Multispectral data by the new generation of high-resolution satellite sensors for mapping phytoplankton blooms in the Mar Piccolo of Taranto (Ionian Sea, southern Italy)

Flavio Borfecchia, Carla Micheli, Tamara Cibic, Vito Pignatelli, Luigi De Cecco, Natalizia Consalvi, Carmela Caroppo, Fernando Rubino, Elena Di Poi, Martina Kralj & Paola Del Negro

Taylor & Francis
Taylor & Francis Group

Multispectral data by the new generation of high-resolution satellite sensors for mapping phytoplankton blooms in the Mar Piccolo of Taranto (Ionian Sea, southern Italy)

Flavio Borfecchia [a], Carla Micheli [b], Tamara Cibic [c], Vito Pignatelli[b], Luigi De Cecco[a], Natalizia Consalvi[d], Carmela Caroppo [e], Fernando Rubino [e], Elena Di Poi[c], Martina Kralj [c] and Paola Del Negro [c]

aENEA, National Agency for New Technologies, Energy and Sustainable Economic Development, Department of Observation and Analysis of Earth and Climate, Research Centre Casaccia, Rome, Italy; bENEA National Agency for New Technologies, Energy and Sustainable Economic Development, Department of Energy Technologies, Laboratory of Biomass and Biotechnologies for Energy, Research Centre Casaccia, Rome, Italy; cIstituto Nazionale di Oceanografia e di Geofisica Sperimentale - OGS, Sezione di Ricerca Oceanografia (OCE), Trieste, Italy; dDepartment Of Ecology and Biology, Laboratory of Experimental Oceanology and Marine Ecology, Tuscia University, Civitavecchia, RM, Italy; eInstitute for Coastal Marine Environment, National Research Council, Taranto, Italy

ABSTRACT

The HR (High-Resolution) EO (Earth Observation) satellite systems Landsat 8 OLI and Sentinel 2 were tested for mapping the frequent phytoplankton blooms and Chl *a* distributions in the sea basin of the Mar Piccolo of Taranto (Ionian Sea, southern Italy), using the sea truth calibration data acquired in 2013. The data were atmospherically corrected for accounting of the aerosol load on optically complexes waters (case II). Various blue-green and additional spectral indices ratios, were then satisfyingly tested for mapping the distribution of Chl *a* and differently sized phytoplankton populations through PLS (Partial Least Square regression) models, regressive statistical models and bio-optical algorithms. The PLS models demonstrated higher robustness for assessing the distribution of all the phytoplankton and Chl *a* except for those related to sub-surface micro-phytoplankton. The distributions obtained via a bio-optical approach (OC3 algorithm and full physically based inversion) showed a general agreement with the previous ones produced by statistical methods. The reflectance signals, captured by OLI and Sentinel 2 sensors in the visible and shorter wavelengths once atmospherically corrected, were found to be useful to map the coastal variability at detailed scale of Chl *a* and different phytoplankton populations, in the optically complexes waters of the Mar Piccolo.

Introduction

The last decades, phytoplankton blooms have been frequently observed in shallow coastal areas of the Mediterranean, where high nutrient loads due to eutrophication stimulate major photoautotrophic abundances (Kennish, Brush, & Moore, 2014). Globally, the increase of anthropogenic impacts and pollution in the marine habitat drives changes in all the ecosystems both in the variation of ecological structures of the aquatic communities (Micheli, Cupido, Lombardi, Belmonte, & Peirano, 2012) and in growth dynamics of plankton (Karuza et al., 2016). Moreover, some species responsible for harmful algal blooms are expanding their distribution on a global scale with a negative impact on the seawater quality, human health and economically relevant activities such as tourism and aquaculture that are concentrated in coastal water environments. To effectively mitigate these trends which are likely to continue (Rhodes, 2011), the sustainable management of the affected coastal areas calls for rapid and accurate methods for assessing and mapping the type and abundance of phytoplankton and Chl *a*, in seawaters (Smayda, 2007). In particular, phytoplankton monitoring requires its effective detection and detailed mapping in the open sea and in coastal and shallow waters where anthropogenic impacts mostly concentrate. For this purpose, both the traditional methods exploiting sea truth data alone (Micheli et al., 2018, 2015) and the EO (Earth Observation)-based innovative techniques using airborne (Borfecchia et al., 1997) and satellite (Borfecchia et al., 2016, 2018b) platforms are necessary to improve the knowledge on ecosystem functioning, in the perspective to support the ecosystem-based sustainable management (Cibic et al., 2016; Micheli et al., 2010). Once suitably corrected for atmospheric noises and calibrated using the in-situ measurements, the satellite EO (Earth Observation) techniques are recognized as effective multiscale tools for monitoring open seas and shallow

CONTACT Flavio Borfecchia ✉ flavio.borfecchia@enea.it; borfecchia@virgilio.it 🖳 ENEA, National Agency for New Technologies, Energy and Sustainable Economic Development, Department of Observation and Analysis of Earth and Climate, Research Centre Casaccia,Via Anguillarese 301/00123, Rome, Italy

Ⓢ Supplemental data for this article can be accessed here.

water quality, including the spatio-temporal changes of marine communities (Blondeau-Patissier, Gower, Dekker, Phinn, & Brando, 2014; Dazhao, Chuqun, Jieqiong, & Dongyang, 2010). In fact, there are well-established algorithms which currently provide operational estimates of global distributions of Chl a and other relevant water parameters concentrations in oligotrophic oceans from satellite ocean colour data (El Hourany, Fadel, Gemayel, Abboud-Abi Saab, & Faour, 2017), at intermediate geometrical resolution (102–103 m). They typically exploit the radiance/ reflectance shift from the blue (440–510 nm) to the green (550–565 nm) wavelengths due to Chl a concentration increases by working in terms of the ratios in these two spectrum ranges (i.e. the SeaWiFS/MODIS OC4/OC3 algorithms). The OC2/3 bio-optical algorithms, based on the blue–green (440–550 nm) spectral band ratio, are within the most common types of bio-optical empirical approaches for retrieving the global Chl a distribution from ocean colour reflectance detected by satellite orbiting sensors (Hu, Lee, & Franz, 2012). These blue–green ratios can vary in response to factors besides chlorophyll concentration, like CDOM (Coloured Dissolved Organic Matter) and SS (suspended matter) contents, which characterize more optically complex waters typical of coastal or inland basins (Blondeau-Patissier et al., 2104). Thus, more effective approaches are required to deal with both atmospheric and water turbidity reflectance noise contributions and more accentuated (compared to open seas) spatial dynamics of coastal water variables of interest. Indeed, the topography of the seabed, the discharge of rivers, artificial coastal infrastructure and coastal activities make a significant contribution to the pattern of shallow water coastal circulation. In this context, due to high spatial dynamics of optically active substances concentration, their effective distribution assessment requires suitable monitoring capabilities in term of both spatial and radiometric resolution. Moreover, in coastal waters (typically of case II), the exploitation of the above-cited algorithms suffers for the limitation in the consolidated atmospheric pre-processing (De Keukelaere et al., 2018; Franz, Bailey, Kuring, & Werdella, 2015; Pahlevan, Lee, Wei, Schaaf, & Schott, 2014). In particular in the retrieval of the aerosol noise parameters distribution, effective under the assumption of null water leaving NIR (Near Infra-Red) radiance (Borfecchia et al., 2018c; Vanhellemont & Ruddick, 2015), typically valid only for open sea and oligotrophic waters (case I). Thus, at present much effort is focused to provide more spectral bands to the new HR (High Resolution) satellite sensors for enabling the development of more sophisticated atmospheric correction schemes and algorithms. In particular, this is required for effective retrieving of water quality variables and

algal bloom proxies in coastal and inland case II waters, where the anthropogenic activities often concentrate (Vanhellemont & Ruddick, 2014), as in our test site within the Mar Piccolo of Taranto (Caroppo, Cerino, Auriemma, & Cibic, 2016).

In the summer 2013, the new Landsat 8 NASA polar satellite was launched with on board the OLI (Operational Land Imager) sensor, characterized by improved radiometry (12 bit) and increased spectral acquisition bands, including the "coastal" one, especially devoted to support the monitoring of shallow waters. These innovative spectral/spatial features of the OLI sensor can provide a new opportunity for effectively mapping phytoplankton and Chl a/phaeopigments optically active water column components, at suitable spatial resolution. In this perspective, one of the goals of this study was to preliminarily test the OLI multispectral HR data at 30 m of a.g.r. (above ground resolution) for detecting and mapping the phytoplankton population and Chl a, within the Mar Piccolo of Taranto, characterized by shallow waters, likely case II (Borfecchia et al., 2018a). This confined sea is located in the Ionian Sea (southern Italy) coastal area of Taranto, which is densely populated and at a high risk of environmental crisis (Cardellicchio, Covelli, & Cibic, 2016), due to the concentration of different impact factors which mainly arise from local anthropogenic activities. Thus the information provided by OLI sensor have been exploited for phytoplankton and Chl a mapping, using some multispectral images taken in 2013 over the Mar Piccolo of Taranto and calibrated by means of near synchronous sea truth data, acquired within the framework of the project "RITMARE (la Ricerca ITaliana per il MARE – Italian Research for the sea). In addition to the above-cited blue–green ratio, other ratios have been introduced based on the coastal (430–457 nm) and red (645–683 nm) bands, trying to properly deal with the optical complexity of the shallow water in our test site, taking advantage of the improved features of the OLI sensor. The statistical multivariate and PLS (Partial Least Square) regressive models were estimated and satisfactorily tested using the sea truth information and ratios obtained from remotely sensed data atmospherically corrected through an innovative method for atmospheric correction on optically complexes waters (Vanhellemont & Ruddick, 2015). The Chl a distribution was then assessed using the usual OC3 algorithm and corrected reflectance data. Finally, a comprehensive approach through a bio-optical model inversion was preliminarily implemented using WASI4 model (Gege, 2014).

Material and methods

Study area

The Mar Piccolo is a shallow, nearly enclosed marine basin that consists of two naturally divided inlets, the

1st and 2nd inlet, with maximum depths of 13 and 10 m and a surface area of 8.28 km² and 12.43 km², respectively (Figure 1). The Mar Piccolo is characterized by scarce hydrodynamic and low water exchange since it is connected with the Mar Grande by only two narrow channels in the 1st inlet: the "Navigabile" and the "Porta Napoli" channels. The sediments load of these basins arises mainly from land runoff, numerous submarine springs, small streams, sewage outfalls and industrial discharges rather than that provided by marine currents (Cibic et al., 2016). The surrounding watershed of 555 km² hosts horticulture (6%) and agriculture activities, mainly cereals (24.1%), olive trees (24.2%) and vineyards (25.4%). These cultivations involve the application of fertilizers (116,800 tons y-1). Since 1960s, Taranto and its coastline have been subjected to the industrialization process that has caused profound environmental changes. The industrial zone is mainly characterized by the presence of the largest steelworks in Europe and navy arsenal in Italy (Military area in Figure 1), a major oil refinery, shipbuilding and other industrial activities that are responsible for severe environmental contamination. The basin is one of the most important plant of mussel farming in Italy (Caroppo et al., 2016; Franzo et al., 2016).

In this context, the sustainable management and safeguard of this vulnerable ecosystem requires effective monitoring at suitable spatial and temporal scales for the characterization of different impact factors and their effects on this particular ecosystem (Petrocelli et al., 2015).

