Cloud-native Seascape Mapping of Mozambique’s Quirimbas National Park with Sentinel-2

Dimitris Poursanidis 1, Dimosthenis Traganos 2, Luisa Teixeira 3, Aurélie Shapiro 4 & Lara Muaves 5

1 Foundation for Research and Technology—Hellas (FORTH), Institute of Applied and Computational Mathematics, N. Plastira 100, Vassilika Vouton, Heraklion 70013, Greece
2 German Aerospace Center (DLR), Remote Sensing Technology Institute, Rutherfordstraße 2, Berlin 12489, Germany
3 Vizzuality, Rua Da Firmeza, 457, Porto 14000-230, Portugal
4 Space + Science, World Wide Fund for Nature (WWF) Germany, Reinhardtstr. 18, Berlin 10117, Germany
5 World Wide Fund for Nature (WWF) Mozambique, Rua Faralay N° 108, Bairro da Somershield, Maputo, Mozambique

Keywords
Seascapes, Seagrasses, Corals, Earth Observation, Google Earth Engine, Sentinel-2, Quirimbas, Cloud-computing

Abstract
The lack of detailed spatial information on coastal resources, notably shallow water coral reefs and associated benthic habitats, impedes our ability to protect and manage them in the face of global climate change and anthropogenic impacts. Here, we develop a semi-automated workflow in the cloud that uses freely available Sentinel-2 data from the European Space Agency (ESA) Copernicus programme to derive information on near-shore coral reef habitats in the Quirimbas National Park (QNP), a recently declared biosphere reserve in northern Mozambique. We use an end-to-end cloud-based framework within the Google Earth Engine cloud geospatial platform to process imagery from raw pixels to cloud-free composites which are corrected for glint and surface artefacts, water column and derived estimated depth and then classified into four benthic habitats. Using independent training and validation data, we apply three supervised classification algorithms: random forests (RF), support vector machine (SVM) and classification and regression trees (CART). Our results show that random forests are the most accurate supervised algorithm with over 82% overall accuracy. We mapped over 105 000 ha of shallow water habitat inside the protected area, of which 18% are dominated by coral and hardbottom; 27.5% are seagrass and submerged aquatic vegetation and another 23.4% are soft and sandy substrates, and the remaining area is optically deep water. We employ satellite-derived bathymetry to assess slope, bathymetric position, rugosity and underwater topography of these habitats. Finally, a spectral unmixing model provides further sub-pixel level information of habitats with the potential to monitor changes over time. This effort provides the first, consistent and repeatable and also scalable coastal information system for an east African tropical marine protected area, which hosts shallow-water ecosystems which are of great significance to local communities and building resilience towards climate change.

Introduction
With a shoreline of over 2700 km, Mozambique hosts a unique number of coastal habitats, including some of the most climate-resilient coral reefs in the world, representing an important opportunity for conservation (Beyer et al., 2018). The western Indian Ocean also features a very high biological diversity: more than 1500 fish species, 200 coral species, 14 mangrove species, 12 seagrass species, 1000 marine algae species, hundreds of species of sponges, and 300 crab species (Richmond, 2000). The region also hosts unique megafauna, including whales, sharks, rays and endangered marine turtles and dugongs (UNEP, 2004). These globally significant marine and coastal habitats provide essential ecosystem services such as carbon sequestration and climate mitigation, and...
essential nurseries for aquatic species to provide food and livelihoods for many (Meleon et al., 2011; Nordlund et al., 2018; Sitee et al., 2010).

The dependence on natural resources in Mozambique is high, with as many as 80% of employment relying on sectors such as agriculture, fisheries and mining (Macamo, 2019). The fishing industry provides a significant contribution to the national GDP, while artisanal fisheries comprise 90% of production and the main source of employment and food sources in coastal communities – where most of the Mozambique’s population reside (Macamo, 2019). Meanwhile, Mozambique is a rapidly growing tourism destination, relying on intact ecosystems and its wealth of biodiversity and wildlife for this economic sector. Despite the value of these coastal ecosystems, increased pressure on marine resources has created significant ecological changes in many parts of the East African coastline. Overfishing has resulted in the decline in great whale populations and valuable fishery species, as well as the degradation of important seagrass beds and coral reef habitats (Sjostedt & Sundstrom, 2013). Many species are heavily over-fished, with destructive methods such as gill nets and dynamite still being used (Obura, Souter, & Linden, 2005), along with under-reported catches putting the entire industry at risk of overexploitation (Jacquet et al., 2010). Demand for building materials such as mangrove poles and corals for lime, along with increasing need for agricultural land have further contributed to habitat destruction (Kideghesho, 2009). All these impacts disturb the ecological balance, reduce the capacity for secure livelihoods and food security for local populations, as severely damaged coral reefs and seagrass beds can not provide critical ecosystem services.

