



Review Article

Image-based seabed classification: what can we learn from terrestrial remote sensing?

Markus Diesing*, Peter Mitchell, and David Stephens

Centre for Environment, Fisheries and Aquaculture Science, Pakefield Road, Lowestoft, Suffolk NR33 0HT, UK

*Corresponding author: tel: +44 1502 524266; fax: +44 1502 513865; e-mail: markus.diesing@cefas.co.uk

Diesing, M., Mitchell, P., and Stephens, D. Image-based seabed classification: what can we learn from terrestrial remote sensing? – ICES Journal of Marine Science, 73: 2425–2441.

Received 1 February 2016; revised 16 May 2016; accepted 12 June 2016; advance access publication 31 July 2016.

Maps that depict the distribution of substrate, habitat or biotope types on the seabed are in increasing demand by marine ecologists and spatial planners, underpinning decision making in relation to marine spatial planning and marine protected area network design. Yet, the science discipline of image-based seabed mapping has not fully matured and rapid progress is needed to improve the reliability and accuracy of maps. To speed up the process we have conducted a literature review of common practices in terrestrial image classification based on remote sensing data, a related discipline, albeit with a larger scientific community and longer history. We identified the following key elements of a mapping workflow: (i) Data pre-processing, (ii) Feature extraction, (iii) Feature selection, (iv) Classification, (v) Post-classification enhancements, and (vi) Evaluation of classification performance. Insights gained from the review served as a baseline against which recent seabed mapping studies were compared. In this way we identified knowledge gaps and propose modifications to the mapping workflow. A main concern in current seabed mapping practice is that a large amount of often correlated predictor features is extracted, creating a multidimensional feature space. To effectively fill this space with an appropriate amount of training samples is likely to be impossible. Hence, it is necessary to reduce the dimensionality of the feature space via data transformation [e.g. principal component analysis (PCA)] or feature selection and remove correlated features. We propose to make dimensionality reduction an integral part of any mapping workflow. We also suggest to adopt recommendations for accuracy assessment originally drawn up for terrestrial land cover mapping. These include the publication of two or more measures of accuracy including overall and class-specific metrics, publication of associated confidence limits and the provision of the error matrix.

Keywords: accuracy, benthic, biotope, classification, habitat, mapping, review, sediment.

Introduction

Categorical maps of seabed characteristics (e.g. substrate, habitat or biotope type) have become an essential tool for various purposes including marine spatial planning and management, design of marine protected area networks, scientific research, assessment of non-living resources, and fisheries resource management among others (Harris and Baker, 2012). As electromagnetic waves in the visible light spectrum (400–700 nm) are quickly attenuated by seawater, optical imaging of the seabed is limited to areas of shallow, and clear water. Acoustic methods have therefore become the main tool for seabed mapping. Recent progress was mainly driven by improvements in multibeam echosounder

technology. In conjunction with ground-truth information collected with various seabed samplers and underwater imaging techniques it is now possible to derive highly detailed maps.

In the context of seabed mapping, the term habitat has acquired various meanings. In its original form, a habitat is a place where an organism lives (Begon *et al.*, 2006) and species distribution modelling is the appropriate methodology to map habitats in that sense. Kostylev *et al.* (2001), in an influential article on benthic habitat mapping, gave a working definition of a habitat as a “spatially defined area where the physical, chemical and biological environment is distinctly different from the surrounding environment”. This definition has recently been revised by

© International Council for the Exploration of the Sea 2016.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Lecours *et al.* (2015) who define benthic habitats as “areas of seabed that are (geo)statistically significantly different from their surroundings in terms of physical, chemical and biological characteristics, when observed at particular spatial and temporal scales”. However, such definitions are largely synonymous with that of a biotope, which might be defined as consisting of the physical environment (habitat) and its distinctive assemblage of conspicuous species (Olenin and Ducrotoy, 2006). It is beyond the scope of this review to settle the question of what a habitat is. For the purpose of this review we are concerned with seabed mapping, which might include any seabed characteristic that can be derived and classified from ground-truth observations and spatially predicted using suitable predictor features and modelling approaches. In that sense, it might include sediments (e.g. Diesing *et al.*, 2014), substrates (e.g. Lucieer *et al.*, 2013), benthoscapes (Brown *et al.*, 2012), habitats (e.g. Che Hasan *et al.*, 2012b), communities (e.g. Ierodionou *et al.*, 2011), biotopes (Buhl-Mortensen *et al.*, 2009), and ecosystems (Costa and Battista, 2013).

Following Brown *et al.* (2011), seabed mapping approaches based on high-resolution acoustic data can be distinguished into expert interpretation, signal-based classification and image-based classification. [These authors actually use the term “segmentation” instead of classification. However, we avoid this term here, as we use segmentation in a specific context related to Geographic Object-Based Image Analysis (GEOBIA) (see chapter 3.4.1)]. Expert interpretation has been the method of choice in the past and is still widely used. However, this approach is highly subjective, not repeatable and often time-consuming. Because of that, we are not dealing with any studies that are solely based on expert interpretation. Signal-based approaches, such as angular range analysis (ARA), usually only utilize backscatter data and are difficult to integrate with other data sets. Here we focus on image-based classification, whereby the “imagery” is taken to mean acoustic backscatter, bathymetry, derivatives of the aforementioned and ancillary data (e.g. oceanographic variables). In that sense, image-based seabed classification has much in common with remote-sensing image classification including terrestrial land cover mapping. The term “land cover” has been defined as the observed bio-physical cover on the earth’s surface (Di Gregorio and Jansen, 1998). Land cover can be readily mapped from images of the Earth’s surface. Initial efforts of land cover mapping were based on aerial photography; however, the modern era of land cover mapping arguably started with the launch of the Landsat 1 satellite in 1972. As such, it has a significantly longer history than image-based seabed mapping and consequently, the science of land cover mapping is much more advanced. This is reflected in the significantly higher number of published studies in the field (Figure 1). It might therefore be beneficial to review the common practice and past experience in land cover mapping processes with a view to improve mapping methodologies and workflows for benthic habitat mapping. Based on such a review, it should be possible to establish best practice in image-based land cover mapping. This will then in turn serve as a baseline against which current practice in seabed mapping can be compared. The objective of this review is to improve current processes in image-based seabed mapping based on experience gained in terrestrial mapping utilizing remote-sensing data.

Method

Initially, we carried out a review of the current literature on land cover mapping based on remotely sensed optical imagery. Due to

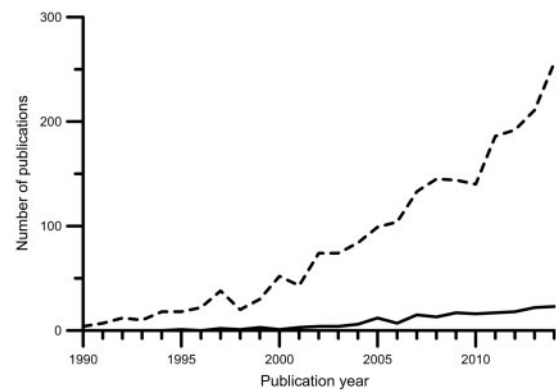


Figure 1. Number of publications per year in the fields of terrestrial land cover (broken line) and benthic habitat mapping (solid line) for the time interval from 1990 to 2014. The literature search was conducted with ISI Web of Science v5.16.1 on 30 September 2015 using the following search criteria: TOPIC: (((marine OR benthic OR seabed) NEAR/1 habitat) NEAR/2 mapping)) and Timespan=1980–2014 for marine studies and TOPIC: ((land NEAR/1 cover NEAR/2 mapping)) and Timespan=1980–2014 for terrestrial studies.

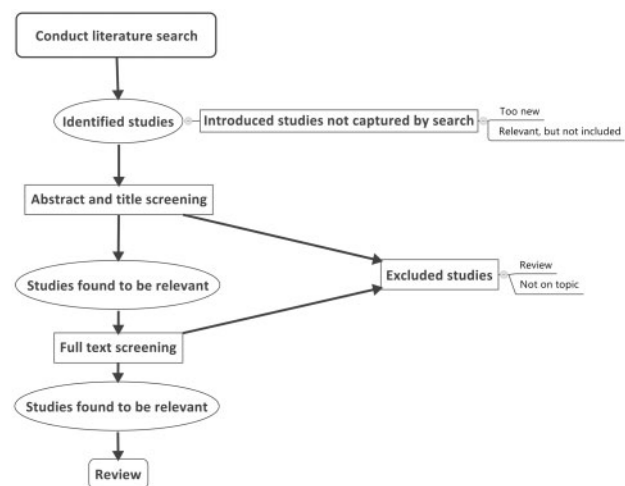


Figure 2. Schematic outlining the workflow of the literature review.

the very large number of published papers (Figure 1), this was mainly conducted as a review of reviews with a focus on classification processes and their common elements. This approach also allowed us to identify relevant studies dealing with specific topics and those were specifically consulted for the sections on feature selection, geographic object-based image analysis (GEOBIA), and accuracy assessment.

As the literature on seabed mapping is much more limited, it was feasible to conduct a structured review of relevant studies. The review approach was adapted from Hughes *et al.* (2014) and is visualized in Figure 2. The review focussed on studies utilizing multibeam echosounder data to make categorical predictions of benthic substrate or habitat utilizing (semi)automated classification methods, i.e. excluding studies that are solely based on expert interpretation. An online literature search was carried out on

30 March 2015 using ISI Web of Science v5.16.1 and the following search criteria:

TOPIC: [(multibeam OR multi-beam) AND (echosounder OR echo-sounder OR sonar OR acoustics)] AND TOPIC: [(seafloor OR habitat OR substrate OR seabed OR benthic) AND mapping] AND TOPIC: (“supervised classification” OR “unsupervised classification” OR classifier OR prediction OR “machine learning” OR “object based”]

This yielded a total of 30 studies. Following an initial screening, a further six studies were introduced. These studies were not included in the search results, as the studies were either too new (in press or just published) or missed for other reasons, but deemed important. The 36 articles were subjected to an abstract and title screening and 7 articles were removed as they were either reviews or not on topic. The remaining 29 articles were obtained as full text copies and after a full text screening a further 9 studies were removed, as they were not on topic. The final set of papers to be reviewed comprised 20 studies spanning the publication years 2004–2015, some of which utilizing more than one classifier (comparative studies).

Land cover mapping process

A universal process for land cover mapping is unlikely to exist; however, a sequence of typical elements of the mapping process (Lu and Weng, 2007) is summarized in Figure 3 and described in the following sections.

Pre-processing

In terrestrial remote sensing, the objective of this step is to present the data in a format from which accurate land cover information can be extracted (Cihlar, 2000). Important image pre-processing steps include radiometric correction of variations in the image resulting from environmental conditions or sensor anomalies, geometric correction to compensate for the Earth's rotation and for variations in the position and attitude of the satellite, terrain correction of relief distortions with the help of digital elevation models (DEMs) and image enhancement to improve the visual interpretability of an image (Purkis and Klemas, 2011). Despite some similarities, acoustic remote-sensing is significantly different from terrestrial remote-sensing and the details of the above-mentioned approaches will not be discussed here.

Acoustic remote-sensing data, either collected with sidescan sonar or multibeam echosounder, need to be radiometrically and geometrically corrected. Geometric corrections change the spatial position of a pixel, while radiometric corrections change the digital number or value assigned to a pixel (Chavez *et al.*, 2002). Further pre-processing steps might include the removal of speckle, i.e. high-frequency noise in the imagery (Blondel and Murton, 1997; Chavez *et al.*, 2002). Variations in backscatter intensity parallel to the vessel's track might be reduced by the application of 2D Fourier filtering (Wilken *et al.*, 2012).

Feature extraction

The term feature is often used to describe predictor variables derived from remote-sensing data to be used for image classification. Many features such as spectral signatures, vegetation indices (e.g. Normalized Difference Vegetation Index, NDVI), transformed images, textural or contextual information, multi-temporal images, multi-sensor images and ancillary data are available for land cover mapping (Lu and Weng, 2007).

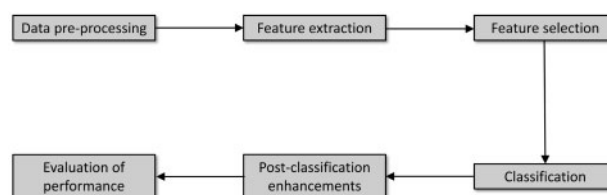


Figure 3. Schematic depicting typical elements of a mapping workflow.

For example, Lucas *et al.* (2011) used NDVI, band ratios, differences and products, relative difference to neighbours, seasonal difference images and other indices for their habitat map of Wales. In seabed mapping, the most widely used feature is acoustic backscatter strength (Brown *et al.*, 2011). As backscatter data are typically collected at one specific frequency (a single band), several useful features successfully applied in terrestrial remote sensing, such as spectral signatures, band ratios, differences, products, and other indices, cannot be applied to image-based seabed mapping. If backscatter data are collected with MBES, then co-registered bathymetric data are available, too. Such data can be used to develop DEMs; a feature that is sometimes used as ancillary data in terrestrial remote sensing.

