

CymoHub: Database for *Cymodocea nodosa* Seagrass Meadows Segmentation using WorldView Satellite Imagery

A. Mederos-Barrera^{*1}, L. Albors², F. Marques², J. Marcello¹, G. Martinez and F. Eugenio¹

¹ Institute of Oceanography and Global Change, University of Las Palmas de Gran Canaria (ULPGC), 35001, Las Palmas de Gran Canaria, Spain.

² Department of Signal Theory and Communications, Technical University of Catalonia (UPC), 08034, Barcelona, Spain.

* Corresponding author. Email: mederosbarrera.antonio@gmail.com

This document presents the procedures and techniques used for the generation of the database for the segmentation of *Cymodocea nodosa* seagrass meadows with WorldView-2 satellite images, where the study area chosen was El Río, Canary Islands, Spain.

Regarding the structure of this document, initially in the first point a summary of the study area can be found. Then, in the second point, information relevant to the in-situ measurement campaign is presented. Subsequently, the generation of the database is explained. Later, information on how to obtain the WorldView multispectral images is presented. Finally, a summary of the database data is presented.

IMPORTANT NOTE: Since Maxar did not grant permission to publish the multispectral images in open access, the masks can be found in the published dataset. Section 4 of this document provides a tutorial on how to obtain images. Please contact the author if any questions arise.

1. Study area

The area selected for the database is in El Río, a strait between the islands of La Graciosa and Lanzarote, Canary Islands, between 20° 13' 29" N and 20° 12' 12" N, and 13° 31' 3" W and 13° 29' 40" W. In this area there have historically existed seagrass meadows of *Cymodocea nodosa*. The importance of the place has been key to establish different conservation areas such as: Spanish Natural Park since 1986, Special Area of Conservation Natura 2000 since 2011 (ZEC-ES7010020), or Special Protection Area for Birds since 2014 (ZEPA- ES0000532) [1, 2, 3]. It should be noted that, in addition to the presence of *Cymodocea nodosa* seagrass meadows, there are also other marine flora of interest in the area, such as the presence of *Caulerpa prolifera*, filamentous red algae and, to a lesser extent, *Dictyota* sp., *maërl* and *Lobophora* sp. The study area can be seen in Figure 1.

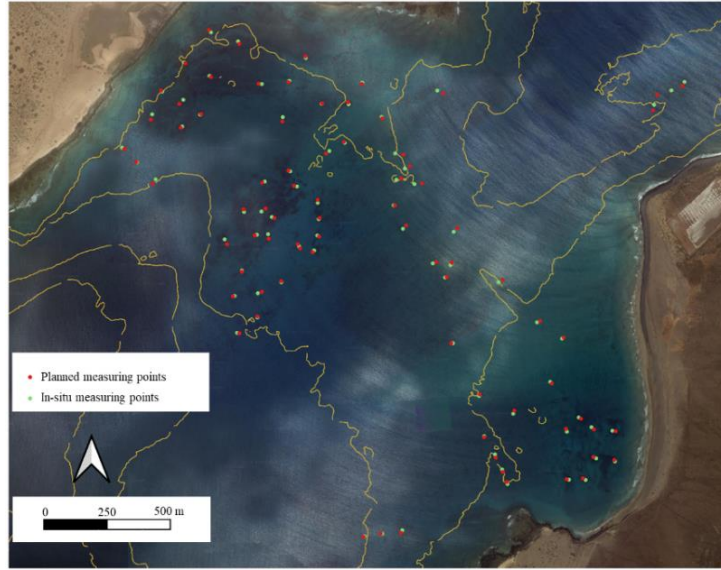


Figure 1. Study area in El Río, strait between the islands of La Graciosa and Lanzarote, Canary Islands.

2. In-situ measurement campaign

To generate the database, an in-situ measurement campaign was initially carried out. For this purpose, the path from the surface to the seabed was recorded at 78 different measurement points. The location of each point was planned according to the distribution of *Cymodocea nodosa* meadows in previous studies [4]. In addition, different equipment was used to capture the videos: an inflatable boat, a GoPro Hero 9 video camera, a mount for the videos to be taken vertically by means of a weight, and a Garmin eTrex 10 GPS geopositioning system. This equipment is shown in Figure 2.