Management approaches to mitigate the pressures in the marine regime have been developed and applied worldwide, including via Marine Protected Areas (MPAs) and Marine Managed Areas (MMAs), which can be implemented to offer a range of ecological, social, cultural and economic benefits (Claudet, 2011). The location, design, characteristics and on-going management of these areas, however, ultimately drive the extent to which the benefits could be achieved in practice. In Mozambique, MPAs and MMAs have been designated, including the Quirimbas National Park (QNP), a recently designated international biosphere reserve (UNESCO, 2018) protecting some of the most resilient reef systems in the region (Hill et al., 2010).

QNP was established in 2002, however, few readily accessible accurate spatial information exist to contribute to a comprehensive baseline for the coastal marine seascape ecosystems to enable informed management practices, detailed zoning and distribution of human activities such as fishing limitations, no-take zones or adaptive management responses to address changes in marine ecosystems. The types of management which benefit from accurate spatial data include the location and designation of temporary closures and sanctuaries for management or recovery of fish resources, more specifically octopus closures; regulating uses in designated tourist areas; and continued monitoring over time to ensure resilient and functioning reef systems which ensure that the main goals of the protected area are being achieved – in this case sustainable supporting local livelihoods. What limited available data exist (e.g. RCRMD, 2015) are either out of date, of insufficient resolution, do not have any comprehensive metadata to assess the status, lack accuracy assessment or are not derived from automated methods, making them difficult to reproduce over time. Other datasets like the recently released Allen Coral Atlas (Lyons et al., 2020) are global products derived from commercial imagery which have limited local validation and accuracy assessment. Although they offer a much improved spatial resolution, and provide valuable geomorphic zone information, this dataset derived from commercial imagery comes with the potential trade-off of a lower temporal resolution, meaning fewer updates which can be delivered or requested, or large datasets which cannot be easily accessed or manipulated in remote locations. Therefore, a complementary data source with simple outputs for protected areas managers is desired for continuous, flexible and adaptable monitoring.

Here, we present the first cloud-based semi-automated approach that uses Copernicus Sentinel-2 optical imagery to map the entire coastal area of Quirimbas National Park in Mozambique, whose reefs possess world-reknowned refugia and environmental variability enabling resilience and potential adaptation of rapid climate change (McClanahan & Muthiga, 2017). Our main aim is to provide consistent mapping of the underwater structure and habitats of the coral reefs, seagrasses and neighbouring underwater shallow-water seascape which can be repeated over time for monitoring, and scaled and expanded to other regions. This information can assist comprehensive conservation activities, management decisions, sustainable development planning for more effective climate change mitigation, resilience and adaptation in the broader region of East Africa providing a crucial starting point for continued operational monitoring.

Many small-scale coral reef habitat mapping studies have relied on high-resolution commercial data, while larger-scales and longer-term monitoring is more appropriate for medium resolution (30 m) from Landsat, which up until 2016 was the dominant free data source (Hedley et al., 2016). The open availability of the Landsat archive since 2008 (Wulder et al., 2012) has provided millions of
scenes covering almost all areas of the world, enabling great progress for seascape mapping. This includes monitoring and change detection to assess the impacts of natural hazards and climate change, which include the increase in frequency and severity of cyclones and associated surges, and coral bleaching due to sea surface temperature increases (Green et al., 1998; Hedley et al., 2016; Liu et al., 2014; Pham et al., 2019). Despite being launched as a terrestrial mission in 2015, the ESA Copernicus Sentinel-2 constellation consisting of two satellites has notably increased spatial and temporal resolution and data availability for a significant number of coral reefs (Hedley et al., 2018). A significant benefit is the minimum mapping unit (MMU), whereas for Landsat is 900 m² as a result of the 30 m square pixel, for Sentinel-2 is decreased to 100 m² via the 10-m resolution (Tobler, 1988). The higher temporal resolution also increases the chances for suitable cloud-free data, stable sea states or clear water. As such, the 5-day time interval and the smaller pixel size allows more effective multi-temporal image composition (Traganos et al., 2018a) and, hence, renders an accurate detection of homogeneous seascape elements such as hard bottom substrates for coral reefs, seagrass meadows and algae/turfs, as the coastal waters are rarely a homogeneous system in the tropics and elsewhere. These data can also be used to evaluate relative bathymetry and underwater structure which inform marine spatial planning including zoning and managing uses of resources (Douvere, 2008). These elements greatly enhance coastal seascape mapping and monitoring, and when accompanied by high-quality in-situ data that match the temporal window of the image composite can be used to assess the trajectories of habitats of interest over time.

