Mapping of coral reefs using hyperspectral CASI data; a case study: Fordata, Tanimbar, Indonesia

L. BERTELS*,†, T. VANDERSTRAETE‡, S. VAN COILLIE‡, E. KNAEPS†, S. STERCKX‡, R. GOOSSENS‡ and B. DERONDE†
†Flemish Institute for Technological Research, Remote Sensing and Earth Observation Processes, Boeretang 200, 2400 Mol, Belgium
‡Ghent University, Geography Department, Krijgslaan 281, 9000 Gent, Belgium

(Received 28 November 2006; in final form 13 April 2007)

Airborne remote sensing with a CASI-550 sensor has been used to map the benthic coverage and the bottom topography of the Pulau Nukaha coral reef located in the Tanimbar Archipelago (Southeast Moluccas, Eastern Indonesia). The image classification method adopted was performed in three steps. Firstly, five geomorphological reef components were identified using a supervised spectral angle mapping algorithm in combination with data collected during the field survey, i.e. benthic cover type, percentage cover and depth. Secondly, benthic cover mapping was performed for each of the five geomorphological components separately using an unsupervised hierarchical clustering algorithm followed by class aggregation using both spectral and spatial information. Finally, 16 benthic cover classes could be labelled using the benthic cover data collected during the field survey. The overall classification accuracy, calculated on the biological diverse fore reef, was 73% with a kappa coefficient of 0.63. A reliable bathymetric model (up to a depth of 15 m) of the Pulau Nukaha reef was also obtained using a semi-analytical radiative transfer model. When compared with independent in-situ depth measurements, the result proved relatively accurate (mean residual error: −0.9 m) and was consistent with the seabed topography (Pearson correlation coefficient: 86%).

1. Introduction

With the growing concern about the future of the world’s coral reefs, research into the possibilities of remote sensing to support the conservation of this vulnerable ecosystem has significantly increased. According to Bryant et al. (1998), 58% of the coral reefs are potentially endangered, mainly by unsustainable human exploitation and pollution (Bryant et al. 1998, Spalding et al. 2001), but, progressively, also by global climate change (Wilkinson 2002). Coral reefs are not only one of the most spectacular marine ecosystems on Earth, displaying an extremely rich biodiversity, they also offer valuable socio-economic resources for the people living along their coasts. As coral reefs are major fishing grounds and attract large numbers of tourists, they generate important contributions to the national income of many countries. This is of vital importance as two thirds of all countries with reef areas are developing countries, one quarter of which even belong to the least developed group (Whittingham et al. 2003, UNDP 2004). In Asia alone, tens of millions (Moberg and

*Corresponding author. Email: luc.bertels@VITO.be
Folke 1999) to over one billion people (Hinrichsen 1997, Bryant et al. 1998, Whittingham et al. 2003) depend on coral reefs at least for part of their livelihood. At the 2002 Johannesburg World Summit on Sustainable Development, the world leaders have recognised that the maintenance of healthy environments, such as coral reefs, is essential to reduce poverty and to improve human health (UN 2003).

The protection and management of the coral reefs surrounding Fordata (Tanimbar Archipelago, Indonesia) are especially important as the livelihoods of the local people largely depend on the natural resources provided by the reefs. Information on benthic cover and the bathymetric structure of a coral reef is not only important to understand its ecological functioning; it is also necessary baseline data for the development of marine protected area (MPA) zoning plans. To protect and conserve the coral reef ecosystem for future generations, scientific research is needed in order to understand how the coral reef ecosystem functions; how it corresponds to natural and/or human-induced changes in its environment; and how its status has and will evolve over time. A bathymetric map is one of the most important basic documents for coral reef studies (Purkis 2005), because, as depth influences the benthic cover of the reef (Green et al. 2000, Stumpf et al. 2003), it gives an insight into the coral reef’s ecology. Moreover, if detailed bathymetric information is available, the slope and aspect of the reef structure can be defined, and correlations with benthic cover type and coverage may be determined (Isoun et al. 2003). Coral reefs, as physical complex structures, interact with and influence their environment. Therefore, knowledge on the depth at which these structures occur will also support the understanding of the biophysical processes in the marine environment (Stumpf et al. 2003). In addition, bathymetric and benthic cover maps are the necessary baseline documents for MPA planning, and are also useful in marine park management. In the case of oil pollution or coral bleaching (Stumpf et al. 2003), where the effects usually decrease with depth, specific prevention and/or remediation activities can be planned more effectively.

Unfortunately, because of the rather shallow depth at which most coral reefs occur and/or their often remote location, it is not always possible to conduct a conventional bathymetric echo-sounding campaign. Therefore, bathymetric information on coral reefs is often lacking, inaccurate, or out of date (Mumby et al. 1998, Stumpf et al. 2003). Remote sensing, in contrary, is not restricted by inaccessibility or shallowness and data are collected synoptically so that large areas can be mapped at once. As a result, compared to ship-borne bathymetric mapping campaigns, it is often more cost effective (Green et al. 2000, Liceaga-Correa and Euan-Avila 2002, Isoun et al. 2003). The major drawbacks of airborne hyperspectral data for routine monitoring, however, are the limited coverage during each flight and the relatively high costs involved. Airborne hyperspectral remote sensing, nevertheless, is a powerful tool for application and algorithm development, given the versatility it offers. Such (automated) algorithms can consequently be applied to data obtained by Unmanned Airborne Vehicles (UAVs) or (micro-)satellite platforms equipped with hyperspectral sensors, such as the Compact High Resolution Imaging Spectrometer/Project for On-Board Autonomy (CHRIS/PROBA) (Sterckx et al. 2005) and Hyperion (Kruse 2003, Kutser et al. 2006). These systems, which have the ability to map larger areas and to revisit the sites at more regular time intervals, augment the utility of remote sensing for coral reef research.

This research aims to develop a monitoring system for coral reefs using airborne hyperspectral CASI data. It comprises in-situ observations and the collection of
information on the bathymetry and benthic cover of the reef structure. This paper describes how bathymetric data has been derived from the CASI data and how a benthic cover map was obtained by using both bathymetric and spectral information. Due to the high number of spectral bands in a hyperspectral dataset, the radiative transfer model describing the relation between reflectance, water column properties, depth and bottom albedo can be solved for each parameter (Paredes and Spero 1983, Hedley and Mumby 2003). Here, the semi-analytical model by Lee et al. (1998, 1999), which describes the subaquatic radiative transfer, is applied to determine depth. This model is interesting as it does not need additional in-situ calibration data and simultaneously provides information on depth, bottom albedo and water column characteristics. During processing, the determination of water column optical properties and bottom albedo, however, did not deliver acceptable values. Alternative techniques had to be used to account for water column and depth effects on the benthic cover reflectance before a detailed benthic cover map of the reef system could be produced.

2. Study area

The coral reefs surrounding Pulau Fordata and Pulau Nukaha (7° 05' S, 132° 00' E), two small islands in the northeast of the Tanimbar Archipelago, Southeast Moluccas, Indonesia, have been chosen as the study area (see figure 1). In this

![Map of Indonesia](image1)

Figure 1. Location of the study area: Pulau Fordata and Pulau Nukaha are both part of the Tanimbar Archipelago situated in the Southeast of the Moluccas province, Indonesia.
paper, the focus is on the Pulau Nukaha reef system. The Tanimbar Archipelago is among the top ten coral reef hotspots identified by the UNEP's World Conservation Monitoring Centre as being exceptionally rich in endemic marine species but facing extreme threat (Roberts et al. 2002). In addition, many small island communities in this area heavily depend on marine resources for their livelihood. As the status of the reefs is declining due to the lack of sustainable management of these resources, natural stress, and influences of people from outside the area, many young people seeking prosperity migrate to other parts of Indonesia or even abroad. This forms a threat to the culture of the Fordatese people and their unique language. If the Pulau Nukaha reef appears to be exceptionally rich and/or relatively pristine in comparison with other parts of Indonesia, it may be designated as a marine protected area (MPA). This would help to preserve the valuable reef resources and secure the livelihoods of the indigenous people.

