

## Uncertainties in land use data

G. Castilla<sup>1</sup> and G. J. Hay<sup>2</sup>

<sup>1</sup>Institute for Regional Development, University of Castilla La Mancha, Albacete, Spain

<sup>2</sup>Department of Geography, University of Calgary, Alberta, Canada

Received: 20 April 2006 – Published in Hydrol. Earth Syst. Sci. Discuss.: 10 November 2006

Revised: 26 October 2007 – Accepted: 19 November 2007 – Published: 29 November 2007

**Abstract.** This paper deals with the description and assessment of uncertainties in land use data derived from Remote Sensing observations, in the context of hydrological studies. Land use is a categorical regionalised variable reporting the main socio-economic role each location has, where the role is inferred from the pattern of occupation of land. The properties of this pattern that are relevant to hydrological processes have to be known with some accuracy in order to obtain reliable results; hence, uncertainty in land use data may lead to uncertainty in model predictions. There are two main uncertainties surrounding land use data, positional and categorical. The first one is briefly addressed and the second one is explored in more depth, including the factors that influence it. We (1) argue that the conventional method used to assess categorical uncertainty, the confusion matrix, is insufficient to propagate uncertainty through distributed hydrologic models; (2) report some alternative methods to tackle this and other insufficiencies; (3) stress the role of metadata as a more reliable means to assess the degree of distrust with which these data should be used; and (4) suggest some practical recommendations.

### 1 Introduction

Land use is an important variable influencing both hydrological and hydrogeological processes. In particular, it affects the volumes of surface runoff and the velocity of flow, which in turn influence infiltration and soil erosion. Additionally, land use is a major factor controlling evapotranspiration, hence it also has an impact on diffuse recharge. Therefore, uncertainties in land use data may propagate through models and diminish the reliability of their predictions. A sound assessment of these uncertainties, if incorporated into decision-

making, would increase the legitimacy of policy decisions based on those predictions, while also fostering greater stakeholder acceptance of the outcome from these decisions.

The goal of this paper is to raise awareness among practitioners who deal with land use data in hydrological studies on i) the uncertainties they bear; ii) the limitations of the conventional methods used to assess these uncertainties; and iii) alternative methods to describe them. After clarifying some concepts on the land use v. land cover and raster v. vector distinctions, the uncertainties in land use data are characterised and separated into positional and categorical uncertainties. Then, the factors that may influence them are examined, and finally, conclusions are presented together with some practical recommendations.

#### 1.1 Land use v. land cover

In this paper, land use is considered a categorical regionalised (i.e., spatially distributed) variable that may adopt as many different values as there are defined classes, the actual value being dependent on location. Land use classes describe the main socio-economic role of a given location, such as residential, industrial, agricultural, forestry, recreational, and conservancy. These roles shape and at the same time are shaped by the pattern of occupation of land, i.e., by the land cover. The latter refers to what is physically on the Earth surface such as vegetation, water or sand. Strictly speaking, land cover should be confined to vegetated and built-up areas. Consequently, classes like bare soil or sand (desert) describe land itself rather than land cover. However, in practise the scientific community is used to describe those situations under the term land cover (FAO, 1997).

The intimate relationship between land use and land cover often fuels confusion between both terms. In fact, they are used interchangeably in many maps, where natural and semi-natural areas are described as land cover, while agricultural and urban areas are described as land use. However, land

---

Correspondence to: G. Castilla  
(gcastill@ucalgary.ca)

use is the function of land cover for human activities, hence they are not synonyms. Furthermore, both domains lack a one-to-one correspondence. For instance, *recreational* is a land use class that may be applicable to different land cover classes like e.g. *water* (an all sports lake), *urban* (a funfair), or *forest* (a peri-urban park). Confusing both terms leads to increased ambiguities and incongruities in class definition; therefore they should be kept separate (Meinel and Hennemerdorf, 2002).

Notwithstanding, this paper deals with common maps currently used by hydrologists, which include both land use and land cover terms in their legend. Hence it is assumed that land use terms appearing in the legend can be inferred from the pattern of occupation of land. This assumption is necessary because the focus of this paper is on those land use data that are assimilated into models at the catchment or river basin scale. These are usually derived from Remote Sensing (RS) data, and in the latter, human activity (people manufacturing, harvesting, shopping or playing) cannot be directly observed, at least with current civil satellites. An overview of land use/land cover mapping can be found in Lins (1996).

## 1.2 Land use data formats and models

Land use data usually come in the form of maps that depict the distribution over a region of the set of land use classes included in the map legend. The latter must consist of a fixed number of mutually exclusive and collectively exhaustive classes (each one represented by a particular *label*), so that any given terrain unit can be assigned a label. Land use maps are typically derived from RS ortho-rectified (i.e. geometrically corrected to some cartographic projection) imagery, either aerial ortho-photos or satellite multispectral ortho-images.