Sea truth and EO data

The sea truth sampling of phytoplankton was carried out during the summer 2013 at 6 stations: four located in the 1st inlet and two in the 2nd inlet, respectively, as shown in Figure 1. At each station, the sampling was carried out at two depths: surface, at 1 m of depth, and 1 m above the bottom, using an acid-rinsed 5L Niskin bottle, equipped with silicon elastic and red silicon O-ring. To support the implementation of calibration/validation procedure (CV), taking into account also the typical uncertainties in localization and of the sea truth variables laboratory estimates, even till 20–25% (Cibic et al., 2018), these latter were considered representative of an area of about 50×50 m around the central coordinates of the station, as detailed in the following. The estimates of abundance and biomass of the three typical size classes of phytoplankton [pico- (0.2–2 μm), nano- (2–20 μm), micro-phytoplankton (20–200 μm)] and the concentration of Chl a and phaeopigments (Chl a degradation products) were derived from the samples collected within the sample area of each station (Figure 1). For Chl a and phaeopigments concentration, 1 L of seawater was filtered and processed as

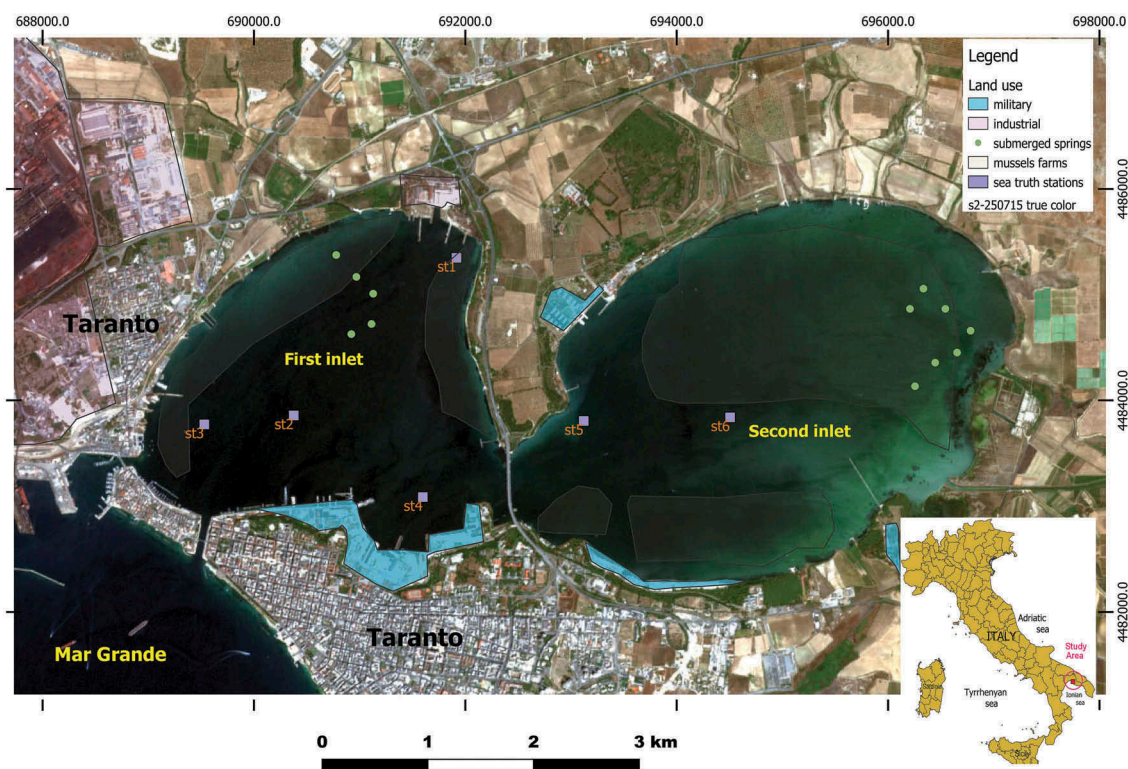**Figure 1.** Distribution of the sea truth areas including the sampling stations of phytoplankton and pheopigments (violet squares). The background consists of the "true colour" image of the Mar Piccolo of Taranto acquired by Sentinel 2 sensor on 25.07.2015 in the UTM 33 WGS 84 geographic projection. The adjacent land uses and relevant elements (mussels farms areas and submerged springs) of the basin are also reported.

described by Kralj et al. (2016). The measurements of Chl *a* and phaeopigments were performed, respectively, before and after acidification with two drops of HCl 1N using JASCO FP 6500 spectrofluorometer. The coefficient of variation for three replicate samples was lower than 5%, and the detection limit, defined as twice the standard deviation of three blank filters treated in the same way as samples, was 0.002 and 0.068 $\mu g\ L^{-1}$ for Chl *a* and phaeopigments, respectively. The procedures were quality controlled and monitored by participation in the QUASIMEME intercalibration program (De Boer, 2016). For the phytoplankton abundance assessment various subsamples (30–120 mL) derived from the main samples were differently pretreated through specific preprocessing, filters and staining then a visual enumeration was carried out using standards methods based on different tools (i.e. a Leica DM 2500 epifluorescence microscope equipped with a 100-W high-pressure mercury burner (HPO 100 W/2) at × 1000 magnification for nanoplankton, an Utermöhl chamber settled with a variable volume of sample (50–100 mL) and inverted microscope, Leitz Labovert FS equipped with phase contrast, at a magnification of ×400 and ×630, for microphytoplankton). While for pico- and nanoplankton, data are reported as mean of three replicates + SD, for microplankton only one replicate per station per depth was counted since the observed volume of seawater (25–50 mL) is considered representative of the sampling site. In agreement with standard approaches for biomass calculation, the cells' biovolume was assessed by visually fitting them through pre-definite classes of geometrical bodies (with different size/shape) that were then exploited for biomass calculation using specific conversion factors (Cibic et al., 2018). Subsamples for the estimate of chlorophyll a (Chl *a*) concentration were filtered onto 47-mm Whatman GF/F filters that were stored at −20°C until laboratory analysis. Chl *a* was extracted overnight (4°C) with 90% acetone from the homogenate filter and determined using Jasco FP 6500 spectrofluorometer. For details on sampling and samples processing, see Karuza et al. (2016) for the different phytoplankton size-classes, and Kralj et al. (2016) for chl *a* and phaeopigments assessment. The point measurements distribution at the six stations is reported in Figure 1 while the list of the measured phytoplankton and Chl *a* variables with the assigned names and related features is reported in Table 1.

Due to requirements in terms of geometrical resolution for effectively capturing the spatial dynamics of these variables in coastal shallow waters, the most used EO techniques exploit the currently available HR (High Resolution) and VHR (Very High Resolution) multispectral aerospatial sensors. In general, they are operating in the visible and NIR spectral ranges from satellite remote platforms (Borfecchia et al., 2018c, 2018d; Pasqualini et al., 2005). The provided typical a. g.r. (above ground resolution) range of HR and VHR sensors is, respectively, of 5–100 m and 0.5–5 m.

The EO satellite systems equipped with VHR sensors, mostly owned by the private enterprises, demonstrated suitable capabilities for monitoring and mapping the coastal shallow waters and *Posidonia oceanica* (PO) ecosystems (Borfecchia et al., 2018a).

Their operational mode is usually based on user request and commercial distribution, which does not involve systematic collection of the Earth's surface data, but only acquisitions on selected and limited areas of interest in the scheduled period based on specific customer requests. On the other hand, due to their poor radiometry (8 bit – 256 levels radiometric range) and low Signal to Noise Ratio, SNR, the older Landsat HR sensors (i.e. TM and ETM+) resulted often inadequate to be widely and usefully exploited for the seabed mapping and for monitoring the Mediterranean shallow waters, frequently characterized by turbidity conditions and optical complexity.

Table 1. Phytoplankton and phaeopigment sea truth data. In the first column, the variable reference name was indicated while the related meaning was reported in the description column.

reference Variable	Unit	Description	Sampling Depth	Population Size
Phapis	Cells L^{-1}	**Surface** concentration of **pico** phytoplankton population cells	surface	pico – 10^{-12} m
Phanas	Cells L^{-1}	**Surface** concentration of **nano** phytoplankton population cells	surface	nano – 10^{-9} m
Phamis	Cells L^{-1}	**Surface** concentration of **micro** phytoplankton population cells	surface	micro – 10^{-6} m
Phapib	Cells L^{-1}	**Bottom** concentration of **pico** phytoplankton population cells	bottom	pico – 10^{-12} m
Phanab	Cells L^{-1}	**Bottom** concentration of **nano** phytoplankton population cells	bottom	nano – 10^{-9} m
Phamib	Cells L^{-1}	**Bottom** concentration of **micro** phytoplankton population cells	bottom	micro – 10^{-6} m
Phbpis	$\mu g\ C\ L^{-1}$	**Surface biomass** concentration of **pico** phytoplankton population	surface	pico – 10^{-12} m
Phbnas	$\mu g\ C\ L^{-1}$	**Surface biomass** concentration of **nano** phytoplankton population	surface	nano – 10^{-9} m
Phbmis	$\mu g\ C\ L^{-1}$	**Surface biomass** concentration of **micro** phytoplankton population	surface	micro – 10^{-6} m
Phbpib	$\mu g\ C\ L^{-1}$	**Bottom biomass** concentration of **pico** phytoplankton population	bottom	pico – 10^{-12} m
Phbnab	$\mu g\ C\ L^{-1}$	**Bottom biomass** concentration of **nano** phytoplankton population	bottom	nano – 10^{-9} m
Phbmib	$\mu g\ C\ L^{-1}$	**Bottom biomass** concentration of **micro** phytoplankton population	bottom	micro – 10^{-6} m
Chls	$\mu g\ C\ a\ L^{-1}$	**Surface Chl *a*** concentration	surface	
Chlb	$\mu g\ C\ a\ L^{-1}$	**Bottom Chl *a*** concentration	bottom	
Phes	$\mu g\ C\ L^{-1}$	Surface Phaeopigments concentration	surface	
Pheb	$\mu g\ C\ L^{-1}$	Bottom Phaeopigments concentration	bottom	

The new generation of Landsat sensors Operational Land Imagery (OLI) and similar satellite HR sensors (i.e. Sentinel 2) are conceived for systematic acquisitions of VIS, NIR (Near Infrared), SWIR (Short Wave Infrared) reflectance data of tiles covering the entire Earth surface. Their acquisition channels have been strengthened in terms of bands number, SNR and radiometry (12 bit), maintaining at the same time the useful ground resolution and revisiting the capability of the previous ones of the Landsat family. In such a way, they might provide effective support to the operative monitoring of coastal shallow and lagoon waters (Pahlevan et al., 2014). In this perspective, various cloud-free Landsat 8 OLI frames acquired in 2013 were preliminarily evaluated; therefore, the multispectral image of 20 June was selected for processing, given its best quality and synchronicity with the sea truth campaign.

Processing methodologies

Atmospheric preprocessing and ratio indices

To preliminarily test the capabilities of Landsat 8 OLI for mapping the distribution of phytoplankton in the water column of the Mar Piccolo, a correction procedure was applied in order to remove atmospheric effects from the TOA (Top Of Atmosphere) spectral responses detected by the remote sensor. The essential pre-processing step of atmospheric correction is required for EO-based quantitative monitoring of marine basins, including the coastal ones, to retrieve the useful signal coming from the water surface as Water Leaving Radiance (WLR) in the different spectral acquisition bands (bi). In addition to that of phytoplankton, WLR includes the contribution from the other optically active components in the water column, as Total Suspended Matter (TSM) and Coloured Dissolved Organic Matter (CDOM), and even that from the sea bottom coverage from seagrasses, macroalgae and various bottom substrates. The radiometric correction for atmospheric effects was carried out through an "image-based" approach using the ACOLITE (Vanhellemont & Ruddick, 2015) recently released software, specifically developed to provide atmospheric correction of the HR remotely sensed data for marine monitoring applications. The "image-based" approach takes advantage of the specific information contained in the same multi-spectral image to be corrected and does not require additional in situ measures simultaneous to satellite overpass (Richter, Schlapfer, & Muller, 2006). This approach, being readily applicable, has been considered suitable for our operational use. In such a way it was possible to take into account the atmospheric noise effects, in particular those more relevant coming from aerosol concentration distribution, using the AOD (Aerosol Optical Depth) parameter that can be estimated from the same EO multispectral data (image-based approach), even in case of optically complex waters (case II). ACOLITE provides the capability to retrieve AOD on a per pixel basis, more appropriate for marine coastal monitoring, where the useful reflectance signals to detect are weak and affected by spatially dependent significant noises (Borfecchia et al., 2018c).