3. Methodology

3.1 Data acquisition

A field survey was conducted in the study area between 29 August and 11 September 2005. Several types of information were assembled in order to calibrate and validate the hyperspectral data. At Fordata, no geodetic point was available to anchor a Global Position System (GPS) base station. A Trimble GeoXT GPS was therefore used to create a geodetic point at Pulau Nukaha. The GPS was logging at a fixed point location for eight hours to assure a very accurate position of the reference point. The measured location was then used to create a base station with the same Trimble GPS. Base station measurements were performed each time field data were collected. Each bathymetric sampling point and the beginning and end locations of the transects were measured using a Leica SR20 GPS that logged for 10 s. These field GPS data were then post-processed with the base station GPS data. This increased the positional accuracy of the field GPS data from 4–5 m to 0.5–2 m. A GPS measurement was taken simultaneously with each bathymetric sampling point giving the highest positional accuracy. The beginning of each transect could be measured with the highest positional accuracy because it was located in shallow water at a depth between 0 and 1 m. The end of each transect was located at a much greater depth and was marked with a buoy. The GPS measurement of the end location had to be corrected for drift caused by current, giving a positional accuracy between 0.5 and 4 m.

Digital underwater photographs were taken along transects to collect benthic cover information. To obtain a general overview of the benthic cover types of the Pulau Nukaha reef, the transects were distributed evenly over the reef. By using the Simrad navigation instrument, which was installed in the main vessel, and a chart of the research area, the rough GPS positions of the previously-defined transect locations were determined. These GPS positions were used to navigate the speedboat to the desired locations for the required subaquatic measurements. In total, 43 transects and 5 point samplings were taken around Fordata and at Pulau Nukaha, these are indicated in figure 4(a) below. The exact positions of the beginning and end locations of these transects were measured with the GPS. The transect line, which had a length of 80 m and was set out at right angles to the reef crest or island. At each mark, a 1 m quadrant was placed and
photographed, resulting in nearly 1500 photos. The photographs were analysed by a coral specialist to derive species distribution and cover percentage.

Bathymetric measurements were performed with a handheld echo-sounder (Plastimo Echotest II). On each sample point, three depth measurements were taken and the results averaged. In this way, nearly 300 points were measured with a depth precision of 10 cm, taking into account a tide correction. For correlating the bathymetric measurements with the CASI data, the depths also needed to be corrected for the tide at CASI data acquisition. The strips covering the Pulau Nukaha reef were collected during a tidal rise of 0.7 m. Due to this relatively small range, a mean tidal height of 1.6 m above datum was used to adjust all depth measurements.

Finally, a sun photometer was used to measure aerosol concentration, optical depth, and water vapour concentration, at the time of the hyperspectral flight. These parameters were used for atmospheric correction of the CASI dataset.

The CASI flight took place on 1 September 2005 between 7 and 11 a.m. local time. The resulting hyperspectral dataset consisted of 16 flight lines with a ground resolution of 2.5 m. In total, 30 spectral bands were recorded, covering the visible and near infrared part of the electromagnetic spectrum: 4 bands with a full width half maximum (FWHM) of 11.28 nm between 430 nm and 500 nm, and 26 bands with a FWHM of 5.64 nm in a spectral range from 500 nm to 850 nm (see table 1).

### 3.2 Data pre-processing

#### 3.2.1 Atmospheric correction

The radiance received by the sensor $L_{rs}$ consists of atmospheric path radiance $L_{atm-path}$, background path radiance $L_{rs,b}$ and water leaving radiation $L_{target}$:

$$L_{rs} = L_{atm-path} + L_{rs,b} + L_{target},$$

(1)

Table 1. CASI band settings showing centre wavelength and bandwidth for the 30 bands that were chosen for benthic cover mapping.

<table>
<thead>
<tr>
<th>Band no.</th>
<th>Centre wavelength (nm)</th>
<th>Band width (nm)</th>
<th>Band no.</th>
<th>Centre wavelength (nm)</th>
<th>Band width (nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>431.2</td>
<td>11.6</td>
<td>16</td>
<td>605.5</td>
<td>5.9</td>
</tr>
<tr>
<td>2</td>
<td>449.9</td>
<td>11.5</td>
<td>17</td>
<td>626.1</td>
<td>5.9</td>
</tr>
<tr>
<td>3</td>
<td>462.9</td>
<td>11.5</td>
<td>18</td>
<td>631.7</td>
<td>6.0</td>
</tr>
<tr>
<td>4</td>
<td>476.0</td>
<td>11.5</td>
<td>19</td>
<td>644.9</td>
<td>6.0</td>
</tr>
<tr>
<td>5</td>
<td>499.2</td>
<td>5.9</td>
<td>20</td>
<td>650.5</td>
<td>6.0</td>
</tr>
<tr>
<td>6</td>
<td>506.7</td>
<td>5.9</td>
<td>21</td>
<td>663.7</td>
<td>6.0</td>
</tr>
<tr>
<td>7</td>
<td>517.8</td>
<td>5.9</td>
<td>22</td>
<td>669.4</td>
<td>5.9</td>
</tr>
<tr>
<td>8</td>
<td>530.9</td>
<td>5.9</td>
<td>23</td>
<td>675.0</td>
<td>6.0</td>
</tr>
<tr>
<td>9</td>
<td>540.2</td>
<td>5.9</td>
<td>24</td>
<td>688.3</td>
<td>6.0</td>
</tr>
<tr>
<td>10</td>
<td>549.5</td>
<td>5.9</td>
<td>25</td>
<td>699.7</td>
<td>6.0</td>
</tr>
<tr>
<td>11</td>
<td>555.1</td>
<td>5.9</td>
<td>26</td>
<td>709.2</td>
<td>6.0</td>
</tr>
<tr>
<td>12</td>
<td>564.4</td>
<td>5.9</td>
<td>27</td>
<td>736.0</td>
<td>6.1</td>
</tr>
<tr>
<td>13</td>
<td>575.6</td>
<td>5.9</td>
<td>28</td>
<td>749.4</td>
<td>6.0</td>
</tr>
<tr>
<td>14</td>
<td>584.9</td>
<td>5.9</td>
<td>29</td>
<td>784.2</td>
<td>6.0</td>
</tr>
<tr>
<td>15</td>
<td>599.9</td>
<td>5.9</td>
<td>30</td>
<td>851.0</td>
<td>6.1</td>
</tr>
</tbody>
</table>
with

$$L_{\text{target}} = \frac{d^*_{\text{direct}}(\tau, \theta_v) R_{\text{app}} E_d(a)}{\pi},$$

and

$$L_{\text{rs},b} = \frac{d^*_{\text{diffuse}}(\tau, \theta_v) A_{\text{app}} E_d(a)}{\pi},$$

where $E_d(a)$ is the down-welling irradiance above the water surface, $d^*_{\text{diffuse}}$ and $d^*_{\text{direct}}$ are the diffuse and direct ground-to-sensor transmittances, $R_{\text{app}}$ is the target apparent reflectance and $A_{\text{app}}$ is the average background apparent reflectance.

The atmospheric correction of the hyperspectral images was performed using the in-house software WATCOR. WATCOR uses the radiative transfer code MODTRAN-4. The at-sensor radiance was converted to apparent reflectance $R_{\text{app}}$. The apparent reflectance $R_{\text{app}}$ could be estimated from the at-sensor radiance $L_{\text{rs}}$ and the background radiance $L_{\text{rs},b}$ (average radiance of surrounding pixels) according to:

$$R_{\text{app}} = \frac{(c_1 + c_2)(L_{\text{rs}} + c_3)L_{\text{rs},b}}{(c_4 + c_5)L_{\text{rs},b}},$$

where $c_1, \ldots, c_5$ are the atmospheric correction parameters. The visibility was derived from the sun photometer measurements.