### 1.2.1 Raster maps

When multispectral satellite imagery is used, it is common to apply a semi-automated classification to the image. This process uses pattern recognition methods to group individual data samples, or *signatures*, into classes. A signature is an  $n$ -element sample where each element usually is the value taken by a given individual pixel in each of  $n$  channels or *bands*. Signatures act as the coordinates of pixels in an  $n$ -dimensional space. The non-uniform arrangement of signatures (which usually tend to cluster into more or less discontinuous regions) within this space is used in the analysis to demarcate the regions of that space occupied by each class of interest. The common digital representation of maps derived from satellite images is a *raster* or *grid* made of square cells, where the value at each cell or *pixel* is the label returned by the automated classifier at that pixel. Eventually, this assignment may change after a post-classification step where a smoothing filter is applied to the raw classified image in order to enhance the accuracy and spatial coherency

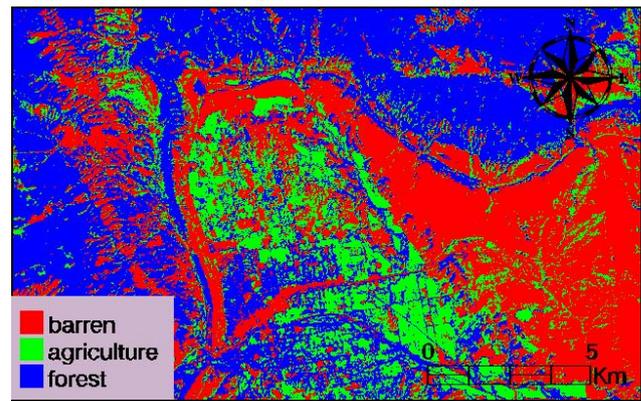


Fig. 1. Fictitious land use map with three classes.

of the resulting map. In others cases, rather than having a single choice for each map cell, each class is mapped separately as a membership function ranging from 0 (no membership to the class) to 1 (full membership). Such would be the case of fuzzy k-means classification, an example of which can be found in Burroughs et al. (2001). Irrespective of the processing sequence followed, the basic (areal) units used in this kind of maps are individual pixels. Thus the conceptual framework underlying this representation is *field-based*, since it considers land use a regionalised variable distributed over geographic space, that is, a *geographic field*. An example of a fictitious land use map, which will be used to illustrate some points, is shown in Fig. 1. It corresponds to a  $19 \times 12$  km region centred at Canon city, Colorado, USA, and it has been derived by applying a *maximum likelihood classifier* (MLC) to a Landsat TM image subset that can be found in the Tutorial Data CD #2 of the ENVI (a popular RS image processing package) distribution.

### 1.2.2 Vector maps

When derived from aerial photography, land use units are usually delineated manually by *photointerpretation*. This process consists in the identification of semantically homogeneous regions in the ortho-photo based on the visual differences that different land use classes create. The usual digital representation of maps derived from ortho-photos is a *polygon vector layer*, i.e., a mosaic of non-overlapping contiguous units or *polygons* representing patches of land whose label, unlike raster cells, is necessarily different than the ones of adjacent units. The term *vector* in this context refers to a data structure that uses the segments connecting a topologically ordered series of points as the basic building blocks for representing the boundaries of polygons. These polygons, being unitary and different from their surroundings, can be regarded as representing geographic objects, thus the conceptual framework underlying this representation is *object-based*.

The present widespread availability of very high resolution (<5 m pixel size) RS imagery is fostering the abandonment of pixel-based methods to derive land use maps, which are being replaced by *geographic object-based imaged analysis* (GEOBIA) methods (Hay and Castilla, 2007). The reason is that, in these images, pixels are too small to be representative of classes whose biophysical description refers to a setting that necessarily encompasses more than a few square metres. GEOBIA methods, unlike the conventional *pixel-based* ones, use image-objects as the basic units of the analysis. The latter are delimited – typically by an image segmentation algorithm- regions of the image that are internally coherent and relatively different from their surroundings (Castilla and Hay, 2007). By using this kind of units, additional features that cannot be obtained from individual pixels (such as those derived from the shape of the regions and their mutual relations) may be included in the analysis. With such enhanced capabilities, GEOBIA has the potential to supersede not only conventional pixel-based methods, but also photointerpretation (Castilla et al., 2007). Therefore it is likely that in the future, most land use data will be derived this way and presented in vector format.

### 1.2.3 Implications of vector to raster conversion

This paper focuses on uncertainties in raster land use maps derived from digital classification of RS data. Raster maps are more commonly used in water-related studies than vector maps, since gridded data can be readily assimilated into distributed hydrologic models. Notwithstanding, we also briefly address uncertainty in land use maps in vector format, since they are occasionally used in this context by converting them to raster before assimilation. In this respect, users should be aware that vector to raster conversion, albeit a straightforward process, has some implications on the reliability of the results. The conversion not only involves a change of data format, but also of conceptual framework. The object-based model has a higher abstraction level than the field-based model (e.g., it uses relational features between objects that are not applicable in the field-model), hence what could be regarded as an error in the latter is simply a necessary generalisation in the former, as explained next.

An important generalisation mechanism inherent to polygon maps is the *minimum mapping unit* (MMU), i.e. the minimum size (or sometimes width, when referred to elongated units) that a land parcel must exceed in order to be represented in the map. Isolated land use units having a size below this threshold are aggregated into the surrounding unit. Afterwards, there is no trace in the map of the smaller units. An exception is the case of *mosaic polygons*, which have a compound label representing a mosaic of patches from different classes, all smaller than the MMU. In these kinds of polygons, the percentage cover from each class may be reported, but information regarding their actual distribution within the polygon is missing as a parsimonious exchange for clarity.

In general, each MMU will yield a different map, and the larger the MMU, the greater the proportion of the map extent catalogued as *mosaic* (Castilla, 2003). The conclusion is that when a polygon map is gridded for use in a model, and the grid cell size is several times smaller than the MMU, then we cannot be certain that all the cells within a given polygon actually belong to the declared land use class, even if the reported accuracy of the polygon map was 100%. The only clues to assess this uncertainty are the MMU and cell size, and the complexity of the mapped landscape.