Due to the fact that primary features are limited to backscatter and bathymetry, it is common practice to calculate secondary features or derivatives. Features derived from bathymetry can be grouped into slope, orientation, curvature/relative position, and terrain variability (Wilson *et al.*, 2007). Several derivatives of backscatter have also been calculated and employed in seabed mapping studies, including standard deviation, roughness, texture (e.g. grey level co-occurrence matrices; Haralick *et al.*, 1973), spatial auto-correlation (Moran's I; Moran, 1950), Hue-Saturation-Intensity (Daily, 1983), Q-values (Preston, 2009), and features extracted from ARA (Fonseca *et al.*, 2009). Table 1 is an attempt to summarize the secondary features employed in the peer-reviewed studies that have explicitly mentioned types of and details on features. From this, it is apparent that the most widely used secondary features are slope, bathymetric position index (BPI), rugosity and some form of curvature. Derivatives of backscatter are less frequently calculated and employed. It is also apparent that frequently derivatives are calculated from a local neighbourhood, i.e. a 3×3 kernel. This might be mainly related to the fact that calculation of derivatives from bordering cells is typically the default setting in the applied software packages, rather than based on scientific reasoning and judgement. It has been argued that a multiscale terrain analysis approach is better suited to capture seabed features at different spatial scales (Wilson *et al.*, 2007) and to account for fuzziness of landscape morphometry (Fisher *et al.*, 2004), but from the above it appears that the possibilities of such an approach are rarely explored.

It follows from the above that the possibilities to extract secondary features from the two primary acoustic data sets are virtually endless, as Table 1, which is non-exhaustive, already contains more than thirty derivatives that could be calculated at various spatial scales. The available choice of features might impact on classification performance and accuracy, as using many features in the hope of capturing as much variability in the data as possible might not only be computationally expensive, but also have a negative effect on classification accuracy (Stephens and Diesing, 2014). It is also well known that several derivatives calculated

Table 1. Secondary features derived from bathymetry and backscatter utilised in benthic habitat mapping studies.

Secondary feature	Kernel size(s)	Examples
<i>from bathymetry</i>		
Slope	3 × 3	Che Hasan <i>et al.</i> (2014)
	3 × 3	Hill <i>et al.</i> (2014)
	3 × 3, 9 × 9, 21 × 21, 49 × 49	Elvenes <i>et al.</i> (2014)
	3 × 3	Ierodiaconou <i>et al.</i> (2011)
	3 × 3	Lucieer <i>et al.</i> (2013)
	3 × 3, 5 × 5, 11 × 11, 31 × 31	Huang <i>et al.</i> (2012)
	3 × 3	Ratray <i>et al.</i> (2009)
	3 × 3	Holmes <i>et al.</i> (2008)
	3 × 3	Diesing <i>et al.</i> (2014)
	?	Brown <i>et al.</i> (2012)
Aspect	3 × 3	Che Hasan <i>et al.</i> (2014)
	3 × 3	Ierodiaconou <i>et al.</i> (2011)
	3 × 3	Holmes <i>et al.</i> (2008)
	?	Brown <i>et al.</i> (2012)
Eastness	3 × 3	Hill <i>et al.</i> (2014)
	3 × 3, 9 × 9, 21 × 21, 49 × 49	Elvenes <i>et al.</i> (2014)
	3 × 3	Lucieer <i>et al.</i> (2013)
	3 × 3	Ratray <i>et al.</i> (2009)
	3 × 3	Stephens and Diesing (2014)
	3 × 3	Diesing <i>et al.</i> (2014)
	3 × 3	Hill <i>et al.</i> (2014)
Northness	3 × 3, 9 × 9, 21 × 21, 49 × 49	Elvenes <i>et al.</i> (2014)
	3 × 3	Lucieer <i>et al.</i> (2013)
	3 × 3	Ratray <i>et al.</i> (2009)
	3 × 3	Stephens and Diesing (2014)
	3 × 3	Diesing <i>et al.</i> (2014)
	10 m, 20 m, 50 m, 100 m	Holmes <i>et al.</i> (2008)
	3 × 3, 5 × 5, 11 × 11, 31 × 31	Huang <i>et al.</i> (2012)
Standard deviation	10 m, 20 m, 50 m, 100 m	Holmes <i>et al.</i> (2008)
	3 × 3	Stephens and Diesing (2014)
	3 × 3	Diesing <i>et al.</i> (2014)
	3 × 3	Che Hasan <i>et al.</i> (2014)
Relief (roughness)	3 × 3	Hill <i>et al.</i> (2014)
	3 × 3	Elvenes <i>et al.</i> (2014)
	3 × 3	Ierodiaconou <i>et al.</i> (2011)
	3 × 3	Lucieer <i>et al.</i> (2013)
	3 × 3, 5 × 5, 11 × 11, 31 × 31	Huang <i>et al.</i> (2012)
	3 × 3	Ratray <i>et al.</i> (2009)
	10 m, 20 m, 50 m, 100 m	Holmes <i>et al.</i> (2008)
	9 × 9, 21 × 21, 49 × 49	Elvenes <i>et al.</i> (2014)
	3 × 3	Hill <i>et al.</i> (2014)
	3 × 3, 9 × 9, 21 × 21, 49 × 49	Elvenes <i>et al.</i> (2014)
Rugosity	3 × 3	Lucieer <i>et al.</i> (2013)
	3 × 3, 5 × 5, 11 × 11, 31 × 31	Huang <i>et al.</i> (2012)
	3 × 3	Ratray <i>et al.</i> (2009)
	10 m, 20 m, 50 m, 100 m	Holmes <i>et al.</i> (2008)
	9 × 9, 21 × 21, 49 × 49	Elvenes <i>et al.</i> (2014)
	3 × 3	Hill <i>et al.</i> (2014)
	3 × 3, 9 × 9, 21 × 21, 49 × 49	Elvenes <i>et al.</i> (2014)
	3 × 3	Lucieer <i>et al.</i> (2013)
	3 × 3, 5 × 5, 11 × 11, 31 × 31	Huang <i>et al.</i> (2012)
	3 × 3	Holmes <i>et al.</i> (2008)
Fractal dimension	3 × 3	Stephens and Diesing (2014)
	3 × 3	Diesing <i>et al.</i> (2014)
	3 × 3	Hill <i>et al.</i> (2014)
	3 × 3, 9 × 9, 21 × 21, 49 × 49	Elvenes <i>et al.</i> (2014)
	3 × 3	Lucieer <i>et al.</i> (2013)
	3 × 3, 5 × 5, 11 × 11, 31 × 31	Huang <i>et al.</i> (2012)
	3 × 3	Holmes <i>et al.</i> (2008)
	3 × 3	Stephens and Diesing (2014)
	3 × 3	Diesing <i>et al.</i> (2014)
	3 × 3	Hill <i>et al.</i> (2014)
Profile curvature	3 × 3, 9 × 9, 21 × 21, 49 × 49	Elvenes <i>et al.</i> (2014)
	3 × 3	Lucieer <i>et al.</i> (2013)
	3 × 3, 5 × 5, 11 × 11, 31 × 31	Huang <i>et al.</i> (2012)
	3 × 3	Holmes <i>et al.</i> (2008)
	3 × 3	Stephens and Diesing (2014)
	3 × 3	Diesing <i>et al.</i> (2014)
	3 × 3	Hill <i>et al.</i> (2014)
	3 × 3, 9 × 9, 21 × 21, 49 × 49	Elvenes <i>et al.</i> (2014)
	3 × 3	Lucieer <i>et al.</i> (2013)
	3 × 3, 5 × 5, 11 × 11, 31 × 31	Huang <i>et al.</i> (2012)
Planar curvature	3 × 3	Holmes <i>et al.</i> (2008)
	3 × 3	Stephens and Diesing (2014)
	3 × 3	Diesing <i>et al.</i> (2014)
	3 × 3	Hill <i>et al.</i> (2014)
	3 × 3, 9 × 9, 21 × 21, 49 × 49	Elvenes <i>et al.</i> (2014)
	3 × 3	Lucieer <i>et al.</i> (2013)
	3 × 3, 5 × 5, 11 × 11, 31 × 31	Huang <i>et al.</i> (2012)
	3 × 3	Holmes <i>et al.</i> (2008)
	3 × 3	Stephens and Diesing (2014)
	3 × 3	Diesing <i>et al.</i> (2014)
Mean curvature	3 × 3, 9 × 9, 21 × 21, 49 × 49	Elvenes <i>et al.</i> (2014)
	3 × 3	Lucieer <i>et al.</i> (2013)
	3 × 3	Holmes <i>et al.</i> (2008)
	?	Brown <i>et al.</i> (2012)
Maximum curvature	3 × 3	Che Hasan <i>et al.</i> (2014)
	3 × 3	Hill <i>et al.</i> (2014)
	3 × 3	Ierodiaconou <i>et al.</i> (2011)
	3 × 3	Ratray <i>et al.</i> (2009)

Continued

Table 1. continued

Secondary feature	Kernel size(s)	Examples
BPI	125 m 4 m, 20 m 150 m, 450 m, 1050 m, 2450 m 75 m 4 m, 20 m 9 m, 15 m, 33 m, 93 m 133 m 200 m, 600 m 21 m, 35 m, 70 m, 175 m 1 × 5, 5 × 10, 10 × 50 10 m, 20 m, 50 m, 100 m	Che Hasan <i>et al.</i> (2014) Hill <i>et al.</i> (2014) Elvenes <i>et al.</i> (2014) lerodiconou <i>et al.</i> (2011) Lucieer <i>et al.</i> (2013) Huang <i>et al.</i> (2012) Rattray <i>et al.</i> (2009) Stephens and Diesing (2014) Diesing <i>et al.</i> (2014) Brown <i>et al.</i> (2012) Holmes <i>et al.</i> (2008)
Hypsometric index	3 × 3	Che Hasan <i>et al.</i> (2014)
Complexity	3 × 3	lerodiconou <i>et al.</i> (2011)
	3 × 3	Rattray <i>et al.</i> (2009)
Moran's I	3 × 3, 5 × 5, 11 × 11, 31 × 31 10 m, 20 m, 50 m, 100 m 5 × 5 5 × 5	Huang <i>et al.</i> (2012) Holmes <i>et al.</i> (2008) Stephens and Diesing (2014) Diesing <i>et al.</i> (2014)
Fuzzy morphometric features (Wood, 1996; Fisher <i>et al.</i> , 2004)	11 × 11, 31 × 31	Huang <i>et al.</i> (2012)
Sobel filter	3 × 3	Stephens and Diesing (2014)
from backscatter		
Standard deviation	Per image object Per image object	Hill <i>et al.</i> (2014) Lucieer <i>et al.</i> (2013)
Roughness	3 × 3 3 × 3	Stephens and Diesing (2014) Diesing <i>et al.</i> (2014)
Hue-Saturation-Intensity	3 × 3 and 11 × 11 3 × 3 3 × 3	Che Hasan <i>et al.</i> (2014) lerodiconou <i>et al.</i> (2011) Rattray <i>et al.</i> (2009)
GLCM Homogeneity	7 × 7 Per image object Per image object 3 × 3, 5 × 5, 11 × 11, 31 × 31	Che Hasan <i>et al.</i> (2014) Hill <i>et al.</i> (2014) Lucieer <i>et al.</i> (2013) Huang <i>et al.</i> (2012)
GLCM Entropy	7 × 7	Che Hasan <i>et al.</i> (2014)
GLCM Correlation	7 × 7	Che Hasan <i>et al.</i> (2014)
GLCM Contrast	Per image object Per image object	Hill <i>et al.</i> (2014) Lucieer <i>et al.</i> (2013)
GLCM Mean	Per image object Per image object	Hill <i>et al.</i> (2014) Lucieer <i>et al.</i> (2013)
GLCM Standard deviation	Per image object Per image object	Hill <i>et al.</i> (2014) Lucieer <i>et al.</i> (2013)
GLCM Variance	3 × 3, 5 × 5, 11 × 11, 31 × 31	Huang <i>et al.</i> (2012)
Moran's I	3 × 3, 5 × 5, 11 × 11, 31 × 31 5 × 5 5 × 5	Huang <i>et al.</i> (2012) Stephens and Diesing (2014) Diesing <i>et al.</i> (2014)
Q-values	?	Brown <i>et al.</i> (2012)
ARA mean	Per image object	Che Hasan <i>et al.</i> (2014)
ARA slope	Per image object	Che Hasan <i>et al.</i> (2014)
ARA skewness	Per image object	Che Hasan <i>et al.</i> (2014)
ARA kurtosis	Per image object	Che Hasan <i>et al.</i> (2014)

from the same primary data are often highly correlated, e.g. slope and measures of terrain complexity or ruggedness (Sappington *et al.*, 2007). It is therefore essential to reduce the number of features (dimensionality) and remove correlated features prior to any classification.

Feature selection

In land cover mapping, features might be selected based on trial and error, previous performance or systematic studies. However,

such approaches become increasingly ineffective, when dealing with a large number of features such as in the case of hyperspectral data (e.g. Demarchi *et al.*, 2014). High dimensionality of data, i.e. a large number of features, may be an obstacle to classification because the number of training samples is often too low to fill the multidimensional space created by these features, a problem known as the curse of dimensionality. The predictive power reduces as the dimensionality increases with a fixed number of training samples (Hughes effect; Hughes, 1968). Dimensionality

reduction methods aim to reduce the dimensionality of the data by reducing redundancy without losing information content.