To map the seascape using these abundant data streams, we exploit an end-to-end cloud-native semi-automated algorithmical framework – within the geospatial platform of Google Earth Engine (Gorelick et al., 2017) – which features the entire open-access image archive of Sentinel-2. The power of cloud computing enables big data processing for creating cloud-free composites, multi-temporal analytics, and efficient machine-learning algorithms calibrated by field data collected by partners on the ground who observed the presence, status and depth of coral reefs, seagrasses and the sandy/soft bottoms. We use a geoprocessing framework designed for submerged vegetation monitoring in temperate waters (Traganos et al., 2018a, Traganos et al., 2018b; Traganos & Reinartz, 2017) and apply it to multiple benthic habitat types in the tropical seascape. This provides the first automated, consistent and expandable assessment for tropical coastal resources in QNP to provide a pre-cyclone baseline. The automated nature of the workflow provides valuable opportunities for repeatable and automated monitoring, which come at a crucial time of political instability and insecurity in the area, resulting in limited accessibility and lack of monitoring resources compounded by the Covid-19 pandemic.

Materials and Methods

Study area

Following the independence of Mozambique in 1975, more than five marine conservation areas have been established by the national government. Among them, the Quirimbas National Park (QNP), in the Province of Cabo Delgado (Figure 1), Northern Mozambique, was created with an intrinsic goal to value and protect the biodiversity and ensure sustainable local livelihoods (MITUR, 2003). In 2018, it was declared a UNESCO international Biosphere Reserve due to its unique terrestrial and marine fauna (UNESCO, 2018). An important aspect of this conservation area is that it follows a “bottom-up” approach, since it was designed in part at the request of communities who, at the time, suffered from human–wildlife conflicts, competition for depleting natural resources, poverty and declining ecosystem services and food sources upon which they are dependent. The QNP is a protected area with a significant local population of 166 000 people living within its boundaries, with 40% in the transition and buffer zone (Mucova et al., 2018). Being the third largest conservation protected area in Mozambique with a significant ecological and economic value, it faces several challenges such as deforestation, poaching, illegal mining, hunting, over-fishing and over-exploitation of resources. All these combined pressures negatively impact biodiversity and resource conservation, further affecting vulnerable local communities and populations.

In situ data

Information collected from snorkel swims, boat and drone surveys was used to create the training data for the classification algorithms, and we were an aggregation of a seascape mapping survey conducted by WWF-Mozambique in September, 2018, and an octopus closure survey conducted in April 2019 (Muaves, 2019). Due to the nature of the different surveys, and the characteristics of typical octopus closures areas (tidal flats), which are exposed reef areas which trap sediment and sand and are increasingly silted and highly reflective like sand are considered as soft substrate. In both surveys, depth information was recorded using a Fishfinders Lucky hand-held portable depth finder to support the derivation of satellite-derived relative bathymetry (SDB). The presence of three major habitat types (hard substrate, vegetation and soft substrate – examples shown in Figure 2) was identified, as well as
the mixture of multiple habitats within an approximate 10 m x 10 m area was assessed either by snorkelers or from the boat using a glass bottom bucket and water-proof camera (GoPro inside a waterproof case) mounted on a 50 cm stick. Habitat classes were identified a priori and according to three major class types: Coral and hard-bottom habitats (hard substrate) include any coral- or rock-dominated surface, live or dead; Seagrass and submerged vegetation (vegetation) comprise all surfaces with at least 30% seagrass cover and underwater flowering plants (Klemas, 2016). Soft and sandy substrates (soft substrate) include all sandy and fine rubble surfaces and may include turf macroalgae. Optically deep areas fall into the deep-water class.

All information was collected in the field using a customized Survey 123 for ArcGIS application, which automatically includes geo-location from the Android phone or tablet, collected in addition to position information recorded at each location using a Garmin 64 s GPS. Drone surveys were also conducted at six locations using a 3DR Solo drone mounted with a GoPro 4 camera with a custom-fitted straight 4 mm lens to avoid fish-eye effect. Surveys were flown with 80% side overlap and 60% forward. Images were geo-located to the drone GPS position obtained from flight logs using GeoSetter 3.4.16 (images are shown here: https://space-science.wwf.de/QNP_drone_survey).

Classification training data distributed for the three habitat types, plus optically deep water (where insufficient light is reflected from the seabed and subsequently measured from the satellite), were added by the digitization of features detected in drone imagery, Google Earth and Google Earth Engine using information from the in situ data (shown in Figure 3) as well as older commercial high-resolution imagery from QuickBird and IKONOS, acquired in 2004 to enhance the distribution of points in all bottom classes (Table 1).

Earth observation image processing

The entire Earth Observation (EO) analysis was performed in the Google Earth Engine (GEE) cloud environment for the analysis of Earth Observation data (Gorelick
et al., 2017), using the workflow of Traganos et al. (2018a, b) adapted to the QNP tropical landscape. Sentinel-2 L1C data were filtered by acquisition dates that coincide with the field surveys, prior to the 2019 cyclone season with adequate cloud-free coverage. We selected all data collected during the dry season months (May to December) for 2017 and 2018 with an overall cloud cover of less than 5%, resulting in a collection of 212 available images to create a best pixel composite. This composite was created by masking clouds using the Sentinel-2 QA60 bitmask, and then taking the median values of the first quintile (20%) of best quality pixels. Next, we performed sun glint removal applying the method of Hedley et al., 2005, and automatic water masking was conducted using the Otsu method (Donchyts et al., 2016; Otsu, 1979). We derived a post-cyclone composite in the same manner using imagery acquired between May 2019 and February 2020.