## 2 Characterising uncertainty in land use data

In this paper, *uncertainty* is an indication of the degree of distrust with which the data should be regarded or used. Consequently, the higher the uncertainty about a given datum, the more likely is that the actual land use of the piece of terrain to which this datum refers to is not the one reported; and at the same time, the more serious are the consequences of wrongly assuming that land use at that location is the one reported. Uncertainty so defined influences the extent to which the predictions made by a model using this data are to be believed, and ultimately the strength with which those predictions may support or justify a given environmental decision. We note that under this definition, uncertainty is not only dependent on the data but also on the user. The same dataset can be regarded with different degrees of distrust depending on how it will be used, meaning that some errors may be relevant for the intended application, while others may have no impact. In the case of hydrological modelling, uncertainty in land use data depends on the sensitivity of model output to varying land use data input. Therefore a sensitivity analysis, typically based on Monte Carlo simulation, may be used to assess the impact of land use data uncertainty on the hydrologic model (e.g., Eckhardt et al., 2003).

Land use may be considered as categorical regionalised variable that describes the main type of activity each land unit (i.e. the footprint of each pixel) is devoted to. As such, land use data uncertainty may be characterised using the integrated framework provided by Brown et al. (2005) (Table 1). Regarding its method of determination, the activity is inferred from the particular combination of recurrent elements (such as trees or buildings) that are typically present in the places where this activity is carried out. This setting yields a particular joint reflectance profile when observed from far-away. Such a profile can in turn be measured (after accounting for atmospheric interactions) by imaging spectrometers mounted on satellites. Hence, land use is ascertained from pixel signatures through a process called *image classification*, which consists in demarcating the regions of the multi-dimensional data space associated with each class of interest  $c_i$  ( $i=1, \dots, m$ ). The classification is carried out with the aid of a set of *discriminant functions*  $g_i$  (one for each of  $m$  classes), such that given a signature  $X$ ,  $g_i(X)$  is greater than the other

**Table 1.** Characterisation, following the method by Brown et al. (2005), of land use data quality and uncertainty.

Variable name	Land use
Method of determination	Semi-automated classification
Data category	D3 (categorical variable that varies in space and time)
Type of empiric uncertainty	M1 (mean classification error derived from a contingency matrix)
Instrument quality	I2 to I3 (instruments well fitted to not well matched, depending on spectral and spatial resolution)
Sampling strategy	S2 (limited number of both training and verification samples)
Overall method	O3 (Reliable method common within discipline)
Longevity of uncertainty information	L1 (change over time)

$g_j$  when  $X$  belong to  $c_i$ . In other words,  $X$  is classified as a member of class  $c_i$  if and only if  $g_i(X) \geq g_j(X)$  for all  $j=1,2,\dots,m$  (Landgrebe, 1999). For example, the map shown in Fig. 1 is the result of applying a maximum likelihood classifier (where the discriminant functions return the probability that the signature belongs to each class assuming that the statistics for each class in each band are normally distributed) to the six optical bands of the Landsat TM image.

Within Brown et al. (2005) framework, land use belongs to data category D3, and its empiric uncertainty is assessed quantitatively through the statistics derived from a contingency table of errors (M1 empirical uncertainty category). The instrument quality, here referring to the remote sensors that record the data from which land use is inferred, is difficult to assess, because these data are only contingently related to the *intensional* definition (i.e., the set of features distinguishing a class from all others) of land use classes. Instead, a surrogate definition, based on spectral rather than biophysical features, is constructed using a set of typical signatures or *training pixels* collected from representative well known locations. For example, the training pixels used for the map of Fig. 1 are those included in the file *classes.roi* of the *can.tm* folder of the ENVI Tutorial Data CD #2. Then the region(s) of the multidimensional data space occupied by each class are demarcated according to this definition. Signatures inside that region(s) constitute the *extensional* definition (the set of instances belonging to it) of the class. The expected result of this indirect method is that the projection of this *extension* onto geographic space, i.e. the set of terrain plots that belong to each class, coincides to a great deal with the one that would have been obtained should the proper *intension* (related to human activities) be applied to exhaustive field observations. The degree of success is later verified from a set of reserved (not used for training) *test pixels* from known locations. Regardless of the particular classifier employed (an overview of the different methods can be found in Richards and Jia, 1999), image classification is considered a reliable and common method for deriving land use data from RS data, so it can be assigned to Brown et al. (2005) category O3.

Finally, regarding the temporal dimension of land use data uncertainty, it can be considered as belonging to category L1 (it changes over time). Land use maps are snapshots taken when primary data (e.g., aerial photographs) were collected. Age decreases the reliability of the information portrayed in the map, since there can be changes that affect land use, such as wildfires, the construction of new infrastructures, and shifts from rural to urban or from agriculture to forestry. The more frequently these changes occur, the more urgent the need for updating. Since the likelihood of changes is not uniform throughout the landscape (e.g., it is higher in the urban-rural buffer), age affects unevenly the reliability of these data. Similarly, it is unusual that all data are coetaneous for a given mapping project, especially field surveys, so again the temporal reliability is likely to vary from one sheet to another. An important consequence of temporal uncertainty is that, as the database is updated, past deductions have to be revised as they may no longer be valid. This is known as the “belief maintenance problem” (Frank, 2003). For example, a conservancy area may have been assessed in an earlier study as having a low erosion risk, but after a wildfire that assessment may not be true anymore.