Two main types of dimensionality reduction methods can be distinguished: Feature transformation methods transform data from the original high-dimensional feature space to a new space with reduced dimensionality. The most common technique of this type of dimensionality reduction methods is the PCA. The main drawback of such methods is that transformed features (e.g. principal components) are often difficult to interpret. Conversely, feature selection algorithms select a subset of features from the original feature set. The objective of feature selection is 3-fold: improving the prediction performance of the predictors, providing faster and more cost-effective predictors and providing a better understanding of the underlying process that generated the data (Guyon and Elisseeff, 2003). Feature selection methods can be categorized into filter-based, wrapper, and embedded techniques. Filter-based methods assess the performance of features independently of the predictor. Filters are generally applied as a pre-processing step before training a classifier. They are usually a fast and generic way of ranking features but do not consider feature interaction (Friedman, 2013). Removing seemingly redundant features based on individual measures of relevance can potentially lead to loss of information due to synergistic interactions between predictor variables (Guyon and Elisseeff, 2003). Wrappers (Kohavi and John, 1997) are linked to a certain classifier and identify relevant features by performing multiple runs of the classifier testing the performance of different subsets of input features. They can be distinguished into forward selection and backward elimination. In the former case, variables are progressively incorporated into larger subsets, whereas in the latter case one starts with the set of all variables and progressively eliminates the least promising ones (Guyon and Elisseeff, 2003). Wrapper methods are generally seen as computationally expensive. Embedded methods perform feature selection in the process of training and are specific to given classifiers (Friedman, 2013).

Removal of correlated features can be achieved via a correlation analysis, whereby only the feature with the highest importance (e.g. established with the Boruta algorithm, Kursa and Rudnicki, 2010) of correlated features is retained. The Marine Geospatial Ecology Tool (MGET, Roberts *et al.*, 2010) provides a means to create a scatterplot matrix plotting predictor features against each other and allowing to assess how correlated they are.

Classification

Classification approaches can be grouped in several ways, e.g. Lu and Weng (2007) distinguish between (i) supervised or unsupervised, (ii) parametric or non-parametric, (iii) hard or soft, and (iv) per-pixel, subpixel, or per-field (including image object) classification. However, the above-mentioned characteristics lead to substantial conceptual overlap; e.g. the same supervised classification might be carried out on pixels and image objects. In the hope of providing more clarity, we therefore distinguish between the unit of analysis and the classification method, similar to Tewkesbury *et al.* (2015), who synthesized image change detection techniques. The unit of analysis might be an image pixel, an image object (a group of contiguous pixels with similar spectral characteristics) or a vector polygon (e.g. a land parcel from a land register database). As the latter is irrelevant to marine applications, it will not be reviewed here. Classification methods can be grouped into supervised and unsupervised methods. Supervised

classification assumes the presence of a priori knowledge of all classes to be mapped.

Unit of analysis

Traditionally, land cover mapping is carried out as a per-pixel approach, whereby classifiers develop a signature by combining the spectra of training-set pixels for a given feature. However, heterogeneity in complex landscapes results in high spectral variation within the same land cover class (Lu and Weng, 2007). This problem has become even more prevalent with the availability of ever higher-resolution images (Blaschke *et al.*, 2014). Under such circumstances, pixel-based approaches often lead to noisy results, as each pixel is grouped into a certain category (Lu and Weng, 2007). Further limitations of pixel-based approaches are that pixels are not true geographical objects, that pixel topology is limited and that texture, context and shape are often neglected (Hay and Castilla, 2006). Burnett and Blaschke (2003) also conclude that the pixel-based approach is usually uni-scale in methodology while landscapes and their land cover exist across several spatial scales. Due to the limitations of pixel-based image analysis, a new field called GEOBIA (Hay and Castilla, 2008) emerged in the early 2000s. Some authors argue that this shift towards object-based approaches has been so profound that GEOBIA might be considered as a new paradigm (Blaschke *et al.*, 2014; Hay and Castilla, 2008).

GEOBIA has been defined as a sub-discipline of geographic information science devoted to developing automated methods to partition remote-sensing imagery into meaningful image objects, and assessing their characteristics through spatial, spectral and temporal scale (Hay and Castilla, 2008). GEOBIA is a special case of OBIA applied to remote-sensing data. The term GEOBIA distinguishes it from other OBIA applications in computer vision, material sciences and biomedical imaging (Blaschke, 2010; Blaschke *et al.*, 2014; Hay and Castilla, 2008).

GEOBIA (as well as OBIA) is a two-step approach consisting of segmentation and classification. Image segmentation has its roots in industrial image processing and was not extensively used in geospatial applications before the 2000s (Blaschke, 2010). Numerous algorithms for image segmentation exist today and are generally divided into four groups: point-based, edge-based, region-based, and combined (Schiewe, 2002). For a review of image segmentation methods see Pal and Pal (1993). Segmentation is the process of complete partitioning of a scene (e.g. remote-sensing image) into non-overlapping regions in scene space (Schiewe, 2002) on the basis of homogeneity (Blaschke *et al.*, 2014). The non-overlapping regions are also called segments or image object primitives, which are different from objects of interest. The latter match real-world objects (e.g. buildings or agricultural land parcels) and object primitives are usually the necessary intermediate step to derive objects of interest (Benz *et al.*, 2004). In image segmentation, regions of minimum heterogeneity given certain constraints have to be found. Criteria for heterogeneity, definition of constraints and the strategy for the sequence of aggregation will determine the final results (Benz *et al.*, 2004).

The creation of image objects increases the number of features that might be used for the subsequent step of image classification. Additional features include image object statistics, texture, shape, topology, and semantics (Benz *et al.*, 2004). In this way, GEOBIA offers possibilities for situations where spectral properties are not unique, but where e.g. shape or neighbourhood relations are

distinct (Blaschke *et al.*, 2014). For example, a lake and a river might have very similar spectral properties and it will be difficult to distinguish these two based on spectral properties alone. Provided the segmentation has resulted in image objects that closely resemble the real-world objects lake and river, it should be relatively straightforward to differentiate the two based on features such as object shape, e.g. length-to-width ratio or asymmetry.

GEOBIA is now widely used in terrestrial remote-sensing. Blaschke *et al.* (2014) estimate that more than 600 peer-reviewed journal articles have been published until September 2013. Refer to Blaschke (2010) for various application examples in terrestrial remote sensing. Conversely, published seabed mapping literature applying an object-based approach is far more limited. Studies can be sub-divided by sensor type: Several studies present results from mapping shallow subtidal habitats such as coral reefs using optical sensors (Knudby *et al.*, 2011; Phinn *et al.*, 2012; Roelfsema *et al.*, 2013; Zhang *et al.*, 2013). Studies applying object-based approaches to acoustic remote-sensing data include Che Hasan *et al.* (2012a,b, 2014), Diesing *et al.* (2014), Hill *et al.* (2014), Lucieer (2008), Lucieer and Lamarche (2011) and Lucieer *et al.* (2013). A recent study has also applied an object-based approach to seabed stills imagery with the aim to extract metrics of seabed heterogeneity (Lacharité *et al.*, 2015).

Classification method

In unsupervised classification, classes to be mapped are not pre-determined. Instead, one attempts to find regularities in unclassified data. In remote-sensing applications, an image is classified based on natural groupings of the spectral properties of the pixels or objects. Typical unsupervised procedures are clustering techniques, e.g. ISODATA (Ball and Hall, 1965) and k-means clustering. Unsupervised classification has been frequently used when mapping land cover over large areas for which there does not exist a sufficient number of ground samples (Cihlar, 2000), while the land cover classes are reasonably straightforward (Franklin and Wulder, 2002). Seabed mapping studies that employed unsupervised classification have been presented by several authors (Lathrop *et al.*, 2006; Brown and Collier, 2008; Blondel and Gomez Sichi, 2009; McGonigle *et al.*, 2009; Brown *et al.*, 2012). One common problem associated with unsupervised classification is the determination of the “correct” or “optimum” number of clusters. A large number of criteria for determining the “optimum” number of clusters exist. For example, Milligan and Cooper (1985) tested 30 criteria for their ability to predict the “correct” number of clusters in a data set. Another potential drawback is that the resultant classification rarely shows an one-to-one relationship with classes derived from ground-truth data (Brown *et al.*, 2012). If the aim is to map habitats according to a classification scheme, then cross-tabulation and aggregation needs to be carried out so that the classified acoustic data reflect the classes found in the ground-truth data.

When ground samples are available in sufficient quantity and quality, and encompassing all land cover types, then supervised classification techniques might be preferable. These are algorithms that “learn” patterns in data to predict an associated discrete class. Terrestrial remote sensing has successfully employed these techniques for optical data for several years and they are being used more regularly in seabed mapping using acoustic data. Supervised machine learning techniques that have been trialled

include maximum likelihood estimation (Buhl-Mortensen *et al.*, 2009; Ierodiaconou *et al.*, 2011; Che Hasan *et al.*, 2012b), k-nearest neighbour (Lucieer *et al.*, 2013; Stephens and Diesing, 2014), various decision trees (Dartnell and Gardner, 2004; Rooper and Zimmermann, 2007; Rattray *et al.*, 2009; Ierodiaconou *et al.*, 2011; Che Hasan *et al.*, 2012a,b; Huang *et al.*, 2012; Stephens and Diesing, 2014), random forest (Che Hasan *et al.*, 2012b, 2014; Diesing *et al.*, 2014; Huang *et al.*, 2012, 2014; Lucieer *et al.*, 2013; Stephens and Diesing, 2014), neural networks (Ojeda *et al.*, 2004; Marsh and Brown, 2009; Huang *et al.*, 2012; Stephens and Diesing, 2014), support vector machines (Che Hasan *et al.*, 2012b; Huang *et al.*, 2012; Stephens and Diesing, 2014) and Bayesian decision rules (Simons and Snellen, 2009; Stephens and Diesing, 2014).

Classification in GEOBIA is typically performed by supervised classification, although it is also possible to carry out unsupervised classification, e.g. k-means clustering, on objects and their properties. Supervised classification requires class-describing information that must be as accurate, representative and complete as possible. Such information on the classes’ characteristics might be acquired by training classifiers (e.g. maximum likelihood, k-nearest neighbour, classification tree, random forest, support vector machine and Bayesian network) with ground samples. Alternatively, the user might describe the class properties based on prior knowledge or knowledge acquired by data exploration. An accurate, representative and complete class description is in most cases effectively impossible. Hence, it can only be a general estimation of the desired class properties (Definiens Imaging, 2004).

Post-classification enhancements

Classifications based on pixels typically lead to “noisy” classification results, often referred to as the “salt and pepper” effect. A majority filter might be applied to generalize the resulting map in such an instance (Lu and Weng, 2007). Further generalization procedures might include the removal of habitat polygons below a previously specified minimum mapping unit by merging them into neighbouring polygons (e.g. Costa and Battista, 2013). Ancillary and contextual data might be applied to modify a classification based on established expert rules. Such knowledge-based enhancements resolve spectral confusion, which occurs when different surface types have similar reflectance properties, and increase the accuracy of the map. The knowledge base might relate to urban context, coastal proximity, soil type, and terrain. For example, littoral sediment might be confused with arable land cover. The distance to the coastline might help resolve such a confusion (Morton *et al.*, 2011). Previous research has indicated the importance of post-classification enhancements in improving the quality of the final map (Harris and Ventura, 1995; Murai and Omatu, 1997).

Evaluation of classification results

Traditionally, land cover maps were qualitatively evaluated based on expert knowledge and estimating map quality was therefore very subjective. This process has been largely superseded by quantitative accuracy assessments to gauge the quality of a produced map. If, however, the aim is to evaluate the suitability of classification algorithms for a specific task, then other criteria such as computational resources, the stability of the algorithm and

robustness to noise might be equally important (DeFries and Chan, 2000).

Accuracy assessment

In land cover mapping the term accuracy refers to how closely a map reflects the actual environment from which it was derived (Foody, 2002). Map accuracy is analysed to determine whether a predictive map is as close to reality as possible given the available resources, and to determine whether it is fit for purpose (Bennett et al., 2013). For example, accuracy assessments allow comparison between maps derived from different datasets and mapping processes (Foody, 2004; Ratray et al., 2009), and are regularly used to inform the confidence of land change maps through time (Olofsson et al., 2013).

In the land cover mapping research literature, accuracy assessments are typically performed by measuring a map's predictive performance against a reference dataset. A number of studies have identified potential improvements to the accuracy assessment process (Stehman, 1997; Foody, 2002; Pontius and Millones, 2011); however, these recommendations have been more slowly adopted in the seabed mapping literature. Due to the sampling methods commonly used by seabed mappers, there are also specific issues to consider when planning a study and during accuracy assessment.

Reference dataset to assess accuracy. Where the quality of a reference dataset is poor the accuracy measure will only indicate the similarity of the map to the reference data, and not necessarily reality (Foody, 2002, 2008). Therefore, selection of appropriate reference data should be carefully considered early in the project development. Ideally, the reference data should be independently sourced from the data used to construct the map (Foody, 2002). However, the costs associated with additional ground-truthing campaigns mean this is rarely an option, particularly in the marine field due to the difficulties of benthic sampling. Therefore, reference data are regularly made up of a subsample of the original dataset that has been withheld from map building.

