Impacts of ignorance on the accuracy of image classification and thematic mapping

Giles M. Foody
School of Geography, University of Nottingham, Nottingham, NG7 2RD, UK

Abstract

Thematic maps are often derived from remotely sensed imagery via a supervised image classification analysis. The training and testing stages of a supervised image classification may proceed ignorant of the presence of some classes in the region to be mapped. This violates the assumption of an exhaustively defined set of classes that is often made in classification analyses. In such circumstances, the overall accuracy of a thematic map produced by the application of a trained classifier will be less than the accuracy of the classification of the test set by the same classifier. This situation arises because the cases of an untrained class can normally only be commissioned into the set of trained classes. Simple mathematical relationships between classification and map accuracy are shown for assessments of overall, user's and producer's accuracy. For example, it is shown that in a simple scenario the accuracy of a thematic map is less than that of a classification, scaling as a function of the abundance of the untrained class(es). Impacts on other estimates made from thematic maps, such as class areal extent, are also briefly discussed. When using a thematic map, care is needed in interpreting and using classification accuracy assessments as sometimes they may not reflect properties of the map well.

1. Introduction

The widespread availability of remotely sensed data in space and time together with their synoptic viewpoint and information content make them an attractive source of data for mapping applications. The maps generated from remote sensing are, and indeed should be expected to be, imperfect. All maps provide a generalization of reality and hence deviate in some way from it. The nature and magnitude of the deviation from reality will vary as a function of many issues such as the features to be mapped and the data and methods used. The basic issues and challenges for mapping are well known. For example, to address fundamental issues connected to cartographic scale for a basic pocket-map the “grandest idea of all” (Carroll, 1893; p 169) would be to map at a 1:1 scale which would be impractical as well as of questionable value. Maps are thus imperfect and partly as a result of this it has been be claimed that they are “the most used and least understood documents of modern civilisation” (L A Brown, 1953, cited in Maling, 1989; p 144). Awareness of the limitations of maps may, however, enhance map interpretation and use. Here, the focus is on thematic maps such as those depicting land cover obtained by popular supervised image classification analyses.

Remote sensing is widely used to generate thematic maps via a supervised image classification analysis. The basic process is very simple and the entire mapping process applied to appropriately pre-processed imagery comprises three stages: training, allocation and testing. In training, pixels of each class of interest are identified in the image to be classified and characterised quantitatively. The latter characterisations are training statistics that are used in the second stage to allocate every pixel in the imagery of the region of interest to be mapped to a class on the basis of their relative similarity to the class characterisations. The accuracy of the allocations made is assessed in the testing stage by comparison of predicted and actual class labels for a sample of pixels drawn from the region of interest that was mapped. There are, of course, many detailed considerations in each stage and variants of this process exist. The training statistics could, for example, arise from spectral libraries, training sites could be from outside the region to be mapped and objects rather than pixels may be used as the fundamental spatial unit but the general nature of the classification analysis remains the same. More critically to this article, there are fundamental assumptions made in a classification analysis.

A range of assumptions are made in a classification analysis. For example, in a conventional ‘hard’ classification a key assumption made about the data is that the pixels are pure (i.e. each represents an area covered by a single class). Unfortunately, pixels are arbitrary spatial units determined mainly by the sensor's properties and can have little relation to natural units on the ground. The major problem arising from deviation from the assumed condition is the presence of mixed pixels; the problem does not disappear in an object-based approach as mixed objects can be common (Costa *et al.*, 2017). The magnitude of the problem is a function of the relationship between the image spatial resolution and the landscape mosaic on the ground. Means to address this type of problem, perhaps via a soft classification analysis or super-resolution mapping exist (Foody, 2004a), and may need to be used for accurate mapping.

Assumptions are also made about the classes in a classification analysis. For example, it is normally assumed that the classes are discrete and mutually exclusive as well as exhaustively defined. Often this is not the case and deviation from the assumed condition can be a source of error and uncertainty in thematic mapping. For example, classes are often not discrete and mutually exclusive. Many classes intergrade. Continuous classes such as found in many semi-natural environments cannot be represented well in a standard 'hard' classification (Foody, 2004a). While a continuum can be divided into a set of classes this is a poor characterisation of reality and neighbouring classes along the continuum may share qualities. The boundaries between these and other classes are not natural but *fiat* and dependent on human decisions (Smith and Mark, 1998: Vogt *et al*., 2012). Many classes may, therefore, be defined in a variety of ways and the process may be inherently political (Comber *et al*., 2005a; 2005b). Again the basic issues are well-known and that many classes are human constructs which can be a definitional challenge rather than natural features is evident in a quote attributed to Wittgenstein: "What is or is not a cow is for the public to decide" (Toulmin, 1953; p51). Care must, therefore, be taken to define classes appropriately and in many studies it is necessary to harmonise legends if meaningful results are to be obtained. Critically, assumptions are made about the classes and deviation from the assumed condition can impact negatively on analyses and hence needs to be addressed. This article is focused on just one of the assumptions often made in supervised classification and how it impacts on the accuracy of class allocations made by a classifier: the classes have been defined exhaustively (i.e. every class that occurs has been included in the analysis). Of central concern to the article is the reference data set used in the testing stage. The latter are typically obtained from fieldwork or interpretation of fine resolution imagery and may be acquired following established guidelines to ensure value (Olofsson *et al*., 2014; Stehman and Foody, 2019). Critically, however, it is suggested that a distinction be made between reference data for the set of classes used in training the classification and reference data that represent all the classes contained within the region of interest that was mapped. In the context of this article, the former may be used to indicate the accuracy of the classification while the latter may be used to indicate the accuracy of the map generated by application of the trained classifier to the imagery of the region of interest to be mapped.

The value of a thematic map is influenced substantially by its quality. There is, therefore, considerable interest in the accuracy of thematic maps produced by a classification analysis. Indeed, an accuracy assessment is viewed as a fundamental component of a mapping programme (Strahler *et al.*, 2006). Many challenges are, however, encountered in an accuracy assessment (Congalton and Green, 2009; Ye *et al*., 2018; Stehman and Foody, 2019). The interpretation of an accuracy assessment may also not always be straightforward and can be complicated by a failure to satisfy underpinning assumptions.