3.2.2 Geometric correction. The georeferencing of the CASI images was difficult due to some unexpected problems. The data supplier, operating the CASI sensor, confused GPS times with UTC times, which resulted in a considerable positional shift and blurry result in the processed imagery. There was no boresight calibration carried out, and the data to perform the calibration was missing, causing a positional shift in the average of about 150 m. Positional errors within and between flight lines were present in the initial datasets, probably due to randomly missing attitude (roll, pitch and yaw) data for each flight track. Correction of the positional shifts was accomplished in several steps: firstly, re-processing the data with the correct GPS times was needed, secondly an algorithm was developed to correct the boresight error, thirdly the random positional errors were manually corrected by selecting tie points in overlapping flight lines, and finally the mosaic image was geopositioned using Landsat and Aster images of the area. It can be expected that all these corrections have had consequences on the acquired classification accuracy, although much attention was paid to them.

3.2.3 Sun glint correction. Sun glint, caused by specular reflection of solar radiation on non-flat water surfaces, can be a serious confounding factor for remote sensing of benthic habitats, especially in images with a spatial resolution less than 10 m (Hedley et al. 2005). A simple and robust algorithm based on the method described by Hedley et al. (2005) was used to remove sun glint from the airborne imagery according to:

$$\text{VIS}_{\text{deglinted}} = \text{VIS}_{\text{initial}} - (\text{NIR} \cdot \text{slope}),$$

where $\text{VIS}_{\text{deglinted}}$ is the sun glint corrected image, $\text{VIS}_{\text{initial}}$ is the initial image, NIR is the near infrared waveband, which, in this particular case is the average of the six last bands (band 25 till band 30), and slope is the regression slope between the blue
waveband (average of band 1 to band 3) and the NIR waveband (average of band 25 to band 30). Taking the average of several consecutive bands avoids random spectral noise, especially in the blue region.

3.3 Bathymetric mapping

As shown in figure 2, the signal recorded by the airborne sensor was not only reflected by the bottom, but had also interacted with molecules and particles in the atmosphere and in the water column.

The changes in direction and intensity of solar radiation can be described by radiative transfer models (Zaneveld et al. 2005). The effects of the interaction with the atmosphere were eliminated during data pre-processing, resulting in above-surface remote sensing reflectance, \( R_{rs}(\lambda) \). Several attempts have already been made

![Diagram of light interactions](www.dmu.dk/rescoman/)

Figure 2. Schematic overview of the interactions of incoming light with the atmosphere, the water column and the object before it reaches the airborne sensor (www.dmu.dk/rescoman/).
to model the radiative transfer in the water column. One of them, the (quasi-)single
scattering theory (Gordon et al. 1975), formulates the sub-surface up-welling
reflectance, $r_{rs}(\lambda)$, as a function of the reflectance contributed by the water column,
$r_{rs}^C(\lambda)$, and the bottom, $r_{rs}^B(\lambda)$:

$$r_{rs}(\lambda) = r_{rs}^C(\lambda) + r_{rs}^B(\lambda).$$  \hfill (6)

These components, in turn, are influenced by the water column inherent optical
properties, water depth, $H$, and bottom albedo, $\rho(\lambda)$:

$$r_{rs}(\lambda) = r_{rs}^{dp}(\lambda)[1 - \exp(-2K H)] + \frac{\rho(\lambda)}{\pi} \exp(-2K H),$$  \hfill (7)

where $r_{rs}^{dp}(\lambda) = r_{rs}(\lambda)$ for optically deep waters, and $K$ is the effective attenuation
coefficient (Maritorena et al. 1994) accounting for absorption, $a(\lambda)$, and back-
scattering, $b_b(\lambda)$, in the water column.

As the CASI-derived reflectance values were above-surface values, the refractive
effects at the water–air interface should not be neglected. Lee et al. (1999) described
the relation between above, $R_{rs}(\lambda)$, and the sub-surface remote sensing reflectance,
$r_{rs}(\lambda)$, as:

$$R_{rs} \approx \frac{0.5r_{rs}}{1 - 1.5r_{rs}}.$$  \hfill (8)

Lee et al. (1998, 1999) and Lee and Carder (2005) consequently elaborated equation
(7) and developed a semi-analytical model describing $r_{rs}(\lambda)$, derived from a nadir-
viewing sensor (Lee and Carder 2005). This model is explicitly invertible for depth,
although reflectance is still expressed as a function of three wavelength-dependent
unknowns: $a(\lambda)$, $b_b(\lambda)$ and $\rho(\lambda)$. As a result, for a CASI dataset with 30 bands, 91
$(3n+1)$ unknowns (where $n$ is the number of bands) need to be solved before the
depth can be estimated. To work out these unknowns, Lee et al. (1998, 1999)
introduced several additional models describing the wavelength-dependent
unknowns. For more details on these models, refer to Lee et al. (1998, 1999) and
Lee and Carder (2005). After an additional 550 nm standardised albedo spectrum
for pure sand was derived from the hyperspectral dataset itself, $r_{rs}$ can be expressed
as a function of five scalar unknowns:

$$r_{rs}(\lambda) = f(B,P,G,X,H),$$  \hfill (9)

where $B$ is the bottom albedo at 550 nm, $P$ is the phytoplankton pigment absorption
coefficient at 440 nm, $G$ is the yellow substance and detritus absorption coefficient at
440 nm, $X$ is the particle backscattering coefficient at 400 nm, and $H$ is the depth.

Subsequently, the model was implemented in ENVI® and run in an optimisation
set-up to determine these five unknowns. Figure 3 shows the result for one pixel. In
figure 3(a), the spectral signature of the pixel derived from the pre-processed CASI
dataset transformed to sub-surface reflectance is shown in black. A first
approximation computed by the model is shown in green. The correlation between
both, clearly, is not yet optimised. The modelled unknown values were iteratively
recalculated until the differences between the modelled spectrum and the original
signature were minimised. The final result of this spectral optimisation is shown in
red. In figure 3(b), the corresponding optimal set of values, derived for this specific
pixel, are outlined. In this way, depth could be estimated for each pixel in the dataset
and a bathymetric map of the study area was obtained.
3.4 Classification

3.4.1 Geomorphological classification. The input for classification was the water column corrected image (shown in figure 4(a)) calculated using the algorithms of Lee et al. (1998, 1999), as described earlier. This image was a mosaic of six flight lines. To determine the inherent dimensionality of the image data and to segregate the

![Image](a)

Figure 3. Example of the implemented optimisation method in ENVI®: (a) the black line shows the spectral signature of the pixel, the green line shows the approximation spectrum, the red line shows the final result after optimisation, and (b) the resulting estimation of the five scalar unknowns after optimal matching.

![Image](b)

Figure 4. (a) A visual representation of the mosaic image composed of six flight lines covering Pulau Nukaha. (b) The geomorphological classified image showing the fore reef, back reef, reef crest, shallow and deep lagoon. Note that the small island and waves are masked out. The small gap in the centre is caused by missing data between adjacent flight lines. (c) Masks are extracted for the different reef parts.
noise from the data, a minimum noise fraction (MNF) transformation was performed on the sun glint corrected dataset. Only the first five MNF bands were used in all further classification steps.

The confounding influence of variable depth on bottom reflectance is cited as one of the most common difficulties with remote sensing of underwater environments (e.g. Cracknell et al. 1987). For example, the spectral signature of a certain coral group at 15 m may be similar to that of an algae cover at 2 m. The success of benthic cover mapping depends on the spectral separability of different habitats in the imagery. Similar spectra may lead to confusion in the classification algorithm and misclassification in the output benthic cover map. As the water column optical properties could not be simultaneously determined with depth, an alternative technique has been used to overcome the effects of the water column on the bottom reflectance. To segregate different benthic cover types with nearly identical reflectance spectra, but located on different reef parts, a geomorphological classification was performed first (Andréfouët et al. 2003). Using the observations made during the field survey, regions of interest (ROIs) could be defined for the different geomorphological reef parts. The fore reef could be separated by setting a threshold value of $-75$ on the first band of the MNF transformed image. A conventional spectral angle mapping algorithm was applied for classification and for separation of the back reef, shallow and deep lagoon. The reef crest was separated by subtracting the fore reef and back reef from the original image. Spectral angle mapping is a technique developed by Kruse et al. (1993) for comparing image spectral data to the reference set data’s vector in $n$-dimensional data space. The criteria used for comparison is the angle in radians between the vectors. Smaller spectral angles represent closer matches to the reference spectra. Since the angle of the spectral data vector is not affected by illumination, the classification method is not skewed by topographical and geometric effects on brightness (Kruse et al. 1993).