### 3 Positional uncertainty

There are two main uncertainties contended with when dealing with land use data, *positional* and *categorical* (usually termed *thematic* within Geographic Information Science literature) uncertainty. This distinction is debatable, since a label disagreement at given location could be interpreted as being due to either positional (e.g. a systematic coordinate error) or categorical error. However, this separation will be followed here, for two reasons. First, it is common in the literature. Second, it is useful, for it distinguishes two types of uncertainty that are associated to two domains very different in nature, namely the cartographic domain and the classification domain.

In raster maps, positional uncertainty relates inversely to the degree of confidence we may have that the actual location of the plot of terrain represented by a given cell corresponds

acceptably to the coordinates of that cell. Hence, positional uncertainty depends mostly on the quality of the geometric correction (ortho-rectification) performed on the satellite image from which the map was derived. Positional accuracy is usually estimated by the Root Mean Square Error (RMSE) of selected points (such as crossroads) clearly identifiable in the image and whose precise coordinates are known from a higher accuracy source (e.g. a finer scale topographic map or differential GPS measurements). RMSE is computed as the square root of the mean of the squared errors, and is calculated combining both  $x$  and  $y$  directions. Such estimation assumes that positional errors are randomly distributed throughout the imaged scene, which may well not be the case, especially in rugged terrain due to relief distortions.

In general, positional uncertainty is far less serious than the categorical one. For example, in a vegetation mapping study, Green and Hartley (2000) calculated positional error introduced by georeferencing, digitising and subjective interpretation, and found that the latter process accounted for 90% of the total error. In practice, RMSE is considered acceptable when it is less than the pixel size, a fact that is referred to as *subpixel accuracy*. In this situation, i.e., when an image has been geocorrected at subpixel accuracy, the true location of the centre of any given pixel can be safely assumed to lie somewhere within a  $3 \times 3$  block of pixels surrounding that point (Goodchild, 1994). For a review on geometric correction of RS images, see Toutin (2004).

In vector maps, positional uncertainty relates inversely to the degree of confidence we can have that the boundary between two given polygons lies in the right place. As previously noted, this uncertainty is inseparable from the categorical one. The reason is that the boundaries being sought and delineated are only those that differentiate the land use classes in the chosen classification scheme (Bie and Beckett, 1973). Therefore the uncertainty attached to boundary placement is proportional to (1) how different the classes separated by the boundary appear in that area; and (2) how abrupt the transition from one class to the other is, i.e., how sharp the boundary is. Since in any given RS image some boundaries are softer than others, uncertainty should be estimated independently for each arc in the map.

This estimation could be made through the definition of a probabilistic epsilon band (Honeycutt, 1987) within which the 'true' boundary between two polygons has a 99% probability of being located. The rationale behind epsilon bands is as follows. Assuming a correct classification of polygons, one could argue that a point precisely on the boundary could equally well belong to either class (Blakemore, 1984). Moving away from the boundary towards the centre of the polygon increases the probability of a correct classification, while at the same time the probability that this location is where the boundary should lie decreases. The manner in which this probability drops off depends again on both boundary and class distinctiveness.

Regrettably, the epsilon band, being boundary-dependent

and arc-specific, is rarely, if ever, computed. Following the example in Green and Hartley (2000), a general procedure for the estimation of the error due to subjective interpretation could be obtained by overlaying several photointerpretations of the same area carried out by different equally-skilled interpreters. After intersecting the vector layers produced by each interpreter, some boundaries will be very consistent, whilst others will vary markedly, resulting in dozens of sliver (i.e., spurious) polygons. The width of the epsilon band corresponding to a given soft boundary would be the mean distance between the inner and outer wrapping lines encompassing the set of sliver polygons existing along that boundary.

Unfortunately, not only is the above procedure hardly feasible within the context of a mapping project, but it is grounded on an unrealistic assumption. That is, given a piece of land independently mapped by several interpreters using a common scale and legend, it assumes that overlaying the resultant maps will resemble an egg-yolk representation (Cohn and Gotts, 1996). In the latter, each polygon is like a fried egg that has a yolk (i.e. a core area free of sliver polygons) and a white (the set of sliver polygons surrounding that core), the white being the epsilon band. Such representation assumes that any two high quality photointerpretations of that piece land would create the same set of geographic objects, but with slightly different boundaries. However, it will not be unusual to find some polygons drawn by interpreter A that are crossed in the middle by an arc delineated by interpreter B, and vice versa.

A more feasible alternative for assessing the positional uncertainty of arcs is to express it as a combined measure of boundary distinctness both from the radiometric and semantic points of view. For a given arc, radiometric distinctness could be estimated as the mean gradient magnitude of pixels crossed by the arc. Semantic distinctness could be equated to the value of some biophysical similarity index between the classes being separated by the arc (see e.g. Orzanco et al., 2004). In addition, positional accuracy could be assessed polygon-wise, preferably for the same sample of polygons that is used to evaluate thematic accuracy, which for the reasons stated in the previous paragraph should consist of non-adjacent polygons. The question to be answered for each polygon in the sample would be: are the boundaries of this polygon delineated in such a way that it can be conceived as representing a coherent land use unit under the view supplied by the classification scheme? The answer could be given qualitatively using a nominal scale, or even quantitatively by computing an epsilon band derived from several interpreters who are separately given the task of improving the delineation of that polygon. A study testing this procedure, which to the best of our knowledge has not been carried out yet, would be desirable.