Determining the correct survey design to develop unbiased and confident accuracy assessments is another commonly discussed issue for land cover and benthic mapping. Random sampling ensures an unbiased estimate of accuracy. However, to sufficiently populate the confusion matrix this may require a very large reference dataset with a minimum of 50 sample sites per class and may require several hundred samples to accurately determine a map's accuracy through random sampling (Congalton, 1991; Carlotto, 2009). Therefore, due to the practical and budgetary constraints of benthic sampling, surveys often target areas of geographic interest or are designed to sample all anticipated habitats (Che Hasan et al., 2012a; Zavalas et al., 2014). Although a random sample of these data may be withheld as the reference dataset, it is important to keep in mind that this is not truly a random sample. Although classification models may be developed from non-randomly targeted samples, the use of these data for accuracy assessment may produce an unrealistic estimation of accuracy (Stehman and Czaplewski, 1998; Foody, 2002). Ultimately, this is a trade-off, and the spatial complexity of the habitat to be mapped may determine the appropriate sampling regime for a site to minimize the amount of bias while ensuring all seabed types are adequately sampled.

As the classification map is compared with a reference dataset, the accuracy assessment is dependent upon the assumption that

the reference data contains no error. This assumption is often optimistic, particularly in the marine environment. Ratray et al. (2014) found large variation in how the same still images of benthic habitat were classified by different trained interpreters, and by the same interpreters at separate times. The differences in interpretation were greater when images were classified into more class levels with only 75% agreement observed for a six-class classification scheme. The differences in classifications would therefore be highly influenced by the classification scheme but would be likely to be interpreted as classification error in the final maps. The distinction between habitat classes in the marine environment are often unclear (Fraschetti et al., 2008), and therefore these types of errors may be more prevalent in benthic mapping than in terrestrial studies.

In determining the quality of a reference dataset, the precision of the testing data location must be considered. If ground truth samples are not accurately georeferenced they may affect the measured map accuracy, particularly in a heterogeneous environment or near habitat boundaries (Stehman and Czaplewski, 1998; Foody, 2002). These types of errors may be more prevalent in the marine realm due to difficulties in accurately positioning benthic samples, even with the use of ultra-short baseline positioning systems (Ratray et al., 2014). Where reference data are not completely accurate, differences between the seabed map and the reference dataset may be incorrectly attributed to false classification, thereby reducing the measured accuracy (Verbyla and Hammond, 1995; Carlotto, 2009).

Accuracy reporting. There has been extensive discussion regarding the relative merit of a number of accuracy assessment measures within the remote sensing literature (Stehman, 1997; Foody, 2002; Liu et al., 2007; Pontius and Millones, 2011). However, all of the most widely recommended and utilized accuracy measures are derived from an error matrix (Foody, 2002; Liu et al., 2007), also known as a confusion matrix or contingency table. The error matrix is a simple cross-tabulation of predicted occurrence versus observed occurrence in the withheld reference dataset. As the error matrix allows all other measures of accuracy, including per-class accuracy, to be calculated, it is recommended to provide a table alongside any other results or accuracy measures (Stehman, 1997; Foody, 2002; Olofsson et al., 2013). Providing additional measures of accuracy to summarize the error matrix is advisable, however, there is still no general consensus as to which measures to include. In addition, different measures may be more suitable depending on the objective of the map, so it is generally recommended to provide multiple measures of map accuracy (Stehman, 1997; Foody, 2002; Liu et al., 2007). It is also important to consider that habitats may change along gradational boundaries and therefore the difference between distinct habitat classes may be somewhat arbitrary (Foody, 2002). Therefore, when considering any accuracy measure it is also important to always consider the significance of the error rather than the value alone.

One of the simplest and most widely used performance metrics is overall accuracy. This is simply the percentage of correctly classified cells from the total number of cells in the error matrix (Congalton, 1991). As this is an estimate of the overall accuracy of the map it should be accompanied by confidence limits (Strahler et al., 2006). In addition to this overall metric, it is often more informative to present class specific measures such as user's and producer's accuracy (Liu et al., 2007; Lyons et al., 2012). User's accuracy is the probability that a pixel on the map

Table 2. Summary of pre-processing methods, extracted features and feature selection methods utilized in the reviewed benthic habitat mapping studies.

References	Year	Primary features		Pre-processing		Secondary features - bathymetry				Secondary feature - backscatter				Oceanographic				Feature selection			
		Bathymetry	Backscatter	Speckle removal	2D Fourier filter	Slope	Orientation	Rel. position	Terrain variability	Multi-scale?	Neighbourhood stats	Spatial auto-correlation	H S I	Texture	ARA	Q-values	Multi-scale?	Maximum orbital velocity	Formal process	Verbal justification	none
Brown <i>et al.</i> (2012)	2012	1	1	0	0	1	1	1	0	1	0	0	0	0	0	1	0	0	0	1	0
Buhl-Mortensen <i>et al.</i> (2009)	2009	1	1	0	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0
Calvert <i>et al.</i> (2015)	2015	1	1	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	1	1	0
Che Hasan <i>et al.</i> (2012a)	2012	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
Che Hasan <i>et al.</i> (2014)	2014	1	1	0	0	1	1	1	1	0	0	1	1	1	1	0	0	0	1	1	0
Che Hasan <i>et al.</i> (2012b)	2012	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
Copeland <i>et al.</i> (2013)	2013	1	1	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
Costa and Battista (2013)	2013	1	1	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0	1	1	0
Darnell and Gardner (2004)	2004	1	1	0	0	0	0	0	1	1	1	0	0	0	0	0	1	0	0	1	0
Diesing <i>et al.</i> (2014)	2014	1	1	0	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	1
Diesing and Stephens (2015)	2015	1	1	1	0	1	1	1	1	1	1	0	0	0	0	0	0	0	1	0	0
Galparsoro <i>et al.</i> (2015)	2015	1	1	0	0	1	0	0	1	0	0	0	1	0	0	0	0	0	0	0	1
Ierodiaconou <i>et al.</i> (2011)	2011	1	1	0	0	1	1	1	1	0	0	0	1	0	0	0	0	0	0	1	0
Lucieer <i>et al.</i> (2013)	2013	1	1	0	0	1	1	1	1	1	1	0	1	0	1	0	0	0	0	0	1
McConigle <i>et al.</i> (2009)	2009	1	1	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1
Neves <i>et al.</i> (2014)	2014	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Plets <i>et al.</i> (2012)	2012	1	1	0	0	1	1	1	1	0	0	0	0	0	0	1	0	0	0	0	1
Rattray <i>et al.</i> (2009)	2009	1	1	0	0	1	1	1	1	0	0	1	0	0	0	0	0	0	0	1	0
Rattray <i>et al.</i> (2015)	2015	1	1	0	0	1	1	1	1	0	0	0	1	0	0	0	1	1	1	0	0
Stephens and Diesing (2014)	2014	1	1	1	0	0	1	1	1	1	1	0	0	0	0	0	0	0	1	0	0
	90%	100%	10%	5%	80%	60%	65%	70%	35%	25%	15%	20%	15%	15%	15%	10%	5%	30%	35%	50%	

Table 3. Summary of classification methods, post-classification enhancements and accuracy assessments utilized in the reviewed benthic habitat mapping studies.

References	Year	Choice of classifier		Classification		Unit of analysis		Classification method		Post-classification enhancements				Accuracy assessment									
		Principal considerations	Not justified	Type	pixel	object	Unsupervised	Supervised	Generalization	KBE rules	manual edits	external	internal	error matrix		Percentage correct	User's accuracy	Producer's accuracy	Balanced error rate	Total agreement	Quantity disagreement	Allocation disagreement	Cramer's V
														included	not included								
Down et al. (2012)	2012	0	0	1	CLUSTER	1	0	1	0	0	0	1	0	1	0	1	0	1	1	0	0	0	0
Buhl-Mortensen et al. (2009)	2009	0	0	1	MLC	1	0	0	1	0	0	0	1	0	1	0	0	0	0	0	0	0	0
Wuert et al. (2015)	2015	0	0	1	ISO Cluster	1	0	1	0	0	0	1	0	0	0	1	0	0	0	0	0	0	1
Wuert et al. (2015)	2015	0	0	1	MLC	1	0	0	1	0	0	0	1	0	0	1	0	0	0	0	1	1	0
Hasan et al. (2012a)	2012	0	1	0	QUEST	0	1	0	1	0	0	1	0	0	1	0	1	1	1	0	0	0	0
Hasan et al. (2014)	2014	0	0	1	Random	1	1	0	1	0	0	1	0	0	1	1	1	0	0	0	0	0	0
Hasan et al. (2012b)	2012	1	0	0	MLC	0	1	0	1	0	0	1	0	0	0	1	1	0	0	0	0	0	0
Hasan et al. (2012b)	2012	1	0	0	QUEST	0	1	0	1	0	0	0	0	1	1	1	0	0	0	0	0	0	0
Hasan et al. (2012b)	2012	0	0	1	Random	0	1	0	1	0	0	1	0	0	1	1	1	0	0	0	0	0	0
Hasan et al. (2012b)	2012	0	0	1	SVM	1	1	0	1	0	0	1	0	0	0	1	1	0	0	0	0	0	0
Copeland et al. (2013)	2013	0	0	1	rule-based	1	0	0	1	0	0	1	0	0	1	1	0	0	0	0	0	0	0
Costa and Battista (2013)	2013	0	0	1	QUEST	0	1	0	1	0	1	0	1	0	1	0	1	1	0	0	0	0	0
Costa and Battista (2013)	2013	0	0	1	QUEST	0	1	0	1	0	1	0	1	0	1	0	1	1	0	0	0	0	0
Costa and Battista (2013)	2013	0	0	1	Decision tree	1	0	0	1	0	0	1	0	0	1	1	0	0	0	0	0	0	0
Gardner (2004)	2014	1	0	0	OBIA	0	1	1	0	0	0	1	1	0	0	1	0	1	1	0	0	0	0
Leising et al. (2014)	2014	1	0	0	Random	1	0	0	1	0	0	1	0	0	1	0	1	1	1	0	0	0	0
Leising et al. (2014)	2014	1	0	0	Random	1	0	0	1	0	0	1	0	0	1	0	1	1	1	0	0	0	0
Leising et al. (2014)	2014	1	0	0	forest	1	0	0	1	0	0	1	0	0	1	0	1	1	1	0	0	0	0
Leising et al. (2014)	2014	1	0	0	Ensemble	1	0	0	1	0	0	1	0	0	0	1	1	0	0	1	0	0	0
Leising et al. (2014)	2014	1	0	0	Ensemble	1	0	0	1	0	0	1	0	0	0	1	1	0	0	1	0	0	0
Stephens (2015)	2015	1	0	0	Ensemble	1	0	0	1	0	0	1	0	0	0	1	1	0	0	1	0	0	0
Galparsoro et al. (2015)	2015	0	0	1	Isodata	1	0	1	0	0	0	1	0	0	0	1	1	1	1	0	0	0	0
Galparsoro et al. (2015)	2015	0	0	1	MLC	1	0	0	1	0	0	1	0	0	0	1	1	1	1	0	0	0	0
Galparsoro et al. (2015)	2015	0	0	1	MLC	1	0	0	1	0	0	1	0	0	0	1	1	1	1	0	0	0	0
Ierodiaconou et al. (2011)	2011	0	0	1	MLC	1	0	0	1	0	0	1	0	0	0	1	1	1	1	0	0	0	0
Ierodiaconou et al. (2011)	2011	1	0	0	QUEST	1	0	0	1	0	0	1	0	0	0	1	1	1	1	0	0	0	0
Ierodiaconou et al. (2011)	2011	1	0	0	QUEST	1	0	0	1	0	0	1	0	0	0	1	1	1	1	0	0	0	0
Ierodiaconou et al. (2011)	2011	1	0	0	QUEST	1	0	0	1	0	0	1	0	0	0	1	1	1	1	0	0	0	0
Ierodiaconou et al. (2011)	2011	1	0	0	QUEST	1	0	0	1	0	0	1	0	0	0	1	1	1	1	0	0	0	0
Ierodiaconou et al. (2011)	2011	1	0	0	QUEST	1	0	0	1	0	0	1	0	0	0	1	1	1	1	0	0	0	0
Ierodiaconou et al. (2011)	2011	1	0	0	QUEST	1	0	0	1	0	0	1	0	0	0	1	1	1	1	0	0	0	0
Ierodiaconou et al. (2011)	2011	1	0	0	QUEST	1	0	0	1	0	0	1	0	0	0	1	1	1	1	0	0	0	0
Ierodiaconou et al. (2011)	2011	1	0	0	QUEST	1	0	0	1	0	0	1	0	0	0	1	1	1	1	0	0	0	0
Ierodiaconou et al. (2011)	2011	1	0	0	QUEST	1	0	0	1	0	0	1	0	0	0	1	1	1	1	0	0	0	0
Ierodiaconou et al. (2011)	2011	1	0	0	QUEST	1	0	0	1	0	0	1	0	0	0	1	1	1	1	0	0	0	0
Ierodiaconou et al. (2011)	2011	1	0	0	QUEST	1	0	0	1	0	0	1	0	0	0	1	1	1	1	0	0	0	0
Ierodiaconou et al. (2011)	2011	1	0	0	QUEST	1	0	0	1	0	0	1	0	0	0	1	1	1	1	0	0	0	0
Ierodiaconou et al. (2011)	2011	1	0	0	QUEST	1	0	0	1	0	0	1	0	0	0	1	1	1	1	0	0	0	0
Ierodiaconou et al. (2011)	2011	1	0	0	QUEST	1	0	0	1	0	0	1	0	0	0	1	1	1	1	0	0	0	0
Ierodiaconou et al. (2011)	2011	1	0	0	QUEST	1	0	0	1	0	0	1	0	0	0	1	1	1	1	0	0	0	0
Ierodiaconou et al. (2011)	2011	1	0	0	QUEST	1	0	0	1	0	0	1	0	0	0	1	1	1	1	0	0	0	0
Ierodiaconou et al. (2011)	2011	1	0	0	QUEST	1	0	0	1	0	0	1	0	0	0	1	1	1	1	0	0	0	0
Ierodiaconou et al. (2011)	2011	1	0	0	QUEST	1	0	0	1	0	0	1	0	0	0	1	1	1	1	0	0	0	0
Ierodiaconou et al. (2011)	2011	1	0	0	QUEST	1	0	0	1	0	0	1	0	0	0	1	1	1	1	0	0	0	0
Ierodiaconou et al. (2011)	2011	1	0	0	QUEST	1	0	0	1	0	0	1	0	0	0	1	1	1	1	0	0	0	0
Ierodiaconou et al. (2011)	2011	1	0	0	QUEST	1	0	0	1	0	0	1	0	0	0	1	1	1	1	0	0	0	0
Ierodiaconou et al. (2011)	2011	1	0	0	QUEST	1	0	0	1	0	0	1	0	0	0	1	1	1	1	0	0	0	0
Ierodiaconou et al. (2011)	2011	1	0	0	QUEST	1	0	0	1	0	0	1	0	0	0	1	1	1	1	0	0	0	0
Ierodiaconou et al. (2011)	2011	1	0	0	QUEST	1	0	0	1	0	0	1	0	0	0	1	1	1	1	0	0	0	0
Ierodiaconou et al. (2011)	2011	1	0	0	QUEST	1	0	0	1	0	0	1	0	0	0	1	1	1	1	0	0	0	0
Ierodiaconou et al. (2011)	2011	1	0	0	QUEST	1	0	0	1	0	0	1	0	0	0	1	1	1	1	0	0	0	0
Ierodiaconou et al. (2011)	2011	1	0	0	QUEST	1	0	0	1	0	0	1	0	0	0	1	1	1	1	0	0	0	0
Ierodiaconou et al. (2011)	2011	1	0	0	QUEST	1	0	0	1	0	0	1	0	0	0	1	1	1	1	0	0	0	0
Ierodiaconou et al. (2011)	2011	1	0	0	QUEST	1	0	0	1	0	0	1	0	0	0	1	1	1	1	0	0	0	0
Ierodiaconou et al. (2011)	2011	1	0	0	QUEST	1	0	0	1	0	0	1	0	0	0	1	1	1	1	0	0	0	0
Ierodiaconou et al. (2011)	2011	1	0	0	QUEST	1	0	0	1	0	0	1	0	0	0	1	1	1	1	0	0	0	0
Ierodiaconou et al. (2011)	2011	1	0	0	QUEST	1	0	0	1	0	0	1	0	0	0	1	1	1	1	0	0	0	0
Ierodiaconou et al. (2011)	2011	1	0	0	QUEST	1	0	0	1	0	0	1	0	0	0	1	1	1	1	0	0	0	0
Ierodiaconou et al. (2011)	2011	1	0	0	QUEST	1	0	0	1	0	0	1	0	0	0	1	1	1	1	0	0	0	0
Ierodiaconou et al. (2011)	2011	1	0	0	QUEST	1	0	0	1	0	0	1	0	0	0	1	1	1	1	0	0	0	0
Ierodiaconou et al. (2011)	2011	1	0	0	QUEST	1	0	0	1	0	0	1	0	0	0	1	1	1	1	0	0	0	0
Ierodiaconou et al. (2011)	2011	1	0	0	QUEST	1	0	0	1	0	0	1	0	0	0	1	1	1	1	0	0	0	0
Ierodiaconou et al. (2011)	2011	1	0	0	QUEST	1	0	0	1	0	0	1	0	0	0	1	1	1	1	0	0	0	0
Ierodiaconou et al. (2011)	2011	1	0	0	QUEST	1	0	0	1	0	0	1	0	0	0	1	1	1	1	0	0	0	0
Ierodiaconou et al. (2011)	2011	1	0	0	QUEST	1	0	0	1	0	0	1	0	0	0	1	1	1	1	0	0	0	0
Ierodiaconou et al. (2011)	2011	1	0	0	QUEST	1	0	0	1	0	0	1	0	0	0	1	1	1	1	0	0	0	0
Ierodiaconou et al. (2011)	2011	1	0	0	QUEST	1	0	0	1	0	0	1	0	0	0	1	1	1	1	0	0	0	0
Ierodiaconou et al. (2011)	2011	1	0	0	QUEST	1	0	0	1	0	0	1	0	0	0	1	1	1	1	0	0	0	0
Ierodiaconou et al. (2011)	2011	1	0	0	QUEST	1	0	0	1	0	0	1	0	0	0	1	1	1	1	0	0	0	0
Ierodiaconou et al. (2011)	2011	1	0	0	QUEST	1	0	0	1	0	0	1	0	0	0	1	1	1	1	0	0	0	0
Ierodiaconou et al. (2011)	2011	1	0	0	QUEST	1	0	0	1	0	0	1	0	0	0	1	1	1	1	0	0	0	0
Ierodiaconou et al. (2011)	2011	1	0	0	QUEST	1	0	0	1	0	0	1	0	0	0	1	1	1	1	0	0	0	0
Ierodiaconou et al. (2011)	2011	1	0	0	QUEST	1	0	0	1	0	0	1	0	0	0	1	1	1	1	0	0	0	0
Ierodiaconou et al. (2011)	2011	1	0	0	QUEST	1	0	0	1	0	0	1	0	0	0	1	1	1	1	0	0	0	0
Ierodiaconou et al. (2011)	2011	1	0	0	QUEST	1	0	0	1	0	0	1	0	0</									

