyses. To be able to capture as many potential environmental patterns as possible (Verfaillie et al., 2009), we initially compiled a wide range of abiotic variables, known to be important for phytoplankton growth (Fig. 2). Finally, we used an independent set of these variables for ecological partitioning of the CS. Despite the objective criteria applied here for the selection of input variables (see Sect. 2.2.2), we inevitably had to make subjective choices due to methodological constraints: the desired level of independency between input variables had to be specified, and representatives of each cluster/set of dependent variables had to be selected. In order to test the sensitivity of our results to the choice between dependent variables, we performed several sensitivity studies. The tests showed minor effects of this subjective selection on the final classification output of the CS (see Sect. 3.2). Investigating the sensitivity of classification to different levels of independence between environmental input variables was beyond the scope of the current study and should be considered in future studies.

In this study, we mainly used remotely sensed physical and biogeochemical input variables for our classification. In addition to the factors investigated here, other factors have been shown to exert a significant influence on the CS ecosystem composition but could not be included in this analysis. These factors include human activities, such as pollution and overfishing (Kopelevich et al., 2004; Zonn, 2005), and other endogenous and exogenous natural factors, such as climate change, sea level fluctuations, mud volcanoes (Zonn, 2005) and especially the invasion of *M. leidyi* and other top-down effects (Shiganova et al., 2001). Since Chl *a* is a proxy for the cumulative effects of both bottom-up and top-down controls, a classification of the CS based on Chl *a* only may yield ecoregions with a higher biological relevance than those obtained here based on physical parameters (Hardman-

Mountford et al., 2008). Since we were interested in mechanistic relationships between physical forcing and biological response, we did not use Chl *a* as an input variable here for the classification.

Our ecoregion validation in terms of their biological composition confirmed to some extent the validity of the assumption of bottom-up control on species distribution patterns and ecosystem structure. However, due to the limited availability of in situ data, our biological validation of the ecoregions remains qualitative at best. The data we used comprises only a few species and predates 2002, thus not allowing us to examine the effects of recent changes in community composition due to the effects of invasive species and food-web interactions between native and non-native species (Shiganova et al., 2011), pollution and other anthropogenic and natural influences. To our knowledge, the species distribution data used here is the only currently available set of habitat maps that covers the entire CS. A comprehensive data synthesis for relevant marine organisms inhabiting the CS, along with in situ Chl *a* concentrations is urgently required (e.g., as in Buitenhuis et al., 2013) and would be essential to further validate ecoregions based on the classification of physical variables. Such a compilation of data would be required in order to quantify the relative importance of invasive species as well as top-down and bottom-up factors for ecosystem structure and functioning in the CS. While our current approach does not incorporate top-down effects (i.e., grazing rates and relationships between predator and prey biomass), climate is likely to drive at least some aspects of lower trophic level dynamics (Day and Roff, 2000; Platt et al., 2005). The comparison of our ecoregions with remotely sensed plankton functional groups based on satellite observations (Raitos et al., 2008; Zwirgmaier et al., 2008; Brewin et al., 2010; Hirata et al., 2011; Alvain et al., 2012) would be of particular interest, as few primary producers could be included in our validation due to the lack or poor spatial coverage of in situ data. However, none of the algorithms used to derive phytoplankton composition from satellite imagery have been validated in the CS. At present, the characterization of zooplankton and higher trophic level biomass and distribution patterns still relies on in situ measurements in routine monitoring programs, which are both costly, time-consuming and provide limited spatial coverage. Thus, the quantitative examination of top-down effects would profit from further systematic long-term ecosystem surveys.

## 5 Conclusions

In this paper, the Caspian Sea was partitioned into 10 distinct ecoregions with similar annual mean climatologies and seasonal amplitudes of physical oceanographic variables using a neural network approach. The biological relevance of these ecoregions was verified using long-term satellite-derived Chl *a* concentrations and a limited set of species

distribution data. The results of the current research can provide policy makers and ecosystem managers with a general picture of the different major ecosystems of the CS surface waters on the annual scale. Researchers may benefit from these results in applications ranging from ecosystem conservation and in the definition of marine protected areas to sampling area selection and/or ecosystem modeling. The approach developed in this paper is flexible in terms of input variables and spatial and temporal resolution, as well as in terms of the extent of the study area and the set of observational data records. Thus, the method can be employed using updated data sets for improving the ecological classification of the CS, but it can also be applied to study finer spatiotemporal dynamics such as interannual, seasonal or diel dynamics, or even to monitor the CS ecosystem in realtime.