The spectral responses captured by OLI sensor in shorter wavelengths have been used in the form of various ratios, more effective to minimize multiplicative noise contributions arising from different factors (i.e. residual atmospheric and water noises). The OLI channels used here with their central wavelength and bandwidth in nanometer (within parenthesis) are: coastal = b_1(442.96, 15.98), blue = b_2(482.04, 60.04), green = b_3(561.41, 57.33), red = b_4(654.59, 37.47), NIR = b_5(864.67, 28.25). In addition to the previously cited blue–green ratio (b_{23}) of the responses derived from the pre-processed images, additional formulation including the new "coastal" and NIR bands have been introduced trying to capture the specific responses of different phytoplankton features and populations in the optically complex waters of the Mar Piccolo. In particular, we used the various sea truth data relative to phytoplankton acquired in 2013, to test the predictive capability of the various channels of the OLI sensor in the form of the following 5 band ratios: $b_{12} = b_1/b_2$, $b_{13} = b_1/b_3$, $b_{23} = b_2/b_3$, $b_{34} = b_3/b_4$, $b_{45} = b_5/b_4$. The first two ratios make use of the new "coastal" channel (b_1) coupled, respectively, with blue (b_2) and green (b_3) channels; the third consists of the aforementioned blue–green ratio and the latter two exploit the red (b_4) and NIR (b_5) channels, more appropriate for providing information related to the water surface and thin layers beneath. The values of these five reflectance ratios derived from the point spectral signatures were appropriately extracted from multispectral pre-processed images in correspondence to the measurement stations and were exploited as independent variables for statistical modelling used here.

The Landsat OLI frame, acquired on 20 June 2013, was preliminarily selected considering its synchronicity with the sea truth campaign and quality in terms of cloud absence. Subsequently, it was preprocessed for atmospheric noise removal based on the estimated AOD load. The station point spectral signatures were then derived for all the OLI first five reflectance bands in output from the ACOLITE code. The available sea truth data of phytoplankton variables had been provided, for each station, as representative of a 50 × 50 m area around the station centre coordinates (station area), while their resulting standard deviation from laboratory assessment was assumed as station measurement error. Thus an area of equivalent dimension, compatible with that seen by

satellite sensor, approximately equal to 3 × 3 pixels (window), centred on the single station centre, was used for the extraction of the point spectral signatures from the multispectral images, to be used for the subsequent modelling processes. Considering that the sea truth measurements were available only as station area averages, first of all these spectral signatures, as mean on the same area, were exploited for the first modelling approaches. Subsequently, the spectral signature of each pixel within the station area was independently accounted in order to allow the application of a more realistic calibration/validation schema to PLS models (see next sections).

Various models have been implemented using PLS approaches by coupling these extracted data with the corresponding phytoplankton measurement, previously recorded at each station. Subsequently, the OC2/3 bio-optical algorithms were applied using the above introduced atmospheric correction schema for Chl *a* mapping while the water column effects were accounted through a preliminary attempt based on a full bio-optical inversion through the WASI4 code. All the continuous distributions of the sea truth variables (i.e. phytoplankton densities, Chl *a* and phaeopigments concentrations/abundance) obtained by the various modelling approaches for the entire area of interest, were processed to provide pseudo-colour maps of six equal-area graduated classes.

PLS modelling & test

The regression PLS models for estimating analytical expression for our phytoplankton variables are based on the Principal Component (PC) preventive decomposition of our spectral ratios independent (X) variables, and this is their main difference with usual multivariate approach, which makes it more robust against the frequent collinearity problems of spectral reflectance variables (Gholizadeh & Robeson Scott, 2016). In case of multiple dependent (Y) variables, characterized by mutual correlation (i.e. our phytoplankton variables), the PC decomposition may be usefully used also for them (Kimberly & Khalid, 2016), using a full decomposition (FD) procedure, as properly explained in the devoted supplementary materials (eq. A.3).

The regression PLS multivariate models have provided an analytical expression for the measured phytoplankton (Phy) and/or pigment Chl *a*/pigment variables in the form of:

$$\text{Phy} = \sum a_i b_{ij} + a_0 \qquad (1)$$

where the b_{ij} are the previously introduced water reflectance ratios, a_i and a_0 are the regression coefficients to be evaluated through calibration best fit for errors minimization using field measurements. The best fit allows assessing also the squared correlation coefficient R^2 (determination coefficient) as indicator of model performance, once adjusted for the specific contribution of each independent variable. In addition to R^2, the average of square errors between the measured and modelled point station values as RMSE (Root Mean Square Error) was considered in model performance assessment.

In general the statistic models based on field measurements are usually implemented through calibration and validation (CV) schema in order to provide a better predictive power. The calibration provide the model parameters by optimal fitting the responses of interest (the phytoplankton and Chl *a* concentrations in our case) using physically or statistically based expression of independent variables (the spectral reflectance data in our case) for a subset of the available measurements. The validation exploits the remaining measurements, different from those used in calibration, for a further tuning by testing the predictive capability of the estimated model using independent data. In such a way a more realistic evaluation of performances and expected errors may be provided in case of subsequent model running. In this framework a sufficient measurements number, likely higher than the six available, is a prerequisite for applying such schema to suitably calibrate and validate the assessed models providing reliable insights about their robustness in term of statistical significance and predictive power.

First of all, a PLS calibration using the sea truth data referring to the six measurement stations coupled with the related point reflectance data extracted from atmospherically corrected multispectral images and averaged on station window was accomplished. The sea truth data were grouped into two set referring to surface and bottom variables that were then PC decomposed according to the formulas of full PLS (eq. A.3) approach (FD), including that of spectral ratios values derived from the extracted reflectance data (auxiliary materials). A selection of the first four latent factors of decomposition allowed us to obtain a satisfying level of mean performance of the PLS models assessed for surface and bottom variables.

Subsequently in order to provide a sufficient numerous sample set to apply the CV schema for PLS models, the spectral reflectance response of each pixel within the station window was considered as independent measure to be coupled with the related concentration station measurements provided by sea truth acquisition (eq. A.4). Here the dependent variables were modelled singularly using the previously decomposed independent spectral variables (CV approach). Although the first FD step provided us with an optimistic evaluation of models performance in term of R^2, it was useful to provide the relative importance of various OLI spectral ratios in modelling bottom and surface dependent variables (i.e. phytoplankton and Chl/phaeopigment concentrations). In the second step, the spectral signatures of the 9 pixels

within the station window have been coupled with the related station measurements of biophysical variables for assessing the PLS models through CV schema. In particular the 60% of randomly selected samples were used for the model calibration while the remaining 40% was devoted to models validation/test. In FD, a fixed decomposition of independent and dependent variables into four latent factors was adopted. The second CV PLS procedure instead included only decomposition of independent variables followed by a backward stepwise selection of most representative latent factors minimizing the RMSE. In the implemented methodology, the first step provided models with too optimistic performance which begun more realistic in the second one with the adoption of the calibration/validation schema. From the point of view of Variable Importance in Projection (VIP), the first step provided us a global performance evaluation in modelling of the different spectral ratios, for the surface and bottom variables, while in the second step a close proxy of VIP was assumed for each sea truth variable as the confidence level derived from the related p-value(t).

Bio-optical algorithms

Since the OC2/3 bio-optical algorithms are successfully used in open ocean case I oligotrophic waters, where the phytoplankton is the primary optically active constituent with limited concentrations (Chl a 0.01–0.5 µg L^{-1}), here they were introduced and tested considering the comparable sea truth data (Figure 3). The general formulation of these algorithms consists in the combination until the fourth power of the above introduced b23 blue-green ratio:

$$\mathrm{Log}_{10}(\mathrm{Chl\ a}) = a_0 + \sum a_i (\mathrm{Log}_{10} b_{23}) \qquad (2)$$

the b23 is the above introduced ratio of the water leaving radiance/reflectance obtained from a suitable atmospheric correction preprocessing of TOA (Top of Atmosphere) Rblue and Rgreen, responses. The ai coefficients are assessed for the different satellite sensors using large sea truth datasets containing the in situ point measurements and related atmospherically corrected radiances/reflectances. Given the availability of the OC3 coefficients for the OLI and Sentinel 2 sensors (https://oceanco lor.gsfc.nasa.gov), these algorithms were applied using reflectance responses obtained through a new atmospheric correction approach based on SWIR channel to retrieve the noise AOD distribution. This approach was imbedded in the ACOLITE freely distributed package (Vanhellemont & Ruddick, 2015), specifically implemented for Landsat 8 OLI and Sentinel 2 HR satellite sensors and case II waters monitoring applications. Using the ACOLITE atmospheric correction, some Chl a distribution maps were produced from

multispectral frames acquired by these HR satellite sensors to be compared with sea truth and modelled data.

Considering the low depth and the optically complex waters of the Mar Piccolo, a further approach was introduced to account for the spectral reflectance contribution of the sea bottom and the water column due to the dissolved components other than phytoplankton (Borfecchia et al., 2018c; Mumby, Clark, Green, & Edwards, 1998). To this end, a complete bio-optical modelling inversion on physical basis was attempted using WASI4 package (Albert & Gege, 2006; Gege, 2014). This software has been implemented to extract information about the optical properties of the water body and eventually of the seabed from the hyper/multispectral WLR (Water Leaving Radiance) signals coming from the aquatic surface. It requires input as atmospherically corrected multispectral images and auxiliary parameters related to typologies of seabed and optically active water column components. The code allows to recover the distribution of the concentration of the optically active main "structures" present in the water column (phytoplankton, CDOM and suspended sediments) and a series of other parameters that influence the WLR signal even in shallow waters, in particular the water depth and the percentage of the various types of seabed cover such as sand, macrophytes and other benthic organisms. The data analysis is based on a semi-analytic inversion of various bio-optical consolidated models (Albert & Gege, 2006) on the basis of typical spectral responses of the considered components, in particular up to 6 typologies of phytoplankton and various types of spectrally different material in suspension.

The reflectance of the bottom is treated as a mixture of 6 types of substrate which reflectance spectra are stored in a database that covers the spectral range from 350 nm to 1000 nm intervals. This database contains all the reflectance spectra used in the code that can be upgraded to adequately represent the real optical properties of the various constituents of coastal waters, the various types of seabed, seagrasses and macroalgae, to characterize the shallow water area of the Mar Piccolo. In order to account for water column additional noise effects, a preliminary exploitation of the WASI4 was attempted. The atmospherically corrected OLI multispectral data of the Mar Piccolo were properly reformatted before being supplied to the input to the WASI4 code. Then we proceeded with the interactive calibration procedure using the point spectral signatures of sampling stations, previously exploited for regressive modelling. To improve convergence in optical inversion various constraints have been introduced, in particular to the concentration ranges of optically active components, based on sea truth measurements and data available

for the study site. In addition to those of sand and sediment, the spectra used for the different seabed coverages, also include those of some typical green macrophytes of inland waters (*Characontraria, Potamogeton perfoliatus*) available in the WASI database. These default values of auxiliary parameters have been exploited for preliminary evaluation of multispectral Landsat data 8 OLI with this bio-optical inversion approach.

Results and discussion

In Figure 2, the results of different measurements relative to the six stations in the Mar Piccolo are reported. In particular, in Figure 2(a), the surface and bottom cells abundances of the three different sizes of phytoplankton are displayed, while Figure 2(b) shows the related biomass values. The Chl *a* and phaeopigments concentrations found at the six measurement stations are reported in Figure 3. As stated above, these concentration measurements derived from sea truth samples were coupled with the related window averaged spectral indices, derived from preprocessed remotely sensed data, for the first PLS modelling calibration step. Here the full PLS approach (FD) allowed us to obtain models for sea surface and bottom variables potentially correlated within each group, with the related global VIP. This data arrangement with the selection of four latent factors for PLS decomposition was devoted to optimize the results in term of global correlation maintaining at the same time a sufficient model generality.