Typically, interest is focused on the properties of the map generated by a classification although in some notable exceptions, such as classifier development, interest may lie in aspects of the classification such as the degree of inter-class separability present. The quality of the map and the ability to discriminate classes in the imagery are intimately related and can often be usefully expressed in terms of thematic accuracy. However, the accuracy with which the set of trained classes is classified by a classifier (referred to here as classification accuracy) may differ from the accuracy with which the entire set of classes present in the region of interest to be mapped is classified via the same classifier (referred to here as map accuracy). The classification exists within the map and hence classification and map accuracy are related but can be different. Thus, while the terms classification accuracy and map accuracy are often used synonymously it may be more appropriate for them to be thought of as relating to different, albeit related, properties. Differences between classification and map accuracy can arise for a variety of reasons. One key reason for differences between classification and map accuracy, and the focus of this article, is that the fundamental assumption that the set of classes has been defined exhaustively which underlies many supervised classification analyses for thematic mapping may be violated.

A wide range of approaches to classification and many different classifiers have been used in thematic mapping from remote sensing (Lu and Weng, 2007; Tso and Mather, 2009; Li *et al*., 2014; Ghamisi *et al*., 2017). In essence, these analyses seek to separate the classes in the feature space provided by the remotely sensed imagery. The concern in this paper is that the analysis may proceed ignorant of the existence in the region to be mapped of one or more classes beyond the set used to train the supervised classifier. If the feature space is partitioned fully by the classifier, cases of such untrained classes must be commissioned into the set of trained classes and hence degrade map accuracy relative to the accuracy of the classification of the set of trained classes. Not all classification analyses are sensitive to this problem. There are, for example, exceptions such as classifiers that partition feature space locally or have the capacity to detect and reject cases from an unknown class (Hudak, 1992; Foody 2004b; Gui et al., 2018). A basic boxcar or parallelepiped classifier, for example, may associate regions of feature space with classes leaving other parts unassociated with any class. A case to be classified that lies within the unassociated area of feature space would be left as ‘unclassified’ or labelled as something such as ‘other’. Similarly some classifiers allow a threshold to be set that allows a proportion of cases atypical of all classes to be left unclassified or labelled as ‘other’. Researchers may also sometimes be able to mask out regions containing classes of no interest to a specific study or, with a focus on a specific class of interest, reduce a study to a binary classification, the class of interest versus others, ensuing that an exhaustive set of classes is used. However, it is common for a classifier that fully partitions feature space to be used and such classifications only are considered in this paper.

This paper aims to explore some key impacts associated with the violation of the assumption of an exhaustively defined set of classes. Specifically, the focus is on the accuracy of a classification and the accuracy of a map for a region containing one or more untrained classes, both obtained from the application of the same trained classifier. In this scenario the mapping is undertaken ignorant of the existence of the untrained class(es). This scenario is common and indeed may be the predominant situation in typical mapping applications. The core aim is to show and explain the effect of such ignorance on classification and map accuracy. This will help address and explain a widely observed but rarely discussed situation in which a thematic map may be evidently less accurate than the (classification) accuracy statement that accompanies it suggests.

2. The problem of ignorance

Ignorance has many dimensions but here the focus is on unawareness. The key concern is on situations in which a supervised image classification analysis is used to produce a thematic map of a region of interest but undertaken in such a way that the analysis is unaware of the existence of one or more thematic classes in the region being mapped. The focus in this paper is entirely on classifiers that fully partition feature space and assume an exhaustively defined set of classes. Particular attention is directed to the relative magnitude of accuracy estimated for a classification and then for a map arising from the application of the same classifier to remotely sensed imagery. Although untrained classes impact on soft classifications (Foody, 2000) these and other issues related to rigorous accuracy assessment (Olofsson *et al*., 2014) are not considered further purely to facilitate a focus on the relative magnitude of classification and map accuracy.

In a supervised image classification, the analyst defines the set of classes to be included in a study. In most situations, this requires training data to be acquired for each and every class. These training data are used to generate training statistics that form spectral signatures which, essentially, characterise the appearance of the classes in the imagery. The latter may then be used to form a thematic map from the imagery via a classification analysis. In the classification, each image pixel (or other suitable spatial unit) in the region to be mapped is allocated to one of the defined set of classes on the basis of their relative spectral similarity. So, for example, a classical maximum likelihood classifier should be trained upon every class and each pixel in the region to be mapped would be allocated to the class with which it had the largest posterior probability of membership. Critically, each pixel can be allocated to only one of the set of defined classes upon which the classifier was trained. This type of approach can be a highly effective and accurate way to classify a remotely sensed image to produce a thematic map. Implicit in the analysis, however, is the assumption that the set of classes has been defined exhaustively (Lu and Weng, 2007).

If a class has been excluded from the training stage, there is no spectral signature for that class and the classifier cannot allocate pixels to that class. Cases of an untrained class can only be allocated to one of the classes that the classifier was trained on. Thus, commission into the set of trained classes can be expected when the set of classes has been defined non-exhaustively (Foody, 2001; 2002). The presence of these misclassifications impacts also on the assessment of the quality of the map that is obtained from the application of the trained classifier to imagery of the region of interest.

Violation of the assumption of an exhaustively defined set of classes must lead to misclassification, with cases of the untrained class(es) commissioned into the set of trained classes. However, if the testing set used in accuracy assessment includes only the set of trained classes such errors may not be observed even though they may exist in the region to be mapped. For example, if the training and testing data sets were acquired at the same time and contain only cases of the set of trained classes then the assessment of accuracy is focused upon only the accuracy with which the set of trained classes are classified. This measure of classification accuracy can be useful but could be misleading in relation to the quality of the thematic map that arises from the application of the classifier to imagery of a region of interest. It is, for example, a potentially poor and misleading assessment of the accuracy with which all classes that exist in the region of study are classified and so is an imperfect measure of map accuracy. If the region of interest contains untrained classes, cases of these classes must be commissioned into the trained set of classes and hence the overall accuracy of the map will be lower than that of the classification as it will contain more erroneous allocations. Map accuracy may, therefore, be incorrectly represented by classification accuracy which will, relative to map accuracy, be optimistically biased. Critically, classification accuracy may not be fully representative of map accuracy as it fails to include information for all classes that exist. It may be helpful, here and more generally, to distinguish between classification and map accuracy. Although the two expressions are often used synonymously, which may be appropriate when the class set has been defined exhaustively, classification accuracy is taken here to be the accuracy with which the set of defined classes used in training has been classified while map accuracy is the accuracy with which a region of interest, including areas of untrained class(es), is mapped.