From the morphologically classified image (figure 4 (b)) different masks (shown in figure 4(c)) were extracted for each geomorphological reef part, i.e. fore reef, reef crest, back reef, shallow lagoon and deep lagoon. These masks were used to segregate the MNF transformed hyperspectral image according to the different geomorphological reef parts.

### 3.4.2 Benthic cover mapping

Benthic cover mapping was performed for each geomorphological reef part separately. Due to the geometric inaccuracy of the hyperspectral dataset obtained, a semi-unsupervised classification scheme was used. This classification algorithm is a semi-unsupervised hierarchical clustering followed by class aggregation, based on spectral and spatial analysis. The distance between two clusters was defined as the average distance for all pairs of objects between each cluster, weighted by the number of objects in each cluster. The distance between clusters was calculated as an Euclidean distance. The objects are the spectra, consisting of the first five bands of the MNF transformed hyperspectral image, extracted using ROIs. Class aggregation was performed to reduce the initial number of classes to a smaller number of meaningful classes that can be labelled using the field data. The classifier used in the clustering and class aggregation algorithm was based on the smallest root mean square error (RMSE) value between the reference spectra and the pixel spectrum. The applied classification algorithm can be divided into three steps.
Firstly, ROIs were defined with the intention to extract the complete spectral variability of each reef part. In the case of the fore reef, the ROIs were the locations of the transects, extended to cover the complete width of the fore reef. Note that most of the transects measured during the survey were located on the fore reef, as this was the biologically most diverse region of the reef. For the other reef parts, several ROIs were randomly chosen, completed with transects from the field survey if existing (back reef and shallow lagoon). In the case of the fore reef, 2236 spectra were extracted by the ROIs (see figure 5(a)), which were grouped by the hierarchical clustering algorithm into 237 clusters (see figure 5(b)).

Secondly, the classes were aggregated to a predefined number that depended on the number of different benthic cover types thought to be expected. This predefined number was set to 20 in the case of the biological diverse fore reef, and was not final as it was corrected in the last, manual, step. The class aggregation was an iterative process based on spatial and spectral analysis. Initially, only the spatial factor was considered, i.e. small classes were merged into larger classes. Several constraints were taken into account: the smallest class found in each iteration step was merged, the larger class needed to have a minimum size and compactness, and the larger class had to be a neighbour of the smallest class. The spectral factor was considered once the size of the smallest class had reached a certain threshold, which in this case was set to 400 pixels. That is, the smaller class was merged into a larger class that resembles the spectral characteristics of the smaller class. This iterative process continued until the predefined number of 20 classes (in the case of the fore reef) was reached.

Finally, some of the aggregated classes were manually fused. Since the aggregation algorithm had its limitations, some classes, containing identical benthic cover types, could not be aggregated by the algorithm. For example, different classes were defined for the sandy bottom found at different depths. These classes were not aggregated, although they could be manually identified as identical benthic cover types. In the case of the fore reef, the 20 classes, obtained after class aggregation, were manually reduced to 12 classes (see figure 6).

Figure 5. (a) A visual presentation of the first three bands from the MNF transformed data (displayed in RGB channels), masked by the fore reef mask. It shows the ROIs (indicated by arrows) used in the semi-unsupervised classification algorithm. (b) Classification result obtained after the first step in the semi-unsupervised classification algorithm. This image contains 237 clusters or classes.
3.4.3 Benthic cover class labelling. To produce a thematic map that accurately represents the distribution of benthic habitats on the reef, accurate class labelling was needed. The nearly 1500 photographs that were acquired during the fieldwork were used to retrospectively add detail to the classification scheme. Species diversity, prevalence and common biotic associations were derived from the photographs. Of particular importance was the identification of key coral species; they were used to define the most important ecological compositions. The geographic position of the transect photographs could not be used because of the geometric inaccuracy of the CASI imagery. Therefore, an interpretation of the photographs and their analysis in combination with the classification result, the geomorphological unit and bathymetric info was performed. Consecutive photographs along the transects could be grouped and assigned to a certain benthic cover class. This analysis finally resulted in classified transects. As the photographs were taken every 2 m along the transects, the length of each transect could be calculated. Subsequently, because the transect locations were measured with the GPS, they could be overlaid on the classified image, albeit with an inaccuracy of several tens to even hundreds of metres. The classes on the classified transects were matched with the underlying classes of the classified image by shifting the transects over the classified image. Interpreting the classes of the shifted transects with the classes of the classified image made it possible to identify and label the benthic cover classes.

3.4.4 Accuracy assessment. Observations during the field survey determined that the fore reef was the most biological diverse part of the Pulau Nukaha reef system. The back reef was mainly homogeneously covered with rubble and calcified rock, while the shallow lagoon consisted mainly of homogeneous unconsolidated carbonate sand, sometimes covered with sparse algae. In the deeper lagoon, patches of corals were found on sandy substrate. Due to the biological diverse fore reef, 90% of the benthic cover data collected during the field survey was collected on this reef part. Consequently, a thorough accuracy assessment of the classification result could only be performed on the fore reef.

Figure 6. Classification result of the fore reef after aggregating the 237 classes to 20 classes followed by manual fusion, reducing the 20 classes to 12 classes.
Due to the geometric inaccuracy of the hyperspectral imagery, it was impossible to use the geographic information of the transects for validation. Nevertheless, an accuracy assessment of the benthic cover map was performed using two methods. Firstly, a standard confusion matrix was applied to the benthic cover map obtained. A relative validation was then performed: a 50% random selection of the ROIs was used to make a first classification result, the remaining 50% was used for a second classification result. The first classification result was considered as the ground truth input file. The second classification result was compared to the predicted cover types of the first classification result to calculate the overall accuracy of the map. The producer’s accuracy is the probability that the predicted class actually presents what is on the ground and is useful for assessing the accuracy of the individual classes (Mumby et al. 1997). Secondly, a comparison was made between the benthic cover distribution percentage obtained by analysing the photo archive and the benthic cover distribution percentage calculated on the classified image. When the transects were randomly sampled, both distribution percentages should match.

4. Results and discussion

4.1 Sun glint correction

Sun glint is readily apparent in the airborne image shown in figure 7(a). The sun glint corrected image is shown in figure 7(b). Sub-aerial features such as land and breaking waves appear as negative values in the ‘deglinted’ image and had to be restored by their initial values if required.

4.2 Bathymetric mapping

The bathymetric map resulting from the implementation of the semi-analytical model on the hyperspectral CASI dataset covering the Pulau Nukaha reef is depicted in figure 8. It should be noted that this map has previously been corrected to tidal datum and a 3 x 3 average filter has been applied to reduce noise. A detailed visual analysis of this result indicated a reasonably good representation of the topographic structure of the reef. Depths could not be determined on those parts of the reef crest where breaking waves cause total reflection of the light at the sea surface.

Figure 7. (a) Hyperspectral track showing sun glint, and (b) the sun glint corrected hyperspectral track.
The two cross sections, indicated by H and V in figure 8, and detailed in figure 9(a) and (b) respectively, clearly illustrate the geomorphological structure of the reef. All characteristics of a mid-ocean reef structure are present: steep drop-offs mark the edge between the reef and the deep ocean floor, while a shallow, broad reef flat encloses a deeper, central lagoon formed by an irregular surface of sand and small coral patches. The W–E profile (figure 9(a)) also crosses the Pulau Nukaha island, which has been given a constant height of 0 m. The model was less consistent over deeper areas. This was partly due to a depth limit of 35 m imposed on the model, but also points to a limitation of the method in deeper waters.