Tables 2 and 3 show the correlation, in terms of cumulated explained variance under form of determination coefficient (R^2), provided by the four latent factors of the PLS models assessed for each target (dependent) sea truth variable of surface and bottom groups. These PLS surface and bottom models achieved respectively more than 97% and 85% of average global correlation (Total Expl.) referring to calibration measurements. In agreement with the higher noise contribution from the water column
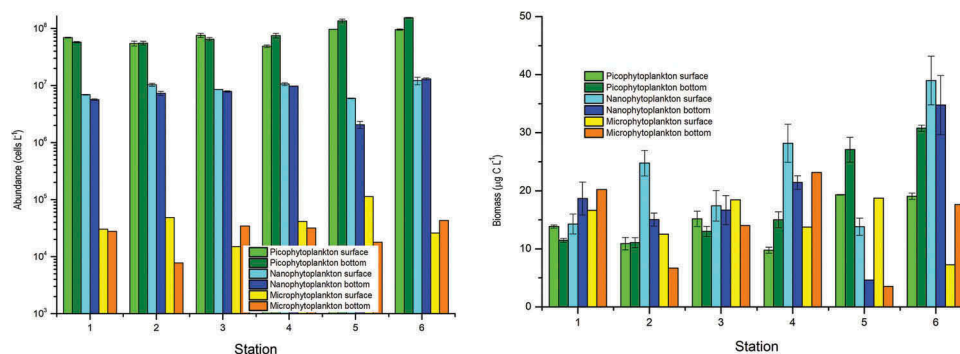

Figure 3. Surface and bottom sea truth data relative to Chl *a* and pheopigments expressed in μg L-1 for the different population size (pico-, nano-, micro-) at the six sampling stations in the Mar Piccolo of Taranto. For the lacking of error bars see the materials & methods chapter.

optically active components, the global correlation of bottom models decreases. Although different performances in term of R^2 characterize each target variable reported in Tables 2 and 3, most of the cumulative correlation coefficients (Factor 4 column) show high values. In particular, the estimated models found effective for all surface variables with the correlation lowest value (0.8748) obtained for Chl *a*. The obtained correlations of the phytoplankton abundance and biomass bottom variables (0.9–0.95) were similar to those of the corresponding surface variables except for micro-phytoplankton that exhibit a cumulated R^2 values dropping to about 0.57. However, the selection of four latent factors for X and Y decomposition (formulas 2 and 3) of the two models allowed to achieve an acceptable effectiveness, also for particular variables of phytoplankton population (i.e. abundance and biomass of micro-phytoplankton) in bottom water layers, more difficult to detect and model.

Tables 4 and 5 contain the coefficients related to global VIP assessed for each independent variable (spectral reflectance ratios) in the respective sea surface and bottom PLS FD models. The variables' importance in the surface model is the highest for b34 and b45 ratios, based

Figure 2. Surface and bottom sea truth data relative to phytoplankton abundance (a) and biomass (b) for the different population size (pico-, nano-, micro-) at the six sampling stations in the Mar Piccolo of Taranto Figure 2.
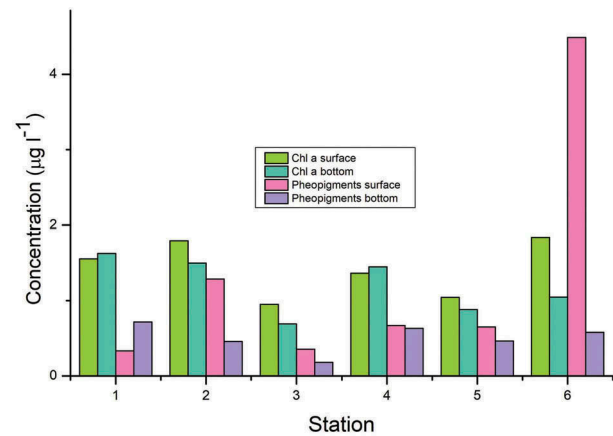
Table 2. Cumulative explained variance (R^2- Factor 4) of surface sea truth variables by the four latent factors of the assessed full PLS decomposition.

	Surface variables			
	Cumulative R^2			
Target	Factor 1	Factor 2	Factor 3	Factor 4
Phapis	0.3065	0.3207	0.3773	0.9889
Phanas	0.3888	0.4087	0.5957	0.9872
Phamis	0.0029	0.3889	0.9308	0.9998
Phbpis	0.3065	0.3207	0.3773	0.9889
Phbnas	0.6403	0.671	0.727	0.9998
Phbmis	0.5616	0.6898	0.6918	0.9814
Chls	0.1417	0.4443	0.5007	0.8748
Phes	0.848	0.9792	0.9794	0.9923
Total Exp.	0.3995	0.5279	0.6475	0.9767

Table 3. Cumulative explained variance (R^2- Factor 4) of bottom sea truth variables by the four latent factors of the assessed full PLS decomposition.

	Bottom Target(s) variables			
	Cumulative R^2			
Target	Factor 1	Factor 2	Factor 3	Factor 4
Phapib	0.6583	0.6808	0.8224	0.9395
Phanab	0.2784	0.487	0.735	0.9885
Phamib	0.1252	0.3492	0.5717	0.5798
Phbpib	0.6583	0.6808	0.8224	0.9395
Phbnab	0.1972	0.6878	0.8031	0.9619
Phbmib	0.0228	0.2132	0.3486	0.5733
Chlb	0.2326	0.3046	0.4352	0.9967
Pheb	0.0218	0.2222	0.5898	0.8545
Total Exp.	0.2743	0.4532	0.641	0.8542

Table 4. Global importance coefficients (VIP) assessed for each independent input variable (ratio) in PLS modelling process of surface variables (the coefficients of the most important variables in the full PLS model are reported in bold characters).

Input	Factor 1	Factor 2	Factor 3	Factor 4
b_{12}	0.5501	0.5074	0.9704	0.8326
b_{13}	0.0866	0.6477	0.6732	0.5626
b_{23}	0.1467	0.8414	0.7598	0.6682
b_{34}	1.4103	1.2536	1.1650	**1.2978**
b_{45}	1.6369	1.4295	1.2925	**1.3637**

The values >1.0 of Factor 4 are reported in bold digits.

Table 5. Global importance coefficients (VIP) assessed for each independent input variable (ratio) in PLS modelling process of bottom variables (the coefficients of the most important variables in the full PLS model are reported in bold characters).

Input	Factor 1	Factor 2	Factor 3	Factor 4
b_{12}	0.7779	0.6241	1.1932	1.0339
b_{13}	0.4878	0.9062	0.8097	0.7273
b_{23}	0.2179	1.0493	0.8999	0.8088
b_{34}	1.4930	1.1635	1.0311	**1.1816**
b_{45}	1.3713	1.1552	1.0235	**1.1627**

The values >1.0 of Factor 4 are reported in bold digits.

on the green, red and NIR reflectance responses. Based on the better penetration of coastal-blue shorter wavelengths, the importance of the related b_{12} ratio independent variable (Table 5) becomes more significant in the PLS bottom models. This first modelling step was mainly devoted to analyse the capability of the OLI spectral ratios to capture the different global features of surface and bottom water column phytoplankton and phaeopigment variables while the subsequent more reliable CV approach previously described was exploited for assessing the separate models for sea truth variables and their final distributions. Here each PLS model corresponding to the specific sea truth variable was assessed independently from the others (eq. 1, supplementary materials), with decomposition of only the independent variables (spectral ratios) based on the number of latent factors minimizing the test RMSE. Figure 4 shows some examples of the CV modelling results obtained respectively for phytoplankton abundance, biomass and phaeopigments sea bottom and surface variables. The upper three graphs of figures show the calibration samples (unfilled circles) and model lines trend in red, while the results of the model validation (test) using different samples (filled circles) are drawn in blue. The sea truth measurements at station level provide x axis coordinates while the corresponding values assessed from the spectral reflectance of pixels within the station area provide the y ones. In the middle part of figure the bar graphs of RMSE depending on the number of latent factor decomposition are reported for calibration (red) and validation (blue) step. As previously specified, a variable number of latent factor decomposition minimizing the validation RMSE (indicated as vertical dashed line in the bar graphs) was assessed for each model. As shown in Figure 4 (lower bar graphs), the minimum calibration RMSE is reached using all five latent factor while in any case the validation RMSE minimization requires a decomposition including less latent factors to increase predicting capability avoiding overtraining. For Phapib and Phbnas models (respectively Figure 4(a, b)) the RMSE minimization was achieved by exploiting respectively two and four factors decomposition (Figure 4(d, e)) while for PheB (Figure 4(c)) the minimizing decomposition includes only one factor (Figure 4(f)). In the lower part of Figure 4 the three bar graphs display the regression coefficients assessed for each of independent variables of the three models with the related p-value (t) class as coloured dot at bar ends. In particular a red dot states for p-value(t)<10–2, orange dot means 5*10–2 > p-value(t) >10–2, yellow dot indicates 10–1*>p-value(t)>5*10–2 and void dot specifies 5*10–1*>p-value(t)>10–1, the bar
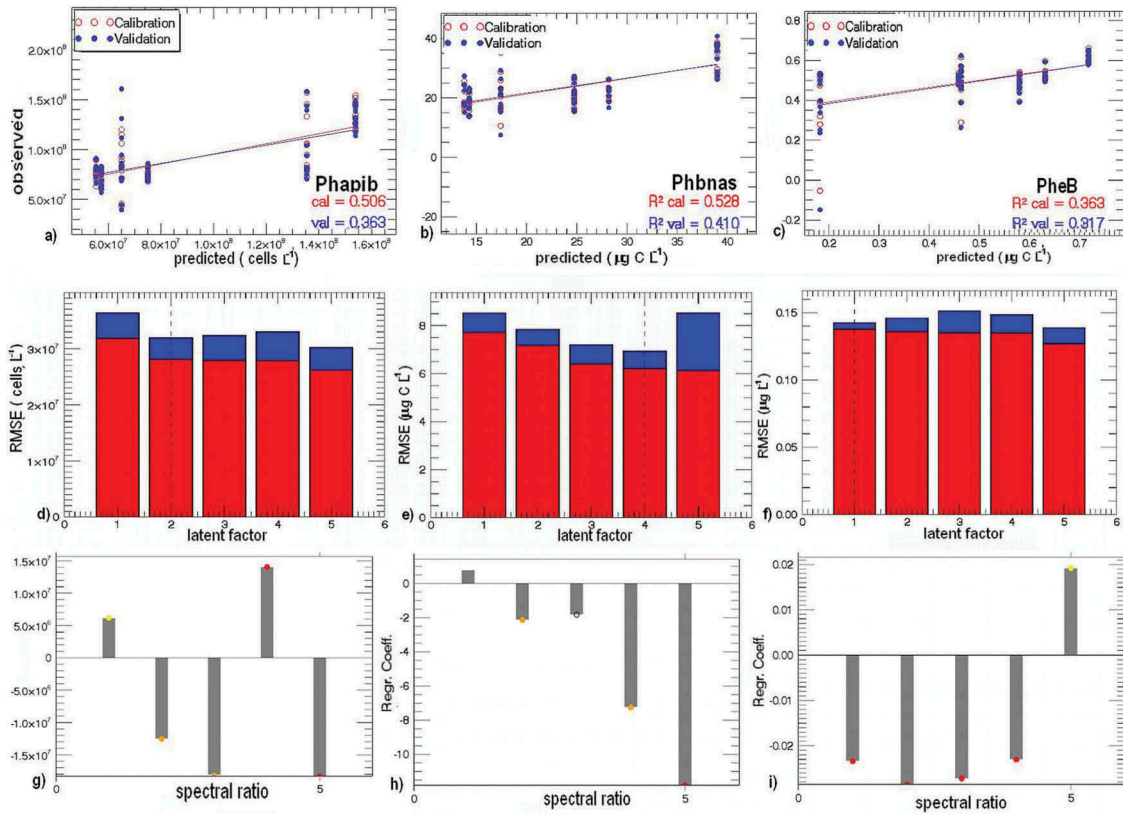
Figure 4. Results of CV (Calibration/Validation) PLS modelling procedure for distributions of sea bottom picoplankton abundance (graphs a, d, g), of nanoplankton surface biomass (graphs b, e, h) and of bottom pheopigments (graphs c, f, i). The upper three graphs show the measured versus modelled sea truth variables for calibration and validation samples, including the related regression lines. The middle graphs display the latent factors selection for test models optimization by RMSE minimizing. The lowest graph report the assessed models coefficients and the related confidence level as p-value(t).

without terminal dot states for p-value(t)>5*10−1. The p-value (t) indicates the confidence level percentage as 100*(1-p-value(t) of coefficient estimate, and can be assumed as proxy of the importance (relative contribution weight) variable in the assessed model of the related independent. The first graph (Figure 4(g)) includes the coefficients of the Phamib model, those referring to b_{13} and b_{34} ratios are positive while the others exhibit negative values. According to their confidence level the most important in model resulted those referring to b_{34} and b_{54} spectral ratios, followed by b_{13} and b_{23} with b_{12} at minimum. The other graphs (Figure 4(h, i)) show the coefficients distribution for Phbnas and PheB variables where the 4th and 5th last ratios, b_{34} and b_{54} maintain highest importance levels in assessed models in agreement with the result of the previous PLS FD modelling.

