

# SEAWEED PRESENCE DETECTION USING MACHINE LEARNING AND REMOTE SENSING

F. Tonion<sup>1,2\*</sup>, F. Pirotti<sup>2</sup>

<sup>1</sup>T.E.R.R.A. S.r.l. Galleria Progresso 5, 30027 San Donà di Piave (VE), Italy - f.tonion@terrasrl.com

<sup>2</sup>Department of Land, Environment, Agriculture and Forestry (TESAF), University of Padova. Via dell'Università 16, 35020 Legnaro (PD), Italy – francesco.pirotti@unipd.it

Commission III, WG III/10

**KEY WORDS:** Machine Learning, Supervised Classification, Seaweed's Detection, Random Forest, Ecosystems Monitoring.

## ABSTRACT:

The human pressure over coastal areas is becoming increasingly relevant, due to the combinations of resource depletion, climate change effects and ocean eutrophication. Coastal ecosystems are so exposed to a huge number of stress factors that endanger their ecosystem services, like carbon uptake and biodiversity maintenance, that can be crucial in facing the effects of climate changes. With a particular focus on seaweeds, these ecosystems are becoming rapidly relevant both for carbon sinks and as a source of high value products, for example thanks to cosmetic and food industries that produce high added values products.

In this contest the capability of conducting efficient monitoring is crucial to monitor environmental dynamics and resources trends. Traditionally seaweed monitoring was carried out with on field surveys that could be based on botanic analysis combined with genetic study, depending on the aims. Recently Remote Sensing techniques, combined with Artificial Intelligence ones, gave a new perspective to seaweed monitoring, introducing tools that are always more efficient.

In this contest the present work aims to test the potentiality of remote sensing and artificial intelligence techniques for seaweed monitoring along the Irish west coast, building the basis for a fully automated tool for monitoring. The results showed that, with a supervised classification approach, it is possible to train Random Forest (RF) to perform very precise classification over the entire West Coast of Ireland. In particular, with all the RF configurations tested the Overall Accuracy (OA) was greater than 98.61, with the best performance obtained with the configuration  $Ntree = 600$  and  $mtry = 2$  that produced an  $OA = 98.87$ .

## 1. INTRODUCTION

Coastal areas are particularly affected by Climate Changes, that cause ocean warming, sea level rise and impacts on ocean's ability to uptake carbon (Dobush et al., 2021). These pressure factors can affect coastal ecosystems and their carbon uptake mechanisms. Focusing in particular on seaweeds, these ecosystems are becoming increasingly more relevant in facing climate change effects as carbon sink (Belgiu and Drăgu, 2016; Krause-Jensen et al., 2018). Seaweeds can be considered part of a set of farming products that have a role in climate-change mitigation (Mozzato et al., 2018; Pagliacci et al., 2020).

In fact different authors considered with increasing interest the importance of Blue Carbon, derived from ecosystems that include tidally influenced freshwater forests, for example, bald cypress forests and Melaleuca forests, which can have huge soil carbon stocks in their soils and which have been greatly reduced in cover (Krauss et al., 2018).

However it must be remarked that kelp and other seaweed beds ecosystems are also being considered as Blue Carbon ecosystems (Lovelock and Duarte, 2019)

However seaweeds provide also a great amount of other ecosystem services, like different author suggest (van den Burg et al., 2022).

It must also be considered the importance of seaweeds production that in the last years, for example in Ireland, is becoming always more interesting, thanks to cosmetic and pharmaceutical industries and cosmetic markets producing high-quality, high-value products.(Mac Monagail and Morrison, 2020).

Considering this general contest, some authors suggest that there are significant knowledge gaps that need to be addressed to anticipate the combined effects of global and local stressors on seaweed communities (Mineur et al., 2015).

In fact traditional monitoring approaches consist mainly of non-destructive monitoring based on field surveys (Terada et al., 2021) or fixed monitoring stations(Barrientos et al., 2020). The surveys, depending on particular aims, can be coupled with botanic or genetic analysis.

However traditional approaches to monitoring generally can be difficult applied to large areas or to multitemporal analysis, mainly for economic costs and for difficult to collect data over large areas.