represents that category on the ground. Producer's accuracy is the probability that a reference pixel has been correctly classified. These measures are inversely proportional to another pair of commonly reported measures, the commission error and omission error respectively (Janssen and van der Wel, 1994). Despite the simplicity of overall accuracy, user's and producer's accuracy, they are generally more informative and widely recommended than other regularly used metrics (Liu *et al.*, 2007; Olofsson *et al.*, 2013).

Overall accuracy has been criticised as being an overly optimistic estimation of accuracy as some classes can be correctly classified as a result of chance. Thus the kappa coefficient (Cohen, 1960) was introduced as an accuracy metric which compensates for chance inflating model accuracy. Despite its widespread use, it has regularly been criticised as an accuracy metric (Liu *et al.*, 2007; Pontius and Millones, 2011), as comparing accuracy relative to randomness is of little interest and provides no indication of how to improve classification. Furthermore, as a metric for comparison between different analyses it is commonly used incorrectly as the maps have not been generated from independent samples (Foody, 2004). Pontius and Millones (2011) suggest two new measures in place of kappa, "quantity disagreement" and "allocation disagreement". These new measures provide more relevant information to researchers, allowing them to examine the source of the map error. Despite kappa's critics, and the availability of more informative metrics, it remains commonly used in seabed mapping (Rioja-Nieto *et al.*, 2013; Zhang *et al.*, 2013; Che Hasan *et al.*, 2014). We are aware of only two studies in the marine literature (Savini *et al.*, 2014; Calvert *et al.*, 2015), that have used these new measures in place of kappa. If map accuracy is being assessed to compare different maps and classifications, or to identify where error is occurring and thereby improve classification, then perhaps it is time to move away from kappa in favour of more informative statistics.

Spatial representation. Although the accuracy measures discussed may allow for comparison between maps they represent the "global" average accuracy, and reveal nothing about the spatial distribution of error. In reality, error is rarely distributed evenly across the map (Steele *et al.*, 1998; Kyriakidis and Dungan, 2001). For example, errors are typically concentrated near class boundaries or in areas of high terrain complexity (Steele *et al.*, 1998; Foody, 2005). This is important, as map users may be particularly interested in certain regions within the map, and therefore the overall accuracy may be of little relevance (Comber *et al.*, 2012). Furthermore, these types of analyses may influence how a map is utilised, for instance informing the design of additional surveys. A number of terrestrial studies have developed approaches to map the distribution of classification error across the study site. These methods include geostatistical approaches (Kyriakidis and Dungan, 2001), developing locally constrained confusion matrices (Foody, 2005), and analysing the spatial variation through geographically weighted regression (Comber *et al.*, 2012). A discussion regarding the relative merit of these different approaches is beyond the scope of this review; however, they all provide an indication of the spatial variation in accuracy and assist in the interpretation of map products.

Few studies in the seabed mapping literature have attempted a spatial representation of accuracy. Ahsan *et al.* (2010) apply bootstrap aggregation to map the confidence of the classification map (Breiman, 1996; Steele *et al.*, 2003; Saatchi *et al.*, 2007). Diesing and Stephens (2015) present a confidence map based on the

degree of agreement between models in a multi-model ensemble approach. While not directly mapping accuracy, these methods produce a map of the relative spatial uncertainty of the prediction, and can be highly informative. There are many scenarios in seabed mapping when it may therefore be beneficial to provide similar figures in combination with other accuracy statistics.

Review of marine habitat mapping studies

The analysed studies and the results of the review are summarized in Tables 2 and 3. Only those studies that employed multibeam echosounder data and carried out a form of supervised or unsupervised classification to spatially predict categorical information on seabed type (sediment, substrate, habitat, biotope etc.) were included. All studies employed multibeam backscatter as a primary feature; however, two studies (Che Hasan *et al.*, 2012a,b) did not utilize multibeam bathymetry.

Pre-processing beyond the standard radiometric and geometric corrections is not common practice in seabed mapping. Only two studies (Stephens and Diesing, 2014; Diesing and Stephens, 2015) employed a form of speckle removal; a Gaussian filter with a 5×5 kernel was used in these cases. One study (Diesing *et al.*, 2014) utilized a 2D Fourier filter to reduce prominent track-parallel intensity variations.

Secondary (derived) features of bathymetry are frequently employed, with seabed slope most often used (80% of studies), followed by terrain variability such as rugosity, roughness, and complexity (70%), curvature and relative position such as BPI (65%), and orientation expressed as northness and eastness (60%). About one third of the studies (35%) calculated bathymetric derivatives at more than one spatial scale. Backscatter derivatives were less frequently calculated. A quarter of the studies utilized neighbourhood statistics and one fifth employed derived HSI layers. Less frequently used were spatial auto-correlation, texture, ARA, and Q-values (all 15% of studies). Again, in the majority of cases these derivatives were calculated at a local scale (3×3 kernel). One study (Rattray *et al.*, 2015) utilized an oceanographic variable (annual maximum wave orbital velocity at the seabed).

Thirty percent of the studies employed a form of data transformation, e.g. PCA (Costa and Battista, 2013; Calvert *et al.*, 2015), or feature selection, such as the Boruta algorithm (Stephens and Diesing, 2014; Diesing and Stephens, 2015), Kruskal-Wallis tests (Calvert *et al.*, 2015), random forest variable importance (Che Hasan *et al.*, 2014) and removal of features based on the variance inflation factor (Rattray *et al.*, 2015). The remaining studies only provided a verbal justification for the selected features or no justification at all.

A wide variety of classification methods has been applied. The most frequently used were classification trees, such as QUEST (Quick Unbiased Efficient Statistical Tree, Loh and Shih, 1997), CRUISE (Classification Rule with Unbiased Interaction Selection and Estimation, Kim and Loh, 2001), and rpart (Therneau and Atkinson, 1997), utilized in 22% of the reviewed cases (8 out of 35 applied classifiers). Random forest and maximum likelihood classifier were employed in 17% of cases each, followed by rule-based classifications (11%). Less frequently used methods included unsupervised clustering (8%), k-nearest neighbour (6%), support vector machines (6%), QTC (6%), Bayesian decision rules (3%), neural networks (3%), and a classifier ensemble (3%). No justification for the choice of the applied classifier is given in more than half of the cases (51%). If a justification is given, past

performance (37%) is more frequently cited than principal considerations (11%). Classifications are most often carried out on pixels as the unit of analysis (71%), while image objects were utilized in 31% of the cases. Note that one study (Che Hasan *et al.*, 2014) utilized both image pixels and objects, hence the percentages do not sum up to 100%. Supervised classifications are the predominant classification method applied in 29 out of 35 cases (83%), as compared with 6 cases that employed unsupervised classification methods.

Post-classification enhancements have rarely been applied in the reviewed cases. Two studies (Costa and Battista, 2013; Diesing *et al.*, 2014) carried out manual edits of the automated classifications. One study (Costa and Battista, 2013) generalized the automated classification based on a pre-defined minimum mapping unit. No study was found that applied knowledge-based rules as a scheme of post-classification enhancements.

More than three quarters (77%) of the cases had a form of accuracy assessment based on an “independent”, i.e. not being used for model training, reference dataset associated with them. The remaining studies had either no accuracy assessment carried out or did not withhold samples for an “independent” test. Occasionally, it was also difficult to judge whether samples were withheld or not. The most frequently reported metrics were overall accuracy (86%) and the kappa coefficient (71%), followed by user’s and producer’s accuracy (each 43%) and the balanced error rate (20%). Newly proposed statistics, such as total agreement, quantity disagreement and allocation disagreement (Pontius and Millones, 2011) were employed in one study (Calvert *et al.*, 2015), as was Cramer’s V (Cramer, 1946). In about two thirds (63%) of cases, the error matrix associated with a classification was not published.