Table 6 shows the results in terms of explained variance under form of R2 of the various models assessed through CV procedure, based on all spectral ratios measurements extracted from the six station windows (69) for each of sea truth variable. In the first two columns the determination coefficient R^2 and related RMSE obtained from calibration based on the randomly selected 60% of available samples are reported. The same parameters derived from the test phase of the calibrated models (with reduced latent factors for RMSE minimizing) with the remaining

Table 6. Table 3 Explained variance (R^2) and RMSE for the sea truth variables obtained by the models assessed using the calibration-validation (test) CV approach and variable PLS decomposition in term of number of latent factors (Nlat). The R^2 of models test superior than 0.3 are reported in bold characters.

Phyto	Calibration		Test			
	R^2	RMSE	R^2	RMSE	DRMSE	Nlat
Phapis	0.376	1.415E + 07	0.289	1.510E + 07	**0.206**	2
Phanas	0.424	1.659E + 06	**0.304**	1.824E + 06	**0.201**	4
Phamis	0.036	3.193E + 04	−0.232	3.609E + 04	0.791	1
Phbpis	0.376	2.829	**0.329**	2.933	**0.200**	2
Phbnas	0.528	6.209	**0.410**	6.942	0.303	4
Phbmis	0.498	2.880	**0.370**	3.225	**0.221**	4
ChlS	0.229	0.306	0.115	0.328	**0.230**	2
PheS	0.646	0.884	**0.562**	0.984	0.758	4
Phapb	0.506	2.809E + 07	**0.363**	3.191E + 07	0.353	2
Phanab	0.483	2.455E + 06	**0.358**	2.737E + 06	0.360	4
Phamib	0.296	9.786E + 03	0.206	1.039E + 04	0.384	4
Phbpib	0.506	5.619	**0.452**	5.919	0.328	2
Phbnab	0.316	7.546	0.121	8.551	0.461	2
Phbmib	0.070	6.704	0.0	7.121	0.501	1
ChlB	0.339	0.283	0.198	0.312	**0.260**	1
PheB	0.363	0.138	**0.317**	0.142	**0.282**	1

The values of R^2 >0.3 or of DRMSE <0.3 are reported in bold digits.

samples, are then included in the table in test columns (3th and 4th column). The fifth column of this table (DRMSE) contains the ratios between the test RMSE of the PLS model and related averages for each sea truth variable in order to provide a dimensionless and variable-independent parameter able to support the

evaluation of different models. The last column (Nlat) shows the number of latent factors minimizing the RMSE of the model test. As expected, the R^2 values for models test are lower than those obtained in calibration and accordingly the related RMSE's are higher for all models. In Table 6, both the R^2 test values superior than 0.3 and the DRMSE values lower than 3.0 are reported in bold. Due to the spectral variability within the station window, the final coefficient of determination R^2 obtained for PLS models via CV approach are generally lower than those provided by the previous FD PLS calibration step, but in general confirmed reduced values for micro sized phytoplankton populations, except for Phbmis variable. In particular for sea surface abundance (Phamis) and bottom biomass (Phbmib) the values are low and negative with high DRMSEs. In agreement with the achievements of the FD procedure, the surface models of phytoplankton abundance assessed through CV method, on average demonstrated better than those related to sea bottom in term of both R^2 and DRMSE test estimates. Analogously superior performances were obtained for all surface variables of plankton biomass respect to bottom ones. Although the model of PheS phaeopigment surface variable shows a good performance in term of test R^2, its DRMSE is the highest, while that of the ChlB variable happens a bit more performing than ChlS surface counterpart, and both show a low DRMSE (bold digit).

The regression coefficients (eq. 1) assessed for the models of all sea truth variables through CV method, are reported in Table 7, where the target (dependent) and input spectral ratios (independent) variables correspond to related items respectively in columns and rows. According to the colour code exploited in coefficients graphs of Figure 4, here the different cell colour indicates the confidence level of the coefficient assessment under form of p-value(t). In particular, five levels classes are reported starting from the highest confidence level (>99%) that correspond to p-value(t)<10^{-2} in underlined digits (HC class), level between 99% and 95% (0.01 < p-value(t)<0.05) in italic-bold digits (MH class), level between 95% and 90% (0.05 < p-value(t)<0.1) in bold digits (M class), level between 90% and 50% (0.1 < p-value(t)<0.5) in italic digits (ML class), level lower than 50% and p-value(t)<0.5 using normal characters (L class). The models coefficients assessed for the surface picoplankton populations with the highest confidence level (HC class), quantify an inverse linear dependence (negative sign) captured by the shorter wavelength (b_{23}, b_{13}) and b_{54} ratios from the abundance and biomass surface concentrations distributions, which instead were found directly proportional (positive sign) to that related to b_{34} HC ratio. The configuration of coefficients sign is maintained also for the bottom distributions of the same picoplankton variables (abundance and biomass) with decreased importance of b_{23} and b_{13} ratios (MH class), and raising of that related to b_{12} for the abundance model. The most important spectral ratios for models of the abundance and biomass distributions related to sea surface and bottom nanoplankton population remain b_{13}, b_{34} and b_{45} (HC and MH classes) while the shorter wavelength ratio b_{13} diminishes its class importance. Respect to that previously estimated for picoplankton population, the assessed HC class regressive coefficient of the b_{34} ratio shows inverted sign for the nano-plankton populations, for both abundance and biomass models of surface and bottom distributions.

Given the previously highlighted general minor reliability of models estimated for micro-plankton populations distributions the better performance was registered for those related to surface abundance and bottom biomass, with higher importance (HC class) of b45 and b23 ratios and opposite coefficient signs.

Table 7. PLS regression equations coefficients (eq. 1) and confidence levels classes of the related estimates assessed for the sea surface/bottom concentrations of different phytoplankton populations and phaeopigments.

	b_{12}	b_{13}	b_{23}	b_{34}	b_{54}	Interc.
Phapis	8.83E + 07	−5.10E± 07	−9.55E± 07	4.10E± 07	−7.79E± 07	1.57E + 08
Phanas	3.08E + 06	*−3.38E + 06*	−1.60E + 06	−1.24E± 07	−3.59E± 07	6.12E + 07
Phamis	1.85E + 03	−2.49E + 04	−3.94E + 04	1.44E + 04	5.78E + 04	3.20E + 04
Phbpis	*17.651,806*	−10.20,684	−19.099419	8.208,494	−15.577,645	31.3880
Phbnas	30.294,509	*−25.29,179*	−27.142,068	*−43.294,663*	−149.84,891	243.9279
Phbmis	−13.04353	12.383,199	**15.817,397**	18.880,735	62.384,964	−84.3384
ChlS	0.709,906	−1.004835	−1.686,013	−0.225,096	*−1.338,495*	4.9326
PheS	*10.719,051*	−5.908,614	−8.529,972	**−3.988,289**	−21.571,232	30.9746
Phapib	**2.35E + 08**	*−1.49E + 08*	*−2.69E + 08*	8.40E± 07	−2.32E± 08	3.92E + 08
Phanab	−5.97E + 06	*−5.05E + 06*	−2.96E + 06	−2.05E± 07	−6.10E± 07	1.06E + 08
Phamib	*−4.96E + 04*	−1.68E + 04	*−2.02E + 04*	−3.36E + 04	−1.48E± 05	2.96E + 05
Phbpib	*47.027888*	−29.71,806	−53.72,498	16.791,095	−46.401,481	78.3782
Phbnab	17.518,799	*−26.9082*	−44.775,911	−0.447,779	−70.07328	139.8884
Phbmib	*−21.39,124*	−5.436,294	−5.507,671	−3.484,277	2.227,989	47.0050
ChlB	−1.814,057	−0.491,344	*−0.505,258*	−0.357,771	0.702,329	3.6438
PheB	−0.895,285	−0.337,856	−0.403,373	−0.137,735	**0.243,884**	2.0056

The digits format states for five confidence level (C) classes of coefficient estimates: C > 99% (p-value(t)<10−2) in underlined digits (HC class), 99% <C < 95% (0.01 < p-value(t)<0.05) in bold-italic digits (MH class), 95% <C < 90% (0.05 < p-value(t)<0.1) in bold digits (M class), 90% <C < 50% (0.1 < p-value(t)<0.5) in italic digits (ML class), C < 50% (p-value(t)<0.5) in normal characters (L class)

Both the assessed models of Chl *a* and phaeopigments surface concentrations show a most relevant dependence by b_{13}, b_{23} and b_{45} ratios (HC and MH classes), with the concordance of negative sign of the related coefficients. The number of the HC and MH classes coefficients related to bottom concentrations of the same dependent variables increases maintaining the negative signs except for that of the b_{45} ratio in Chl *a* model.

The spectral ratios reflectance of phytoplankton populations dispersed in water mainly depend on their capacity to absorb and reflect the solar radiance in involved wavelengths through water column, dependent on the vegetal photosynthesis and tissues properties in term of cell dimension and pigments contents. Thus, the different behaviour in terms of sign and importance level of the obtained regression coefficients of models is mainly linked to the specific phytoplankton population considered. Even if with varying values, the sign configuration of the coefficients related to the principal ratios (HC and MH classes) of models for the same population is generally maintained not only for abundance and biomass variables, but also for surface and bottom distribution (Table 7). In addition, the models assessed for sea bottom distributions of microplankton present an increase of the importance of the ratios in shorter wavelengths. The opposite sign of b_{34} ratio coefficients related to pico- and nanoplankton populations might denote diverse spectral responses due to their specific structural constituents interacting with the OLI sensor wavelength. Although the sign and importance (HC and MH confidence classes coefficients) of b_{13} and b_{34} ratios contributions in the Chl *a* surface PLS CV model, have been maintained in the corresponding counterparts of bottom model, in this former the influence of the b_{45} ratio is changed (positive coefficient) with all others showing importance increase, including the b_{12} ratio, in agreement with the general trend. The phaeopigment models show a similar behaviour with more reliable estimates of contribution from the different ratios for the sea bottom distributions (on the contrary respect to abundance and biomass models). Globally from the above results, the most important contributions in PLS modelling of all the phytoplankton and phaeopigment/Chl *a* variables at sea surface and bottom, were found those of the b_{45} and b_{13} ratios, confirming the importance of the OLI coastal band (b_1) that in this context performed better than the blue one.

In general, according with the sea truth measurements, the pseudo-colour maps of distribution obtained from the surface and bottom CV models indicate the larger blooms concentration in the second inlet with diverse patterns for different phytoplankton populations and pigments (Figures 5

and 6). In particular, the surface and bottom distributions of pico and nano-plankton abundance (Figure 5(a, b)) and biomass (Figure 6(a)), derived from the most reliable models (high R^2 and low DRMSE) look quite similar, with maxima located along the lower right and left corners of the second inlet. Specular trends were found for the microplankton abundance (less consistent) and biomass surface distributions (Figure 6(b)), with respective concentration minima to the lower right borders of the second inlet where there are maxima of the other populations (Figures 5(a) and 6(a)). Instead, the bottom distributions assessed for microplankton abundance and biomass (less reliable) are in agreement with those estimated for pico and nanoplankton populations, with maxima distributed at lower right borders of the second inlet.