For this reason Remote Sensing based seaweed monitoring is becoming a relevant technique. Some authors in particular tested positively the potentiality of multi sensors monitoring approach to seaweed monitoring (Xing et al., 2019).

\* Corresponding author

Other authors reviewed the potential of Remote Sensing techniques applied to seaweed monitoring, considering also Species Distribution Models and aerial imagery, underwater imagery and lidar data (Bennion et al., 2019).

The combination of Satellite data with Artificial Intelligence techniques, considering both Machine Learning and Deep Learning, can represent a strong innovation for the monitoring phase, with incredibly positive results, reported by different authors. (Gonzalez-Rivero et al., 2020; Houskeeper et al., 2022)

Focusing in particular on the Random Forest (RF) algorithm that was introduced in 2001 (Breiman, 2001) this model is based on an ensemble of decision trees (forest) that grows through training towards best combinations. An ensemble consist of a set of individual trained classifier (decision trees), which are combined for classify new instances (Kulkarni, 2013).

Today the RF algorithm finds a large number of applications thanks to its velocity of computation especially while handling large datasets, for example from Sentinel/Landsat imagery (LaRocque et al., 2020) or even laser scanning datasets (Pirotti et al., 2019, 2014).

Considering this contest, the present work aims to:

- build the basis for a fully automated tool for monitoring seaweeds, considering in particular the west coast of Ireland;
- test the potentiality of combining remote sensing data and Artificial Intelligence for seaweeds monitoring.

## 2. MATERIALS AND METHODS

### 2.1 Study area

The study area considered for this work consists of the western part of the Ireland coasts, including all the areas within 750 m from the shoreline, as it is shown in the following picture.



Figure 1. Figure placement and numbering.

The study area represented in the previous picture is extended approximately 5670 km<sup>2</sup>, distributed along more than 3000 km of shoreline.

### 2.2 Satellite data

Considering the kind of analysis and the phenomena studied, it was decided to acquire Sentinel-2 data; the choice was based on following points:

- **Spatial Resolution.** The spatial resolution of Sentinel-2 data is relatively high, with a resolution of 10 m in the visible and NIR. This resolution is ideal to handle a classification of seaweed presence at national scale.
- **Radiometric Resolution.** Having 13 bands that range from visible to far infrared, Sentinel-2 data are particularly suitable to study environmental phenomena, using an AI based approach.
- **Temporal Resolution.** The temporal resolution of approximately 5 days (considering both Sentinel-2A and Sentinel-2B) is ideal for seaweed monitoring, also considering possible changes due to wind, currents and sea storms.

Considering meteorological conditions and cloud absence over a large part of the study area, it was decided to acquire satellite data for 28/05/2021; in detail the following Sentinel-2 tiles were acquired.

ID	Sentinel-2 tile
1	T29UMA_20210528T115359
2	T29UMT_20210528T115359

3	T29UMU_20210528T115359
4	T29UMV_20210528T115359
5	T29UNA_20210528T115359
6	T29UNB_20210528T115359

**Table 1.** Sentinel-2 tiles acquired.

### 2.3 Satellite data processing

All the processing operations computed over satellite data were done with a dedicated algorithm developed with R Studio programming language. Each band with the same spectral resolution of the satellite tiles acquired were merged together and cropped, maintaining only the raster pixels inside the study area. Then all bands with rougher spatial resolution were resampled to 10 m resolution, in order to compute spectral indexes among bands with different spatial resolution. After that, different indexes were randomly computed in order to define a sufficient number of features to train the models, that was assumed equal to nine; the features created included indexes like Normalized Difference Vegetation Index (NDVI), Blue-Normalized Difference Vegetation Index (BNDVI) and Normalized Difference Water Index (NDWI).

### 2.4 Dataset creation

The first step of data analysis phase was the definition of target classes for the classification, in order to distinguish both seaweeds presence and all other relevant classes within the study area that are not seaweeds; in detail the defined classes were:

- 1) Seaweeds. This is the main target class, and includes different kinds of seaweeds.
- 2) Water. This class includes sea water without the presence of seaweeds.