Before exploring the issue further with a focus on the relationship between classification and map accuracy, it may first be useful to determine if this is a real problem worthy of concern rather than an unimportant detail that can be ignored. A relatively superficial assessment of the problem, which aimed simply to show its existence and potentially non-trivial magnitude, was gained through a search of literature using *Google Scholar* (15 July 2020). To facilitate focusing on studies in which the problem may arise, the focus was on a type of study in which the set of classes might conceivably have been defined non-exhaustively. The study scenario selected was for the mapping of crops. Specifically, a search for 'Landsat crop map classification accuracy' was undertaken. The aim was to find articles reporting results for the mapping of crops in a region of sufficient size to include a range of non-crop classes. It would be possible to imagine a study, for example, including all the crops that are grown in a study area and maybe some additional classes such as grasses and forests but ignoring, deliberately or accidentally, other classes that exist in the region of study such as urban areas and water. It is also perfectly possible for an analyst to have successfully defined all of the thematic classes that fall within the region of interest but still encounter cases of an untrained class. For example, transient features such as clouds or floods can obscure the ground surface and could, therefore, represent an untrained class within the region of study. A total of 64,400 outputs was returned from this search and the first 50 were examined.

In many cases, the articles identified in the search did not provided sufficient information for the purposes of this investigation. For example, the exact test site used was sometimes unclear. In some articles, an existing thematic data set was used to mask out classes beyond the set of interest (e.g. Li et al., 2015) and, if the mask is accurate, in such instances classification and map accuracy may be the same. However, there were two studies that were described well and appeared to offer the potential for the problem of a non-exhaustively defined set of classes to arise. Furthermore, as the region of interest to be mapped in each study was well defined, it was possible to locate it within a reference land cover map that had an exhaustive class set produced at a time close to that of the mapping reported in the papers. Thus, for each of the two papers, a reference land cover classification that had an exhaustive set of classes was available to compare against the results reported. The two papers of interest were Skakun *et al*. (2016) and Sonobe *et al.* (2017) and their mapping was compared against that in the 30m resolution global land cover map FROMGLC version 2. The latter is one of a series of Finer Resolution Observation and Monitoring of Global Land Cover (FROMGLC) maps produced and based on data for approximately the year 2015 (Li *et al*., 2020). This map provided a representation of the test sites at a broadly comparable time which should be at an appropriate scale, spatially and thematically, to relate to the maps reported in the selected papers. While the exact definition of the test sites and the time period between the map products are likely to be sources of error, the core aim here was simply to show that classes beyond the set defined in training existed and gain some indication of their abundance.

In their article, Skakun *et al.* (2016) show a land cover map and discuss its accuracy. Their interest was focused on 8 crop classes but they did include some other classes, notably water, grass and forest in the analysis. However, other classes, such as urban, may exist and this was assessed with the FROMGLC version 2 map. The latter showed that approximately 2.28% of the test site was impervious cover, an untrained class. This situation fits well with the discussion above, where an analysis may focus on the classes of interest plus other common classes in the region but still fail to include all classes. It is also one in which it would be feasible to imagine that reference data acquired by, for example, a randomised sample design might also fail to include any cases of an untrained class, as rare, and hence the analysis and interpretations could proceed ignorant of their presence. The other article, Sonobe *et al.* (2017), was focused on 6 crop classes and presents a land cover map of the specific region of interest. The latter, however, included a range of classes that were not present in the training of the classifications. In comparison to the FROMGLC map, it is evident that the set of defined classes occupied approximately 61.73% of the region of interest that is mapped. That is, approximately 38.27% of the study region was covered by cases of one or more untrained classes; the latter includes a relatively large urban area, Memuro, in the region mapped which appears to have been labelled mostly as belonging to the maize class. Critically, a large proportion of the region of interest mapped, over a third of its total area, was comprised of untrained classes the cases of which were commissioned into the set of trained classes. The accuracy of the thematic map produced would have to be substantially less than the high classification accuracies reported (up to 94.5% overall accuracy) as the impact of the cases of untrained classes would need to be addressed in the assessment of map accuracy.

The key issue of relevance to this article is that regions mapped may contain untrained classes. The cases of these untrained classes are commissioned into the set of trained classes and the magnitude of the potential problem can be sizeable. It must be stressed that the untrained classes may not have impacted negatively on key aspects of the studies discussed above and no criticism is suggested, indeed these papers deserve credit for providing sufficient details to allow the analysis to be undertaken. Critically, however, untrained classes do occur in mapping studies, their effect on map accuracy is sometimes overlooked and yet could be very substantial. Indeed the concern relates to the issue of applicability that is flagged as a grand challenge in image classification (Small, 2021). The applicability problem is in some ways similar to model overfitting with a classifier's applicability to a region beyond that it was developed in a function of geographical differences. Here, the key differences of concern is that the region a classifier is applied to may contain classes not present in that upon which the classifier was developed. A variety of actions can be taken to try and reduce the problem of untrained classes. Researchers could, for example, seek to identify and mask out pixels that are atypical of all trained classes and hence potentially members of an untrained class. Alternatively, a classifier that does not fully partition feature space or which allows a proportion of cases to be left ‘unclassified’ could be used. Perhaps more appropriately, effort could simply be invested in ensuring that an exhaustively defined set of classes is used. However, the problem can arise unknowingly and hence it may be useful to know the relationship between classification and map accuracy.

3. The relationship between classification and map accuracy

Taking the term classification accuracy to relate to the accuracy with which the set of trained classes is classified and map accuracy to be that of a thematic map that arises from the application of a classifier to imagery of the region of interest, which could include untrained classes, a simple exploration of the relationship between classification and map accuracy was undertaken. This exploration is based first on a basic discussion of the situation for a theoretical example and then a scenario based on real data is presented.

Table 1 shows a basic confusion matrix used in accuracy assessments. In the matrix illustrated, the actual class of membership obtained from a reference data set is shown in the columns of the matrix. The predicted class of membership obtained from a classification analysis is shown in the rows of the matrix. Purely for ease of presentation it will be assumed that the testing set used to form this matrix, and all other matrices presented, was acquired by simple random sampling. The cases that lie along the main diagonal of the matrix are those that have been correctly allocated (i.e. pixels that have been labelled as belonging to the class that they actually do belong to). Conversely, those cases that lie in off-diagonal elements of the matrix are misclassifications (i.e. pixels allocated to a class that is different to the actual class of membership). The overall accuracy of the classification, *O*, may be quantified using equation 1.

$$O = \frac{\sum_{i=1}^{t} n_{ii}}{n} \qquad (1)$$

where *n* indicates the total number of cases drawn from the set of *t* thematic classes on which the classifier had been trained.