## Appendix A

Table A1. Acronyms.

CS	Caspian Sea
SCB	South Caspian Basin
MCB	Middle Caspian Basin
NCB	North Caspian Basin
SCB-C	South Caspian Basin Coastal
SCB-OS	South Caspian Basin Offshore
MCB-C	Middle Caspian Basin Coastal
MCB-OS	Middle Caspian Basin Offshore
MCB-T	Middle Caspian Basin Transition
NCB-T	North Caspian Basin Transition
NCB-RO	North Caspian Basin River Outflows
NCB-WS	North Caspian Basin Western Shelf
NCB-ES	North Caspian Basin Easternmost Shelf
NCB-UF	North Caspian Basin Ural Furrow

**The Supplement related to this article is available online at doi:10.5194/bg-11-6451-2014-supplement.**

**Acknowledgements.** F. Fendereski thanks the Iranian Ministry of Science and the Environmental Physics Group at the Institute for Biogeochemistry and Pollutant Dynamics at the Swiss Federal Institute of Technology for the funding for this project. M. Vogt, Z. Lachkar and N. Gruber acknowledge funding from ETH Zurich. F. Fendereski greatly thanks Trevor Platt for providing the possibility of working with ETH Zurich. We would like to thank the Environmental Physics Group of ETH Zurich, especially Kay Steinkamp, for their help with data processing in MATLAB.