Validation of the result was complicated by the problematic geometrical correction. To overcome this, the CASI flight lines were resampled to the georeferenced Landsat 7 ETM+ scene dating from 22 December 2001. This reduced the spatial resolution from 2.5 to 28.5 m, which allowed the in-situ depth measurements to be linked to the modelled depths. As a result, 295 independent measurements could be used to assess the accuracy of the bathymetric model. The validation results are shown in figure 10 and table 2. The spread of points in figure 10 suggests a close match between measured depth and calculated depth, at least for the shallower parts of the reef. Calculated depths appeared to be less accurate below 15 m. The Pearson product correlation coefficient of 86% (table 2) also points to a good correspondence between the modelled topography and the actual seabed. The mean residual error was relatively small (~0.9 m), although it indicates a general overestimation of depth.
In addition, the CASI-derived result was compared to an alternative bathymetric model of the Pulau Nukaha reef derived from the Landsat 7 ETM+ dataset. The latter bathymetric model was obtained by the application of the maximum depth of penetration zone mapping method (Jupp 1988), which is a generally accepted method to derive bathymetric information out of a multispectral dataset (Green et al. 2000). As seen in figure 11, the same structure and depth variations as in the CASI-derived result were obtained, although the maximum depth was limited to 15 m. The statistical validation of the Landsat 7 ETM+-derived bathymetric model (using dependent field collected data as the in-situ measurements that also needed to be

Figure 9. (a) W–E cross section, and (b) N–S cross section of the bathymetric model of the Pulau Nukaha reef. X axis is metres, y axis is relative pixel distance.

Figure 10. Correlation between measured depths and CASI-derived depths resampled to 28.5m.
used to calibrate the model) resulted in a comparable Pearson product correlation coefficient (87%), although the mean residual error (−1.7 m) was almost double. The scatter plot between measured depth and Landsat 7 ETM+-derived depths shown in figure 12 also indicates more deviation from the optimal fit.

Finally, a difference map (figure 13) was made by subtracting the Landsat 7 ETM+-derived model from the resampled CASI-derived result. The most important differences occurred on the deeper parts of the reef and on the outer rim of the shallow reef flat, as shown by figure 12. The first difference originated from the disparity between the depth limitations imposed on each method: the method of Jupp (1988) was restricted to 15 m, while the semi-analytical model defined depths up to 35 m. The deeper areas, therefore, were less well represented on the Landsat 7 ETM+ result. The differences on the shallow reef flat were caused by an inherent shortcoming of the Jupp (1988) method, which overestimates depth over dark substrates. The outer reef flat border is covered by a dark substrate containing encrusting algae and coral. Seemingly, this problem did not arise with the Lee and

<table>
<thead>
<tr>
<th></th>
<th>Pear...</th>
<th>Mean residual error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson product correlation coefficient</td>
<td>86%</td>
<td>−0.9 m</td>
</tr>
</tbody>
</table>

Figure 11. Bathymetric map of the Pulau Nukaha reef as derived from the Landsat 7 ETM+ dataset by applying the maximum depth of penetration zone mapping method of Jupp (1988).
Figure 12. Correlation between measured depths and Landsat 7 ETM+-derived depths.

Figure 13. Differentiation between resampled CASI-derived bathymetric map and Landsat 7 ETM+-derived model.
Carder (2005) approach, where the depth calculated for the outer rim was comparable to the depths of the adjacent sand and rubble covered reef flat.

The validity of the model thus holds when compared to a bathymetric model obtained using the maximum depth of penetration zone mapping method (Jupp 1988) on a Landsat 7 ETM+ dataset. The actual CASI-derived model might prove to be even more accurate taking into account the generalisation of depth in 28.5 by 28.5 m pixels, a levelled depth estimate used to correct for tidal changes during data acquisition, and potential errors in the in-situ measurements, resulting from the heterogeneous seabed and the instability of the boat due to waves and currents.

This paper described one of the first applications of the Lee et al. (1998, 1999) model to derive depth on a coral reef environment using recent airborne, hyperspectral data. Lee et al. (2001) have already tested the model using AVIRIS data on the shallow (0–4.6 m) Tampa Bay estuary (Florida, USA), which is composed of sand and seagrass beds. The model was found valid for this shallow and turbid environment, although a differentiation needed to be made according to the substrate. Such a separation could not be made in this study as the substrate is very complex and information on its composition had not yet been derived. However, as is indicated by the result and its validation, this is not considered a problem. McIntyre et al. (2006) have compared depths derived using this model to AVIRIS data in deeper waters (10–15.5 m) against high resolution multibeam bathymetry. The modelled depth has been proven to generally underestimate multibeam bathymetry with differences ranging between 0 and 3.3 m and a mean difference of 0.9 m. Below 14 m, differences progressively increased with depth. This is in accordance with the present work, in which the bathymetric model also performs less for depths deeper than 15 m, where it generally overestimates the depth. This may indicate a general depth limitation of the model at a maximum of approximately 15 m, probably related to the turbidity of the water. From this depth onwards, the reflectance of the seabed becomes largely absorbed so that the contribution from the water column overclasses this weak signal and uncertainties in the parameterisation of the radiative transfer functions (Lee and Carder 2005) become more prominent. While depth estimations may still be acquired below this maximum depth, their accuracy significantly decreases.

All evidence nevertheless points to a generic bathymetric method that, in contrast to more conventional (semi-)empirical approaches, does not need a large set of ground truth data to calibrate the model and even performs in turbid waters (Lee et al. 2001, Lee and Carder 2005).

4.3 Classification

As indicated by Green et al. (2000), there is no ‘best’ classification method, i.e. the different methods should be judged in terms of their cost and accuracy. The area in the present research was very remote, difficult to reach, large, and, as could be observed during the field survey, biological very rich and diverse. In addition, the reefs at this location have never before been observed and studied with remote sensing. The use of ecological methods for deriving habitat classes requires an extensive and representative dataset for which the collection would have been very time consuming and costly. Therefore, the use of a semi-supervised classification scheme applied to the hyperspectral dataset, in combination with the benthic cover
data that was collected during the field survey and which was used for class labelling, was a logical and obligatory choice.

When the classification strategy is applied to all five geomorphological reef parts, i.e. hierarchical clustering, class aggregation and manual fusion, a total of 36 classes were found. Since identical classes existed in the five different geomorphological reef parts (e.g. sandy bottom), the total of 36 classes could be manually reduced to 16 classes (see table 3).

Adding classes for the island, waves and deep ocean water results in a total of 19 classes found on the final benthic cover map for the Pulau Nukaha reef system (figure 14). The Nukaha island is the only structure rising above the water surface and could be easily distinguished from the reef structures. Waves braking at the reef crest have a high reflectance value and obstruct the detection of bottom reflectance. Therefore, they are classified as a separate class. The ‘deep water’ class resulted from the geomorphological and bathymetric information and was defined from a depth of 265 m. The ‘sand’ class is found in all geomorphological reef parts. This ‘coral sand’

<table>
<thead>
<tr>
<th>Geomorphological reef unit</th>
<th>Number of spectra extracted</th>
<th>Number of clusters obtained</th>
<th>Number of classes after cluster aggregation</th>
<th>Number of classes after manual fusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fore reef</td>
<td>2236</td>
<td>237</td>
<td>20</td>
<td>12</td>
</tr>
<tr>
<td>Reef crest</td>
<td>288</td>
<td>36</td>
<td>10</td>
<td>6</td>
</tr>
<tr>
<td>Back reef</td>
<td>758</td>
<td>31</td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td>Shallow lagoon</td>
<td>2134</td>
<td>162</td>
<td>20</td>
<td>6</td>
</tr>
<tr>
<td>Deep lagoon</td>
<td>963</td>
<td>52</td>
<td>12</td>
<td>4</td>
</tr>
<tr>
<td>Total</td>
<td>6379</td>
<td>518</td>
<td>72</td>
<td>36 → 16</td>
</tr>
</tbody>
</table>

Figure 14. Benthic cover map of the Pulau Nukaha coral reef. The small island is the only structure that rises above the water surface. As well as sand and the reef crest, 11 coral classes and 3 algae classes could be distinguished.
consists of particles originating from skeletal material of marine organisms, i.e. coral, molluscs, calcareous algae, etc., and have a relative high reflectance. The ‘reef crest’ was the only geomorphological structure that was not subdivided into classes. Due to the strong current and high wave energy, no organism except encrusting algae could flourish at this location. As well as the 11 coral groups, mainly found on the fore reef and deep lagoon, 3 algae classes were also defined.