- 3) Rocks. This class includes rock out of the water.
- 4) Fine Sediment. This class includes fine sediments out of the water.
- 5) Vegetation. This class comprehends vegetation elements outside sea water.
- 6) Clouds. This class includes cloud cover areas within the study area; this class helps only in masking cloud covered areas in order to not produce misclassification once the process will be fully automated.

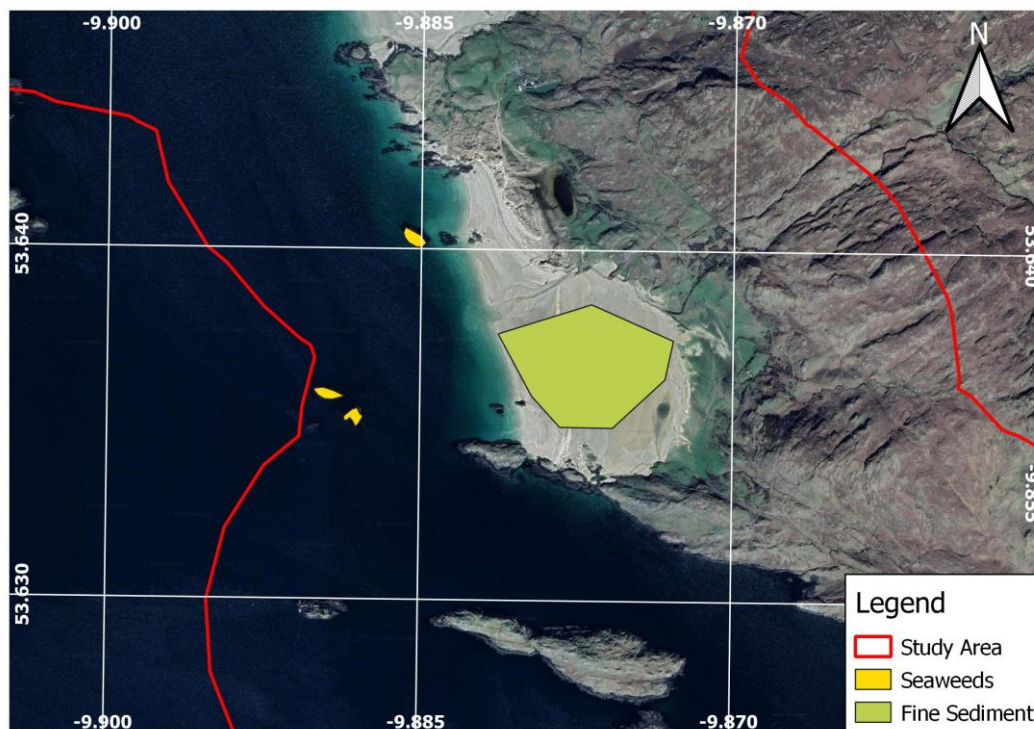
For each of the previous classes, different polygons, to be used as training and validation areas, were created. The creation of polygons along the study area was based on Sentinel-2 true colour images, ground surveys information and spectral indexes calculated.

The result of this process was the creation of an ensemble of polygons for each of the target classes. The following table provides an overall analysis of polygons created.

ID	Class Description	Number of Polygons	Area (m <sup>2</sup> )
1	Water	14	4.16*10 <sup>6</sup>
2	Rock	41	96.138*10 <sup>3</sup>
3	Fine Sediment	27	2.710*10 <sup>6</sup>
4	Vegetation	26	2.628*10 <sup>5</sup>
5	Seaweeds	82	8.027*10 <sup>5</sup>
6	Cloud	24	7.311*10 <sup>6</sup>

**Table 2.** Polygons created for supervised classification

The following picture shows some training and validation polygons within the study area.