Figure 2. Equipment used in the in-situ measurement campaign: (a) Inflatable boat, (b) GoPro Hero 9 camera, (c) Mount for the verticality of the videos, and (d) Garmin eTrex 10 GPS.

From the analysis of the videos, it was possible to observe different types of substrates: rocky platform, stony ground, rock clogged with sediments, sand and mixed bottom (sand and rock). Figure 3 shows examples of the different substrates captured. Regarding the areas with rocky substrates, a total of 33 measurement points were explored, where red filamentous algae communities were observed to a greater extent, followed by *Dictyota* sp. and *Lobophora* sp. In addition, on the sediment-ridden rocky bottoms, the presence of red filamentous algae and *Dictyota* sp. was noted. On the mixed substrate of sand and rock, diverse algal comodels formed by *maërl*, *Dictyota* sp., red filamentous algae and *Lobophora* sp. were observed. Finally, with respect to the sandy substrate, a predominance of *Cymodocea nodosa* meadows was observed, in addition to the presence of *Caulerpa prolifera* and *Dictyota* sp. The presence of *Cymodocea nodosa* together with *Caulerpa prolifera* in sandy bottoms is noteworthy, as well as the notable presence of filamentous red algae in rocky or mixed areas. Figure 4 shows examples of captures of the predominant marine flora.

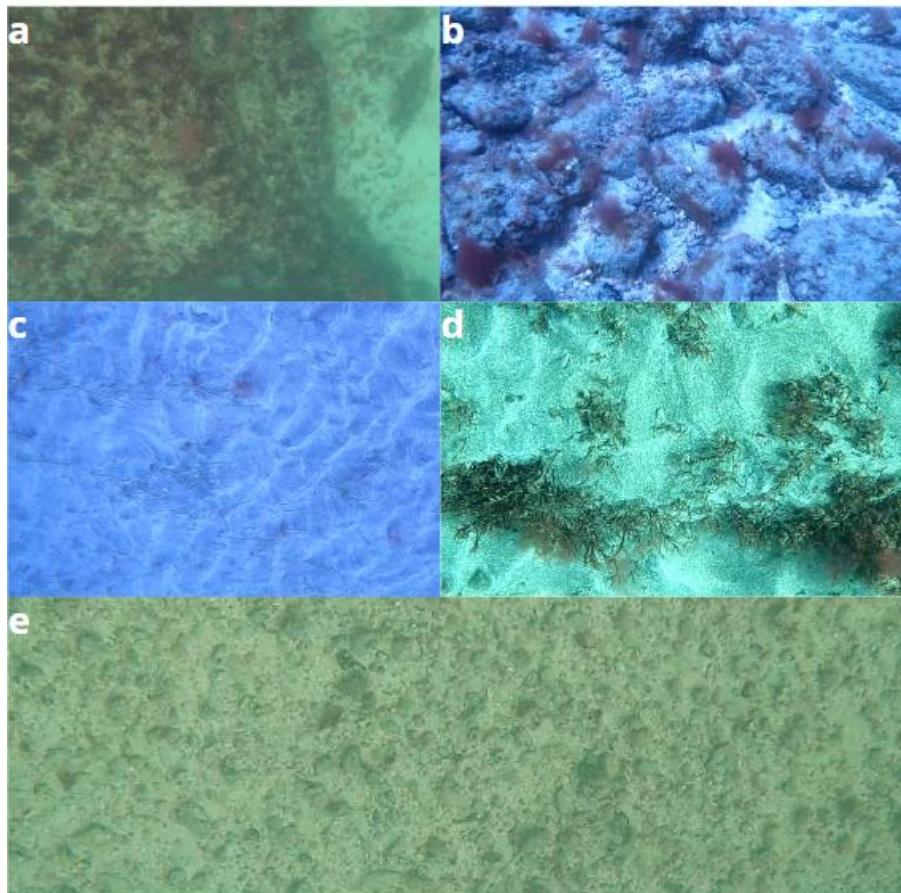


Figure 3. Detected substrates: (a) Rocky platforms, (b) Stony ground, (c) Sand, (d) Sediment clogged rock, and (e) Mixed substrates (sand and rocks).