4.4 Class labelling

Class labelling was carried out by interpreting the photo archive in combination with the classification result, geomorphological information and bathymetric information. In addition to a brief description of the different benthic cover classes, table 4 gives the typical assemblage composition, the most likely geomorphological unit and depth where the class is found and areal coverage of the class. The pictures in figure 15 give a typical impression of those classes.

Analysis of benthic habitats in shallow aquatic environments using hyperspectral remote sensing techniques is complicated by the confounding effects of the overlying water column and the air–water interface. At-sensor measurements over water are a function of water properties, water depth, bottom characteristics, surface waves, illumination conditions and atmospheric influences. The typical land-based atmospheric correction algorithms are therefore not sufficient for the aquatic environment. The WATCORN correction algorithm is an adequate and proven technique for atmospheric correction above water, while the sun glint correction algorithm by Hedley et al. (2005) successfully removes sunlight scattering caused by waves. Despite the poor geometric accuracy of the hyperspectral dataset, a satisfying benthic coverage map could be produced. The applied method, i.e. hierarchical clustering and class aggregation based on spatial and spectral characteristics, is an innovative technique that was validated successfully although some manual interference was necessary.

4.5 Classification accuracy assessment

The overall accuracy of the fore reef classification, obtained by the confusion matrix, was 73% with a kappa statistic of 63% calculated for 10 different bottom types, as shown in table 5. The producer’s accuracy was the highest for deep sand (93%), coral group 5 (deep fore reef with healthy coral) (81%) and coral group 11c (encrusting algae on calcified rock) (84%). Deep sand and coral group 5 cover large continuous areas. Coral group 11c, found on shallow fore reef, has unique spectral properties due to the presence of calcified rock. Coral group 4, rubble found on the deep fore reef (32%) and coral group 2b (38%) had the lowest producer’s accuracies. Coral group 4 is spectrally very similar to coral group 5, deep fore reef with healthy coral, due to missing ‘green’ and ‘red’ wavelengths. Coral group 2a and coral group 2b, both found on the fore reef, are very similar and therefore they are merged into one class, coral group 2, shallow fore reef with hard coral on calcified rock, in the final benthic cover map. The coral groups 11a, 11b and 11c are located on shallow fore reef. To simplify the final benthic cover map, these groups were merged into one coral group 11, shallow fore reef with calcified rock covered with encrusting algae and small coral patches.

A second validation method was obtained by comparing the distribution of benthic cover classes derived from photograph analysis and the distribution of
Table 4. Benthic cover classes and their description.

<table>
<thead>
<tr>
<th>Benthic class</th>
<th>Typical assemblage composition</th>
<th>Comments</th>
<th>Geomorphological unit and depth</th>
<th>Proportion of study area (%)</th>
<th>Total area (km²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coral group 1</td>
<td><em>Sarcophyton</em> spp., <em>Lobophyton</em> spp.</td>
<td>Dense soft coral associations are generally found on sandy substrate. Different gorgonian species can be present. Coverage of 70%–100%.</td>
<td>Fore reef</td>
<td>4.2</td>
<td>1.14</td>
</tr>
<tr>
<td>Coral group 2</td>
<td><em>Acropora clathrata</em>, <em>Acropora donei</em>, <em>Acropora hyacinthus</em>, <em>Acropora latistella</em>.</td>
<td>Colonies of different table and corymbose <em>Acropora</em> spp. on calcified rock. Coverage is generally 60%–80%. Patches of soft corals, mainly <em>Sarcophyton</em> spp., can be present. The bare rock is mostly covered with encrusting algae.</td>
<td>Fore reef</td>
<td>0.7</td>
<td>0.19</td>
</tr>
<tr>
<td>Coral group 3</td>
<td><em>Acropora clathrata</em>, <em>Acropora hyacinthus</em>, <em>Acropora digitifera</em>, <em>Sarcophyton</em> spp.</td>
<td>Widely spaced patches of table and digitate <em>Acropora</em> spp. on calcified rock. Coverage is 20%–50%. Patches of <em>Sarcophyton</em> spp. and gorgonians are frequently present. Rubble often accumulates in grooves and dips.</td>
<td>Fore reef</td>
<td>1.9</td>
<td>0.52</td>
</tr>
<tr>
<td>Coral group 4</td>
<td><em>Acropora</em> spp., Gorgonians.</td>
<td>Hard coral skeletons disintegrated into piles of rubble are found on the deep flat parts of the fore reef. They have a coverage of 80%–100%. Isolated patches of <em>Acropora</em> spp. and gorgonians can be present.</td>
<td>Fore reef, &lt; -15 m</td>
<td>1.5</td>
<td>0.41</td>
</tr>
<tr>
<td>Coral group 5</td>
<td><em>Acropora</em> spp., <em>Sarcophyton</em> spp., <em>Lobophyton</em> spp., Gorgonians.</td>
<td>Healthy patches of hard and soft corals often intermingled with gorgonians are found on calcified rock located at the deep steeper parts of the fore reef. Coverage of 70%–100%.</td>
<td>Fore reef, &lt; -15 m</td>
<td>7.6</td>
<td>2.07</td>
</tr>
<tr>
<td>Benthic class</td>
<td>Typical assemblage composition</td>
<td>Comments</td>
<td>Geomorphological unit and depth</td>
<td>Proportion of study area (%)</td>
<td>Total area (km²)</td>
</tr>
<tr>
<td>--------------</td>
<td>--------------------------------</td>
<td>----------</td>
<td>---------------------------------</td>
<td>----------------------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>Coral group 6</td>
<td><em>Acropora clathrata</em>, <em>Acropora hyacinthus</em>, <em>Acropora palifera.</em></td>
<td>The parallel ridges of <em>Acropora palifera</em> growing on calcified rock are typical for this coral group. They are accompanied by other <em>Acropora</em> spp., mainly table species. Coverage of 70%–90%.</td>
<td>Fore reef $\pm -2 \leftrightarrow -7$ m</td>
<td>0.4</td>
<td>0.11</td>
</tr>
<tr>
<td>Coral group 7</td>
<td><em>Acropora</em> spp., <em>Seriatopora</em> spp., <em>Heliopora</em> spp., <em>Porites</em> spp., <em>Sarcophyton</em> spp., <em>Lobophyton</em> spp.</td>
<td>The unconsolidated carbonate sand floor of the deep lagoon part is covered with patches of coral associations, composed of all kinds of hard and soft coral species. The diameter of the patches varies from a few metres to tens of metres. Coverage 80%–100%.</td>
<td>Lagoon $\pm -2 \leftrightarrow -9$ m</td>
<td>1.7</td>
<td>0.46</td>
</tr>
<tr>
<td>Coral group 8</td>
<td><em>Asterospicularia</em> spp., <em>Xenia</em> spp.</td>
<td>The unconsolidated carbonate sand floor of the deep lagoon part is covered with patches of mainly soft coral species. The diameter of the patches varies from a few metres to tens of metres. Coverage 60%–90%.</td>
<td>Lagoon $\pm -2 \leftrightarrow -9$ m</td>
<td>2.3</td>
<td>0.63</td>
</tr>
<tr>
<td>Coral group 9</td>
<td><em>Acropora</em> spp., <em>Porites</em> spp.</td>
<td>Unconsolidated rubble and dead coral covered with encrusting algae. Calcified rock covered with widely spaced small plate and table <em>Acropora</em> spp., occasionally <em>Porites</em> boulder corals appear. Coverage is generally 10%–40%.</td>
<td>Back reef $\pm -2$ m</td>
<td>7.5</td>
<td>2.04</td>
</tr>
<tr>
<td>Benthic class</td>
<td>Typical assemblage composition</td>
<td>Comments</td>
<td>Geomorphological unit and depth</td>
<td>Proportion of study area (%)</td>
<td>Total area (km²)</td>
</tr>
<tr>
<td>---------------</td>
<td>--------------------------------</td>
<td>----------</td>
<td>---------------------------------</td>
<td>----------------------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>Coral group 10</td>
<td><em>Sarcophyton</em> spp., <em>Lobophyton</em> spp., <em>Millepora</em> spp.</td>
<td>Dense soft coral associations on sandy substrate alternate with wide open sand planes. Different gorgonian and hard coral species can be present. Coverage of 50%–100%. Very large branching colonies of <em>Millepora</em> spp., often larger than 2 m diameter, can be found as well.</td>
<td>Fore reef ± 7 ↔ ± 15 m</td>
<td>0.4</td>
<td>0.11</td>
</tr>
<tr>
<td>Coral group 11</td>
<td><em>Acropora</em> spp., <em>Sinularia</em> spp., <em>Cladiella</em> spp.</td>
<td>Sparse small table and digitate <em>Acropora</em> spp. on calcified rock often accompanied by soft coral patches and encrusting soft corals. Coverage 50%–80%. Bare rock is mostly covered with encrusting algae.</td>
<td>Fore reef ± 2 m</td>
<td>1.4</td>
<td>0.38</td>
</tr>
<tr>
<td>Algae group 1</td>
<td><em>Acanthophora</em> spp., <em>Turbinaria</em> spp.</td>
<td>Lithified rubble and calcified rock covered with turf algae and moderately dense stands of macro algae. Coverage 80%–100%.</td>
<td>Back reef ± 2 m</td>
<td>10.8</td>
<td>2.94</td>
</tr>
<tr>
<td>Algae group 2</td>
<td><em>Polysiphonia</em> spp., <em>Halodule</em> spp.</td>
<td>Unconsolidated carbonate sand sparsely covered with seagrass and turf algae. Coverage 10%–30%.</td>
<td>Lagoon ± 0.5 ↔ ± 2 m</td>
<td>5.3</td>
<td>1.44</td>
</tr>
<tr>
<td>Algae group 3</td>
<td><em>Polysiphonia</em> spp., <em>Halodule</em> spp.</td>
<td>Unconsolidated carbonate sand sparsely covered with seagrass and turf algae. Coverage 30%–60%.</td>
<td>Lagoon ± 0.5 ↔ ± 2 m</td>
<td>8.3</td>
<td>2.26</td>
</tr>
<tr>
<td>Reef crest</td>
<td></td>
<td>Lithified calcified sediment, mostly covered with encrusting algae. Grooves filled with carbonate sand can be present. Coverage 100%.</td>
<td>Reef crest ± 0.5 ↔ ± 2 m</td>
<td>0.3</td>
<td>0.08</td>
</tr>
<tr>
<td>Sand</td>
<td></td>
<td>Unconsolidated carbonate sand. Coverage 100%.</td>
<td></td>
<td>45.7</td>
<td>12.44</td>
</tr>
</tbody>
</table>
Figure 15. Images of the different benthic cover types showing the typical class composition.
Table 5. Accuracy assessment of the fore reef classification.