**Figure 2.** Example of different classes training and validation polygons.



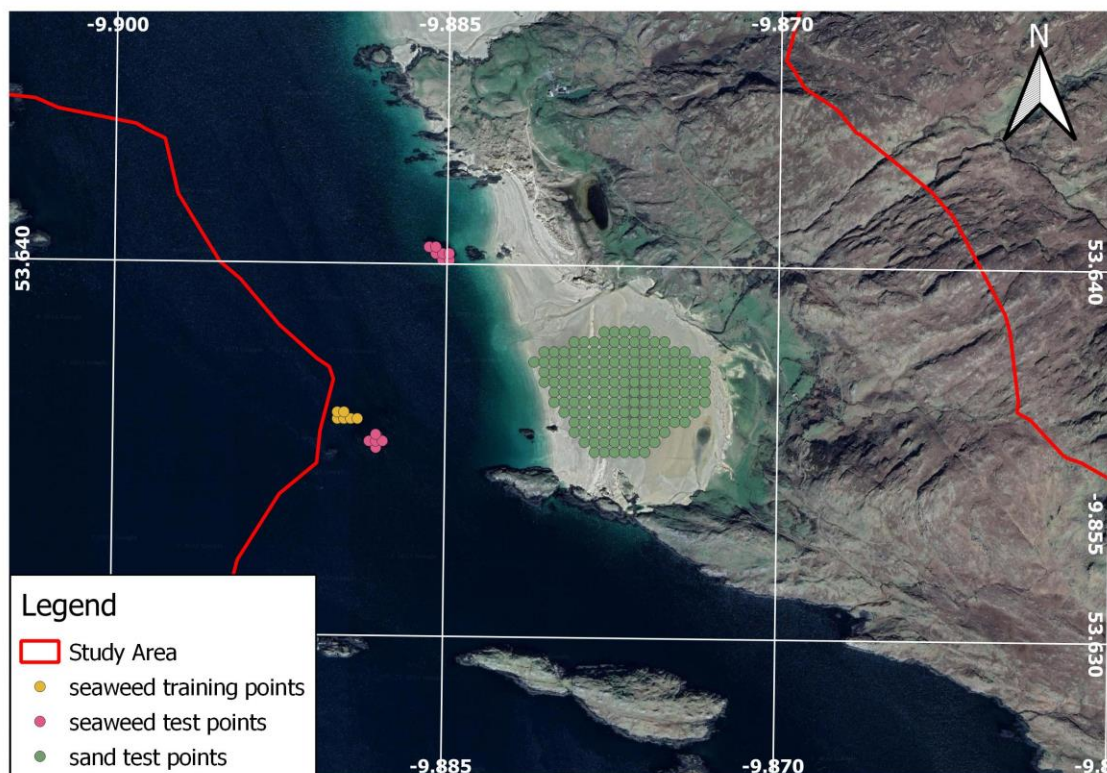
As it is possible to notice from the previous table, the polygon numerosity of different classes is different. The reason for this difference is connected with the characteristics of each class. In fact dealing with classes like “water” it was possible to create bigger polygons within the study area; oppositely dealing with the class “seaweeds” it was necessary to define small polygons to avoid the consideration of not “seaweeds” area within the “seaweeds” class.

For both the training and validation datasets, a regular points grid was created within each polygon. The following table provides the overall information of the numerosity of the created points both in training and validation datasets.

ID	Class Description	Number of Points Training	Number of Points Validation
1	Water	214	386
2	Rock	203	237
3	Fine Sediment	2305	2299
4	Vegetation	253	220
5	Seaweeds	1127	872
6	Cloud	136	1927

**Table 2.** Polygons created for supervised classification

The following picture shows some training and validations points within the study area.



**Figure 3.** Example of different classes training and validation points.

### 2.5 Artificial Intelligence Models - Tuning

For all the points of the training and validation datasets the Machine Learning (ML) model tuning phase aimed to identify the best configuration of the Random Forest (RF) model for data analysis.

The methodology adopted was an iterative one, in order to find the best configuration of parameters. The RF parameters object of the iterative calibration were:

- Number of Tree (Ntree) value. An integer number that represents the number of decision trees created to perform the analysis. The range of variation of this parameter values from 100 to 1500 in a sequence by 100.

- Variable Split to create each tree (Mtry). An integer number representing the “height” of each tree, that is the number of variable split to create each single tree of the forest. The range of variation of Mtry was assumed from 2 to 8.

The iterative procedure consists of the test of each combination of the Ntree and Mtry values within the fields of variation. For each RF configuration tested the Overall Accuracy was calculated (percentage of correctly classified points on the test dataset).