In practice, positional accuracy is estimated through the RMSE of sample points along vector arcs that correspond to sharp boundaries in the image from which the map was derived. This estimation is biased towards human-influenced

**Table 2.** Sample confusion matrix for the fictitious map of Fig. 1.

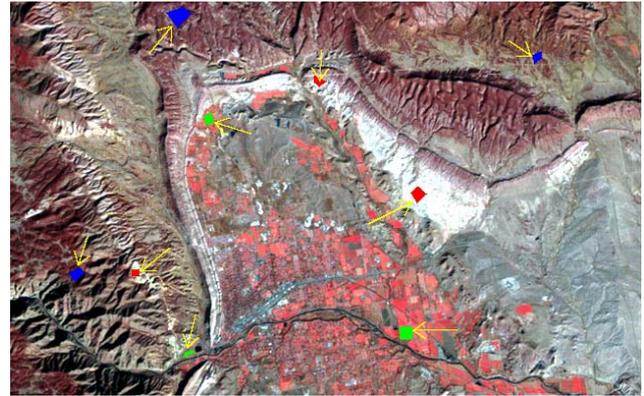
Class	agriculture	forest	barren	Total	User Acc. %
agriculture	275	8	1	284	96.83
forest	27	459	2	488	94.06
barren	1	6	273	280	97.50
Total	303	473	276	1052	
Prod. Acc.%	90.76	97.04	98.91		<b>95.72</b>

features, such as the edge between a woodlot and a paddock, since ‘natural’ boundaries are less distinct. Therefore, the RMSE method is not suitable to assess how well polygon boundaries represent landscape structure, rather it is an indication of the steadiness of the interpreter’s hand (and of the visualisation scale they used).

#### 4 Categorical uncertainty

Categorical uncertainty is inversely related to the degree of confidence we can have about whether the plot of terrain corresponding to a given map unit is actually devoted to the use indicated in the map. This uncertainty is commonly assessed using a contingency table of agreement between predicted and observed values, typically referred to as the *confusion matrix* (Table 2). We note that *accuracy*, the term commonly used in the RS literature, may be considered the antonym of *uncertainty* in this context, i.e., the more accurate the map is, the less uncertainty it has. However, as noted earlier, the actual relation between accuracy estimates and uncertainty as a measure of distrust will depend on the particular application. For a review on accuracy assessment of land cover maps, see Foody (2002). For an example of an accuracy assessment of a national land cover map, see Stehman *et al.* (2003), which is based on the methodological framework put forward by Stehman and Czaplewski (1998), which in turn divides the accuracy assessment into three components: (1) *sample design*; (2) *response design*; and (3) *analysis*.

(1) *Sample design* is the protocol used to determine the number, location, spatial support and nature (e.g. aerial photos or field plots) of the sample units that will populate the confusion matrix; (2) *response design* is the protocol for assigning a label to each sample unit, including the procedures to collect the information used in the assignment; and (3) *analysis* is the protocol for deriving accuracy statistics from the confusion matrix. Unfortunately, this framework is not worked out explicitly or reported in most maps. Without such details, it is difficult for a user to appreciate how close the accuracy estimates are to the “true” map accuracy, and how robust or repeatable they are. A good example on how to develop a sound accuracy assessment framework, includ-

**Fig. 2.** Location of ground truth pixels used to construct the confusion matrix of Table 2.

ing some useful information on key decisions that have to be made, can be found in Wulder *et al.* (2006a).

In practice, the confusion matrix is computed from a subset of pixels from known areas that were not used as training pixels, and it compares for each land use class the predicted class with the actual one on the ground. There are a number of methods to measure accuracy from this table, the simplest being the percent correctly classified, usually called *overall accuracy* (95.72% in the example). The recommended accuracy threshold, below which the resulting map should be discarded for operational purposes, is 85% (Anderson *et al.*, 1976). Another common measure is the *kappa index*, similar to the former but it ranges from 0 to 1 and is not biased by chance agreement (i.e. it takes into account the expected rate of agreement between predicted and actual datasets based on chance alone). Kappa values over 0.75 indicate very good correspondence between the two datasets, while values below 0.50 indicate poor correspondence. Individual class accuracy may be reported either from the map user’s or producer’s perspective, or both (Story and Congalton, 1986). User’s accuracy refers to *commission errors* (i.e., pixels with an assigned class that in reality do not occur at them), whereas producer’s accuracy refers to *omission errors* (i.e., pixels of a given class that were wrongly assigned to a different class). The commission/omission distinction is especially relevant for binary maps (with only two classes, e.g., change v. no-change), as it enables the assessment of the *specificity* (absence of commission error) and *sensitivity* (absence of omission error) of the map to e.g. varying change thresholds. Such assessment may be carried out through Receiver Operating Characteristic (ROC) analysis (Zweig and Campbell, 1993).

Regarding vector maps, the comparison between predicted and actual land use class should consider the polygon as a whole. This becomes troublesome if the validation method consists of field surveys, because it is difficult to infer the polygon label from plot or transect data due to the inevitable

heterogeneity of polygon interiors. This difficulty may be tackled using field plots larger than the minimum mapping unit (MMU), but this may hinder the cost-effectiveness of the sampling design, or simply be unfeasible when the MMU is larger than say 1 ha. Also, if there is considerable variability in polygon size, care should be taken when selecting the sampling design (Stehman and Czaplewski, 1998), since error estimates should be referred to total area rather than to the number of polygons.