<table>
<thead>
<tr>
<th>Bottom type</th>
<th>Deep sand</th>
<th>2a</th>
<th>2b</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>10</th>
<th>11a</th>
<th>11b</th>
<th>11c</th>
<th>accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deep sand</td>
<td>93</td>
<td>6</td>
<td>1</td>
<td>15</td>
<td>3</td>
<td>17</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>93</td>
</tr>
<tr>
<td>2a</td>
<td>0</td>
<td>38</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>12</td>
<td>38</td>
</tr>
<tr>
<td>2b</td>
<td>0</td>
<td>22</td>
<td>73</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>20</td>
<td>15</td>
<td>6</td>
<td>0</td>
<td>73</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0</td>
<td>25</td>
<td>51</td>
<td>1</td>
<td>2</td>
<td>26</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>51</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>15</td>
<td>32</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>32</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>11</td>
<td>64</td>
<td>81</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>81</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>32</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>52</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>52</td>
</tr>
<tr>
<td>11a</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>76</td>
<td>20</td>
<td>15</td>
<td>0</td>
<td>76</td>
</tr>
<tr>
<td>11b</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>54</td>
<td>1</td>
<td>0</td>
<td>54</td>
</tr>
<tr>
<td>11c</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>8</td>
<td>84</td>
<td>84</td>
<td>63</td>
</tr>
</tbody>
</table>
benthic cover classes obtained from the classification result (see figure 16). Again, only the fore reef was considered. From the photo archive, almost 650 photographs from 14 transects covering the fore reef, were manually analysed and classified according to the 8 classes that were obtained by the image classification procedure. The distribution percentage of the classified image was calculated using more than 260 000 pixels. The calculated cover percentage of coral group 6 and 10 are quasi-identical for both calculated distributions (‘photo analysis’ and ‘classified image’) and are 4% and 6%, and 5% and 7% respectively. Coral group 1, soft corals in shallow water, was found mainly at the northern and western location of the small island. These corals flourish there because they take advantage of the protection of the small island against strong waves and currents that can occur in the area as a result of a strong eastern wind. This means that this benthic cover type is not equally distributed over the reef and therefore it is ‘over-sampled’ by the transects, i.e. a cover percentage of 3% is calculated in the ‘photo analysis’ and 0.2% in the ‘classified image’. The width of the fore reef can vary from 150 m to more than 1000 m. The transects that were set out in the field always start as close as possible against the reef crest and have a limited length, i.e. ± 80 m. Sometimes, two consecutive transects were set out, resulting in a total length of ± 160 m. This means that the transects do not always cover the complete width of the fore reef and therefore underestimate the cover percentage of the benthic classes at greater depths, i.e. coral group 5, healthy hard and soft coral found in deep fore reef. Therefore, coral group 5 is not considered in the distribution analysis. Due to the limited transect length, coral group 11, fore reef with encrusting algae and small coral patches found in shallow water, is overestimated and coral group 4, hard and soft coral found on the deeper fore reef part (−15 m), is underestimated in the ‘photo analysis’, 37% and 20%, and 8% and 25% respectively. The total cover percentage of coral groups 2 and 3, is identical in both distributions and is 43% and 42% respectively, although the individual cover percentage varies. These two benthic classes comprise the same bottom type, i.e. hard coral mainly on calcified rock, minor soft coral can be present. They can only be distinguished by the depth range where they appear, and this is more difficult to deduce from the photographs, i.e. coral group 2 (−2 m ← −7 m) is overestimated, 19% versus 11% respectively, and coral group 3 (−7 m ← −15 m) is underestimated, 24% versus 31% respectively.

Figure 16. Accuracy assessment by comparison between benthic cover type distribution of the photo archive and the benthic cover type obtained after image classification.
A valuable and meaningful accuracy assessment is not possible when there is insufficient reference data available. During the field survey, it was found that the back reef, shallow and deeper lagoon were very homogeneous. Only on the fore reef was a high biological diverse coral reef system found. Therefore, most field observations and measurements concentrated on the fore reef, while minor field observations were made on the back reef, shallow and deeper lagoon. As a consequence, a reliable accuracy assessment could only be performed on the fore reef. However, visual examination of the benthic cover classification on the back reef, shallow and deeper lagoon, also points to an acceptable result. The shallow and deeper lagoons are mainly covered by carbonate sand, and the deeper lagoon contains sparsely distributed coral patches. The back reef is mainly covered by unconsolidated or lithified rubble and calcified rock, possibly covered with algae. This is in accordance with the descriptive field observations of those reef parts that were made by the divers.

4.6 General remark on coral reef classification accuracy

Comparing the classification accuracies between different coral reef studies is not feasible most of the time. Firstly, the reef structures at different locations generally have different geomorphological and biological characteristics, leading to different identifiable benthic cover classes. Those classes have different spectral signatures that cause different separability properties, finally influencing the classification accuracies. Secondly, the coral biodiversity in the Pacific Ocean, especially the Indonesian Archipelago, is one of the highest in the world, and a high coral species distribution was indeed observed during the field survey at Pulau Nukaha. It was also found that the coral benthic cover at this location was very patchy. Therefore, the definition of unique benthic cover classes, containing only one dominant species, was fairly impossible. As a consequence of the benthic cover heterogeneity, the spectral separability between some of the benthic cover classes was poor and this of course has its consequence on the classification accuracy. Therefore, studies conducted at different locations are not entirely comparable due to disparity in the areas mapped by each. For example, Mumby et al. (1998) used CASI hyperspectral data for classifying 9 benthic habitats on a Caribbean coral reef. They achieved an accuracy of 81%, much higher than the 73% that was obtained for the 10 benthic classes in the present study.