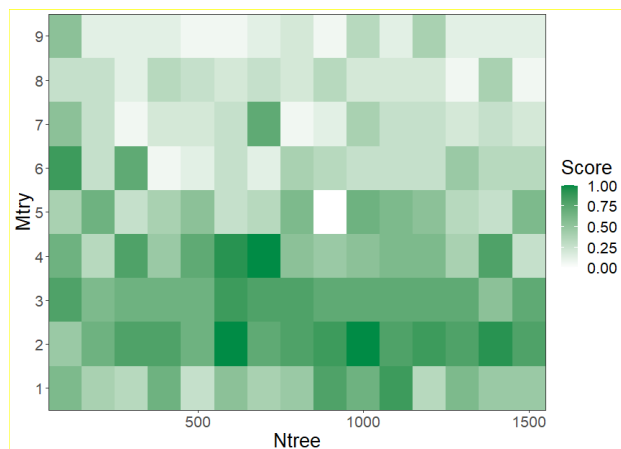
Eventually, it was maintained the model configuration that ensured the highest value of accuracy; for the selected configuration it was created a confusion matrix and different

error metrics were computed (Overall Accuracy, F measure, K index and Cramer V index).

### 3. RESULTS

The tuning procedure of RF tested 135 different combinations of parameters, considering the range of variation of Ntree and Mtry described in the previous chapter. In general the RF tuning procedure showed, in all the combinations tested, a high level of Overall Accuracy (OA), ranging from a minimum of 98.61% to a maximum of 98.87%.

The following picture provide a graphical representation of the variation of OA score during RF tuning procedure.



**Figure 4.** Variation of OA score with different RF configurations

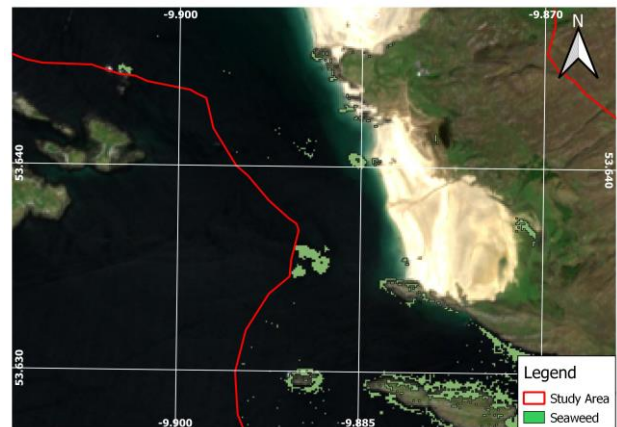
As it is possible to notice from the previous picture, the best RF configuration was obtained with Ntree value equal to 600 and Mtry value equal to 2; considering the best configuration the following table provides the confusion matrix of the classification performed, with the indication of the F measure by class.

ID	1	2	3	4	5	6	F measure
1	385	0	0	0	7	0	98.97
2	0	204	0	0	15	3	88.70
3	0	3	2299	0	0	4	99.85
4	0	0	0	220	2	0	99.56
5	1	30	0	0	847	3	96.80
6	0	1	0	0	1	1920	99.77

**Table 3.** Confusion Matrix

Considering the previous confusion matrix, the K index was equal to 98.43% and the Cramer V was equal to 96.88%.

Figure 5 and figure 6 give graphical representations of some areas classified as seaweed within the study area.



**Figure 5.** Particular of Seaweed areas detected – Uggool beach.



**Figure 6.** Particular of Seaweed areas detected – Blacksod Bay.

### 4. DISCUSSION

The results showed that predictions over the test dataset were extremely precise. In general, all the tested models were characterized by Overall Accuracy higher than 98.61%.

The best results were generally obtained with lower Mtry values (the height of each tree) and a higher number of trees.

Having such good performances both with lower and higher mtry numbers, with best results at lower values, is an indicator that overfitting is not occurring. As a matter of fact even with the simplest combination tested (low number of trees and low mtry value) gave extremely positive results.

Such good performances are mainly due to the features engineering phase: having meaningful variables, which can be interpreted with a supervised approach, simplify the Artificial Intelligence data analysis phase. The features engineering phase can be a key phase for large scale applications of Remote Sensing techniques.