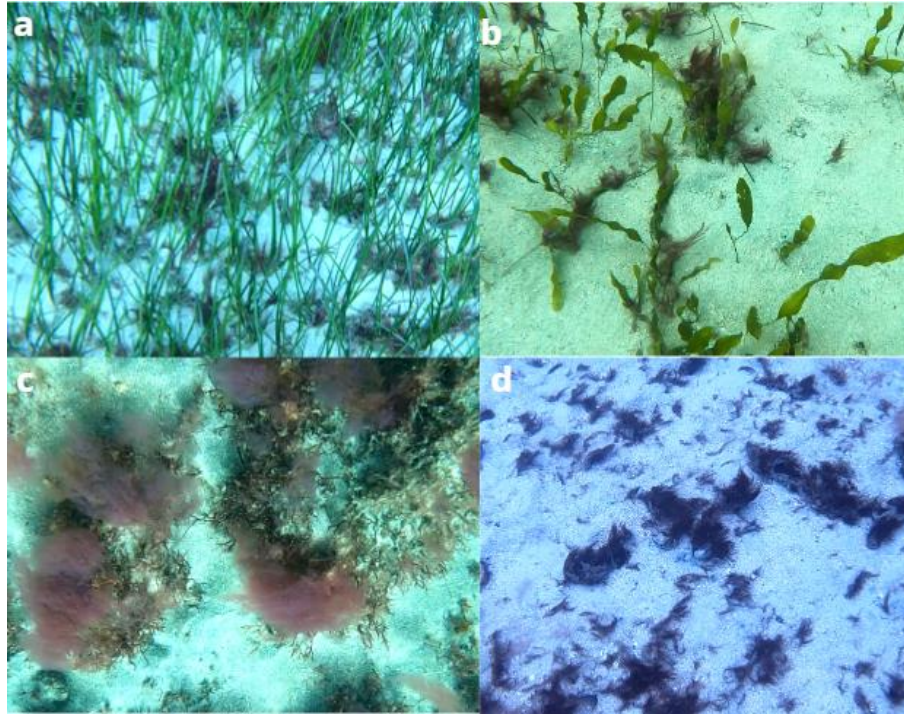


Figure 4. Flora communities: (a) *Cymodocea nodosa*, (b) *Caulerpa prolifera*, and (c) (d) filamentous red algae.

3. Database generation

The WorldView-2 image of January 22, 2022, with a spatial resolution of 2 m was used to generate the database. It should be noted that the image was preprocessed to suppress unwanted effects in the capture of the satellite image. The preprocessing used were [5]: water mask using the Normalized Difference Water Index (NDWI), radiometric corrections to convert digital numbers in radiances at the top of the atmosphere (ToA), atmospheric corrections with the Second Simulation of a Satellite Signal in the Solar Spectrum (6S) model to obtain the reflectance at the bottom of the atmosphere (BoA), and the Hedley glint correction to suppress the noise produced by the waves. With the preprocessed image, four different types of seabeds were chosen, that can be detected in the satellite image given the spatial and spectral resolution. The seabeds chosen were: land, with 0 values in mask images; filamentous red algae, with values 1; *Cymodocea nodosa* (with the presence of *Caulerpa prolifera*), with values 2; rock without vegetation, with values 3; and sand, with value 4.

Subsequently, a manual segmentation of the WorldView-2 image was performed with the information from the videos. 51 image crops were obtained with a dimension of 70x70 pixels, or 140x140 meters, maximizing the surface of these images. These images, with the mask images, make up the database. It should be noted that these images also consider the land pixels, which have been previously masked, so they are of zero value. Figure 5 shows the preprocessed WorldView-2 image, as well as the cropped images corresponding to the datasets.



Figure 5. WorldView-2 image of January 22, 2022, with the corresponding database images.

As can be seen in Figure 5, the images are divided by zones, which correspond to different types of bottoms:

- *Zone 1*: Rock without vegetation.
- *Zone 2*: Filamentous red algae and sand.
- *Zone 3*: Filamentous red algae and sand.
- *Zone 4*: Sand. This zone is of special mention, since deep water pixels are considered as sand.
- *Zone 5*: *Cymodocea nodosa* and sand.
- *Zone 6*: Filamentous red algae and sand.
- *Zone 7*: Filamentous red algae and sand.
- *Zone 8*: *Cymodocea nodosa* and sand.
- *Zone 9*: Filamentous red algae and sand.