Most of the surface concentrations patterns found for Chl *a* (Figure 5(c)) and phaeopigments maintain a similar configuration of those estimated for the pico and nano plankton population, with maxima located in the same borders portion of the second inlet except of that in the lower left which is a local minimum for Chl *a* (Figure 5(c)).

Their bottom distributions instead look enough specular of their surface counterparts with a location of minima in correspondence to the maxima of surface concentration (Figure 6(c)).

Figure 7 includes Chl *a* maps generated by means of OC3 bio-optical algorithms using atmospherically corrected data remotely sensed by Landsat 8 OLI and Sentinel 2 sensors in different dates. Figure 7(a, b) refers to Chl *a* maps derived from Landsat 8 OLI acquired respectively on 2 July 2014 and on 20 June 2013 while the map of Figure 7(c) was obtained from Sentinel 2 data acquired on 25 July 2015. The multispectral data obtained in 2013, synchronous with the sea truth acquisition have also been used for PLS modelling, whereas the other two in any case refer to the same summer period of contiguous years, so they have been considered for comparing purposes. In agreement with previous results obtained through PLS modelling, the Chl *a* maximum concentrations (red shades), evidenced in the three maps, are located close to the coast of the second inlet. The OC3 map dated in 2013 (Figure 7(b)) exhibits a local minimum (blue shades) in the lower left corner of the second inlet in agreement with the corresponding CV PLS modelled distribution (Figure 5(c)). In Figure 8, the mean surface Chl *a* values obtained from the sea truth measurements and assessed with the different approaches from the remotely sensed data were reported for each of the six stations. Here, considering the synchronous estimates using the Landsat OLI data remotely sensed in June 2013, the results of OC3/2 bio-optical algorithms globally tend to overestimate the sea truth
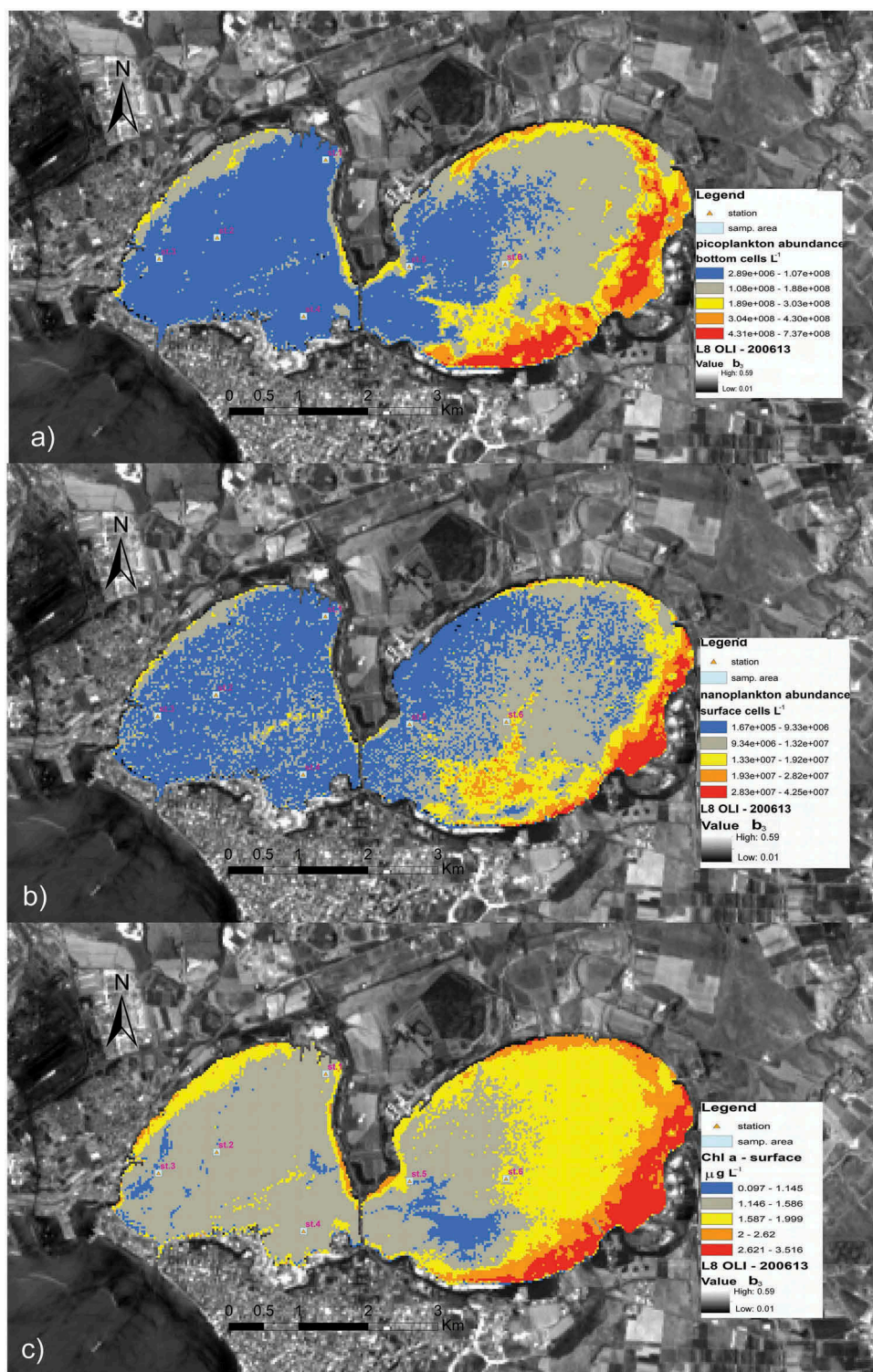
Figure 5. Maps of bottom picoplankton (a), surface nanoplankton abundance (b) and surface Chl *a* (c), assessed using, respectively, Phapis, Phanas and ChlS models derived through CV PLS approach.

measurements more than those obtained via PLS modelling, which in any case benefited from the point calibration. The differences are maintained within 0.5 µg L^{-1} at the stations with more accentuated bathymetry while are particularly increased at the Stations 3 and 5 where the sea bottom contribution might have increased the reflectance noise. Although the OC3 estimations using OLI data acquired in September 2013 and in July 2014 seem to be globally

comparable with sea truth data, due to the more homogeneous seasonal period, the agreement of the latter is higher. The Sentinel 2 OC3 assessments also demonstrated a suitable capacity to capture the Chl *a* summer distribution patterns with plausible uncertainties at the measurement stations.

The map synthetizing the preliminary results obtained from the WASI4 full bio-optical inversion based on default parameters and spectra of the
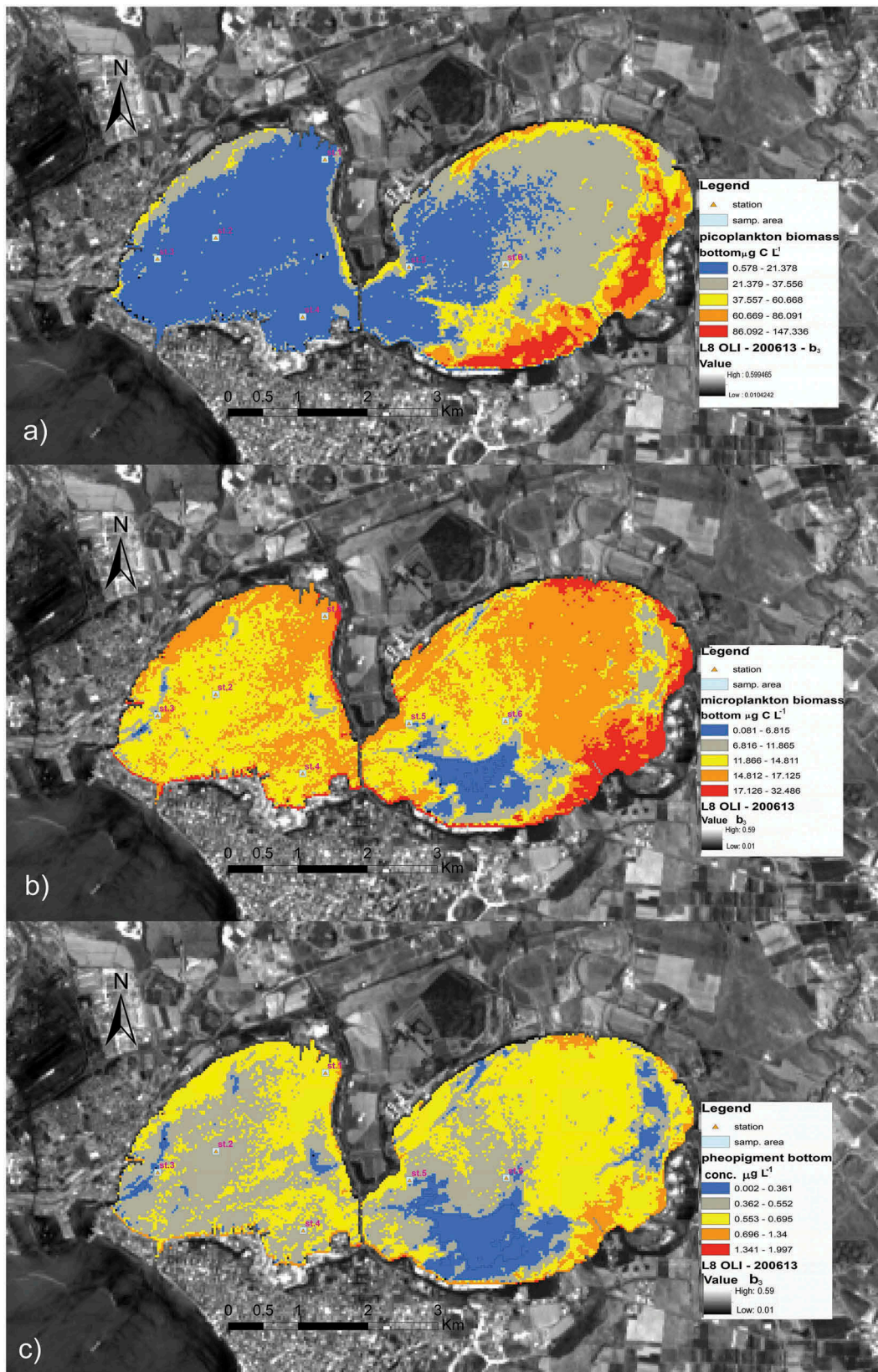
Figure 6. Maps of bottom picoplankton (a), surface microplankton biomass (b) and bottom pheopigments (c), assessed using, respectively, Phbpis, Phbmis and PheB models derived through CV PLS approach.