We derive a relative bathymetry and depth-invariant index following the log-linear transformed linear model (Lyzenga, 1978; Lyzenga, 1981) resulting in a relative estimation of depths (m) and three-band reflectance image derived from ratios which are independent of the water column (Traganos et al., 2018b). We quantitatively validated the satellite-derived bathymetry models through the metrics of $R^2$, Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) using all 683 available data points.

**Figure 2.** Representative photos of the classification scheme. Top row: hard substrate; middle row: vegetation and bottom row: soft substrate.
As we lack an independent dataset for validation of the depth retrieval, we use the satellite-derived depth as relative depth layer to enhance the benthic classification.

To maximize the data available for benthic habitat classification, we added two additional bands to the data stack, which are the first and second principle components layers derived from the sun-glint corrected image. A 3x3 boxcar convolution filter is applied to the stack before classification to remove any artefacts or anomalies by a low-pass smoothing. The bands used in the classification included the coastal aerosol, blue, green and red (bands 1, 2, 3 and 4 of S2 L1C), as well as the three depth invariant bands, the relative bathymetry and two principle components layers.

We derived bathymetric slope in degrees, rugosity and bathymetric position index (BPI) using the NOAA Benthic Terrain Modeler extension for ArcGIS (Walbridge et al., 2018). The broad-scale bathymetric position was calculated using an inner radius of 25 and an outer radius of 50 pixels. We use these outputs to evaluate relative depth and position of the benthic habitat classification and to provide auxiliary data products for underwater topography of the reef environment.

We applied three supervised classification methods to the image stack: Random Forests (RF) machine learning algorithm (Breiman, 2001), classification and regression tree (CART; Breiman et al., 2017) and support vector machine (SVM; Zhang et al., 2001). The resulting classified habitat maps have four broad classes: hard substrate, submerged vegetation, soft substrate and deep water. These were determined based on the characteristics of the seascape, the degree of feasibility and efficiency of field data collection.

The training data were randomly split into 70% for training and 30% for validation to assess training and

Table 1. QNP in situ and training data: field survey data from 2018, 2019 and the desktop-added points (image interpretation in conjunction with drone and underwater photos).

<table>
<thead>
<tr>
<th>Class</th>
<th>2018 Number</th>
<th>2019 Number</th>
<th>Desktop points</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>%</td>
<td>%</td>
<td>%</td>
</tr>
<tr>
<td>Soft substrate</td>
<td>182 21%</td>
<td>446 67%</td>
<td>426 33%</td>
</tr>
<tr>
<td>Vegetation</td>
<td>518 60%</td>
<td>69 10%</td>
<td>490 38%</td>
</tr>
<tr>
<td>Hard Substrate</td>
<td>145 17%</td>
<td>140 21%</td>
<td>320 25%</td>
</tr>
<tr>
<td>Deep water</td>
<td>18 2%</td>
<td>10 2%</td>
<td>44 3%</td>
</tr>
<tr>
<td>Total points</td>
<td>863</td>
<td>665</td>
<td>1280</td>
</tr>
</tbody>
</table>

Figure 3. Distribution of field data collected during the September 2018 expedition were used to train the analysis of Earth Observation data, which included habitat classification identified from snorkel, boat (glassbottom bucket) and drone surveys.
classification accuracy. We evaluate the results of the coastal habitat map by calculating overall (OA), producer (PA) and user accuracy (UA) of each class and estimate habitat area based on weight-adjusted accuracies according to Olofsson et al. (2013).

Using the same training dataset we apply a spectral unmixing algorithm (Adams, Smith, & Johnson, 1986) provided by the .unmix function in Google Earth Engine applied to the deglinted Sentinel-2 image. We use a random sample (70%) of the “pure” (single, dominant habitat) endmembers identified selected at various depths. We then interpret a continuous measure of % contribution of the four habitat classes, essentially providing sub-pixel estimates of habitats. We apply the same unmixing approach to the post-cyclone Sentinel-2 de-glinted composite and assess the percent change in each fraction. We use the remaining data of mixed and pure classes to validate the presence of multiple habitats at one location.

Results

EO Processing

The outputs from EO processing resulted in composites with cloud and glint removal, satellite-derived relative bathymetry (SDB) and associated derivatives for water column correction, followed by the habitat classifications. The raw multi-temporal image mosaic, the de-glinted and water column corrected outputs are shown in Figure 4. Due to the nature of the deglinting algorithm and the multi-temporal analytics, surface artefacts and waves are removed in the deglinted image, while the water column corrected image shows coral reefs and seagrass habitats with similar reflectance independent of their depth.