Discussion

We have reviewed the current practice in terrestrial remote-sensing with a view to inform and improve standards for seabed mapping. The rationale was that land cover mapping, while similar to image-based seabed mapping in many respects, has a significantly longer history and larger user community and can therefore inform the development of best practices in seabed mapping. Whilst both disciplines have many similarities, significant differences with regard to data availability, temporal and spectral resolution exist between satellite-based remote sensing and acoustic datasets. Whereas the terrestrial land surface is almost continuously mapped and monitored through a wide array of satellite-based sensors at varying spatial and spectral resolutions, large parts of the seafloor remain unsurveyed at spatial resolutions similar to those achieved with satellite sensors. It has been estimated that only 5–10% of the seabed is mapped with a resolution comparable to that on land (Wright and Heyman, 2008) and that >50% of the ocean floor is more than 10 km away from a depth sounding (Sandwell *et al.*, 2014). Global maps of bathymetry estimated from marine gravity measurements are available (Smith and Sandwell, 1997; Sandwell *et al.*, 2014), but lack the resolution that is required for detailed benthic habitat mapping. Backscatter data, arguably the most important feature for seabed mapping, are even more restricted in availability. To complicate matters further, there exist no standards for the collection and processing of backscatter to date. This means that backscatter data collected during different surveys are practically not comparable. Another significant drawback is the lack of spectral resolution of backscatter data: With only one band available, we lack the ability to define and exploit spectral signatures, band ratios

and indices (such as the NDVI in land cover mapping). The potential for improved seabed classification using an array of three multibeam echosounders with frequencies spaced about an octave apart has recently been demonstrated (Hughes-Clarke, 2015); however no commercial multifrequency multibeam echosounders for seabed mapping exist to date. Clearly, progress on that front would benefit the seabed mapping community. Conversely, spatial resolutions achieved with sonar systems are comparable to high- and very high-resolution satellite data products in the case of continental shelf applications or when multibeam echosounders are brought close to the seabed, e.g. mounted on remotely operated vehicles or automated underwater vehicles. Although the limitations outlined above present real barriers to progress in seabed mapping, these differences do not mean that improvements in current image-based seabed mapping practice can’t be made by adapting common land cover mapping practice. We will therefore discuss in which areas image-based seabed mapping practice can be improved and make suggestions how this could be achieved.

Despite the fact that backscatter data are typically noisy, there is little evidence for the application of smoothing filters or speckle removal in the reviewed literature. Furthermore, track-parallel variations in backscatter intensity are commonplace, although less so in the case of multibeam echosounder as compared with sidescan sonar data; yet there is only one instance of the use of a 2D Fourier filter (Diesing *et al.*, 2014). The reluctance to use certain data pre-processing methods might stem from the fact that it is often difficult to establish how much filtering is appropriate, as such filters not only suppress noise but also affect the signal. Speckle, for example, is difficult to distinguish from real signals at the limit of the resolution of the sonar (Blondel, 2009). Studies that systematically investigate the benefits and limitations of various filtering methods across a range of settings might therefore be desirable.

The extracted features, apart from the primary features, utilized in the classification process are typically derived from bathymetry and to a lesser extent backscatter. In only one instance (Rattray *et al.*, 2015) has an oceanographic variable been used. The influence of environmental variables, such as hydrodynamics, temperature, salinity, light, and productivity on habitats is widely acknowledged (Brown *et al.*, 2011; Kostylev and Hannah, 2007); however, such datasets are seldom available at spatial scales comparable to multibeam echosounder data and the variables might also exhibit high temporal variability, which might explain the paucity in their usage. Further on, the choice of the model, the hindcast period, the model resolution and the extracted statistics are likely to influence prediction performance differently, but our understanding regarding this is very limited. Systematic studies that investigate the influence of such factors on prediction performance and accuracy would therefore be an important step towards a better choice of oceanographic features.

Secondary features derived from bathymetry and backscatter are most frequently calculated at a local scale (3×3 kernel). The advantages of multiscale terrain analysis have been highlighted in the past (Wilson *et al.*, 2007). It is also well known that it is not possible to select a single fixed scale (image resolution and neighbourhood size) that will perfectly capture all landscape elements of interest (e.g. MacMillan and Shary, 2009). The importance of spatial scale and geographic context in benthic habitat mapping have been reviewed elsewhere (Lecours *et al.*, 2015) and are not repeated here.

It is frequently the case that a suite of features is utilized with little or no consideration regarding the relevance of these features

for the spatial predictions made. At the same time, the number of training samples is generally low as it is often costly and logistically difficult to obtain seabed samples. This means that a formal process of either data transformation or feature selection is advisable. Yet, the literature review presented here demonstrated that this was rarely the case and no form of either data transformation or feature selection was applied in studies prior to 2013. We strongly argue in favour of a formal step of dimensionality reduction as part of the standard image-based seabed mapping workflow. A main consideration for the choice of method will be whether interpretability of predictor features is desirable and necessary. If this is not the case, then well-known and implemented methods of data transformation such as PCA might be appropriate. Otherwise, there is a choice between different classes of feature selection methods. It appears that filter-based methods, although easy to implement, computationally fast and independent of the selected predictor, run the risk of potentially losing information due to synergistic interactions between predictor variables (Guyon and Elisseeff, 2003). Their independence of predictors might make filter-based methods more suitable for comparative studies with the aim to investigate the performance of various classifiers. Otherwise, wrappers, such as the Boruta algorithm used in conjunction with the popular random forest classifier (Breiman, 2001), and embedded methods might be preferable. Feature selection should be followed by the removal of correlated features, which can be identified with a scatterplot matrix as provided by the MGET tool (Roberts *et al.*, 2010).

Several principal arguments have been made in favour of image objects as the unit of analysis, especially in the case of high to very-high resolution data (Burnett and Blaschke, 2003; Hay and Castilla, 2006, 2008; Blaschke *et al.*, 2014). GEOBIA has been successfully trialled on marine acoustic datasets as well (Lucieer, 2008; Che Hasan *et al.*, 2012b, 2014; Lucieer *et al.*, 2013; Diesing *et al.*, 2014; Hill *et al.*, 2014; Lucieer and Lamarche, 2011). However, no marine study exists that has systematically compared pixels and image objects as units of analysis whilst utilizing the same input datasets and classifiers. The assumption that image objects yield better classification results must therefore be seen as unproven although plausible, until the effect of the unit of analysis on classification performance has been assessed in a structured way.

The application of a certain classification scheme is often mandatory in seabed mapping, which might explain the preference for supervised classification approaches observed in the reviewed studies. However, the classes of frequently used classification schemes were often derived without consideration of what can be mapped acoustically. The relationships between the class of interest and acoustic response are often not well understood and show large class overlap (see boxplots in Lucieer *et al.*, 2013; Diesing *et al.*, 2014; Neves *et al.*, 2014; Stephens and Diesing, 2014; Calvert *et al.*, 2015). The most likely causes are limited spectral resolution as discussed above and insufficient class descriptions associated with shortcomings in sampling design (Clements *et al.*, 2010) and classification of samples (Rattray *et al.*, 2014). The limitations in spectral resolution might be alleviated by the application of ARA (Fonseca *et al.*, 2009). Che Hasan *et al.* (2014) have demonstrated how backscatter angular response and multibeam bathymetry data might be integrated for benthic habitat mapping. Unsupervised classification has been frequently used when mapping land cover over large areas for which ground samples are not available in a sufficient amount (Cihlar, 2000). Unsupervised

classification for seabed mapping might therefore be an alternative when the amount of samples is too small to make supervised classification viable. One important advantage of unsupervised classification is that concerns about the location and representativeness of the sample data are much reduced because clusters are homogeneous by definition (Cihlar, 2000). Therefore, it might be advisable to further explore the benefits of unsupervised classification in seabed mapping. A major problem with this approach is the effect of user-specified parameters, such as the number of clusters and allowable dispersion around a cluster mean, on the classification results. This limitation might be circumvented by “hyperclustering”, i.e. producing a large number of clusters, typically 100–400. Subsequently, the number of clusters is reduced by merging steps based on statistical measures or iteratively by the analyst (Cihlar, 2000). Unsupervised classification of remote sensing data might also be utilized to stratify the collection of ground samples, which then in turn could be utilized in a supervised classification. Such an approach could help increase the spectral separability of classes of interest (Franklin and Wulder, 2002).

There is little evidence for the use of post-classification enhancement methods in the reviewed literature on image-based seabed mapping. This is probably due to the fact that this science discipline is still maturing and has not yet become fully operational. The current focus seems to be placed on the performance of various classifiers with an emphasis on providing objective results. Hence, any post-classification alterations based on expert knowledge, be it by manual editing or in a more formalized rule-based way, are avoided. Once, seabed mapping moves towards being fully operational, it might be expected that knowledge-based enhancements will be implemented more frequently. This will, however, also require a growing knowledge base including geodatabases containing relevant contextual information.

Although it is probably impossible to specify a single, all-purpose measure of classification accuracy (Foody, 2002), it should nevertheless be feasible to draw up minimum requirements based on recommendations made in the literature. These include: (i) the publication of two or more measures of accuracy (Stehman, 1997; Foody, 2002) including overall and class-specific metrics (Strahler *et al.*, 2006), (ii) publication of associated confidence limits (Stehman, 1997; Strahler *et al.*, 2006), and (iii) the provision of the error matrix (Stehman, 1997; Foody, 2002). More research into ways of visualizing spatial variability of accuracy is also encouraged, as such depictions are expected to be of great use for managers and decision makers.

This review has focussed on the processes and workflow elements of an image-based mapping approach. It can be expected that more can be learned from terrestrial remote-sensing and image classification practice that is of relevance for seabed mapping in relation to marine management. This might include how a link between abiotic variables and species is established in terrestrial studies to develop biotope maps at a scale that is relevant to management. This is likely a field where transfer of experience to seabed mapping would be fruitful.

Summary

We have reviewed the current state of terrestrial land cover mapping and identified elements of a mapping workflow, which served as a baseline against which the current practice in image-based seabed mapping was assessed. This has led to the identification of knowledge gaps, which could be addressed with the design of specific comparative studies and recommendations for

improvements, which should be implemented in mapping workflows. This would allow thematic seabed maps to be used more effectively in marine management and decision making.

Funding

This study was funded by the Cefas Research and Development Fund.