original database included in the code release, is shown in Figure 9. In agreement with the results of the previous modelling approaches, the maxima of phytoplankton distribution in the Mar Piccolo of Taranto obtained by WASI4 code were approximately localized in the same portion of the second inlet, i.e. in proximity of the coast. Even if the WASI bio-optical model does not allow distinguishing the different phytoplankton population size and despite a salt and pepper diffuse noise in the
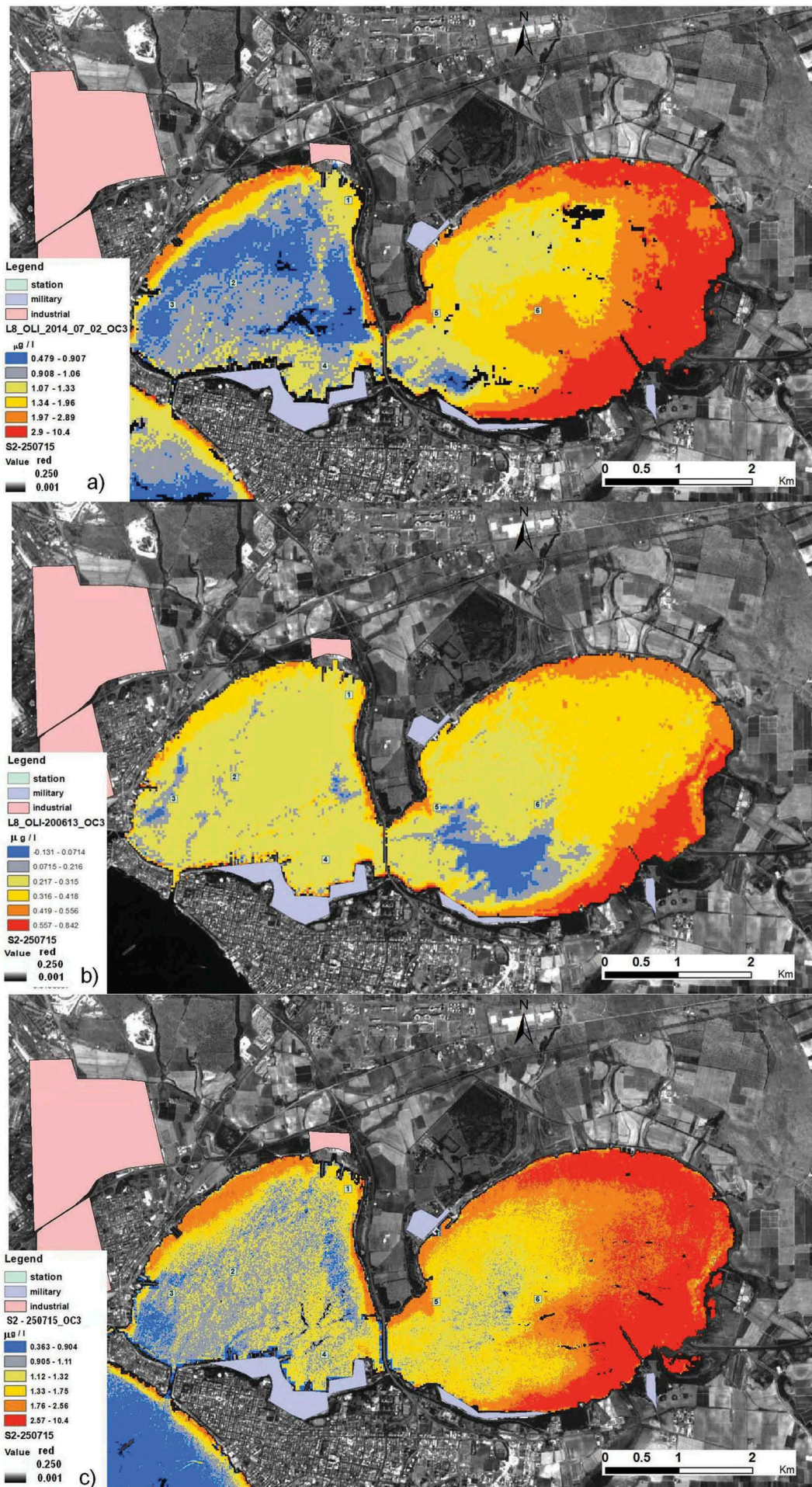
Figure 7. Maps of Chl *a* distributions assessed trough OC3 bio-optical algorithm using atmospherically corrected data of Landsat 8 OLI (a, b) and Sentinel 2 (c), acquired, respectively, on 2 July 2014 (a), on 20 June 2013 (b) and on 25 July 2015 (c).
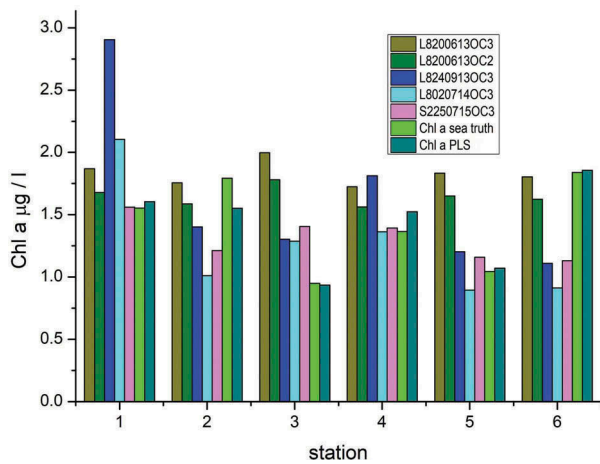
Figure 8. Chl *a* concentration in correspondence of the six sea truth measurement stations estimated using the OC3/OC2 bio-optical algorithm and PLS models using Landsat 8 OLI and Sentinel 2 data.

resulting map, there is a general agreement between the WASI maps and the other ones; this issue is encouraging since there is room for improvement through a more refined exploitation of the code capabilities. In this perspective, our goal is to advance the exploitation of the full optical inversion provided by WASI4 code on the basis of a suitable customization of input parameters provided by available measurements acquired on the area of interest and/or provided by a specific sea truth campaign.

The CV PLS modelling of surface and bottom phytoplankton populations, phaeopigment and Chl *a* distributions through Landsat 8 OLI atmospherically corrected reflectance responses reached a satisfying level with high level of agreement with the sea truth point measurement stations. In this context, given the unavailability of MODIS Ocean

Colour products for coastal shallow water and lagoons, the comparison with MODIS Chl *a* maps (1 km of geometric resolution) derived from for the open sea portion of Mar Grande (Figure 1, lower left corner), evidence a compatibility of estimates within 0.05 µg L^{-1} (Dazhao et al., 2010; Dierssen, 2009).

These results demonstrate the capabilities of the Landsat 8 OLI orbiting sensor to usefully capture the different spectral responses from coastal distribution of phytoplankton, Chl *a* and phaeopigment surface and bottom concentrations. In particular, its atmospherically corrected green-red-NIR spectral ratios, through these PLS implemented methods was more suitable for modelling the surface distribution of abundance and biomass of the phytoplankton populations. Furthermore, the contribution of coastal-blue ratio demonstrated more useful to capture the subsurface concentration of the same phytoplankton population. The implemented PLS approach seemed less effective for modelling the large-sized bottom population of micro-phytoplankton. Although the PLS approach demonstrated the minor importance of the usual blue–green ratio (b_{23}) for Chl *a* and phaeopigments distribution modelling, a general agreement was found with the Chl *a* distributions assessed through OCX bio-optical algorithms and the WASI4 full bio-optical inversion. In any case, differences between the Landsat 8 OLI detection capability of different surface and bottom features of phytoplankton population require additional studies and sea truth information to be better explained in terms of spectral and/or dimensional characteristics of phytoplankton cells. Further, in these future studies attention should be paid to exclude noise effects from other optically active
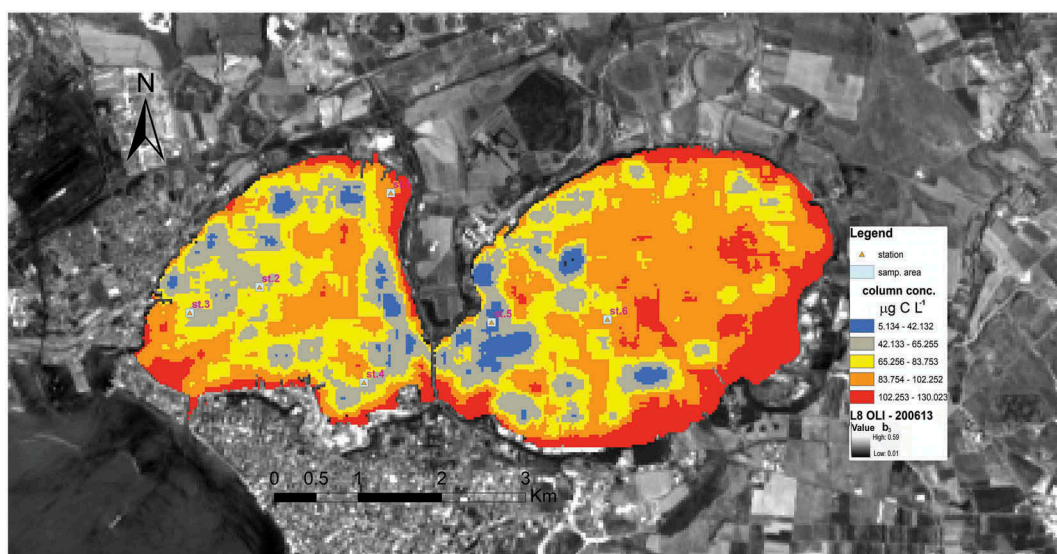

Figure 9. Phytoplankton concentration distribution on the whole column in the Mar Piccolo di Taranto obtained by atmospherically corrected Landsat 8 OLI data of June 2013, using WASI4 model.

components, dissolved in the water column crossed by detected signals, and from the sea bed cover reflectance contribution.

The results of this study demonstrate that the spectral/radiometric improvements of the OLI sensor increase the overall efficiency in its marine applications, in particular for the effective characterization of phytoplankton populations of shallow waters, including coastal lagoons. These achievements also indicate a good opportunity to apply in this type of marine areas, like the Mar Piccolo, the new generation HR multispectral satellite sensors (i.e. Sentinel 2) which have recently had a remarkable development. In particular, the introduction of the new "coastal" band, was advantageous for monitoring phytoplankton that in turn, plays a significant role in the primary productivity and in the carbon storage of coastal productive ecosystems. It is expected to improve the results obtained in this study with a customization of the WASI4 system database based on the typical spectra of the sea bed covers and dissolved optically active components present in the area of interest and involved in the bio-optical modelling.

Conclusion

The main objective of this study was to preliminary evaluate the new features of Landsat 8 OLI sensor for extensive and systematic monitoring of phytoplankton and its blooms in the Mar Piccolo of Taranto. Here, as in other Italian coastal water environments, several anthropogenic activities give rise to major load of pollutants, nutrients and suspended sediments, with reduction in water transparency and considerable development of macroalgal and potentially harmful phytoplankton species for aquaculture activities. The spatial and temporal dynamics of phytoplankton and Chl a concentrations need suitable synchronization and distribution of sea truth samples collection to capture an effective snapshot of their status, useful also to provide a basic input for calibration of remotely sensed data acquired during the reference overpasses of satellite. Given these constraints, collecting a significant number of useful samples with these characteristics is very hard and usually requires human and instrumental resources frequently unavailable. In this context, the implemented CV procedure allowed to usefully exploit the limited number of our measurement stations as best trade off compatible with more robust statistic approach and also with the limits in the accessibility of various subareas of the basins occupied by military infrastructures and mussels farms. The OLI multispectral data, in form of selected ratio indices corrected for atmospheric noises, allowed us to assess the distributions of phytoplankton and Chl a concentration at different depths of entire basin, in acceptable agreement with sea truth data. This CV PLS original approach,

mainly based on the exploitation of spectral responses variability within the station sea truth sampling area, provided also the different importance for the spectral ratios in assessing the phytoplankton distributions. In addition to red/NIR (b_{45}) for surface variables, the OLI coastal/green (b_{13}) ratio demonstrated more performant than usual blue/green (b_{23}) for retrieving the phytoplankton and Chl distributions (surface/bottom) in these coastal optically complexes waters. The implemented EO-based methodology can support the prediction of the available biomass growth at suitable temporal and spatial scale according to the seasonal cycles. Since many coastal shallow waters are highly productive but sensitive and vulnerable to planktonic blooms, they require effective multiscale tools based on integration of HR remote sensing and in situ techniques. In this context, the valence of this remote sensing based monitoring approach characterized by early detection and repeating HR mapping capabilities is twofold: an effective tool both for early warning in case of harmful phytoplankton blooms and for mapping their biomass distribution, a prerequisite for sustainable management interventions.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Flavio Borfecchia http://orcid.org/0000-0002-2388-8247
Carla Micheli http://orcid.org/0000-0001-7277-8063
Tamara Cibic http://orcid.org/0000-0003-1519-4891

Carmela Caroppo http://orcid.org/0000-0002-0572-1048

Fernando Rubino http://orcid.org/0000-0003-2552-2510

Martina Kralj http://orcid.org/0000-0003-1544-5515

Paola Del Negro http://orcid.org/0000-0003-2465-4896

References

Albert, A., & Gege, P. (2006). Inversion of irradiance and remote sensing reflectance in shallow water between 400 and 800 nm for calculation of water and bottom properties. *Applied Optics*, 45, 2331–2343.