Satellite-derived relative depth was estimated up to 15 m for optically clear waters (Figure 5), with MAE of 1.21 m, RMSE of 1.61 m and an $R^2$ of 0.62. This output shows the entire potential shallow reef shelf throughout the protected area, and around the atolls. The lagoon bathymetry was also retrieved, showing underwater channels and coastal features. Additional derivatives from bathymetry include slope, rugosity and BPI which show the areas of relatively homogenous flat surfaces in the lagoons compared to those with more complex topography (Figure 5). The BPI discerns shallow reef flats from slopes and deeper flat zones typical for the lagoon areas around the islands and along the mainland shore.

Benthic Habitats

Training accuracy evaluating the random sample of the training dataset and the validation accuracy using an independent sample of training points are shown in Table 2. Support vector machine had a training accuracy of 100% as expected for a machine learning approach which might be over-fitted. It does, however, produce the lowest validation accuracy. CART has the next highest training accuracy and produces a map with slightly more speckly in the seagrass habitat. Random forest has the highest validation accuracy and was selected for the final classification.

The QNP classification map developed using random forest classification has an overall accuracy of 84.6% (Figure 6, Table 3). Coral is the least accurate class, being most often confused with soft substrate and to a lesser extent vegetation. Soft substrate had the highest producer accuracy meaning low omission errors, while vegetation has the highest user accuracy. In comparison, the SVM classification greatly underestimates hard substrates, and overestimates soft substrates which showed a 50% user and producer accuracy, which is especially low considering the small number of overall classes. The CART classification showed highest user accuracies for soft substrates, and all producer accuracies between 70 and 80%, however, the overall accuracy was under 77% and lower than random forests, and deep water is overestimated compared to the other classifications. Based on the accuracy assessment of the Random Forest classifier, except for optically deep waters, all other three classes are neither overestimated nor underestimated following their balanced producer and user accuracies. Here, vegetation is the most dominant habitat, followed by soft substrate as shown in Table 4. The mapped habitats have unique depth ranges and topographic position (Table 5). Soft substrate is found generally on shallower, flatter, smoother underwater surfaces in comparison to the other habitats, while seagrass shows highest rugosity and hard bottom generally at deeper depths.

The spectral unmixing results are shown in Figure 7, with three bands representing the unique classes as a proportion from 0 to 1, where the sum of all bands in one pixel is 1. The zoomed areas show the presence of mixed habitats. We note sand mixed with seagrass on the outermost edges of the atolls and some areas of seagrass and corals in the southern half of the protected area. Given the difficulties in acquiring detailed quantitative data on sub-pixel habitat presence, we use the multiple habitat types identified in the field and map these onto the spectrally unmixed image (Figure 8). The presence of vegetation and soft substrate in the field is generally represented in the mixed image, however, the observed presence of hard substrate does not appear to coincide as well with the unmixed fractions. The cluster of observed areas which were identified in the field as only soft substrate in fact have a significant vegetation fraction, which is expected as these are often mixed, with seagrass found on
sandy substrates. We recommend better validation data tailored to assessing mixed habitats and their proportions, such as high-resolution image classifications. We use the pre- and post-unmixing fractions to demonstrate a method to assess change due to the severe cyclone season (Figure 9). We identify primarily major decreases in coral fractions in Matemo, which are accompanied by increases in soft substrate, which could be indicative of sedimentation and correspond to local reports of large-scale coral cover loss but cannot be directly verified.

**Discussion**

Well-informed and effective conservation management in the coastal zone requires an up-to-date state of knowledge and comprehensive data concerning the resources to be managed. In particular, the coastal marine seascape, its distribution of major habitats and underwater morphology are all absolute prerequisites to conservation activities for these assemblages, their context and distribution, not only presence or absence (Purkis et al., 2019). Accurate and reliable spatial data are required for active and efficient management of marine protected areas, and more recently applied to restoration activities. The baseline requirements to manage coastal ecosystems include the typology and structure of the seascape environment, dynamics through time, its state of health and/or conservation status and a suitable monitoring system to support adaptive management or interventions as needed. In QNP, there has been relatively little available spatial data for marine resource management, although it is a highly valuable and resilient reef system of global importance.
These data are fundamental to shaping policies and decision-making, notably related to fisheries management and zoning and are now more critical than ever, particularly in countries facing challenges to sustainable management of coastal resources in the face of climate change and instabilities (Diop et al., 2012) and the long-term human impacts that have drastically altered coral reef systems and associated biodiversity until today (Mcclenachan et al., 2017). Our herein presented benthic habitat mapping effort assesses over 100 000 ha of underwater shallow habitats classified into soft substrates, coral and hardbottom and seagrasses with over 80% accuracy.

Figure 5. top left: Satellite-derived relative depth (up to 15 m); top right: slope in degrees; bottom left: rugosity; bottom right: broad scale bathymetric position index (BPI).