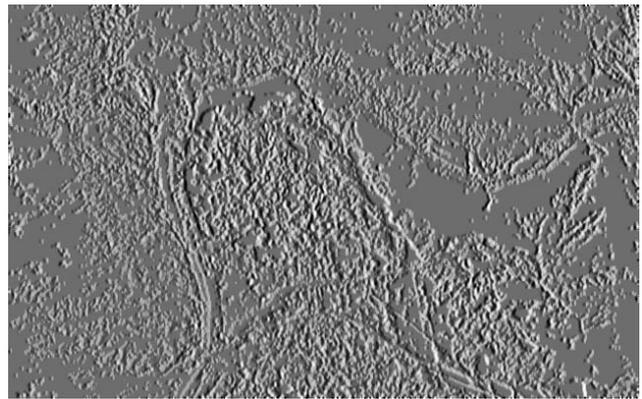
#### 4.1 Problems of the conventional approach

The main drawback of the confusion matrix is that it only captures the average error over the entire mapped area, whereas the likelihood of misclassification may vary markedly from one place to another (Goodchild, 1994). For example, in a recent study on spatially constrained confusion matrices derived from the same image, Foody (2005) reported that while the global accuracy for the whole image was estimated to be 84%, local estimates varied from 53% to 100%. We note that in addition to the often biased spatial distribution of errors, usually there are also significant differences in error rates among the classes (Davis and Simonett, 1991), albeit this aspect is well displayed in the confusion matrix.

The selection of the sample pixels used in the construction of the matrix may also optimistically bias the accuracy estimates (this actually occurs in the example used throughout this paper), since they are usually collected in blocks of contiguous pixels rather than individually (Fig. 2). Blocks usually correspond to homogeneous areas far from boundaries between different land use units. In this way, mixed pixels, which are prone to be misclassified, are systematically excluded from the sample (Plourde and Congalton, 2003). The block sampling procedure also violates the independency assumption of statistical sampling, because near pixels are usually correlated and therefore tend to show similar values.

Another aspect that the usual confusion matrix neglects is the relative seriousness of the misclassification. In many maps, the errors observed in a classification are between relatively similar classes and often these may be unimportant, while other errors may be highly significant (Foody, 2002). For example, it is more serious to confuse a *lake* with a *forest* than a forest with a *sparse woodland*. However, both have the same weight in the computation of conventional accuracy estimates, which treat all errors as equally serious.

It is also worth noting another flaw regarding the source of “ground truth”, that is, the reference data upon which the classification results are validated. Many times, because of financial and logistic constraints, ground truth do not come from field surveys but from interpreted aerial photos of higher resolution, or even from previously compiled maps of greater level of detail, available for some limited parts of the new map. When a finer map is used as reference, the comparison may be misleading, due to the likely differ-



**Fig. 3a.** Uncertainty landscape derived from the confusion matrix of Table 2.



**Fig. 3b.** Uncertainty landscape derived from MLC classifier.

ent conceptual and averaging filters that each map applies to the territory, especially if the areal units of the reference map are polygons (Wulder et al., 2006b). For example, Finke et al. (1999) compared the CORINE landcover map with the Landcover database of the Netherlands and concluded that the former contains considerable errors, reporting that 69% of the area covered by (semi)natural vegetation was misclassified in a combined soil/vegetation map. Such disagreement probably comes from different map legends, spatial support and minimum mapping unit, rather than to sheer ‘error’ in the less detailed map. Indeed, mapping errors are “forcible deviations between a representation and actual circumstances” (Chrisman, 1991). But actual circumstances are described at a given scale of observation using a given set of concepts – the map legend. Therefore, in order to estimate error in a map by means of a more detailed map, the latter should use the same concepts as the former; and prior to comparison, it should be resampled to the same resolution as the former. Otherwise the map uncertainty may be overestimated, as the fine scale map inherently has more spatial information.

The sensitivity of analytical results to the type of the areal units from which data are collected has been conceptualized

by Openshaw and Taylor (1981) as the Modifiable Areal Unit Problem, or MAUP, which is akin to the Change Of Support Problem (COSP) identified in Geostatistics (Cressie, 1996). It arises from the fact that these units are arbitrarily defined and eventually modified to form larger units. Therefore, if the areal units are arbitrary and modifiable, then the soundness of any model based upon them may be rightly questioned. MAUP was identified in the context of socio-economical geography, but it has been also found in Landscape Ecology and Remote Sensing. Marceau (1999) gives a comprehensive review on the issue.

#### 4.2 Alternative approaches

To sum up, despite the fact that confusion matrices are widely used as the standard accuracy assessment for RS-derived maps, they are clearly insufficient for propagating land use uncertainties through a hydrologic distributed model, as they do not take into account the spatial distribution of errors. To illustrate this point, Fig. 3a displays a shaded relief from the *uncertainty landscape* (where the altitude of each point is proportional to the accuracy of the class it has been assigned to) that could be inferred from the confusion matrix of Table 2. This image conspicuously differs from the shaded relief of Fig. 3b, which corresponds to a landscape where the altitude of each point is proportional to the estimated probability that the point actually belongs to the class assigned by the *maximum likelihood classifier* (MLC). In both landscapes, areas of higher uncertainty correspond to pits or canyons. The landscape from the confusion matrix is less rugged than the one from the MLC, and rugged terrain is distributed differently in each image. The main reason is that Fig. 3a was derived using global estimates, whereas Fig. 3b used local estimates.