On a per-class basis, accuracy may be calculated from the user's and producer's perspectives depending on the type of misclassification error of concern (Story and Congalton, 1986; Olofsson *et*

	Class 1	Class 2	Class 3	Class 4	Σ
Class 1	n_{11}	n_{12}	n_{13}	n_{14}	$n_{1\cdot}$
Class 2	n_{21}	n_{22}	n_{23}	n_{24}	$n_{2\cdot}$
Class 3	n_{31}	n_{32}	n_{33}	n_{34}	$n_{3\cdot}$
Class 4	n_{41}	n_{42}	n_{43}	n_{44}	$n_{4\cdot}$
Σ	$n_{\cdot 1}$	$n_{\cdot 2}$	$n_{\cdot 3}$	$n_{\cdot 4}$	n

Table 1. The confusion matrix used in accuracy assessment. The columns show the actual class as indicated in a reference data set and the rows the class predicted in the classification analysis. All other confusion matrices in this paper show the same layout.

al., 2014). If interest is focused on commission errors, the focus is on the rows of the confusion matrix and the user's accuracy for a class, *U*, may be calculated from:

$$U_i = \frac{n_{ii}}{n_{i\cdot}} \tag{2}$$

Alternatively, if interest is focused on omission errors, the focus is on the columns of the confusion matrix and the producer's accuracy, *P*, may be calculated from:

$$P_i = \frac{n_{ii}}{n_{\cdot i}} \tag{3}$$

Other measures of accuracy and additional properties, such as class areal extent, may be estimated from the confusion matrix and may be impacted by the presence of one or more untrained classes. For example, the area of a class is also often of interest and can be estimated with regard to the classified area and expressed as a percentage from:

$$A_{c,i} = 100\frac{n_{i\cdot}}{n} \tag{4}$$

where the subscript *c,i* indicates belonging to class *i* in the classification output. Alternatively, area could be estimated based on the reference data and expressed, as a percentage, using:

$$A_{r,i} = 100\frac{n_{\cdot i}}{n} \tag{5}$$

where the subscript *r,i* indicates belonging to class *i* in the reference data; this approach to area estimation is often recommended as good practice (Olofsson *et al*., 2014). As area estimation is common in remote sensing studies this issue will be briefly touched upon. Issues such as adjusting estimates for different sample designs and the fitting of confidence intervals are not considered here to aid focus on the impact of an untrained class on the relative magnitude of accuracy estimates.

The classifier can only allocate cases to the classes upon which it was trained and only the trained set of classes can be represented in the confusion matrix. In the context of this paper such a matrix may be used to estimate classification accuracy. Such a matrix would commonly arise if, for example, the reference data for training and accuracy assessment were acquired at the same time and divided into separate training and testing sets. If, for example, the reference data were collected following a stratified by the actual class design it is possible the analyst would have no knowledge of the existence of the untrained class(es). The reference data could feasibly be collected such that the required number of cases for each class were acquired one class at a time until data had been acquired for the full set of desired classes for use in training the classifier. The reference data set could then be divided into independent training and testing sets. The imagery might then be classified and the testing set used to evaluate the accuracy of the classification. Such an analysis could be useful but it is not a good indication of the accuracy of a map generated by the application of the same trained classifier to a region of interest that contains one or more untrained classes. Cases of an untrained class would act to degrade the accuracy of the thematic map.

Had the data for accuracy assessment been acquired following a stratified by mapped class design a different situation to that outlined above may be obtained. Critically, it would be expected that some of the sites selected off the map for use in the accuracy assessment would be found to be members of the untrained class; the number of sites representing an untrained class would be a function of its abundance in the region mapped and the size of the sample selected. Similarly, had the sample of cases for accuracy assessment been acquired following another approach, such as a simple random sample of pixels drawn from the map, it would be expected that some cases of the untrained class would be included and the analyst, therefore, becomes aware of its presence. It may be that the researcher decides to simply ignore such cases as they do not relate to the classes of interest and cannot be entered into the confusion matrix. Many challenges are encountered in an accuracy assessment and classes as well as sample cases are sometimes deliberately excluded. For example, in the validation of the IGBP DISCover land cover map, two classes (snow and ice, and water) were excluded because of difficulties in acquiring suitable reference data and if the set of interpreters labelling sampled cases could not agree a label for it case it was excluded from the assessment reducing the size of the sample used in the validation (Scepan, 1999). Such actions may at times be necessary and, while not ideal, can still support a useful accuracy assessment if the work is fully and transparently documented (Stehman and Foody, 2019). However, what should ideally happen is that the untrained class(es) should become apparent and action to include it (them) in the accuracy assessment made.

Cases of an untrained class cannot be inserted into the planned confusion matrix for classification accuracy assessment as this would represent only the set of trained classes. On becoming aware of an untrained class it, thus, becomes appropriate to add a column to the confusion matrix to represent it (Table 2). Although measures of accuracy can be derived for non-square confusion matrices (e.g., Finn, 1993) it would be simple to also add an empty row to the matrix for this class to allow standard accuracy assessments from a square confusion matrix; the row is empty because no cases can be allocated to the untrained class as the classifier was not trained upon it. Thus both the row and column for the untrained class become apparent and need to be included in the assessment of map accuracy; multiple rows and columns would be added if there was more than one untrained class. Critically, for the simple scenario under discussion, an expanded confusion matrix emerges. As some elements of the confusion matrix associated with the untrained class take on a value >0 while others are set at 0, the presence of an untrained class may impact on some aspects of accuracy assessment and other analyses based on the confusion matrix.

At this point it may be helpful to visualise the confusion matrix for classification and map accuracy assessment when an untrained class is present. In keeping with the discussion thus far, Table 2 shows key properties for the simple scenario under discussion when a classifier is trained on a set of 4 classes but applied to imagery to produce a map of a region that contains an additional class (class 5). Accuracy assessments are based on a simple random sample of cases drawn from the region of interest. Critically, cases selected for inclusion in the accuracy assessment may include members of the untrained class. As these cannot be inserted into the anticipated 4x4 confusion matrix and may not fit the core focus of a study, such cases may (wrongly) be ignored. Only the sub-set of the sample of cases that relate to the set of trained classes would normally be used to indicate classification accuracy. This simple scenario is used throughout the discussion in this paper. It is, for example, similar to the scenario used in the assessment of the accuracy of the IGBP DISCover map (Scepan, 1999): cases are selected at random from the map and those cases not confidently associated with a trained class are ignored reducing the sample size for the accuracy assessment. The accuracy of the map arising from the application of the trained classifier to the imagery for the region of interest, however, should account for the untrained class and be based upon the full sample of cases selected for accuracy assessment.