The outputs are very useful for efficient and effective fisheries management and support of local livelihoods and programs such as temporal closures which are important management tools for coral reef ecosystems (Friedlander, 2015).

An effective baseline study should underlie any establishment of MPAs and include the mapping and quantification of the spatio-temporal distribution of the habitats to conserve using replicable methods for status monitoring. As such, remote sensing plays an increasingly important role in the monitoring and management of coastal seascape, including the mapping and monitoring of coral reefs, seagrass meadows and other shallow aquatic environments (Foo & Asner, 2019). Ongoing advances in the development of satellite imagery, cloud computing, machine learning and associated technologies are continuing to improve our ability to accurately derive information on the seascape composition (habitats and species), water properties (nutrients and sedimentation) and water depth, which are important for assessing the ecosystem health of a largely shallow-water MPA. However, given the physical complexity and inherent variability of the aquatic environment, most of the remote-sensing models used to address these challenges require localized input parameters to be effective and are thereby limited in geographic scope.

Although there have been considerable efforts to assess biodiversity in East Africa (Richmond, 2000), QNP has lacked quality, detailed coastal seascape maps since its establishment. Available data are not entirely able to meet the requirements of protected area managers to ensure sustainable fisheries and tourism activities. Our habitat classification, bathymetry and underwater terrain maps indicate a diverse distribution of habitats distributed throughout the seascape, with extensive seagrass beds located at river mouths and bordering mangroves in relatively flat, shallow near shore lagoons. Sand and soft substrates dominate the shallower zones near the atolls, with

<table>
<thead>
<tr>
<th></th>
<th>RF</th>
<th>CART</th>
<th>SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training accuracy</td>
<td>98.6%</td>
<td>96.9%</td>
<td>100%</td>
</tr>
<tr>
<td>Validation Accuracy</td>
<td>82.2%</td>
<td>76.9%</td>
<td>53.1%</td>
</tr>
</tbody>
</table>

Table 2. Training and validation accuracies for the three classification methods.

Figure 6. Classification outputs from three different classifiers: RF (left), CART (middle), SVM (right).
reef lining the outward edges of the atolls, extending to the lagoon areas in the northern and southern parts of the park. This information provides the first holistic view of benthic cover that protected area managers are tasked to conserve for the future. Knowing where habitats exist, their relative depth, structure and pattern are the first step in assessing coral reef resilience, exposure to extreme events, accessibility by humans and potential management or restoration strategies to avoid ecosystem collapse (Bland et al., 2017).

The relative bathymetry dataset shows underwater topography in far greater detail than best available information in nautical maps or charts which are out of date and limited in resolution in shallow waters. The underwater features, and bathymetric structure, notably rugosity are critical drivers for fish communities and biodiversity (Dustan, Doherty, & Pardede, 2013; Wedding et al., 2019). These data also directly enhance the benthic habitat mapping classification as certain habitats and substrates tend to occur in unique underwater zones and knowing relative depth helps account for effects of a varying water column (Eugenio, Marcello, & Martin, 2015). These data also contribute to the baseline information requirements for designating potential fishing areas, temporal closures, use zones, but also can be utilized to evaluate major changes in depths due to cyclones or storm events which might cause extensive sedimentation or changes in the seafloor.

To support these conservation efforts, we opted to use four major discernible and ecosystem important classes for our approach defined primarily by their substrate, which is an important determinant of the ecology of the reef ecosystem as these habitats associate with certain functional groups of species or life cycles (Osuka et al., 2018); while the changes between these classes can be an indicator of degradation (Bellwood et al., 2004). A simple classification scheme was selected to provide unambiguous classes whose presence can be easily identified in situ, while maximizing potential accuracy from a medium resolution sensor (Hochberg & Atkinson, 2003). Despite the accuracy values being within the generally accepted range for management activities, more typologies including a macroalgae class could potentially support a greater number of management activities such as the detection of bleaching or dead coral, or use macroalgae cover as an indicator of reef health (Roff & Mumby, 2012). Classifications define homogenous classes, however, we found that this is not often the case in situ, and within the 10-m Sentinel-2 pixel size, there is in fact a high likelihood of finding mixed coral and rubble, vegetation and sandy seabeds. Our discrete classification results owe to the fact that we could produce a clear satellite image composite with minimal water quality and natural artefacts, and a reference dataset with an adequate horizontal and vertical distribution of habitat classes. The Sentinel-2 dataset also allowed us to employ a spectral unmixing algorithm to define sub-pixel benthic habitats, also enabled by clear water image composition, although this approach is more often applied to hyperspectral imagery (Hedley et al., 2004) and might benefit from additional non-linear techniques to address different water