Edited by: C. Klaas

## References

- Alvain, S., Loisel, H., and Dessailly, D.: Theoretical analysis of ocean color radiances anomalies and implications for phytoplankton groups detection in case I waters, *Opt. Express*, 20, 1070–1083, 2012.
- Araujo, M. B.: Biodiversity hotspots and zones of ecological transition, *Conserv Biol.*, 16, 1662–1663, 2002.
- Bailey, R. G. (Ed.): *Ecosystem geography*, Springer, New York, 1996.
- Bakun, A. and Parrish, H. R.: Comparative studies of coastal pelagic fish reproductive habitats: the anchovy (*Engraulis anchoita*) of the southern Atlantic, *ICES J. Mar. Sci.*, 48, 343–361, 1991.
- Barale, V.: The European and Marginal Seas: An Overview, in: *Remote Sensing of the European Sea*, edited by: Barale, V. and Gade, M., Springer, the Netherlands, 3–22, 2008.
- Bredin, K. A., Gerriets, S. H., and Van Guelpen, L.: Distribution of rare, endangered and keystone marine vertebrate species in Bay of Fundy Seascapes, Report to the Gulf of Maine Council on the Marine Environment, Atlantic Canada Conservation Data Centre, 73 pp., 2001.
- Brewin, R. J. W., Lavender, S. J., Hardman-Mountford, N. J., and Hirata, T.: A spectral response approach for detecting dominant phytoplankton size class from satellite remote sensing, *Acta Oceanol. Sin.*, 29, 14–32, 2010.
- Buitenhuis, E. T., Vogt, M., Moriarty, R., Bednaršek, N., Doney, S. C., Leblanc, K., Le Quéré, C., Luo, Y.-W., O'Brien, C., O'Brien, T., Peloquin, J., Schiebel, R., and Swan, C.: MAREDAT: towards a world atlas of MARine Ecosystem DATA, *Earth Syst. Sci. Data*, 5, 227–239, doi:10.5194/essd-5-227-2013, 2013.
- Campbell, J. W.: The lognormal distribution as a model for bio-optical variability in the sea, *J. Geophys. Res.*, 100, 13237–13254, 1995.
- CEP: Belyaeva, V. N. (Ed.): *Caspian biodiversity database*, Caspian Environment Programme, Baku, Azerbaijan, 2002.
- Chon, T.: Self-Organizing Maps applied to ecological sciences, *Ecol. Inform.*, 6, 50–61, 2011.
- Day, J. and Roff, J. C.: Planning for representative marine protected areas: A framework for Canada's oceans, Report prepared for World Wildlife Fund Canada, Toronto, 148 pp., 2000.
- De'ath, G. and Fabricius, K. E.: Classification and regression trees: A powerful yet simple technique for ecological data analysis, *Ecology*, 81, 3178–3192, 2000.
- Devred, E., Sathyendranath, S., and Platt, T.: Delineation of ecological provinces using ocean colour radiometry, *Mar. Ecol.-Prog. Ser.*, 346, 1–13, 2007.
- Dice, L. R.: Measures of the amount of ecologic association between species, *Ecology*, 26, 297–302, 1945.
- Dietrich, G. (Ed.): *General Oceanography: An Introduction*. John Wiley and Sons Ltd., New York, 1963.
- Ecosystem Principles Advisory Panel (EPAP): *Ecosystem based management: A report to congress by the Ecosystem Principles Advisory Panel*, 54 pp., 1996.
- Fendereski, F., Salmanmahiny, A., Hosseini, S. A., and Fazli, H.: The First Attempt to Test Correlation between MODIS Chlorophyll *a* Concentrations and In Situ Measurements in The Southern Caspian Sea, in: *Proceedings of the 4th International Symposium on Oceans from Space*, European Commission EUR 24324 EN, Venice, Italy, 26–30 April 2010, 89–90, 2010.
- Frades, I. and Matthiessen, R.: Overview on Techniques in Cluster Analysis, in: *Bioinformatics Methods in Clinical Research, Methods in Molecular Biology*, edited by: Matthiessen, R., Humana Press, New York, USA, 593, 81–107, 2010.
- Ginzburg, A. I., Kostianoy, A. G., and Sheremet, N. A.: Sea Surface Temperature Variability, in: *The Caspian Sea environment: The Handbook of Environmental Chemistry*, edited by: Kosarev, A. N. and Kostianoy, A. G., Springer, Berlin, 59–81, 2005.
- Gonzalez, A. I., Grana, M., Anjou, A. D., and Albizuri, F. X.: A sensitivity analysis of the self organizing map as an adaptive one-pass non-stationary clustering algorithm: the case of color quantization of image sequences, *Neural Process. Lett.*, 6, 77–89, 1997.
- Gregg, E. J. and Bodtker, K. M.: Adaptive classification of marine ecosystems: Identifying biologically meaningful regions in the marine environment, *Deep-Sea Res. Pt. I*, 54, 385–402, 2007.
- Hardman-Mountford, N. J., Hirata, T., Richardson, K. A., and Aiken, J.: An objective methodology for the classification of ecological pattern into biomes and provinces for the pelagic ocean, *Remote Sens. Environ.*, 112, 3341–3352, 2008.
- Hirata, T., Hardman-Mountford, N. J., Brewin, R. J. W., Aiken, J., Barlow, R., Suzuki, K., Isada, T., Howell, E., Hashioka, T., Noguchi-Aita, M., and Yamanaka, Y.: Synoptic relationships between surface Chlorophyll *a* and diagnostic pigments specific to phytoplankton functional types, *Biogeosciences*, 8, 311–327, doi:10.5194/bg-8-311-2011M, 2011.
- Hoepffner, N. and DoWell, M. D.: Assessing the dynamics of ecological provinces in the European Seas, *European Commission-Joint Research Centre*, 43 pp., 2005.
- Hong, Y., Li, Q., Jiang, J., and Tu, Z.: Learning a mixture of sparse distance metrics for classification dimensionality reduction, *ICCV proceedings of the 2011 international conference on computer vision*, USA, 906–913, 2011.
- Ibrayev, R. A., Ozsoy, E., Schrum, C., and Sur, H. İ.: Seasonal variability of the Caspian Sea three-dimensional circulation, sea level and air-sea interaction, *Ocean Sci.*, 6, 311–329, 2010, <http://www.ocean-sci.net/6/311/2010/>.
- Irwin, L. L., Rock, D. F., and Rock, S. C.: Habitat selection by northern spotted owls in mixed-coniferous forests, *J. Wildl. Manage.*, 76, 200–213, 2012.