It should be noted that Data Augmentation has been applied exclusively to rotations, with 90°, 180°, and 270°. These rotations have been chosen to avoid a modification of the spectral and spatial information. Therefore, there are a total of 204 images in the database.

Finally, some examples of image and masks pairs are presented in Figure 6.

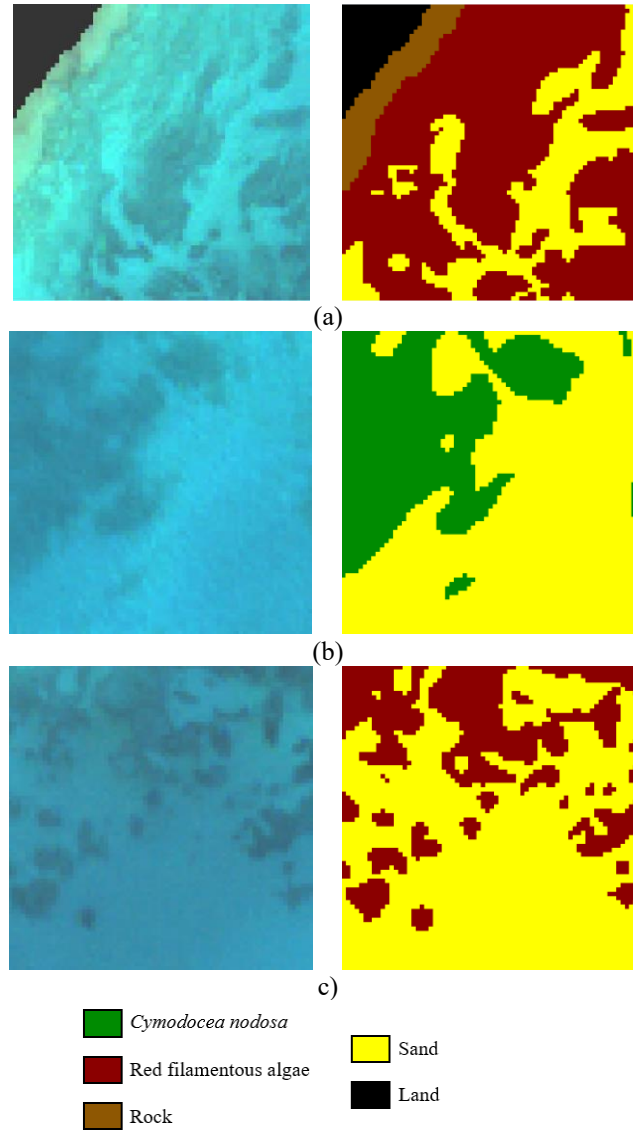


Figure 6. Examples of image (left) and mask (right) pairs without Data Augmentation: (a) “2” example, (b) “25” example, (c) “8” example.

4. Imaging acquisition

Since Maxar did not give permission to publish the images, only the masks were published. However, here is a summary of how to acquire individual images. To do so, you must first acquire the WorldView-2 image from January 22, 2022, between the coordinates (WGS 1984 UTM Zone 28N):

- Top: 3233785.582270 m.
- Bottom: 3231493.680970 m.
- Left: 643357.856035 m.
- Right: 646339.792135 m.

The resulting imaging will have a dimension of 1452 columns, 1116 rows, and 8 channels.

To generate the zones, you can choose the following indices by zone:

- Zona 1: [914, 976]
- Zona 2: [915, 1046]
- Zona 3: [844, 1046]
- Zona 4: [844, 976]
- Zona 5: [773, 976]
- Zona 6: [1062, 866]
- Zona 7: [1027, 796]
- Zona 8: [1097, 796]
- Zona 9: [1263, 854]
- Zona 10: [1193, 854]
- Zona 11: [1263, 784]
- Zona 12: [1192, 784]
- Zona 13: [1169, 714]
- Zona 14: [1092, 523]
- Zona 15: [1022, 475]
- Zona 16: [951, 475]
- Zona 17: [882, 421]
- Zona 18: [1418, 2]
- Zona 19: [1344, 113]
- Zona 20: [1415, 73]
- Zona 21: [1344, 41]
- Zona 22: [541, 538]
- Zona 23: [540, 467]
- Zona 24: [501, 397]
- Zona 25: [573, 396]
- Zona 26: [505, 326]
- Zona 27: [576, 325]
- Zona 28: [543, 253]
- Zona 29: [614, 254]
- Zona 30: [647, 325]
- Zona 31: [398, 534]
- Zona 32: [388, 605]
- Zona 33: [262, 555]
- Zona 34: [322, 481]
- Zona 35: [194, 248]
- Zona 36: [231, 177]
- Zona 37: [265, 248]
- Zona 38: [336, 248]
- Zona 39: [302, 177]
- Zona 40: [373, 177]
- Zona 41: [443, 178]
- Zona 42: [260, 106]
- Zona 43: [330, 106]
- Zona 44: [380, 36]
- Zona 45: [309, 35]
- Zona 46: [401, 106]
- Zona 47: [834, 340]

- Zona 48: [807, 457]
- Zona 49: [668, 405]
- Zona 50: [908, 279]
- Zona 51: [979, 395]

As can be seen, since all images correspond to 70x70 pixels, only the position of the upper left corner is indicated in [row, column] format.

As mentioned above, if you have any questions, please contact the author.

5. Database information

This section summarizes the information on the database data:

- *Number of images*: 204 with their respective mask images (51 without Data Augmentation).
- *Number of classes*: 5 (land, with values of 0 in the mask images; filamentous red algae, with values 1; *Cymodocea nodosa*, with values of 2; unvegetated rock, with values 3; and sand, with values 4).
- *Spatial resolution of images*: 2 m.
- *Size of images*: 70x70 pixels or 140x140 m.
- *Depth of images*: 8 WorldView-2 channels.
- *Data Augmentation*: rotations of 90°, 180°, and 270°.
- *Type of archives*: NumPy (.npy) and MatLab (.mat).

In addition, if the database is to be used in supervised learning, the following images are recommended for validation: 6, 10, 9, 7, 13, 17, 19, 25, 29, 34, 37, 42, 46, and 50. The rest of the images are recommended for training. Bathymetric information has been used for their choice and class overfitting has been avoided. However, other distributions may also be suitable. Furthermore, in the case of using Data Augmentation, it is recommended to use the images with Data Augmentation in the same training or validation dataset. For example, if image 6 is used for validation, it is recommended to use the image with the rotations (90°, 180° and 270°) in the validation.

Finally, a folder is available in the database with different subfolders. The structure and information of the folders and subfolders is shown below:

- “Data” folder with the images applying Data Augmentation (rotations of 90°, 180°, and 270°).
 - “Masks”:
 - “MAT”.
 - “NPY”.

References

- [1] L. Brown and M. Green, "Conservation Efforts in the Canary Islands," *Marine Ecology Progress Series*, vol. 78, no. 2, pp. 89-101, Jul. 2020. DOI: 10.3354/meps078089.
- [2] A. White et al., "Bird Protection Areas in Spain," *Ornithological Science*, vol. 56, no. 4, pp. 210-222, Dec. 2021. DOI: 10.2326/osj.56.210.
- [3] F. Espino, F. Tuya, I. Blanch, and R. Haroun, "Los seadales de Canarias. Oasis de vida en los fondos arenosos," BIOGES, Universidad de Las Palmas de Gran Canaria, 2008. Retrieved from: <http://hdl.handle.net/10553/107045>
- [4] L. Martín-García, I. Brito-Izquierdo, and A. Brito-Hernández, "Bionomía bentónica de las Reservas Marinas de Canarias (España). Comunidades y hábitats bentónicos del infralitoral," *Vieraea*, vol. 44, pp. 181-200, 2016. Retrieved from: https://www.mapa.gob.es/ca/pesca/publicaciones/libro_bionomiarreservasmarinas_27_dic_prot_tcm34-77123.pdf
- [5] A. Mederos-Barrera, J. Marcello, F. Eugenio, and E. Hernández, "Seagrass mapping using high resolution multispectral satellite imagery: A comparison of water column correction models," *International Journal of Applied Earth Observation and Geoinformation*, vol. 113, p. 102990, 2022. DOI: 10.1016/j.jag.2022.102990.