### Table 3. Accuracy assessment for Random Forest Classification.

<table>
<thead>
<tr>
<th>Validation data</th>
<th>Soft Substrate</th>
<th>Vegetation</th>
<th>Hard Substrate</th>
<th>Deep Water</th>
<th>Total Points</th>
<th>Producer Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Map Data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Soft Substrate</td>
<td>98</td>
<td>6</td>
<td>9</td>
<td>1</td>
<td>114</td>
<td>86%</td>
</tr>
<tr>
<td>Vegetation</td>
<td>9</td>
<td>127</td>
<td>14</td>
<td>2</td>
<td>152</td>
<td>83.5%</td>
</tr>
<tr>
<td>Hard Substrate</td>
<td>9</td>
<td>13</td>
<td>62</td>
<td>0</td>
<td>84</td>
<td>73.8%</td>
</tr>
<tr>
<td>Deep Water</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>9</td>
<td>10</td>
<td>90%</td>
</tr>
<tr>
<td>Total Points</td>
<td>116</td>
<td>147</td>
<td>85</td>
<td>12</td>
<td>350</td>
<td><strong>84.6%</strong></td>
</tr>
<tr>
<td>User Accuracy (%)</td>
<td>84.5%</td>
<td>86.4%</td>
<td>72.9%</td>
<td>75%</td>
<td>overall: 84.6%</td>
<td></td>
</tr>
</tbody>
</table>

### Table 4. Final area calculations and per cent composition for QNP based on the Random Forest classifier, and applying area-weighted accuracy.

<table>
<thead>
<tr>
<th>Class</th>
<th>Area (ha)</th>
<th>Area (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Soft substrate</td>
<td>24 720 ± 1183</td>
<td>23.4%</td>
</tr>
<tr>
<td>Vegetation</td>
<td>29 073 ± 1278</td>
<td>27.5%</td>
</tr>
<tr>
<td>Hard Substrate</td>
<td>19 413 ± 1319</td>
<td>18.3%</td>
</tr>
<tr>
<td>Deep water</td>
<td>32 610 ± 346</td>
<td>30.8%</td>
</tr>
</tbody>
</table>

### Table 5. Benthic habitat types with varying bathymetric indicators.

<table>
<thead>
<tr>
<th>Class</th>
<th>Mean relative depth (m)</th>
<th>Mean Slope (degrees)</th>
<th>Mean Baythymetric Position (unitless)</th>
<th>Mean Rugosity index (unitless)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Soft substrate</td>
<td>4.18</td>
<td>0.93</td>
<td>-17.52</td>
<td>1775</td>
</tr>
<tr>
<td>Seagrass</td>
<td>5.2</td>
<td>1.22</td>
<td>2293</td>
<td></td>
</tr>
<tr>
<td>Coral/Hard Bottom</td>
<td>6.99</td>
<td>1.08</td>
<td>-10.76</td>
<td>2010</td>
</tr>
<tr>
<td>Deep water</td>
<td>12.85</td>
<td>1.32</td>
<td>26.52</td>
<td>74179</td>
</tr>
</tbody>
</table>

D. Poursanidis et al.  Seascapes Mapping with Sentinel-2 in the Cloud

© 2020 The Authors. Remote Sensing in Ecology and Conservation published by John Wiley & Sons Ltd on behalf of Zoological Society of London 11
depths (Hedley & Mumby, 2003). There is a great value in fuzzy classifications, to accompany thematic maps, providing additional detail for mixed and heterogeneous environments. Identify areas which potentially support unique fish assemblages and require additional assessments – this is even more important at spatial resolutions which are larger than the fine-scale habitats of interest for mapping and management.

Northern Mozambique is among the target areas of other mapping efforts, notably the Allen Coral Atlas project (Lyons et al., 2020; https://allenatlasproject.org) aiming to map the global extent of reefs. The Coral Atlas is certainly able to discern habitats in more detail in comparison to Sentinel-2, and also has three additional classes, but has the disadvantage of a higher financial cost, the time and effort to pre-process large volumes of data prior to downstream data analysis, and the very low signal-to-noise ratios (Li et al., 2019). The geomorphological datasets, associated satellite imagery and benthic maps are, however, a great contribution to calibrating and improving a Sentinel-2–based workflow by refining habitat class locations. The 5-day revisit and stable spectral parameters can enable new datasets derived from this effort over time, including relative bathymetry produced at finer time intervals or to respond directly to local needs when they arise.

Given the increasing availability of free data from the Copernicus Sentinel-2 constellation, we see a great potential in consistent, long-term monitoring. The benefits of frequent observations and higher resolution than Landsat allow the creation of optimum surface and water column–corrected reflectance image composites suitable for optically shallow coastal aquatic remote sensing for desired time frames, removing obstacles such as clouds, cloud shadows, turbid waters and sunglint. The use of machine learning algorithms and cloud processing allow for a nearly automated workflow which improves with new calibrations via reference data. The automated aspect of the process means that repeated assessments may be performed over different temporal scales providing results as consistently as possible with minimal user interference. While the classification workflow shown here can be used for monitoring
habitat classes, we have also presented a potential approach to detect sub-pixel changes and trends in mixed habitats using spectral unmixing, which can potentially assess disturbances from cyclone Kenneth which delivered a direct hit to Quirimbas in April of 2019 (Figure 9). Our baseline dataset was developed for a crucial time period before a significant cyclone season in 2019, which was later compounded by recent political instability and insecurity in the area, and the covid-19 pandemic which has greatly reduced access and eliminated most of the protected area enforcement capabilities. Preliminary reports have indicated major damage from the cyclones, and simultaneously little capacity on the ground for collection of additional data in 2020. Given the highly automated nature of our cloud-native geoprocessing framework and the stability, consistency of the Sentinel-2 sensors, we have several options to assess changes, either by evaluating major changes in benthic habitats either via the random forest supervised or by changes in the sub-pixel proportions of the spectral unmixed product.