- Jain, A. K. and Dubes, R. C.: Algorithms for Clustering Data. Prentice-Hall advanced reference series, Prentice Hall, Inc., Upper Saddle River, NJ, USA, 1988.
- Kara, A. B., Wallcraft, A. J., Metzger, E. J., and Gunduz, M.: Impacts of freshwater on the seasonal variations of surface salinity and circulation in the Caspian Sea, *Cont Shelf Res.*, 30, 1211–1225, 2010.
- Karpinsky, M. G.: *Pseudosolenia calcar-avis* (Bacillariophyta, Centropheceae) in the Caspian Sea, *Russ. J. Biol. Invas.*, 1, 81–86, 2010.
- Karpinsky, M. G., Katunin, D. N., Goryunova, V. B., and Shiganova, T. A.: Biological features and resources, in: *The Caspian Sea environment: The Handbook of Environmental Chemistry*, edited by: Kosarev, A. N. and Kostianoy, A. G., Springer, Berlin, 191–210, 2005.
- Kavanaugh, M. T., Hales, B., Saraceno, M., Spitz, Y. H., White, A. E., and Letelier, R. M.: Hierarchical and dynamic seascapes: A quantitative framework for scaling pelagic biogeochemistry and ecology, *Prog Oceanogr.*, 120, 291–304, 2013.
- Kideys, A. E., Roohi, A., Eker-Develi, E., Melin, F., and Beare, D.: Increased chlorophyll levels in the southern Caspian Sea following an invasion of jellyfish, *Res. Lett. Ecol.*, 2008, 1–4, 2008.
- Kohonen, T. (Ed.): *Self-Organizing Maps*, Springer-Verlag, Berlin Heidelberg, 3rd Edn., 2000.
- Kopelevich, O. V., Burenkov, V. I., Ershova, S. V., Sheberstov, S. V., and Evdoshenko, M. A.: Application of SeaWiFS data for studying variability of bio-optical characteristics in the Barents, Black and Caspian Seas, *Deep-Sea Res. Pt. II*, 51, 1063–1091, 2004.
- Kopelevich, O. V., Burenkov, V. I., and Sheberstov, S. V.: Case studies of optical remote sensing in the Barents Sea, Black Sea and Caspian Sea, in: *Remote Sensing of the European Sea*, edited by: Barale, V. and Gade, M., Springer, the Netherlands, 53–66, 2008.
- Kosarev, A. N.: Physico-geographical conditions of the Caspian Sea, in: *The Caspian Sea environment: The Handbook of Environmental Chemistry*, edited by: Kosarev, A. N. and Kostianoy, A. G., Springer, Berlin, 5–31, 2005.
- Kosarev, A. N. and Kostianoy, A. G. (Eds.): *The Caspian Sea environment: The Handbook of Environmental Chemistry*, Springer, Berlin, 2005.
- Kosarev, A. N. and Tuzhilkin, V. S.: *The Caspian Sea Climatic thermohaline fields*, Moscow: State Oceanographic Institute, 96 pp., 1995 (in Russian).
- Kramer, C. Y.: Extension of multiple range tests to group means with unequal numbers of replications, *Biometrics*, 12, 307–310, 1956.
- Lachkar, Z. and Gruber, N.: A comparative study of biological production in eastern boundary upwelling systems using an artificial neural network, *Biogeosciences*, 9, 293–308, doi:10.5194/bg-9-293, 2012.
- Landis, J. R. and Koch, G. G.: The measurement of observer agreement for categorical data, *Biometrics*, 33, 159–174, 1977.
- Leloup, J. A., Lachkar, Z., Boulanger, J., and Thiria, S.: Detecting decadal changes in ENSO using neural networks, *Clim. Dynam.*, 28, 147–162, 2007.
- Leonov, A. V. and Stygar, O. V.: Mathematical modeling of organogenic material biotransformation processes for studying the conditions of water eutrophication in the Caspian Sea surface layer, *Water Resour.*, 28, 535–552, 2001.