References

- Ahsan, N., Williams, S. B., Jakuba, M., Pizarro, O., and Radford, B. 2010. Predictive habitat models from AUV-based multibeam and optical imagery. *In* OCEANS 2010 MTS/IEEE SEATTLE, pp. 1–10. IEEE. <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5663809> (Accessed 23 July 2016).
- Ball, G. H., and Hall, D. J. 1965. *Isodata: A Method of Data Analysis and Pattern Classification*. Stanford Research Institute, Menlo Park, United States. 60 pp.
- Begon, M., Townsend, C. R., and Harper, J. L. 2006. *Ecology: From individuals to Ecosystems*. Blackwell Publishing, Malden, MA, USA. 738 pp.
- Bennett, N. D., Croke, B. F. W., Guariso, G., Guillaume, J. H. A., Hamilton, S. H., Jakeman, A. J., Marsili-Libelli, S., *et al.* 2013. Characterising performance of environmental models. *Environmental Modelling and Software*, 40: 1–20.
- Benz, U. C., Hofmann, P., Willhauck, G., Lingenfelder, I., and Heynen, M. 2004. Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. *ISPRS Journal of Photogrammetry and Remote Sensing*, 58: 239–258. <http://www.sciencedirect.com/science/article/pii/S0924271603000601>.
- Blaschke, T. 2010. Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, 65: 2–16.
- Blaschke, T., Hay, G. J., Kelly, M., Lang, S., Hofmann, P., Addink, E., Queiroz Feitosa, R. *et al.* 2014. Geographic Object-Based Image Analysis - Towards a new paradigm. *ISPRS Journal of Photogrammetry and Remote Sensing*, 87: 180–191. <http://www.sciencedirect.com/science/article/pii/S0924271613002220>.
- Blondel, P. 2009. *The Handbook of Sidescan Sonar*. Springer, Berlin, Heidelberg, New York. 316 pp.
- Blondel, P., and Gomez Sichi, O. 2009. Textural analyses of multi-beam sonar imagery from Stanton Banks, Northern Ireland continental shelf. *Applied Acoustics*, 70: 1288–1297.
- Blondel, P., and Murton, J. B. 1997. *Handbook of Seafloor Sonar Imagery*. John Wiley and Sons, Chichester. 314 pp.
- Breiman, L. 1996. Bagging predictors. *Machine Learning*, 24: 123–140.
- Breiman, L. 2001. Random Forests. *Machine Learning*, 45: 5–32.
- Brown, C. J., and Collier, J. S. 2008. Mapping benthic habitat in regions of gradational substrata: An automated approach utilising geophysical, geological, and biological relationships. *Estuarine Coastal and Shelf Science*, 78: 203–214.
- Brown, C. J., Sameoto, J. A., and Smith, S. J. 2012. Multiple methods, maps, and management applications: Purpose made seafloor maps in support of ocean management. *Journal of Sea Research*, 72: 1–13.
- Brown, C. J., Smith, S. J., Lawton, P., and Anderson, J. T. 2011. Benthic habitat mapping: a review of progress towards improved understanding of the spatial ecology of the seafloor using acoustic techniques. *Estuarine Coastal and Shelf Science*, 92: 502–520.
- Buhl-Mortensen, P., Dolan, M., and Buhl-Mortensen, L. 2009. Prediction of benthic biotopes on a Norwegian offshore bank using a combination of multivariate analysis and GIS classification. *ICES Journal of Marine Science*, 66: 2026–2032.
- Burnett, C., and Blaschke, T. 2003. A multi-scale segmentation/object relationship modelling methodology for landscape analysis. *Ecological Modelling*, 168: 233–249. <http://www.sciencedirect.com/science/article/pii/S030438000300139X>.
- Calvert, J., Strong, J. A., Service, M., McGonigle, C., and Quinn, R. 2015. An evaluation of supervised and unsupervised classification techniques for marine benthic habitat mapping using multibeam echosounder data. *ICES Journal of Marine Science: Journal Du Conseil*, 72: 1498–1513. <http://icesjms.oxfordjournals.org/content/72/5/1498.abstract>.
- Carlotto, M. J. 2009. Effect of errors in ground truth on classification accuracy. *International Journal of Remote Sensing*, 30: 4831–4849.
- Chavez, P. S., Isbrecht, J., Galanis, P., Gabel, G. L., Sides, S. C., Soltesz, D. L., Ross, S. L., *et al.* 2002. Processing, mosaicking and management of the Monterey Bay digital sidescan-sonar images. *Marine Geology*, 181: 305–315. <http://www.sciencedirect.com/science/article/pii/S0025322701002730> (Accessed 30 September 2015).
- Che Hasan, R., Ierodiaconou, D., and Laurenson, L. 2012a. Combining angular response classification and backscatter imagery segmentation for benthic biological habitat mapping. *Estuarine, Coastal and Shelf Science*, 97: 1–9.
- Che Hasan, R., Ierodiaconou, D., Laurenson, L., and Schimel, A. 2014. Integrating multibeam backscatter angular response, mosaic and bathymetry data for benthic habitat mapping. *Plos One*, 9: e97339.
- Che Hasan, R., Ierodiaconou, D., and Monk, J. 2012b. Evaluation of four supervised learning methods for benthic habitat mapping using backscatter from multi-beam sonar. *Remote Sensing*, 4: 3427–3443. <http://www.mdpi.com/2072-4292/4/11/3427>.
- Cihlar, J. 2000. Land cover mapping of large areas from satellites: status and research priorities. *International Journal of Remote Sensing*, 21: 1093–1114. <http://www.tandfonline.com/doi/abs/10.1080/014311600210092>.
- Clements, A. J., Strong, J. A., Flanagan, C., and Service, M. 2010. Objective stratification and sampling-effort allocation of ground-truthing in benthic-mapping surveys. *ICES Journal of Marine Science*, 67: 628–637. <http://icesjms.oxfordjournals.org/content/67/4/628.abstract>.
- Cohen, J. 1960. A coefficient of agreement of nominal scales. *Educational and Psychological Measurement*, 20: 37–46.
- Comber, A., Fisher, P., Brunsdon, C., and Khmag, A. 2012. Spatial analysis of remote sensing image classification accuracy. *Remote Sensing of Environment*, 127: 237–246.
- Congalton, R. G. 1991. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, 37: 35–46. <http://www.sciencedirect.com/science/article/pii/S003442579190048B>.
- Copeland, A., Edinger, E., Devillers, R., Bell, T., LeBlanc, P., and Wroblewski, J. 2013. Marine habitat mapping in support of Marine Protected Area management in a subarctic fjord: Gilbert Bay, Labrador, Canada. *Journal of Coastal Conservation*, 17: 225–237.
- Costa, B. M., and Battista, T. A. 2013. The semi-automated classification of acoustic imagery for characterizing coral reef ecosystems. *International Journal of Remote Sensing*, 34: 6389–6422.
- Cramer, H. 1946. *Mathematical Methods of Statistics*. Princeton University Press, Princeton. 282 pp.
- Daily, M. 1983. Hue-saturation-intensity split-spectrum processing of seasat radar imagery. *Photogrammetric Engineering and Remote Sensing*, 49: 349–355.
- Dartnell, P., and Gardner, J. V. 2004. Predicting seafloor facies from multibeam bathymetry and backscatter data. *Photogrammetric Engineering and Remote Sensing*, 70: 1081–1091.
- Definiens, I. 2004. *eCognition Professional 4 User Guide*. Definiens Imaging GmbH, München. 485 pp.
- DeFries, R., and Chan, J. C. W. 2000. Multiple criteria for evaluating machine learning algorithms for land cover classification from satellite data. *Remote Sensing of Environment*, 74: 503–515.

- <http://www.sciencedirect.com/science/article/pii/S0034425700001425> (accessed 13 August 2015).
- Demarchi, L., Canters, F., Cariou, C., Licciardi, G., and Chan, J. C. W. 2014. Assessing the performance of two unsupervised dimensionality reduction techniques on hyperspectral APEX data for high resolution urban land-cover mapping. *ISPRS Journal of Photogrammetry and Remote Sensing*, 87: 166–179. <http://www.sciencedirect.com/science/article/pii/S0924271613002347> (accessed 7 December 2015).
- Di Gregorio, A., and Jansen, L. J. M. 1998. Land Cover Classification System (LCCS): Classification Concepts and User Manual. 157 pp.
- Diesing, M., Green, S. L., Stephens, D., Lark, R. M., Stewart, H. A., and Dove, D. 2014. Mapping seabed sediments: comparison of manual, geostatistical, object-based image analysis and machine learning approaches. *Continental Shelf Research*, 84: 107–119. <http://www.sciencedirect.com/science/article/pii/S0278434314001629>.
- Diesing, M., and Stephens, D. 2015. A multi-model ensemble approach to seabed mapping. *Journal of Sea Research*, 100: 62–69. <http://www.sciencedirect.com/science/article/pii/S1385110114001865>.
- Elvenes, S., Dolan, M. F. J., Buhl-Mortensen, P., and Bellec, V. K. 2014. An evaluation of compiled single-beam bathymetry data as a basis for regional sediment and biotope mapping. *ICES Journal of Marine Science*, 71: 867–881.
- Fisher, P., Wood, J., and Cheng, T. 2004. Where is Helvellyn? Fuzziness of multi-scale landscape morphometry. *Transactions of the Institute of British Geographers*, 29: 106–128.
- Fonseca, L., Brown, C., Calder, B., Mayer, L., and Rzhonov, Y. 2009. Angular range analysis of acoustic themes from Stanton Banks Ireland: a link between visual interpretation and multibeam echosounder angular signatures. *Applied Acoustics*, 70: 1298–1304. <http://www.sciencedirect.com/science/article/pii/S003682X08002028>.
- Foody, G. M. 2002. Status of land cover classification accuracy assessment. *Remote Sensing of Environment*, 80: 185–201.
- Foody, G. M. 2004. Thematic map comparison: evaluating the statistical significance of differences. *Classification Accuracy. Photogrammetric Engineering and Remote Sensing*, 70: 627–633.
- Foody, G. M. 2005. Local characterization of thematic classification accuracy through spatially constrained confusion matrices. *International Journal of Remote Sensing*, 26: 1217–1228.
- Foody, G. M. 2008. Harshness in image classification accuracy assessment. *International Journal of Remote Sensing*, 29: 3137–3158.
- Franklin, S. E., and Wulder, M. A. 2002. Remote sensing methods in medium spatial resolution satellite data land cover classification of large areas. *Progress in Physical Geography*, 26: 173–205.
- Fraschetti, S., Terlizzi, A., and Boero, F. 2008. How many habitats are there in the sea (and where)? *Journal of Experimental Marine Biology and Ecology*, 366: 109–115.
- Friedman, A. 2013. Automated Interpretation of Benthic Stereo Imagery. University of Sydney. 236 pp.
- Galparsoro, I., Agrafojo, X., Roche, M., and Degrendele, K. 2015. Comparison of supervised and unsupervised automatic classification methods for sediment types mapping using multibeam echosounder and grab sampling. *Italian Journal of Geosciences*, 134: 41–49.
- Guyon, I., and Elisseeff, A. 2003. An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3: 1157–1182.
- Haralick, R. M., Shanmugam, K., and Dinstein, I. 1973. Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics, SMC*, 3: 610–621.
- Harris, P. M., and Ventura, S. J. 1995. The integration of geographic data with remotely sensed imagery to improve classification in an urban area. *Photogrammetric Engineering and Remote Sensing*, 61: 993–998.
- Harris, P. T., and Baker, E. K. 2012. Why Map Benthic Habitats? *In* Seafloor Geomorphology as Benthic Habitat, pp. 3–22. Ed. by P. T. Harris and E. K. Baker. Elsevier. <http://www.sciencedirect.com/science/article/pii/B9780123851406000013> (accessed 23 December 2015).
- Hay, G. J., and Castilla, G. 2006. Object-based image analysis: strengths, weaknesses, opportunities and threats (SWOT). http://www.isprs.org/proceedings/XXXVI/4-C42/Papers/01_OpeningSession/OBIA2006_Hay_Castilla.pdf.
- Hay, G. J., and Castilla, G. 2008. Geographic Object-Based Image Analysis (GEOBIA): A new name for a new discipline. Springer, Berlin, Heidelberg.
- Hill, N. A., Lucieer, V., Barrett, N. S., Anderson, T. J., and Williams, S. B. 2014. Filling the gaps: Predicting the distribution of temperate reef biota using high resolution biological and acoustic data. *Estuarine, Coastal and Shelf Science*, 147: 137–147. <http://www.sciencedirect.com/science/article/pii/S0272771414001474>.
- Holmes, K. W., Van Niel, K. P., Radford, B., Kendrick, G. A., and Grove, S. L. 2008. Modelling distribution of marine benthos from hydroacoustics and underwater video. *Continental Shelf Research*, 28: 1800–1810.
- Huang, Z., Nichol, S. L., Siwabessy, J. P. W., Daniell, J., and Brooke, B. P. 2012. Predictive modelling of seabed sediment parameters using multibeam acoustic data: a case study on the Carnarvon Shelf, Western Australia. *International Journal of Geographical Information Science*, 26: 283–307.
- Huang, Z., Siwabessy, J., Nichol, S. L., and Brooke, B. P. 2014. Predictive mapping of seabed substrata using high-resolution multibeam sonar data: A case study from a shelf with complex geomorphology. *Marine Geology*, 357: 37–52. <http://www.sciencedirect.com/science/article/pii/S0025322714002205>.
- Hughes, G. 1968. On the mean accuracy of statistical pattern recognizers. *Information Theory, IEEE Transactions on*, 14: 55–63.
- Hughes, K. M., Kaiser, M. J., Jennings, S., McConnaughey, R. A., Pitcher, R., Hilborn, R., Amoroso, R. O., *et al.* 2014. Investigating the effects of mobile bottom fishing on benthic biota: a systematic review protocol. *Environmental Evidence*, 3: 23. <http://www.environmentalevidencejournal.org/content/3/1/23> (accessed 29 September 2015).
- Hughes-Clarke, J. E. 2015. Multispectral acoustic backscatter from multibeam, improved classification potential. *In* United States Hydrographic Conference 2015, pp. 1–18. National Harbor, MD, USA.
- Ierodiaconou, D., Monk, J., Rattray, A., Laurenson, L., and Versace, V. L. 2011. Comparison of automated classification techniques for predicting benthic biological communities using hydroacoustics and video observations. *Continental Shelf Research*, 31: S28–S38.
- Janssen, L. L. F., and van der Wel, F. J. M. 1994. Accuracy assessment of satellite derived land-cover data: a review. *Photogrammetric Engineering and Remote Sensing*, 60: 419–426.
- Kim, H., and Loh, W. Y. 2001. Classification trees with unbiased multiway splits. *Journal of the American Statistical Association*, 96: 589–604.
- Knudby, A., Roelfsema, C., Lyons, M., Phinn, S., and Jupiter, S. 2011. Mapping fish community variables by integrating field and satellite data, object-based image analysis and modeling in a traditional fijian fisheries management area. *Remote Sensing*, 3: 460–483. <http://www.mdpi.com/2072-4292/3/3/460>.
- Kohavi, R., and John, G. H. 1997. Wrappers for feature subset selection. *Artificial Intelligence*, 97: 273–324. <http://www.sciencedirect.com/science/article/pii/S000437029700043X>.
- Kostylev, V. E., and Hannah, C. G. 2007. Process-driven characterization and mapping of seabed habitats. *In* Mapping the Seafloor for Habitat Characterization, pp. 171–184. Ed. by B. J. Todd and H. G. Greene. Geological Association of Canada Special Paper 47.