Cloud-based infrastructures and frameworks for regional- or continental-scale mapping have demonstrated powerful impact for conservation in the terrestrial realm (Hansen et al., 2013) but recent efforts have been targeting the coastal zone (Lyons et al., 2020; Murray et al., 2012). Disk space and bandwidth are no longer barriers in the quest for large-scale mapping efforts, allowing scientists to tailor better methods and apply computation-heavy algorithms such as machine learning. The designed and adapted cloud-native workflow can be rapidly updated by changing the temporal window to update the coastal seascape maps of habitat and bathymetry, ideally calibrated and validated with updated and suitable field data. The use of a cloud computing infrastructure like the Google Earth Engine, and the ability to make the developed code available to local scientists and coding novices is an important step towards the simplification of the use of such tools for the management of an MPA, the creation of baseline maps for conservation prioritization and zonation of the desired area, and the detection of changes after natural hazards. With this effort we aim to implement new baselines for higher temporal resolution monitoring in the long term.

The significant advances of cloud computing, public satellite data archives and automated artificial intelligence frameworks have given birth to efforts pertaining to the mapping and monitoring of the entire coastal seascape ecosystem like the present one, the aforementioned Allen Coral Atlas project, tidal flat monitoring (Murray et al. 2019), the German Aerospace Center funded Global Seagrass Watch project, and Global Mangrove Watch (Bunting et al., 2018). Leveraging cloud-native geoprocessing frameworks for regional-to-continental to global-scale mapping, all of these efforts are demonstrating their value and impact towards effective coastal seascape inventories which will highlight priority areas of resilience or sensitivity for protection, restoration and conservation, enhancing the capacity of countries to measure and monitor their natural resources. As global data become more available, it should however, not deter from efforts to provide locally validated and calibrated datasets.

This seascape mapping effort contributes to a larger overall goal of mapping the entire coastal ecosystem in the region and its essential components, which include corals, seagrasses but also coastal mangroves. When
present together, these elements have been shown to provide better coastal protection and resilience to the impacts of climate change (Guannel et al., 2016). A national mangrove mapping effort also using Sentinel-2 (Shapiro, 2018) has shown that overall mangroves are increasing in Quirimbas, which lends additional support to this relatively intact and important natural resource providing significant ecosystem service benefits in the face of climate change, and warrants long-term protection (Beyer et al., 2018).

The availability of these kinds of seascape datasets can support sustainable development and international financing mechanisms. The East Africa Seascape is still relatively unknown compared to other large reef areas of the world, with few coordinated attempts to create datasets at national scales in support of conservation, protection, climate change adaptation and Nationally Determined Contributions (NDCs); which are at the heart of the Paris Agreement and long-term climate goals. As “blue carbon” from seagrasses is increasingly recognized for potential carbon stock and sequestration (Fourqurean et al., 2012; United Nations Environment Programme, 2020) countries can adapt strategies to reduce national emissions through coastal management and restoration. The International Coral Reef Initiative (ICRI) recently endorsed the inclusion of coral reefs and related ecosystems within the CBD post-2020 Global Biodiversity Framework, of which a number of indicators for priority development will be derived from remote sensing, and the most efficient approach is likely to use Copernicus data and cloud computing (ICRI, 2020).

Regarding the near future of our efforts, we aim to scale up the geoprocessing framework and the related observations to the regional extent of four East African countries (Mozambique, Tanzania, Kenya, Madagascar) to comprehensively map the coastal seascape including seagrasses, corals and mangroves and potentially include additional benthic classes to discern macroalgae from other vegetation, when training data are available. Such scalability can empower the measurement and accountability of blue
carbon inventories which will in turn support conservation and national climate change policy agendas for the four concerned countries; and could potentially serve as good practices to more countries, which feature these blue carbon habitats, for data-driven and effective ecosystem-based adaptation to climate change, both nationally and globally.

Acknowledgements

We wish to acknowledge the dedicated and enthusiastic teams of Quirimbas, including the national park staff, local community assistants and WWF-Mozambique for coordinating field expeditions and crucial data collection. Additional support from BMZ/Bengo, WWF Oceans Practice Innovation Funds and Conservation X Labs.

References