	Class 1	Class 2	Class 3	Class 4	Class 5	$^{C}\Sigma$	$^{M}\Sigma$
Class 1	n_{11}	n_{12}	n_{13}	n_{14}	n_{15} (≥0)	$n_{1\cdot}$	$m_{1\cdot}=(n_{1\cdot}+n_{15})$
Class 2	n_{21}	n_{22}	n_{23}	n_{24}	n_{25} (≥0)	$n_{2\cdot}$	$m_{2\cdot}=(n_{2\cdot}+n_{25})$
Class 3	n_{31}	n_{32}	n_{33}	n_{34}	n_{35} (≥0)	$n_{3\cdot}$	$m_{3\cdot}=(n_{3\cdot}+n_{35})$
Class 4	n_{41}	n_{42}	n_{43}	n_{44}	n_{45} (≥0)	$n_{4\cdot}$	$m_{4\cdot}=(n_{4\cdot}+n_{45})$
Class 5	n_{51} (=0)	n_{52} (=0)	n_{53} (=0)	n_{54} (=0)	n_{55} (=0)	$n_{5\cdot}$ (=0)	$m_{5\cdot}=(n_{5\cdot}+n_{55})=0$
$^{C}\Sigma$	$n_{\cdot 1}$	$n_{\cdot 2}$	$n_{\cdot 3}$	$n_{\cdot 4}$		n	
$^{M}\Sigma$	$m_{\cdot 1}=n_{\cdot 1}$	$m_{\cdot 2}=n_{\cdot 2}$	$m_{\cdot 3}=n_{\cdot 3}$	$m_{\cdot 4}=n_{\cdot 4}$	$m_{\cdot 5}$ (>0)		$m=n+m_{\cdot 5}$

Table 2. Basis of confusion matrices for classification and map accuracy assessment when an untrained class may occur. The superscript prefixes C and M indicate that the total value shown relates to the classification or map respectively. The example shown is for the situation in which a classifier has been trained on 4 classes but there is also one additional, untrained, class (class 5) that lies within the imagery of the region of interest to be converted into a thematic map via the trained classifier (see text for further discussion). If the analysis is ignorant of the presence of the untrained class the row and column for class 5 and associated marginal values, the elements highlighted in grey, are unobserved and play no role in the calculation of classification accuracy. Classification accuracy is based entirely in the information for only the trained classes (classes 1-4). The classifier's application to the imagery to produce a map for the region of interest should mean that cases of the untrained class become apparent. These cases cannot be inserted into the 4x4 confusion matrix used to assess the classification accuracy and necessitate adding an additional row and column (for class 5). Note that as cases cannot be allocated to the untrained class the new row for the untrained class is empty, all elements contain 0 cases. Note that as a result of adding the empty row the column totals for classes 1-4 remain constant. The column associated with the untrained class does, however, contain cases; the total number of cases of the untrained class encountered is $m_{\cdot 5}$. The cases of the untrained class can only be commissioned into the set of trained classes and thus the top four elements of the column for the untrained class may have a value >0 if the relevant class is confused with the untrained class. As a consequence of the commission errors associated with cases of the

untrained class, the row marginal values can be equal to or larger than those used in the calculation of classification accuracy.

If the analyst is unaware of the untrained class, the elements shaded grey in the confusion matrix representing the data for the untrained class (Table 2) would be absent and estimates of properties such as classification accuracy obtained from the confusion matrix ignorant of its presence. The untrained class does, however, exist and may be made apparent yielding the expanded, 5x5, confusion matrix that includes it. The former matrix (4x4) indicates only the accuracy with which the set of trained classes can be classified from the specific sample of data acquired (i.e. classification accuracy). This differs from the accuracy of the map if untrained classes are contained as they must be commissioned into the set of trained classes. To distinguish between the accuracy of the classification and that of the map obtained by its application to an image of the region of interest the superscript prefixes C and M will be used to indicate the focus on the classification and the map respectively. For the simple scenario in which a random sample of cases was used to form the testing set for accuracy assessment, the accuracy of the map of the study region obtained through the application of the trained classifier may be obtained from

$$^{\mathrm{M}}O = \frac{^{\mathrm{M}}\sum_{i=1}^{t} n_{ii}}{m} \tag{6}$$

where m is the number of cases sampled from the mapped region. Note that the summation can still be made over the t trained classes as n_{55}=0; alternatively the equation could be written to sum over all classes but the result would be the same. Note also that the sample size is larger than that used in the assessment of classification accuracy, n, by the inclusion of the cases of the untrained class, with, in the example under discussion, $m = n + m_{\cdot 5}$. In the assessment of classification accuracy the $m_{\cdot 5}$ cases of the untrained class may have been (wrongly) dropped from the study as not fitting the focus of the study on the trained classes or because of some other concern such as uncertainty in the labelling. It may be preferable to keep the cases of the untrained class and include them as an ‘other’ class which can be added to the confusion matrix. This would ensure that the set of classes is exhaustive.

Estimates of user’s accuracy for classes in the map may be obtained from:

$$^{\mathrm{M}}U_i = \frac{n_{ii}}{m_{i\cdot}} \tag{7}$$

Similarly, estimates of producer’s accuracy may be obtained from:

$$^{\mathrm{M}}P_i = \frac{n_{ii}}{m_{\cdot i}} \tag{8}$$

Finally, the areal extent of a class may be estimated using the marginal values $m_{\cdot i}$ or $m_{i\cdot}$ and m rather than $n_{\cdot i}$ or $n_{i\cdot}$ and n in equation 4 or 5 as appropriate to the study.