Blondeau-Patissier, D., Gower, F.R.J., Dekker, G.A., Phinn, R.S., & Brando, E.V. (2014). A review of ocean color remote sensing methods and statistical techniques for the detection, mapping and analysis of phytoplankton blooms in coastal and open oceans. *Progress in Oceanography*, 123, 123–144. doi:10.1016/j.pocean.2013.12.008

Borfecchia, F., Cimbelli, A., De Cecco, L., Della Rocca, A. B., Martini, S., Barbini, C., … Ribezzo, A.S. (1997). *Integrated remote sensing mission in the Venice Lagoon*. International Society for Optics and Photonics Proc. SPIE2959, Satellite Remote Sensing III (pp. 162–170). 1996, Taormina, Italy.

Borfecchia, F., Consalvi, N., Micheli, C., Carli, F., Cognetti De Martiis, S., Gnisci, V., … Marcelli, M. (2018c). Landsat 8 OLI satellite data for mapping of the Posidonia oceanica and benthic habitats of coastal ecosystems. *International Journal of Remote Sensing*, 39 (18), 1–18. doi:10.1080/01431161.2018.1528020

Borfecchia, F., Micheli, C., Belmonte, A., De Cecco, L., Sannino, G., Bracco, G., … Struglia, M.V. (2016). *Impact of ISWEC sea wave energy converter on posidonia oceanica meadows assessed by satellite remote sensing in the coastal areas of Pantelleria island*. EGU General Assembly Conference, Geophysical Research Abstracts, Vol. 18, EGU2016-16787, Vien, Austria.

Borfecchia, F., Micheli, C., De Cecco, L., Sannino, G., Struglia, M.V., Di Sarra, A.G., … Mattiazzo, G. (2018b). *Mapping the seagrass and coastal habitats of Mediterranean Islands using the new HR satellite multi-spectral sensors. Id. 015*. ESA 2th MWBS (Mapping Water Bodies from Space) international conference, 27-28 March 2018 ESA-ESRIN, Frascati Rome.

Borfecchia, F., Micheli, C., De Cecco, L., Sannino, G., Struglia, M.V., Di Sarra, A.G., … Mattiazzo, G. (2018d). *Mapping the seagrass and coastal habitats of Mediterranean Islands using the new HR satellite multi-spectral sensors. Id. 015*, ESA 2th MWBS (Mapping Water Bodies from Space) conference, 27-28 March 2018, ESA-ESRIN, Frascati, Rome

Borfecchia, F., Rubino, F., Cibic, T., Caroppo, C., Cecere, E., De Cecco, L., … Micheli, C. (2018a). Satellite mapping of macro-algae and phytoplankton communities in the Mar Piccolo of Taranto (Ionian Sea, southern Italy), a confined marine basin heavily impacted by anthropogenic activities. *20th EGU General Assembly (EGU2018-6606-13)*. Retrieved from http://adsabs.harvard.edu/abs/2018EGUGA.20.6606B

Cardellicchio, N., Covelli, S., & Cibic, T. (2016). Integrated environmental characterization of the contaminated marine coastal area of Taranto, Ionian Sea (southern Italy). *Environmental Science and Pollution Research*, 23, 12491–12494. doi:10.1007/s11356-016-6836-5

Caroppo, C., Cerino, F., Auriemma, R., & Cibic, T. (2016). Phytoplankton dynamics with a special emphasis on harmful algal blooms in the Mar Piccolo of Taranto (Ionian Sea, Italy). *Environmental Science and Pollution Research*, 13, 12691–12706. doi:10.1007/s11356-015-5000-y

Cibic, T., Bongiorni, L., Borfecchia, F., Di Leo, A., Franzo, A., Giandomenico, S., & Del Negro, P. (2016). Ecosystem functioning approach applied to a large contaminated coastal site, the study case of the Mar Piccolo of Taranto (Ionian Sea). *Environmental Science and Pollution Research*, 23(13), 12739–12754. doi:10.1007/s11356-015-4997-2

Cibic, T., Cerino, F., Karuza, A., Fornasaro, D., Comici, C., & Cabrini, M. (2018). Structural and functional response of phytoplankton to reduced river inputs and anomalous physical-chemical conditions in the Gulf of Trieste (northern Adriatic Sea). *Science of the Total Environment*, 636, 838–853. doi:10.1016/j.scitotenv.2018.04.205

Dazhao, L., Chuqun, C., Jieqiong, G., & Dongyang, F. (2010). Remote sensing of chlorophyll-a concentrations of the Pearl River estuary from MODIS land bands. *International Journal of Remote Sensing*, 31(17–18), 4625–4633. doi:10.1080/01431161.2010.485212

De Boer, J. (2016). Quality in scientific research. In commemoration of Dr. David Wells. Editorial. *Chemosphere*, 154, A1eA2. doi:10.1016/j.chemosphere.2016.04.034

De Keukelaere, L., Sterckx, S., Adriaensen, S., Knaeps, E., Reusen, I., Giardino, C., … Vaiciut, D. (2018). Atmospheric correction of Landsat-8/OLI and Sentinel-2/MSI data using iCOR algorithm: Validation for coastal and inland waters. *European Journal of Remote Sensing*, 51(1), 525–542. doi:10.1080/22797254.2018.1457937

Dierssen, M.H. (2009). Perspectives on empirical approaches for ocean color remote sensing of chlorophyll in a changing climate. In D.M. Karl (Ed.), *PNAS Perspective*. Honolulu: University of Hawaii. 107(40), 17073-17078, approved August 24, 2010.

El Hourany, R., Fadel, A., Gemayel, E., Abboud-Abi Saab, M., & Faour, G. (2017). Spatio-temporal variability of the phytoplankton biomass in the Levantine basin between 2002 and 2015 using MODIS products. *Oceanologia*, 59, 153–165. doi:10.1016/j.oceano.2016.12.002

Franz, B.A., Bailey, S.W., Kuring, N., & Werdella, P.J. (2015). Ocean color measurements with the Operational Land Imager on Landsat-8, implementation and evaluation in SeaDAS. *Journal of Applied Remote Sensing*, 9(096070), 1–16. *Remote Sensing.SPIE DigitalLibrary.org*. doi:10.1117/1.JRS.9.096070.

Franzo, A., Auriemma, R., Nasi, F., Vojvoda, J., Pallavicini, A., Cibic, T., & Del Negro, P. (2016). Benthic ecosystem functioning in the severely contaminated Mar Piccolo of Taranto (Ionian Sea, Italy), focus on heterotrophic pathways. *Environmental Science and Pollution Research*, 23(13), 12645–12661. doi:10.1007/s11356-015-5339-0

Gege, P. (2014). WASI4, A software tool for regionally optimized analysis of imaging spectrometer data from deep and shallow waters. *Computer and Geosciences*, 62, 208–215. doi:10.1016/j.cageo.2013.07.022

Gholizadeh, H.A., & Robeson Scott, M. (2016). Revisiting empirical ocean-colour algorithms for remote estimation of chlorophyll-a content on a global scale. *International*

Journal of Remote Sensing, *37*(11), 2682–2705. doi:10.1080/01431161.2016.1183834

Hu, C., Lee, Z., & Franz, B. (2012). Chlorophyll a algorithms for oligotrophic oceans, a novel approach based on three-band reflectance difference. *Journal of Geophysical Research*, *117*, C 01011. doi:10.1029/2011JC007395

Karuza, A., Caroppo, C., Monti, M., Camatti, E., Di Poi, E., Stabili, L., & Del Negro, P. (2016). End to end planktonic trophic web and its implications for the mussel farms in the Mar Piccolo of Taranto (Ionian Sea, Italy). *Environmental Science and Pollution Research*, *23*(13), 12707–12724. doi:10.1007/s11356-015-5621-1

Kennish, M.J., Brush, M.J., & Moore, K.A. (2014). Drivers of change in shallow coastal photic systems, an introduction to a special issue. *Estuaries & Coasts*, *37*(1), 3–19. doi:10.1007/s12237-014-9779-4

Kimberly, R., & Khalid, A. (2016). Application of a partial least-squares regression model to retrieve chlorophyll-a concentrations in coastal waters using hyper-spectral data. *Ocean Science Journal*, *51*(2), 209–221. doi:10.1007/s12601-016-0018-8

Kralj, M., De Vittor, C., Comici, C., Relitti, F., Auriemma, R., Alabiso, G., & Del Negro, P. (2016). Recent evolution of the physical-chemical characteristics of a site of national interest – the Mar Piccolo of Taranto (Ionian Sea) – and changes over the last 20 years. *Environmental Science and Pollution Research*, *23*(13), 12675–12690. doi:10.1007/s11356-015-5198-8

Micheli, C., Borfecchia, F., De Cecco, L., Belmonte, A., Piermattei, V., & Marcelli, M. (2018). *Effects of anthropogenic pressure on posidonia oceanica (L) delile (1813) ecosystems.* 20th EGU General Assembly (EGU2018-6606-13). Vien(Austria)

Micheli, C., Borfecchia, F., De Cecco, L., Martini, S., Ceriola, G., Bollanos, S., & Campbell, G. (2010). Seagrass monitoring by remote sensing in the context of biodiversity conservation. *Rapports Commission Internationale Marina Mediterranea*, *39*, 778–779.

Micheli, C., Cupido, R., Lombardi, C., Belmonte, A., & Peirano, A. (2012). Changes in genetic structure of posidonia oceanica at monterosso al mare (Ligurian Sea) and its resilience over a decade (1998–2009). *Environmental Management*, *50*, 598–606. doi:10.1007/s00267-012-9917-3

Micheli, C., D'Esposito, D., Belmonte, A., Peirano, A., Valiante, L.M., & Procaccini, G. (2015). Genetic diversity and structure in two protected Posidonia oceanica meadows. *Marine Environmental Research*, *109*, 124–131. doi:10.1016/j.marenvres.2015.06.016

Mumby, P.J., Clark, C.D., Green, E.P., & Edwards, A.J. (1998). Benefits of water column correction and contextual editing for mapping coral reefs. *International Journal of Remote Sensing*, *19*(1), 203–210. doi:10.1080/014311698216521

Pahlevan, N.Z., Lee, J., Wei, B., Schaaf, C., & Schott, J.R. (2014). On-orbit radiometric characterization of OLI (Landsat 8) for applications in aquatic remote sensing. *Remote Sensing of Environment*, *154*, 272–284. doi:10.1016/j.rse.2014.08.001

Pasqualini, V., Pergent-Martini, C., Pergent, G., Agreil, M., Skoufas, G., Sourbes, L., & Tsirika, A. (2005). Use of SPOT 5 for mapping seagrasses, an application to Posidonia Oceanica. *Remote Sensing of Environment*, *94*, 39–45. doi:10.1016/j.rse.2004.09.010

Petrocelli, A., Cecere, E., Portacci, G., Micheli, C., De Cecco, L., Martini, S., & Borfecchia, F. (2015). Preliminary mapping of the alien seaweed Hypnea Cornuta (Rhodophyta, Gigartinales) in the Mar Piccolo of Taranto (Southern Italy, Mediterranean Sea). *Biologia Marina Mediterranea*, *22*(1), 44–45.

Rhodes, L. (2011). World-wide occurrence of the toxic dinoflagellate genus Ostreopsis Schmidt. *Toxicon*, *57*(3), 400–407. doi:10.1016/j.toxicon.2010.05.010

Richter, R., Schlapfer, D., & Muller, A. (2006). An automatic atmospheric correction algorithm for visible/NIR imagery. *International Journal of Remote Sensing*, *27*, 2077–2085. doi:10.1080/01431160500486690

Smayda, T.J. (2007). Reflections on the ballast water dispersal – Harmful algal bloom paradigm. *Harmful Algae*, *6*(4), 601–622. doi:10.1016/j.hal.2007.02.003

Vanhellemont, Q., & Ruddick, K. (2014). Turbid wakes associated with offshore wind turbines observed with Landsat 8. *Remote Sensing of Environment*, *145*, 105–115. doi:10.1016/j.rse.2014.01.009

Vanhellemont, Q., & Ruddick, K. (2015). Advantages of high quality SWIR bands for ocean colour processing, Examples from Landsat-8. *Remote Sensing of Environment*, *16*, 89–106. doi:10.1016/j.rse.2015.02.007