In order to tackle this deficiency of the confusion matrix, there are two previously proposed alternatives that may be tried. One is the general error model proposed by Goodchild et al. (1992). In this model, each pixel is associated with an  $m$ -component vector of probabilities giving the probability that the pixel belongs to each class 1 through  $m$ . The classes allocated to the pixels in the map represent one realization of a stochastic process defined by these vectors. That is, over a large number of realizations, the proportion of times a pixel is assigned to each class will converge on each class's probability at that pixel. In addition, within any given realization, the outcomes in neighbouring pixels are correlated, so that the model also includes parameters describing the level of spatial dependence. Sample realizations can be obtained as the outcome of a classification performed using a randomly selected subset of training pixels. The parameters of the model can then be calibrated by adjusting them so that the range of outcomes matches reasonably the range observed in reality. An example of its application can be found in Horttanainen and Virrantaus (2004). Despite being an interesting alternative, the crux of this method is how to define the parameter

that controls the level of spatial dependency, since the latter varies markedly throughout the image.

Another alternative would be the computation of spatially constrained confusion matrices to characterize the spatial variation of accuracy throughout the scene (Foody, 2005). Given a set of  $n$  predefined locations of interest situated well apart from each other,  $n$  confusion matrices could be computed from the  $k$  nearest samples to each location. This approach assumes that enough samples are available around the locations of interest, which may not be always the case. Also, different choices in the selection of these locations and the number of samples per location will in all likelihood produce different accuracy estimates. However, it is a simple and inexpensive means of extending the conventional approach with information on how classification accuracy varies across the mapped landscape. Further details can be found in Foody (2005).

The problem of accounting for the relevance of different classification errors could be tackled by using a nominal scale to evaluate map v. field comparisons, ranging from "absolutely irrelevant" to "absolutely severe" (e.g., Gopal and Woodcock, 1994). This scale would easily allow weighing the degree of disagreement between the map and field observations according to the intended use of the map, and hence would enable users to convert disagreement into distrust, i.e., relate accuracy to uncertainty. For example, confusing two land use classes that have similar parameterisation in a given hydrologic model would not lead to a significantly different model output. Therefore this error could be labelled as "absolutely irrelevant" and have a low weight in the uncertainty analysis. The weight itself could be determined by the relative impact of this error on the model response, or at a higher level, by the economic loss due to wrong decisions based on erroneous data (e.g., Naesset, 1996).

## 5 Factors that influence land use data uncertainty

The main factors influencing the reliability of raster land use data can be identified as (1) the quality of the image(s) used as input for classification; (2) the quality of the training pixels used to define quantitatively the classes; (3) the degree of correspondence between the proper definition of classes and their radiometric definition. The factors affecting vector land use maps derived from phointerpretation are briefly discussed at the end of this section.

### 5.1 Image quality

Satellite images are measurements, distributed at a fixed ground sampling interval (equal to the pixel size), of regionalised variables. These variables are usually related to some electromagnetic property of the Earth surface and the atmosphere, such as the radiance recorded by optic sensors at several bands of the spectrum. The latter depends not only on

the reflectance of the surface, but also on atmospheric conditions and on incidence and viewing angles. So the clearer the atmosphere and the flatter the terrain, the more can be expected that reflectance estimates derived from the recorded values are equally good for all the pixels in the image. Assuming that the image either fulfils these requirements or has been adequately corrected for atmospheric and relief effects, each pixel can then be considered as a sample introduced in a desktop spectrometer. The resulting signature is then compared to the ones of selected samples (training pixels) of each material that can be found in the imaged landscape. After comparison (i.e. classification), the material having the most similar signature(s) to the one under analysis is selected as the class to which that pixel belongs.

This discourse can be extended to cases where signatures are not spectral, like e.g. crop classification using multitemporal radar images. The key point is that the set of images used in the analysis allows for a good discrimination between classes, i.e. that no two signatures from different class are similar. In other words, in order for the classification to be successful, signatures should clump into clusters in the  $n$ -dimensional data space, where each cluster is composed of signatures of a prevailing class, and where clusters from different classes are separated by quasi-empty space. Conversely, the greater the overlap between two given classes, the higher the probability that they will be confused (Schowengerdt, 1997). Therefore a high quality multiband image in this context is one where signatures are segregated in the data space according to their class, enabling an accurate classification.

A related issue is, given a data set, what is the best combination of images, in type and number, which can be used to map land use (e.g., Mausel et al., 1990). This is a classic problem of pattern recognition, called *feature selection* (where feature stands for *band*), consisting of two inter-related parts: feature extraction (the transformation and/or combination of the original images/bands into new ones) and feature reduction (the reduction of the dimensionality of the data set by selecting the smallest subset of bands providing an acceptable discriminative power). Feature selection is generally considered a process of mapping the original measurements into more effective features. Unfortunately, in many applications, the important features are nonlinear functions of original measurements. Since there is no general theory to generate mapping functions and to find the optimum one, feature selection becomes very much problem oriented (Fukunaga, 1972).

## 5.2 Representativeness of training pixels

Turning to training pixels, the accuracy of image classifications depends heavily on their quality, even more than on the actual classifier used (Buttner et al., 1989). Moreover, the same classifier is likely to produce different results on the same image when trained with a different set of training pix-

els (Smits et al., 1999). As a consequence, the result is prone to reflect inconsistencies in the selection of training samples. Thus “good” training pixels must be fully representative of their respective class, so that a good number of instances of the set of typical signatures of that class are included. This implies e.g. that they should be well distributed across the scene, as there may be “local varieties” of the material, where each variety may conform a separate cluster in the data space.