- Longhurst, A. (Ed.): *Ecological Geography of the sea*, Academic Press, San Diego, 1998.
- Maritorena, S., D'Andon, O. H. F., Mangin, A., and Siegel, D. A.: Merged Satellite Ocean Color Data products using a bio-optical model: characteristics, benefits and issues, *Remote Sens. Environ.*, 114, 1791–1804, 2010.
- May, R., Dandy, G., and Maier, H.: Review of input variable selection methods for artificial neural networks, in: *Artificial neural networks-Methodological advances and biomedical applications*, edited by: Suzuki, K., In-Tech, Vukovar, Croatia, 19–44, 2011.
- McDermid, G. J., Franklin, S. E., and LeDrew, E. F.: Remote sensing for large-area habitat mapping, *Prog. Phys. Geog.*, 29, 449–474, 2005.
- Nezlin, P. N.: Patterns of seasonal and inter-annual variability of remotely sensed chlorophyll, in: *The Caspian Sea environment: The Handbook of Environmental Chemistry*, edited by: Kosarev, A. N. and Kostianoy, A. G., Springer, Berlin, 143–157, 2005.
- Niiler, P.: Deepening of the wind-mixed layer, *J. Mar. Res.*, 33, 405–422, 1975.
- Nybaken, J. W. (Ed.): *Marine biology: an ecological approach*, Benjamin Cummings, San Francisco, California, fifth Edn., 2000.
- Platt, T. and Sathyendranath, S.: Spatial structure of pelagic ecosystem processes in the global ocean, *Ecosystems*, 2, 384–394, 1999.
- Platt, T., Bouman, H., Devred, E., Fuentes-Yaco, C., and Sathyendranath, S.: Physical forcing and phytoplankton distribution, *Sci. Mar.*, 69, 55–73, 2005.
- Raftery, A. E. and Dean, N.: Variable Selection for Model-Based Clustering, *J. Am. Stat. Assoc.*, 101, 168–178, 2006.
- Raitsos, D. E., Lavender, S. J., Maravelias, C. D., Haralabous, J., Richardson, A., and Reid, P. C.: Identifying four phytoplankton functional types from space: An ecological approach, *Limnol. Oceanogr.*, 532, 605–613, 2008.
- Reynolds, C. S. (Ed.): *The Ecology of Phytoplankton*. Cambridge University Press, USA, 2006.
- Richardson, L. L. and LeDrew, E. F. (Eds.): *Remote Sensing of Aquatic Coastal Ecosystem Processes*, Science and Management Applications, Springer, the Netherlands, 2006.
- Robinson, I. (Ed.): *Measuring the Oceans from Space – the Principles and Methods of Satellite Oceanography*, Springer-Praxis Publishing Ltd., Chichester, UK, 2004.
- Roff, J. C., Taylor, M. E., and Laughren, J.: Geophysical approaches to the classification, delineation and monitoring of marine habitats and their communities, *Aquat. Conserv.*, 13, 77–90, 2003.
- Saraceno, M., Provost, C., and Lebbah, M.: Biophysical regions identification using an artificial neuronal network: A case study in the South Western Atlantic, *Adv. Space Res.*, 37, 793–805, 2006.
- Saravi Nasrollahzadeh, H., Bin Din, Z., Foong, S. Y., and Makhloogh, A.: Trophic status of the Caspian Sea based on water quality parameters and phytoplankton diversity, *Cont. Shelf Res.*, 28, 1153–1165, 2008.
- Shiganova, T. A.: Biotic Homogenization of Inland Seas of the Ponto-Caspian, *Annu. Rev. Ecol. Syst.*, 41, 103–125, 2010.
- Shiganova, T. A.: Review of the status of invasive species, with special focus on the most invasive species *Mnemiopsis leidyi* (A. Agassiz, 1865) and their effects on the Caspian ecosystem, Regional consultant on invasive species to the CaspECO, P. P. Shir-