4.7 Classification remark

Kutser and Jupp (2006) have shown that living corals are spectrally not separable from each other on a species level. In addition, many authors (Hochberg and Atkinson 2000, 2003, Hochberg et al. 2003, Kutser et al. 2003) have pointed out that living corals are even harder to distinguish from other benthic cover types, based on their spectral signatures. In addition, the coral communities of the Pulau Nukaha reef system (especially those found on the fore reef) are very diverse, which means that no large homogeneous areas of a single or a few coral species can be found. Instead, aggregates of many different species exist close to each other. This means that the depicted classes in this work are actually not representing unique benthic cover communities rather site-specific aggregates of many different species.

The generic application of the present approach is therefore limited by the use of site specific optical differences, and an approach based on bottom type specific
(location-independent) spectral signatures would expand the possibilities. Such an alternative, however, requires an extensive spectral library (e.g. Sandidge and Holyer 1998, Acharya et al. 2002, Kutser et al. 2003) covering all benthic cover types present in the region of interest at all possible depths and under most prevalent water column and atmospheric (Kutser et al. 2006) conditions. Such a spectral library could either be developed based on extensive field measurements, on modelled spectra (e.g. using Hydrolight), or derived from an existing library. No optical field measurements of any parameter were however available and the present authors did not have access to an existing ‘coral reef spectral library’. The development of a modelled spectral library, on the other hand, would have been very time-consuming, especially without quantitative knowledge of the parameters to be modelled.

As long as a publicly available ‘coral reef spectral library’ that has been developed and tested based on numerous, globally distributed in-situ spectral measurements does not exist, other, image-based, methods are still valuable. At least the present approach has provided an informative benthic cover map of the Pulau Nukaha reef that was previously not available, and that it may show other scientists or managers how to overcome water column effects without inherent optical property (IOP) data and derive benthic cover information without a spectral library.

5. Further research

5.1 General

Even though an acceptable bathymetric map has been obtained using the semi-analytical model of Lee et al. (1998, 1999), it failed to correctly estimate the water column optical properties and thus to correct the seabed reflectance for water column and depth effects. If depth and water column optical properties could be simultaneously derived, the classification of benthic cover would probably become more straightforward. Further research into the implementation of the semi-analytical model is thus required. The benthic cover map should also be combined with the bathymetric model to determine the value of the Pulau Nukaha reef as a marine protected area (MPA) and, eventually, to demark a multi-use MPA zoning scheme to prepare the establishment of an integrated coastal zone management plan. This could prove to be an ideal support for the livelihood and culture of the local communities. In addition, the approach presented here should be further tested using both airborne and spaceborne hyperspectral data, as the latter would make the bathymetric and benthic cover mapping less costly and more generally applicable.

5.2 Recommendations

Collecting airborne hyperspectral remote sensing data over aquatic environments is a complex task. Numerous factors must be considered when planning a mission in order to collect a proper hyperspectral dataset and to avoid confounding problems when processing the data afterwards (Myers and Miller 2005). A comprehensive overview of imaging spectroscopy of water is given by Dekker et al. (2001). Based on the experience gained with this study, the problems encountered during the field survey and the hyperspectral data processing, some general recommendations for the collection of airborne hyperspectral data over aquatic environments can also be made:
Essential, although it looks trivial, is a reliable flight operator who has experience of collecting hyperspectral data, ensures a calibrated sensor, and delivers the pre-processed and optimal georeferenced data in a reasonable timeframe. Concluding a contract with all commitments and arrangements is not a needless luxury.

To minimize data pre-processing, e.g. sun glint correction, the airborne remote sensing mission should take place in the most optimal conditions. Time of year is an important factor to ensure optimal weather and atmospheric conditions (minimum wind, waves, haze, clouds). Time of day is also important. In the morning and early afternoon, air moisture content is generally low and visibility is best. As coral reefs grow at low latitudes, solar noon must be avoided as it causes direct reflection from the horizontal water surface (sunspot effects) into the field of view (FOV) of the sensor. A rule of thumb is that solar zenith angles of 30° to 60° are optimal and that flight paths should be flown at 0° and 180° headings with respect to the solar azimuth (Dekker et al. 2001).

The collection of in-situ data is essential for tuning the data processing models and assessing the accuracy of the derived products. Since coral reef study sites are often remote areas, it is cost efficient and advantageous for the quality of the final products to spend a sufficient amount of time in the field gathering the required data. Firstly, the collection of a comprehensive dataset, covering the total variability of bottom types, will ameliorate the calibration of the data processing and the validation of the final result. Secondly, it allows the field crew to become better acquainted with the study area such that more knowledge-based rules can be incorporated into the geomorphological and benthic cover mapping.

The formal involvement of a coral specialist (biologist or ecologist) is desired both to assist in the field work and to process and quantitatively analyse the benthic cover data collected. Their knowledge increases the significance of the class labels attached to the benthic cover classification.

Although the collection of in-situ spectroscopy data and IOP determination are not considered strictly necessary, such data are still regarded as a valuable surplus to the subsequent data processing. For example, knowledge of the IOP of the water column would allow bottom albedo values (i.e. bottom reflectance corrected for the water column effects) to be derived, which would optimize, and simplify the subsequent bottom type classification.

6. Conclusions

Benthic cover mapping was performed in two steps. Firstly, five geomorphological reef components were identified using a conventional spectral angle mapping algorithm in combination with the observations made during the field survey that took place in September 2005. Secondly, benthic cover mapping was performed for each of the five geomorphological components separately using a hierarchical clustering algorithm. Both classification algorithms were applied on the minimum noise fraction (MNF) transformed dataset. The clustering algorithm was performed with the spectral information obtained from the different benthic cover types, yielding more than 500 clusters. Subsequently, these clusters were aggregated using both spectral and spatial information, resulting in a set of 16 benthic cover classes. These classes were finally labelled using the data collected during the field survey,
i.e. benthic cover type, percentage cover and depth. The overall classification accuracy, calculated on the biological diverse fore reef, was 73% with a kappa coefficient of 0.63.

Bathymetric measurements using conventional echo-sounding techniques are difficult to implement in shallow reef waters. Therefore, a semi-analytical radiative transfer model has been applied to the hyperspectral dataset. The model has been implemented as an iterative optimisation process to derive the depth directly from the hyperspectral data. When compared with independent in-situ depth measurements, the result proves relatively accurate (mean residual error: $-0.9\,\text{m}$) and consistent with the seabed topography (Pearson correlation coefficient: 86%). As an accurate bathymetric model of the Pulau Nukaha reef has been obtained up to a depth of 15m (below this depth, errors progressively increase), the semi-analytical model applied appears to be an outstanding methodology to derive detailed bathymetric information using hyperspectral data, without requiring a large set of field collected data.

This study presented the methods and algorithms used on hyperspectral CASI data in order to characterise the benthic cover and bathymetric topography of the Pulau Nukaha coral reef system. In summary, although some problems arose during data pre-processing, the applied classification method and the application of the semi-analytical model by Lee et al. (1998, 1999) to the airborne, hyperspectral CASI dataset have resulted in a valuable benthic cover map and bathymetric map of the Pulau Nukaha reef. This is confirmed both by visual inspection of the result, as well as by statistical validation.

Acknowledgements

The authors wish to thank the Belgian Science Policy Office (BelSPO) for providing funding under the STEREO projects UN/67/013 and UN/02/014. We also wish to thank Marina Frederick (BPPT), Yudi Wahyudi (BPPT), Hartanto Sanyaya (BPPT), Freddy Leatemia (LIPI) and the crew from the Majestic for their support during the field campaign. We would also like to thank Herb Ripley (HDI), Joost Vandenabeele (BelSPO) and in memoriam John Trippett for performing the hyperspectral flight campaign. We would also like to thank Sam Purkis (NCRI) for his support of this work.

References