It should be evident that, for the very basic scenario outlined, simple relationships may exists between classification and map accuracy and of other properties estimated from the confusion matrix such as the areal of extent of classes based on classification or map matrices. These relationships can also be described mathematically, forming essentially simple laws that relate the interpretation from a classification and that from a map. Moreover, the explanation for these relationships lies in the effect of the cases of the untrained class(es) that violate the assumption of an exhaustively defined set of classes. As such a simple theory to explain the observed relationships can be offered. For the basic scenario being considered, four key relationships are especially apparent and relevant to typical remote sensing projects:

(i) Overall map accuracy ≤ overall classification accuracy. The overall accuracy of the classification is obtained from equation 1 (calculated over 4 classes). Specifically, in the example shown in Table 2,

$$ {}^{\mathrm{C}}O = \frac{{}^{\mathrm{C}}\sum_{i=1}^{t} n_{ii}}{n} \tag{9} $$

The only difference in the calculation of overall map accuracy relative to overall classification accuracy is the inclusion of the cases of the untrained class, class 5. In the calculation of overall map accuracy, the numerator in the equation remains unchanged from that in the assessment of classification accuracy as its calculation for map accuracy is based on the addition of no new correctly allocated cases since n_{55}=0; as class 5 has not been included in training no cases, correctly or incorrectly, can be allocated to it. The denominator of equation 1 for map accuracy assessment does, however, differ from that used in the calculation of classification accuracy as the inclusion of the cases of the untrained class(es) increases the total number of cases such that $m = n + m_{\cdot 5}$. Because of this situation,

$$ \left({}^{\mathrm{M}}O = \frac{\sum_{i=1}^{t} n_{ii}}{m} \right) \leq \left({}^{\mathrm{C}}O = \frac{\sum_{i=1}^{t} n_{ii}}{n} \right) \tag{10} $$

since $m > n$ and the relationship between map and classification accuracy for the scenario under discussion takes the form:

$$ {}^{\mathrm{M}}O = {}^{\mathrm{C}}O \frac{n}{m} \tag{11} $$

The magnitude of the difference between classification and map accuracy is thus dependent on the difference between n and m which is a function of the abundance of the untrained class(es). Critically, overall map accuracy scales with classification accuracy by a function of n/m. That is, map accuracy is the classification accuracy weighted by the proportion of the region of interest covered by the trained classes. This can be illustrated further with the confusion matrices based on a real data set shown below.

(ii) User's accuracy from a map ≤ user's accuracy from a classification. User's accuracy is a measure of commission error and as cases of untrained classes are commissioned into the set of trained classes it follows that the accuracy estimated for a map can only be decreased relative to that estimated from a classification if there is confusion between the specific class of interest and the untrained class. Thus, as the numerator of the equation to calculate user's accuracy remains constant whether considering map or classification accuracy but the denominator changes it follows that

$$\left({}^{\mathrm{M}}U_i = \frac{n_{ii}}{m_{i\cdot}}\right) \leq \left({}^{\mathrm{C}}U_i = \frac{n_{ii}}{n_{i\cdot}}\right) \tag{12}$$

because $m_{i\cdot} \geq n_{i\cdot}$ due to commission errors associated with the untrained class. Note that the scaling of map and classification accuracy is class specific and by $n_{i\cdot}/m_{i\cdot}$ for the scenario considered with

$${}^{\mathrm{M}}U = {}^{\mathrm{C}}U \frac{n_{i\cdot}}{m_{i\cdot}} \tag{13}$$

(iii) Producer's accuracy from a map = producer's accuracy from a classification. Note from the producer's perspective, both the numerator and denominator for the calculation of map accuracy remain unchanged from that used for classification accuracy after the inclusion of an untrained class; a pixel cannot be omitted from a trained class to be commissioned into the untrained class. As the row associated with the untrained class is, therefore, full of 0 values (Table 2), it follows that,

$$\left({}^{\mathrm{M}}P_i = \frac{n_{ii}}{m_{\cdot i}}\right) = \left({}^{\mathrm{C}}P_i = \frac{n_{ii}}{n_{\cdot i}}\right) \tag{14}$$

because $m_{\cdot i} = n_{\cdot i}$ (Table 2).

(iv) Area estimates from a map may differ from those obtained from a classification. The detail of the difference is in part dependent on the perspective adopted and whether area estimation is based on equation 4 or 5. For the scenario under consideration, if area is estimated with regard to the actual class, which is often the recommended approach (Olofsson *et al*., 2014), then the area estimated from a map < area estimates from a classification if an untrained class is present. This is because the marginal value associated with a trained class remains the same ($m_{\cdot i} = n_{\cdot i}$) but the total number of cases, used as the denominator, increases from *n* to *m*. Since the untrained class must occupy space within the region mapped it follows that the proportional cover of the other classes must decline and in a manner determined by the difference between *n* and *m*. If area was to be estimated from the allocated class labels the situation is different. Here, if the class of interest was confused with the untrained class and commissions cases of it, the total number of cases allocated to the class will increase ($m_{i\cdot} > n_{i\cdot}$).and hence its apparent area may increase. If, however, the class was not confused with the untrained class, the number of allocations remains constant ($m_{i\cdot} = n_{i\cdot}$) but, as the total number of cases rises from *n* to *m*, the area estimated from a map will decrease relative to that estimated from a classification. Hence, for area estimation based on the allocated class labels, it is possible for the area estimates to show a class specific response to the untrained class, with the direction and magnitude of any change dependent on the degree to which the class of interest is confused with the untrained class.

To help illustrate the effect of an untrained class and give a guide to the magnitude of the impacts of ignorance on the estimation of classification and map accuracy it may be helpful to explore the issues with a real data set. In previous studies, airborne thematic mapper (ATM) data have been used to classify and map crop classes in the vicinity of the village of Feltwell, Norfolk, UK (Foody and Arora, 1997). Based on the experience of these studies, attention focused on the classification of data acquired in three ATM bands, 0.60-0.63, 0.69-0.75 and 1.55-1.75 μm, with a discriminant analysis using training and testing sets designed to fit with standard recommendations for size (Foody and Arora, 1997). The training sets comprised 100 cases for each defined class and the testing set comprised 320 randomly selected cases.

Attention focussed on 6 land cover classes: sugar beet (SB), wheat (Wh), barley (Ba), carrots (Ca), Potatoes (Po) and grass (Gr). It was assumed that this was an exhaustively defined set of classes, although, as in other studies, it is likely that other classes (e.g. impervious, water *etc.*) were actually present. However, it was possible to select one class to be excluded from the training of the classifier and hence represent an untrained class. As this is most likely to occur with relatively rare classes, the grass class, which was least abundant in the region of interest, was not included initially in the training set. A classification with a discriminant analysis was undertaken, trained on the 500 cases of the 5 classes selected for training and then applied to the 320 cases in the test set. Cases of grass, the untrained class, in the test set were (i) ignored as not fitting the goal of the study and enabling classification accuracy for the 5 trained classes to be assessed or (ii) used to add a row and column to the confusion matrix to allow map accuracy assessment from a 6x6 matrix. The key features of both the 5x5 and 6x6 matrices that arise are shown in Table 3. Critically, the highlighted row and column associated with the untrained class are not observed in the assessment of classification accuracy (5x5 matrix) but are in the assessment of map accuracy (6x6 matrix).