- Kostylev, V. E., Todd, B. J., Fader, G. B. J., Courtney, R. C., Cameron, G. D. M., and Pickrill, R. A. 2001. Benthic habitat mapping on the Scotian Shelf based on multibeam bathymetry, surficial geology and sea floor photographs. *Marine Ecology Progress Series*, 219: 121–137.
- Kursa, M., and Rudnicki, W. 2010. Feature selection with the Boruta Package. *Journal of Statistical Software*, 36: <http://www.jstatsoft.org/v36/i11/paper/>.
- Kyriakidis, P. C., and Dungan, J. L. 2001. A geostatistical approach for mapping thematic classification accuracy and evaluating the impact of inaccurate spatial data on ecological model predictions. *Environmental and Ecological Statistics*, 8: 311–330.
- Lacharité, M., Metaxas, A., and Lawton, P. 2015. Using object-based image analysis to determine seafloor fine-scale features and complexity. *Limnology and Oceanography: Methods*, 13: 553–567. <http://doi.wiley.com/10.1002/lom3.10047> (accessed 17 December 2015).
- Lathrop, R. G., Cole, M., Senyk, N., and Butman, B. 2006. Seafloor habitat mapping of the New York Bight incorporating sidescan sonar data. *Estuarine Coastal and Shelf Science*, 68: 221–230.
- Lecours, V., Devillers, R., DC, S., VL, L., and CJ, B. 2015. Spatial scale and geographic context in benthic habitat mapping: review and future directions. *Marine Ecology Progress Series*, 535: 259–284. <http://www.int-res.com/abstracts/meps/v535/p259-284/>.
- Liu, C., Frazier, P., and Kumar, L. 2007. Comparative assessment of the measures of thematic classification accuracy. *Remote Sensing of Environment*, 107: 606–616.
- Loh, W. Y., and Shih, Y. S. 1997. Split selection methods for classification trees. *Statistica Sinica*, 7: 815–840.
- Lu, D., and Weng, Q. 2007. A survey of image classification methods and techniques for improving classification performance. *International Journal of Remote Sensing*, 28: 823–870.
- Lucas, R., Medcalf, K., Brown, A., Bunting, P., Breyer, J., Clewley, D., Keyworth, S., *et al.* 2011. Updating the Phase 1 habitat map of Wales, UK, using satellite sensor data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66: 81–102. <http://www.sciencedirect.com/science/article/pii/S0924271610000870>.
- Lucieer, V., Hill, N. A., Barrett, N. S., and Nichol, S. 2013. Do marine substrates ‘look’ and ‘sound’ the same? Supervised classification of multibeam acoustic data using autonomous underwater vehicle images. *Estuarine, Coastal and Shelf Science*, 117: 94–106. <http://www.sciencedirect.com/science/article/pii/S0272771412004246>.
- Lucieer, V. L. 2008. Object-oriented classification of sidescan sonar data for mapping benthic marine habitats. *International Journal of Remote Sensing*, 29: 905–921. <http://www.informaworld.com/10.1080/01431160701311309>.
- Lucieer, V., and Lamarche, G. 2011. Unsupervised fuzzy classification and object-based image analysis of multibeam data to map deep water substrates, Cook Strait, New Zealand. *Continental Shelf Research*, 31: 1236–1247.
- Lyons, M. B., Phinn, S. R., and Roelfsema, C. M. 2012. Long term land cover and seagrass mapping using Landsat and object-based image analysis from 1972 to 2010 in the coastal environment of South East Queensland, Australia. *ISPRS Journal of Photogrammetry and Remote Sensing*, 71: 34–46. International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS).
- MacMillan, R. A., and Shary, P. A. 2009. Landforms and landform elements in geomorphometry. In *Geomorphometry: Concepts, Software, Applications*. (Developments in Soil Science 33), pp. 227–254. Ed. by T. Hengl and H. I. Reuter. Elsevier Science, Amsterdam.
- Marsh, I., and Brown, C. 2009. Neural network classification of multibeam backscatter and bathymetry data from Stanton Bank (Area IV). *Applied Acoustics*, 70: 1269–1276.
- McGonigle, C., Brown, C., Quinn, R., and Grabowski, J. 2009. Evaluation of image-based multibeam sonar backscatter classification for benthic habitat discrimination and mapping at Stanton Banks, UK. *Estuarine Coastal and Shelf Science*, 81: 423–437.
- Milligan, G. W., and Cooper, M. C. 1985. An examination of procedures for determining the number of clusters in a data set. *Psychometrika*, 50: 159–179.
- Moran, P. 1950. Notes on continuous stochastic phenomena. *Biometrika*, 37: 17–23.
- Morton, D., Rowland, C., Wood, C., Meek, L., Marston, C., Smith, G., Wadsworth, R., *et al.* 2011. Final Report for LCM2007 - the new UK land cover map. Countryside Survey Technical Report No 11/07. NERC/Centre for Ecology & Hydrology. <http://www.countrysidesurvey.org.uk/outputs/land-cover-map-2007-final-report>.
- Murai, H., and Omatu, S. 1997. Remote sensing image analysis using a neural network and knowledge-based processing. *International Journal of Remote Sensing*, 18: 811–828.
- Neves, B. M., Du Preez, C., and Edinger, E. 2014. Mapping coral and sponge habitats on a shelf-depth environment using multibeam sonar and ROV video observations: Learmonth Bank, northern British Columbia, Canada. Deep Sea Research Part II: Topical Studies in Oceanography, 99: 169–183. <http://www.sciencedirect.com/science/article/pii/S0967064513002178> (accessed 9 October 2015).
- Ojeda, G. Y., Gayes, P. T., Van Dolah, R. F., and Schwab, W. C. 2004. Spatially quantitative seafloor habitat mapping: example from the northern South Carolina inner continental shelf. *Estuarine Coastal and Shelf Science*, 59: 399–416.
- Olenin, S., and Ducrottoy, J. P. 2006. The concept of biotope in marine ecology and coastal management. *Marine Pollution Bulletin*, 53: 20–29. <http://www.sciencedirect.com/science/article/B6V6N-4JMW58F-3/2/3a2e1083036418216f306f53e603693a>.
- Olofsson, P., Foody, G. M., Stehman, S. V., and Woodcock, C. E. 2013. Making better use of accuracy data in land change studies: estimating accuracy and area and quantifying uncertainty using stratified estimation. *Remote Sensing of Environment*, 129: 122–131.
- Pal, N. R., and Pal, S. K. 1993. A review on image segmentation techniques. *Pattern Recognition*, 26: 1277–1294. <http://www.sciencedirect.com/science/article/pii/003132039390135J>.
- Phinn, S. R., Roelfsema, C. M., and Mumby, P. J. 2012. Multi-scale, object-based image analysis for mapping geomorphic and ecological zones on coral reefs. *International Journal of Remote Sensing*, 33: 3768–3797.
- Plets, R., Clements, A., Quinn, R., Strong, J., and Edwards, H. 2012. Marine substratum map of the Causeway Coast, Northern Ireland. *Journal of Maps*, 8: 1–13.
- Pontius, R. G., and Millones, M. 2011. Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. *International Journal of Remote Sensing*, 32: 4407–4429.
- Preston, J. 2009. Automated acoustic seabed classification of multibeam images of Stanton Banks. *Applied Acoustics*, 70: 1277–1287. <http://www.sciencedirect.com/science/article/pii/S0003682X08001801>.
- Purkis, S., and Klemas, V. 2011. *Remote Sensing and Global Environmental Change*. Wiley-Blackwell, Chichester. 367 pp.
- Rattray, A. J., Ierodiaconou, D., Laurenson, L., Burq, S., and Reston, M. 2009. Hydro-acoustic remote sensing of benthic biological communities on the shallow South East Australian continental shelf. *Estuarine, Coastal and Shelf Science*, 84: 237–245. Elsevier Ltd.
- Rattray, A. J., Ierodiaconou, D., Monk, J., Laurenson, L., and Kennedy, P. 2014. Quantification of spatial and thematic uncertainty in the application of underwater video for benthic habitat mapping. *Marine Geodesy*, 37: 315–336.

- Rattray, A. J., Ierodiaconou, D., and Womersley, T. 2015. Wave exposure as a predictor of benthic habitat distribution on high energy temperate reefs. *Frontiers in Marine Science*, 2: 1–14.
- Rioja-Nieto, R., Barrera-Falcón, E., Hinojosa-Arango, G., and Riosmena-Rodríguez, R. 2013. Benthic habitat β -diversity modeling and landscape metrics for the selection of priority conservation areas using a systematic approach: Magdalena Bay, Mexico, as a case study. *Ocean and Coastal Management*, 82: 95–103.
- Roberts, J. J., Best, B. D., Dunn, D. C., Treml, E. A., and Halpin, P. N. 2010. Marine Geospatial Ecology Tools: an integrated framework for ecological geoprocessing with ArcGIS, Python, R, MATLAB, and C++. *Environmental Modelling and Software*, 25: 1197–1207. <http://www.sciencedirect.com/science/article/pii/S1364815210000885> (accessed 24 July 2015).
- Roelfsema, C., Phinn, S., Jupiter, S., Comley, J., and Albert, S. 2013. Mapping coral reefs at reef to reef-system scales, 10s–1000s km², using object-based image analysis. *International Journal of Remote Sensing*, 34: 6367–6388.
- Rooper, C. N., and Zimmermann, M. 2007. A bottom-up methodology for integrating underwater video and acoustic mapping for seafloor substrate classification. *Continental Shelf Research*, 27: 947–957.
- Saatchi, S. S., Houghton, R. A., Dos Santos Alvalá, R. C., Soares, J. V., and Yu, Y. 2007. Distribution of aboveground live biomass in the Amazon basin. *Global Change Biology*, 13: 816–837.
- Sandwell, D. T., Müller, R. D., Smith, W. H. F., Garcia, E., and Francis, R. 2014. New global marine gravity model from CryoSat-2 and Jason-1 reveals buried tectonic structure. *Science*, 346: 65–67. <http://www.sciencemag.org/content/346/6205/65.abstract>.
- Sappington, J. M., Longshore, K. M., and Thompson, D. B. 2007. Quantifying landscape ruggedness for animal habitat analysis: a case study using bighorn sheep in the Mojave Desert. *Journal of Wildlife Management*, 71: 1419–1426.
- Savini, A., Vertino, A., Marchese, F., Beuck, L., and Freiwald, A. 2014. Mapping cold-water coral habitats at different scales within the Northern Ionian Sea (central Mediterranean): an assessment of coral coverage and associated vulnerability. *PLoS One*, 9.
- Schiewe, J. 2002. Segmentation of high-resolution remotely sensed data - concepts, applications and problems. Ottawa. http://www.ecognition.com/sites/default/files/383_358.pdf.
- Simons, D. G., and Snellen, M. 2009. A Bayesian approach to seafloor classification using multi-beam echo-sounder backscatter data. *Applied Acoustics*, 70: 1258–1268.
- Smith, W. H. F., and Sandwell, D. T. 1997. Global sea floor topography from satellite altimetry and ship depth soundings. *Science*, 277: 1956–1962.
- Steele, B. M., Chris Winne, J., and Redmond, R. L. 1998. Estimation and mapping of misclassification probabilities for thematic land cover maps. *Remote Sensing of Environment*, 66: 192–202.
- Steele, B. M., Patterson, D. A., and Redmond, R. L. 2003. Toward estimation of map accuracy without a probability test sample. *Environmental and Ecological Statistics*, 10: 333–356.
- Stehman, S. V. 1997. Selecting and interpreting measures of thematic classification accuracy. *Remote Sensing of Environment*, 62: 77–89. <http://www.sciencedirect.com/science/article/pii/S0034425797000837>.
- Stehman, S. V., and Czaplewski, R. L. 1998. Design and Analysis for Thematic Map Accuracy Assessment - an application. *Of Satellite Imagery. Remote Sensing of Environment*, 64: 331–344.
- Stephens, D., and Diesing, M. 2014. A comparison of supervised classification methods for the prediction of substrate type using multibeam acoustic and legacy grain-size data. *Plos One*, 9: e93950.
- Strahler, A. H., Boschetti, L., Foody, G. M., Friedl, M. A., Hansen, M. C., Herold, M., Mayaux, P., *et al.* 2006. Global land cover validation: recommendations for evaluation and accuracy assessment of global land cover maps. Luxembourg, 51. pp.
- Tewkesbury, A. P., Comber, A. J., Tate, N. J., Lamb, A., and Fisher, P. F. 2015. A critical synthesis of remotely sensed optical image change detection techniques. *Remote Sensing of Environment*, 160: 1–14. <http://www.sciencedirect.com/science/article/pii/S0034425715000152> (accessed 24 February 2015).
- Therneau, T., and Atkinson, E. 1997. An Introduction to Recursive Partitioning Using the rpart Routine. Section of Biostatistics, Mayo Clinic, Rochester. 52 pp. <http://www.mayo.edu/hsr/techrpt/61.pdf>.
- Verbyla, D. L., and Hammond, T. O. 1995. Conservative bias in classification accuracy assessment due to pixel-by-pixel comparison of classified images with reference grids. *International Journal of Remote Sensing*, 16: 581–587.
- Wilken, D., Feldens, P., Wunderlich, T., and Heinrich, C. 2012. Application of 2D Fourier filtering for elimination of stripe noise in side-scan sonar mosaics. *Geo-Marine Letters*, 32: 337–347.
- Wilson, M. F. J., O'Connell, B., Brown, C., Guinan, J. C., and Grehan, A. J. 2007. Multiscale terrain analysis of multibeam bathymetry data for habitat mapping on the continental slope. *Marine Geodesy*, 30: 3–35.
- Wright, D. J., and Heyman, W. D. 2008. Introduction to the special issue: marine and coastal gis for geomorphology, habitat mapping, and marine reserves. *Marine Geodesy*, 31: 223–230.
- Zavalas, R., Ierodiaconou, D., Ryan, D., Rattray, A. J., and Monk, J. 2014. Habitat classification of temperate marine macroalgal communities using bathymetric LiDAR. *Remote Sensing*, 6: 2154–2175.
- Zhang, C., Selch, D., Xie, Z., Roberts, C., Cooper, H., and Chen, G. 2013. Object-based benthic habitat mapping in the Florida Keys from hyperspectral imagery. *Estuarine, Coastal and Shelf Science*, 134: 88–97. <http://www.sciencedirect.com/science/article/pii/S0272771413004253>.

Handling editor: Steven Degraer