- show institute of oceanology Russian Academy of Sciences, Russia, 126 pp., 2011.
- Shiganova, T. A., Kamakin, A. M., Zhukova, O. P., Ushivtsev, V. B., Dulimov, A. B., and Musaeva, E. I.: An invader in the Caspian Sea ctenophore *Mnemiopsis* and its initial effect on pelagic ecosystem, *Oceanology*, 41, 517–524, 2001.
- Shiganova, T. A., Dumont, H. J., Sokolsky, A. F., Kamakin, A. M., Tinenkova, D., and Kurasheva, E. K.: Population dynamics of *Mnemiopsis leidyi* in the Caspian Sea, and effects on the Caspian ecosystem, in: *Aquatic invasions in the Black, Caspian, and Mediterranean Seas*, edited by: Dumont, H. J., Shiganova, T. A., and Niermann, U., Springer, the Netherlands, 71–111, 2004.
- Shiganova, T. A., Musaeva, E. I., Pautova, L. A., and Bulgakova, Yu. V.: The problem of invaders in the Caspian Sea in the context of the findings of new zoo- and phytoplankton species from the Black Sea, *Biol. Bull.*, 32, 65–74, 2005.
- Sim, J. and Wright, C. C.: The kappa statistic in reliability studies: use, interpretation, and sample size requirements, *Phys. Ther.*, 85, 257–268, 2005.
- Sørensen, T.: A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons, *Biol. Skrif.*, 5, 1–34, 1948.
- Spalding, M. D., Agostini, V. N., Rice, J., and Grant, S. M.: Pelagic provinces of the world: A biogeographic classification of the world's surface pelagic waters, *Ocean Coast. Manage.*, 60, 19–30, 2012.
- The MathWorks, Inc. MATLAB. Version 5.3, Natick, Massachusetts, 1998.
- Thomalla, S. J., Fauchereau, N., Swart, S., and Monterio, P. M. S.: Regional scale characteristics of the seasonal cycle of chlorophyll in the Southern Ocean, *Biogeosciences*, 8, 2849–2866, doi:10.5194/bg-8-2849, 2011.
- Thrush, S. F. and Dayton, P. K.: What Can Ecology Contribute to Ecosystem-Based Management?, *Annu. Rev. Mar. Sci.*, 2, 419–441, 2010.
- Tuzhilkin, V. S. and Kosarev, A. N.: Thermohaline structure and general circulation of the Caspian Sea waters, in: *The Caspian Sea environment: The Handbook of Environmental Chemistry*, edited by: Kosarev, A. N. and Kostianoy, A. G., Springer, Berlin, 33–57, 2005.
- Tuzhilkin, V. S., Katunin, D. N., and Nalbandov, Y. R.: Natural chemistry of Caspian Sea waters, in: *The Caspian Sea environment: The Handbook of Environmental Chemistry*, edited by: Kosarev, A. N. and Kostianoy, A. G., Springer, Berlin, 83–108, 2005.
- UNEP: Stolberg, F., Borysova, O., Mitrofanov, I., Barannik, V., and Egtesadi, P.: Caspian Sea, GIWA Regional assessment 23, University of Kalmar, Kalmar, Sweden, 2006.
- Uriarte, E. A. and Martin, F. D.: Topology preservation in SOM, *Int. J. Mathemat. Comput. Sci.*, 1, 19–22, 2005.
- Valavanis, V. D. (Ed.): *Essential fish habitat mapping in the Mediterranean*, Springer, the Netherlands, 2008.
- Verfaillie, E., Degraer, S., Schelfaut, K., Willems, W., and Lancker, V. V.: A protocol for classifying ecologically relevant marine zones, a statistical approach, *Estuar. Coast Shelf Sci.*, 83, 175–185, 2009.
- Vichi, M., Allen, J. I., Masina, S., and Hardman-Mountford, N. J.: The emergence of ocean biogeochemical provinces: A quantitative assessment and a diagnostic for model evaluation, *Global Biogeochem. Cy.*, 25, 1–17, 2011.
- Wheater, C. P. and Cook, P. A. (Eds.): *Using Statistics to Understand the Environment*, The Taylor & Francis e-Library, London and New York, 2005.
- Zacharias, M. A. and Roff, J. C.: A hierarchical ecological approach to conserving marine biodiversity, *Conserv. Biol.*, 14, 1327–1334, 2000.
- Zar, J. H. (Ed.): *Biostatistical Analysis*, fourth ed. Prentice Hall, Upper Saddle River, NJ, 1999.
- Zonn, I. S.: Environmental issue of the Caspian, in: *The Caspian Sea Environment: The Handbook of Environmental Chemistry*, edited by: Kosarev, A. N. and Kostianoy, A. G., Springer, Berlin, 223–242, 2005.
- Zwirlmaier, K., Jardillier, L., Ostrowski, M., Mazard, S., Garczarek, L., Vulot, D., Not, F., Massana, R., Ulloa, O., and Scanlan, D. J.: Global phylogeography of marine *Synechococcus* and *Prochlorococcus* reveals a distinct partitioning of lineages among oceanic biomes, *Environ. Microbiol.*, 10, 147–161, 2008.