However, to build a fully automated seaweed distribution monitoring tool, the training and validation datasets must be significantly enlarged to ensure the representativity at the annual scale, including phenological variations of seaweeds (growth, colors...etc.).

Moreover, once building automated tools for the entire coast of a country like Ireland some considerations must be considered the particular meteorological conditions, with the sky often covered by clouds; this can significantly restrict the periods of the year where it is possible to collect useful satellite data. Considering in particular the case study of the Irish west coast it was observed that the period for useful satellite data collection was generally those from late spring to early autumn.

Eventually it must also be considered that monitoring seaweeds at national scale is a relevant work that can be affected by multiple sources of disturbances (for example turbidity, meteorological conditions and human activity).

Based on all these considerations the results of the work are extremely positive and relevant; the significant precision of Machine Learning models observed on a single day can be a solid base for future multitemporal works.

For this reason, further improvement of the present work can include a multitemporal analysis of seaweed distributions during an entire year, enlarging significantly the training and validation datasets, to ensure representativity along the entire year similar to what was done in (Vaglio Laurin et al., 2016).

## 5. CONCLUSION

The study has compared different Random Forest configurations for seaweed distributions monitoring along the entire west coast of Ireland. The approach to this task was with a supervised classification, based on different spectral indexes calculated on Sentinel-2 bands.

Results show that all the configurations of RF gave particularly good results with extremely content errors. In conclusion, this study shows that ML can be an efficient tool to predict seaweed presence and give practical support to all those interested in having a smart and fast monitoring tools for coastal habitat.