	SB	Wh	Ba	Ca	Po	Gr	$^{C}\Sigma$	$^{M}\Sigma$	^{C}U	^{M}U
SB	86	3	0	0	0	0	89	89	96.6	96.6
Wh	3	91	6	1	2	0	103	103	88.3	88.3
Ba	1	2	45	0	0	6	48	54	93.7	83.3
Ca	0	0	0	29	1	9	30	39	96.6	74.3
Po	7	0	0	3	23	2	33	35	69.6	65.7
Gr	0	0	0	0	0	0	0	0	-	-
Σ	97	96	51	33	26	17	303	320		
^{C}P	88.6	94.7	88.2	87.8	88.4	-				
^{M}P	88.6	94.7	88.2	87.8	88.4	-				

Table 3. Confusion matrices for the analyses of the ATM data. When the grass class is unknown the row and column associated with it and associated marginal values, highlighted in grey, are unobserved and are not part of the calculations of accuracy. However, when the grass class is manifest the highlighted elements are included in the accuracy assessment. The overall, user's and producer's accuracies are also shown (as %). Note ^{C}O=90.4%, ^{M}O=85.6% and n/m=0.9468.

In Table 3 it is evident that of the 320 cases selected for accuracy assessment, 17 were actually members of the untrained grass class. When these cases were ignored, the classification accuracy assessment was based on the allocations for the remaining 303 cases. The cases of the untrained class were, however, manifest in the assessment of map accuracy as they appear as commission errors and map accuracy assessment is based on the full set of 320 cases.

The impact of the untrained class on the accuracy assessments is evident. When the analysis was ignorant of the grass class, the classification accuracy was estimated to be 90.4%. Knowledge of the untrained class's presence and confusion with the set of trained classes results in a map accuracy of 85.6%; the classification accuracy scaled by n/m.

On a per-class basis, it is evident that the presence of the untrained class was associated with an increase in commission error for some classes. For example, the largest error was associated with the carrots class, which commissioned 9 cases of the untrained class. As a result, the accuracy of the classification for the carrots class could change when the analysis was aware of the grass class. Specifically, the user's accuracy for carrots declined from 96.6% to 74.3%; scaling as a function of $n_{i\cdot}/m_{i\cdot}$. The producer's accuracy for each of the trained classes, however, remained the same, unaffected by knowledge of the untrained class as no case could be allocated to the untrained class and the row for the class full of 0s. Thus, for example, the producer's accuracy for the carrot class obtained from both the classification and map was 87.8%. Finally, the presence of the untrained class impacts on other properties such as area estimation. For area calculated with regard to the reference class, all area estimates (%) are reduced, scaling by n/m. For example, the carrot crop changes from 10.9% to 10.3% and the untrained class itself covers 5.3% of the region. For area estimates made relative to the classification labels, extent can increase if cases of the untrained class were commissioned by the class or could decline if not. For example, the sugar beet class was not confused with the untrained class and inclusion of the untrained class in the analysis results in the area estimation dropping from 29.3% to 27.8%. Conversely, with the carrot class, which did commission cases of the untrained class, the area estimate rises from 9.9% to 12.1% when its presence is included in the analysis.

Ideally, once the presence of the untrained class became apparent the analysis should have been repeated with it included in the training stage to ensure the satisfaction of the exhaustively defined set of classes assumption. Repeating the classification with grass included in the training set, now comprising 600 cases, yielded a classification which could be summarised in Table 4. Note that the inclusion of the grass class impacted on overall and per-class estimates of accuracy. Note also that the inclusion of the grass class impacts upon the training of the classifier. Since the grass class is now included in the training set it will influence the fitting of the class decision boundaries that separate the classes and this does impact on the accuracy of the classification. For example, it is evident that the inclusion of the grass class in training actually results in 1 additional case of the sugar beet class being classified correctly. Conversely, 1 less case of wheat is correctly classified when grass is included in the training set (Tables 3 and 4). These differences highlight that the class set defined for use in training the classifier also influences the classification and its accuracy.

	SB	WW	Ba	Ca	Po	Gr	Σ	*U*
SB	87	3	0	0	0	0	90	96.6
WW	3	90	6	1	2	0	102	88.2
Ba	0	2	45	0	0	0	47	95.7
Ca	0	1	0	29	0	1	31	93.5
Po	7	0	0	3	23	2	35	65.7
Gr	0	0	0	0	1	14	15	93.3
Σ	97	96	51	33	26	17	320	
P	89.6	93.7	88.2	87.8	88.4	82.3		

Table 4. Confusion matrix for the classification of the ATM data when all 6 classes included in training.

4. Conclusions

Thematic maps are commonly produced from remotely sensed imagery through the application of a supervised classifier. The reference data used to form the training and testing sets to respectively develop the classifier and evaluate the class allocations produced are typically acquired from imagery of the region of interest to be mapped. However, reference data for only a sub-set of the classes that exist within the region of interest may sometimes be acquired which violates the assumption of an exhaustively defined set of classes made with many classification methods. As a result, some parts of the region of interest to be mapped belong to a class beyond the set upon which the classifier was trained. When the analysis is ignorant of the existence of a class, the accuracy of the classification, assessed with the testing set of cases, may be a misleading guide to the accuracy of the thematic map produced by the application of the same trained classifier to the imagery of the region of interest. Since the cases of an untrained class can only be commissioned into the set of trained classes by most classifiers it follows that the overall accuracy of the map must be less than the accuracy of the classification as it must contain more incorrectly labelled cases. Similarly, on a per-class basis, the user's accuracy for a class will be less than suggested from the classification accuracy assessment if the class is confused with an untrained class due to increased commission error. Producer's accuracy for the set of trained classes, however, is unaffected by the presence of an untrained class. Other measures estimated from the classification confusion matrix, such as class areal extent, may also be impacted by the presence of untrained class(es). Simple relationships to scale map and classification accuracy were illustrated for a basic scenario which highlight that the magnitude of any difference between map and classification accuracy is a function of the abundance of the untrained class(es). Given interest is typically on the map, researchers may need to take care when interpreting and using classification accuracy statements.