## REFERENCES

- Barrientos, S., Barreiro, R., Cremades, J., Piñeiro-Corbeira, C., 2020. Setting the basis for a long-term monitoring network of intertidal seaweed assemblages in northwest Spain. *Mar. Environ. Res.* 160, 105039.
- Belgiu, M., Drăgu, L., 2016. Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* <https://doi.org/10.1016/j.isprsjprs.2016.01.011>
- Bennion, M., Fisher, J., Yesson, C., Brodie, J., 2019. Remote sensing of kelp (Laminariales, Ochrophyta): monitoring tools and implications for wild harvesting. *Rev. Fish. Sci. Aquac.* 27, 127–141.
- Breiman, L.E.O., 2001. Random Forests, in: *Machine Learning*. Kluwer Academic Publishers, pp. 5–32. <https://doi.org/10.1023/A:1010933404324>
- Dobush, B.-J., Gallo, N.D., Guerra, M., Guilloux, B., Holland, E., Seabrook, S., Levin, L.A., 2021. A new way forward for ocean-climate policy as reflected in the UNFCCC Ocean and Climate Change Dialogue submissions. *Clim. Policy* 1–18.
- Gonzalez-Rivero, M., Beijbom, O., Rodriguez-Ramirez, A., Bryant, D.E.P., Ganase, A., Gonzalez-Marrero, Y., Herrera-Reveles, A., Kennedy, E. V., Kim, C.J.S., Lopez-Marcano, S., 2020. Monitoring of coral reefs using artificial intelligence: A feasible and cost-effective approach. *Remote Sens.* 12, 489.
- Houskeeper, H.F., Rosenthal, I.S., Cavanaugh, Katherine C, Pawlak, C., Trouille, L., Byrnes, J.E.K., Bell, T.W., Cavanaugh, Kyle C, 2022. Automated satellite remote sensing of giant kelp at the Falkland Islands (Islas Malvinas). *PLoS One* 17, e0257933.
- Krause-Jensen, D., Lavery, P., Serrano, O., Marbà, N., Masque, P., Duarte, C.M., 2018. Sequestration of macroalgal carbon: the elephant in the Blue Carbon room. *Biol. Lett.* 14, 20180236.
- Krauss, K.W., Noe, G.B., Duberstein, J.A., Conner, W.H., Stagg, C.L., Cormier, N., Jones, M.C., Bernhardt, C.E., Graeme Lockaby, B., From, A.S., 2018. The role of the upper tidal estuary in wetland blue carbon storage and flux. *Global Biogeochem. Cycles* 32, 817–839.
- Kulkarni, V.Y., 2013. Random Forest Classifiers: A Survey and Future Research Directions 36, 1144–1153.
- LaRocque, A., Phiri, C., Leblon, B., Pirotti, F., Connor, K., Hanson, A., 2020. Wetland Mapping with Landsat 8 OLI, Sentinel-1, ALOS-1 PALSAR, and LiDAR Data in Southern New Brunswick, Canada. *Remote Sens.* 12, 2095. <https://doi.org/10.3390/rs12132095>
- Lovelock, C.E., Duarte, C.M., 2019. Dimensions of blue carbon and emerging perspectives. *Biol. Lett.* 15, 20180781.
- Mac Monagail, M., Morrison, L., 2020. The seaweed resources of Ireland: a twenty-first century perspective. *J. Appl. Phycol.* 1–14.
- Mineur, F., Arenas, F., Assis, J., Davies, A.J., Engelen, A.H., Fernandes, F., Malta, E.-J., Thibaut, T., Van Nguyen, T., Vaz-Pinto, F., 2015. European seaweeds under pressure: Consequences for communities and ecosystem functioning. *J. sea Res.* 98, 91–108.
- Mozzato, D., Gatto, P., Defrancesco, E., Bortolini, L., Pirotti, F., Pisani, E., Sartori, L., Mozzato, D., Gatto, P., Defrancesco, E., Bortolini, L., Pirotti, F., Pisani, E., Sartori, L., 2018. The Role of Factors Affecting the Adoption of Environmentally Friendly Farming Practices: Can Geographical Context and Time Explain the Differences Emerging from Literature? *Sustainability* 10, 3101. <https://doi.org/10.3390/su10093101>
- Pagliacci, F., Defrancesco, E., Mozzato, D., Bortolini, L., Pezzuolo, A., Pirotti, F., Pisani, E., Gatto, P., 2020. Drivers of farmers' adoption and continuation of climate-smart agricultural practices. A study from northeastern Italy. *Sci. Total Environ.* 710, 136345. <https://doi.org/10.1016/j.scitotenv.2019.136345>
- Pirotti, F., Laurin, G., Vettore, A., Masiero, A., Valentini, R., 2014. Small Footprint Full-Waveform Metrics Contribution to the Prediction of Biomass in Tropical Forests. *Remote Sens.* 6, 9576–9599. <https://doi.org/10.3390/rs6109576>
- Pirotti, F., Zanchetta, C., Previtali, M., Della Torre, S., 2019. DETECTION OF BUILDING ROOFS AND FACADES FROM AERIAL LASER SCANNING DATA USING DEEP LEARNING. *ISPRS - Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* XLII-2/W11, 975–980. <https://doi.org/10.5194/isprs-archives-XLII-2-W11-975-2019>
- Terada, R., Abe, M., Abe, T., Aoki, M., Dazai, A., Endo, H., Kamiya, M., Kawai, H., Kurashima, A., Motomura, T., 2021.

Japan's nationwide long-term monitoring survey of seaweed communities known as the "Monitoring Sites 1000": Ten-year overview and future perspectives. *Phycol. Res.* 69, 12–30.

Vaglio Laurin, G., Hawthorne, W., Chiti, T., Di Paola, A., Cazzolla Gatti, R., Marconi, S., Noce, S., Grieco, E., Pirotti, F., Valentini, R., 2016. Does degradation from selective logging and illegal activities differently impact forest resources? A case study in Ghana. *iForest - Biogeosciences For.* e1–e9. <https://doi.org/10.3832/ifor1779-008>

van den Burg, S.W.K., Termeer, E.E.W., Skirtun, M., Poelman, M., Veraart, J.A., Selnes, T., 2022. Exploring mechanisms to pay for ecosystem services provided by mussels, oysters and seaweeds. *Ecosyst. Serv.* 54, 101407.

Xing, Q., An, D., Zheng, X., Wei, Z., Wang, X., Li, L., Tian, L., Chen, J., 2019. Monitoring seaweed aquaculture in the Yellow Sea with multiple sensors for managing the disaster of macroalgal blooms. *Remote Sens. Environ.* 231, 111279.