Acknowledgements

I am grateful to Dr Gavin Long for help with the estimation of class areal extents from the FROMGLC map which was downloaded from http://data.ess.tsinghua.edu.cn/. I am also grateful to the editors and three referees for their thoughtful and constructive comments on the original manuscript that helped to enhance the article.

References

Carroll, L., 1893. *Sylvie and Bruno Concluded*. Macmillan, London.

Comber, A., Fisher, P. and Wadsworth, R., 2005a. What is land cover?. *Environment and Planning B: Planning and Design*, 32(2), 199-209.

Comber, A., Fisher, P. and Wadsworth, R., 2005b. You know what land cover is but does anyone else?… an investigation into semantic and ontological confusion. *International Journal of Remote Sensing*, 26(1), pp.223-228.

Congalton, R.G. and Green, K., 2009. *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*, second edition, CRC press, Boca Raton.

Costa, H., Foody, G.M. and Boyd, D.S., 2017. Using mixed objects in the training of object-based image classifications. *Remote Sensing of Environment*, 190, 188-197.

Finn, J.T., 1993. Use of the average mutual information index in evaluating classification error and consistency. *International Journal of Geographical Information Science*, 7(4), 349-366.

Foody, G.M., 2000. Estimation of sub-pixel land cover composition in the presence of untrained classes. *Computers and Geosciences*, 26(4), 469-478.

Foody, G.M., 2001. Thematic mapping from remotely sensed data with neural networks: MLP, RBF and PNN based approaches. *Journal of Geographical Systems*, 3(3), 217-232.

Foody, G.M., 2002. Hard and soft classifications by a neural network with a non-exhaustively defined set of classes. *International Journal of Remote Sensing*, 23(18), 3853-3864.

Foody, G.M., 2004a. Sub-pixel methods in remote sensing. In Jong S. M. D. and Meer, F. D. V. (eds) *Remote Sensing Image Analysis: Including the Spatial Domain*. Springer, Dordrecht, 37-49.

Foody, G.M., 2004b. Supervised image classification by MLP and RBF neural networks with and without an exhaustively defined set of classes. *International Journal of Remote Sensing*, 25(15), 3091-3104.

Foody, G.M. and Arora, M.K., 1997. An evaluation of some factors affecting the accuracy of classification by an artificial neural network. *International Journal of Remote Sensing*, 18(4), 799-810.

Ghamisi, P., Plaza, J., Chen, Y., Li, J. and Plaza, A.J., 2017. Advanced spectral classifiers for hyperspectral images: A review. *IEEE Geoscience and Remote Sensing Magazine*, 5(1), 8-32.

Gui, R., Xu, X., Wang, L. and Pu, F., 2018. A generalized zero-shot learning framework for PolSAR land cover classification, *Remote Sensing*, 10(8), 1307.

Hudak, M.J., 1992. RCE classifiers: Theory and practice. *Cybernetics and System*, 23(5), 483-515.

Li, M., Zang, S., Zhang, B., Li, S. and Wu, C., 2014. A review of remote sensing image classification techniques: The role of spatio-contextual information. *European Journal of Remote Sensing*, 47(1), 389-411.

Li, Q., Wang, C., Zhang, B. and Lu, L., 2015. Object-based crop classification with Landsat-MODIS enhanced time-series data. *Remote Sensing*, 7(12), 16091-16107.

Li, W., Dong, R., Fu, H., Wang, J., Yu, L. and Gong, P., 2020. Integrating Google Earth imagery with Landsat data to improve 30-m resolution land cover mapping. *Remote Sensing of Environment*, 237, p.111563.

Lu, D. and Weng, Q., 2007. A survey of image classification methods and techniques for improving classification performance. *International journal of Remote Sensing*, 28(5), 823-870.

Maling, D. H., 1989, *Meaurements from Maps: Principles and Methods of Cartometry*, Pergammon press, Oxford.

Olofsson, P., Foody, G.M., Herold, M., Stehman, S.V., Woodcock, C.E. and Wulder, M.A., 2014. Good practices for estimating area and assessing accuracy of land change. *Remote Sensing of Environment*, 148, 42-57.

Scepan, J., 1999. Thematic validation of high-resolution global land-cover data sets. *Photogrammetric Engineering and Remote Sensing*, 65, 1051-1060.

Skakun, S., Kussul, N., Shelestov, A.Y., Lavreniuk, M. and Kussul, O., 2016. Efficiency assessment of multitemporal C-band Radarsat-2 intensity and Landsat-8 surface reflectance satellite imagery for crop classification in Ukraine. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(8), 3712-3719.

Small, C., 2021. Grand challenges in remote sensing image analysis and classification. *Frontiers in Remote Sensing*, (in press).

Smith, B., and Mark, D. M., 1998. Ontology and geographic kinds. In T. K. Poiker and N. Chrisman (eds.), *Proceedings. 8th International Symposium on Spatial Data Handling (SDH'98)*, Vancouver: International Geographical Union, 1998, 308–320.

Sonobe, R., Yamaya, Y., Tani, H., Wang, X., Kobayashi, N. and Mochizuki, K.I., 2017. Mapping crop cover using multi-temporal Landsat 8 OLI imagery. *International Journal of Remote Sensing*, *38*(15), 4348-4361.

Stehman, S.V. and Foody, G.M., 2019. Key issues in rigorous accuracy assessment of land cover products. *Remote Sensing of Environment*, 231, p.111199.

Story, M. and Congalton, R.G., 1986. Accuracy assessment: a user's perspective. *Photogrammetric Engineering and Remote Sensing*, 52(3), 397-399.

Strahler, A.H., Boschetti, L., Foody, G.M., Friedl, M.A., Hansen, M.C., Herold, M., Mayaux, P., Morisette, J.T., Stehman, S.V. and Woodcock, C.E., 2006. Global land cover validation: Recommendations for evaluation and accuracy assessment of global land cover maps. *European Communities, Luxembourg*, 51(4).

Tso, B. and Mather, P., 2009. *Classification Methods for Remotely Sensed Data*, CRC Press, Boca Raton.

Toulmin, S., 1953. *The philosophy of science. London: Hutchinson.*

Vogt, L., Grobe, P., Quast, B. and Bartolomaeus, T., 2012. Fiat or bona fide boundary—a matter of granular perspective. *PLoS One*, *7*(12), p.e48603.

Ye, S., Pontius Jr, R.G. and Rakshit, R., 2018. A review of accuracy assessment for object-based image analysis: From per-pixel to per-polygon approaches. *ISPRS Journal of Photogrammetry and Remote Sensing*, 141, 137-147.