## 5.3 Correspondence between biophysical and radiometric class definitions

Classes are defined quantitatively through the signatures of training pixels. In doing so, it is assumed that there is a *bijection* between location in data space and location in the categorical space defined by the classification scheme of the map. That is, if a signature is located in a region of the data space that belongs to the class *forest*, it is expected that the plot of terrain from which that signature was extracted is “densely covered by trees”. Conversely, the signature of any given place densely covered by trees is expected to lie within some cluster of the data space that has been allocated to the *forest* class. Such correspondence does not depend solely on the quality of both the image and training pixels, but also on the very definition and number of classes. The more adapted the set of classes to the structure of the data space, the better the correspondence and therefore the accuracy. Hence, if we include in the map legend two classes that share the same tracts of the data space, the classification results will be poor. In general, the higher the number of classes, the higher the number of both attributes (bands) and training samples (pixels) required for a good classification.

In any case, the correspondence cannot be perfect, as classes must fulfil some conflicting requirements. On the one hand, classes must be meaningful for users and meet their needs, covering exhaustively all the possible land uses that can be found in the mapped region. On the other hand, classes must be separable to an adequate degree in the data space. In fact, there will always be some settings that lie in between the definition of two classes. For example, a given area may have such a tree density that it cannot be considered a *forest*, but neither a *sparse woodland*. If such areas are common in the region, it would be advisable to create a new class, e.g., *open forest*. However, since it can be expected that the signatures of the new class overlap with those of the former classes, it is unlikely that the accuracy of the map is increased.

In a similar way, there are always pixels that are crossed by an edge separating different land patches, the so called *mixed pixels*. Their signature consists of a mixture of two (or more) classes and may be located in tracts of data space occupied by other classes. For example, a Landsat pixel situated between a corn field and a bare field has a mixed signature that may be confused with the signature of a class having a low green cover fraction, such as *sparse woodland*. The

abundance of mixed pixels depends on the resolution of the image and the complexity of the landscape (Markham and Townshend, 1981), so that the odds of a correct classification decreases with decreasing patch size and increasing heterogeneity (Smith et al., 2003). Since, in addition to other factors, the problem of mixed pixels is inversely related to the problem of the spectral heterogeneity of classes, it is impossible to achieve a 100% accuracy. For instance, the proportion of mixed pixels may be reduced by decreasing the pixel size, but at the expense of increasing intraclass variability. In short, classes should be defined in such a way that can be distinguished with the satellite data used to map them. Attempting to include classes that consists of instances with radiometric signatures very different among them (e.g. the class *urban* in Landsat imagery) and similar to the ones of other classes (e.g. *wheat* and *barley*) will result in a poor and inconsistent map.

#### 5.4 Factors affecting vector land use data

In the case of polygon land use maps derived from photointerpretation, in addition to the quality of the image, the uncertainty of the derived map will rely on the quality of the interpretation. This in turn depends on the skills and experience of the interpreters, and on the time allocated for interpretation (and even on the interpreter's mood during that time). Since this quality may change from sheet to sheet due to different interpreters performing this task, it is of utmost importance to standardise observational techniques (e.g. digitising scale) and criteria (through e.g. a photointerpretation key consisting of several examples for each class) among interpreters (Lillesand et al., 2003).

## 6 Conclusions

The *confusion matrix* is the standard means for assessment of categorical uncertainty in RS-derived land use raster maps. An example of an accuracy assessment protocol based on this matrix can be found in Stehman et al. (2003). Unfortunately, this matrix does not provide information on the spatial distribution of errors, and hence it cannot be used to propagate uncertainty through distributed hydrologic models. There are two previously proposed alternatives to tackle this deficiency. One is the error model of Goodchild et al. (1992), an applied example of which can be found in Horttanainen and Virran-taus (2004). The other is the local characterisation of classification accuracy through spatially constrained confusion matrices (Foody, 2005).

In addition to neglecting the spatial distribution of errors, conventional procedures do not account for the relative importance of different misclassification errors in relation to the intended use of the data. Hence, the estimates they provide cannot be operationally used as a measure of the degree of distrust the data deserve. This could be tackled by weigh-

ing different types of errors according to an ordinal scale that measures the relative impact of the error on the particular application. In any case, despite the apparent objectivity of quantitative estimates derived from any given method, it is important that they are interpreted with care, since there are many factors that may result in a misleading interpretation drawn from an apparently objective statement (Foody, 2002).

For this reason, quantitative analytical results must be complemented with qualitative insights on how reliable a map is. The *whole picture* could be gained by a thorough inspection of (well documented) *metadata*, i.e., information describing the land use map itself. Metadata can provide users with a sense of the amount of distrust with which the data should be used. In order to make such intuitive assessment, metadata should not only describe comprehensively the material – images and ancillary information- used in the compilation of the map, but the methods, including the location of training and verification samples (Stehman and Czaplewski, 1998). Within this scope, the International Metadata Standard for Geographic Information ISO 19115 defines more than 300 metadata elements structured into 14 packages, most of which can be applied optionally. Metadata are usually stored in XML format, which can be accessed with standard text editors. If for a given map the package related to *Data\_Quality\_Information* is not empty, then the user may have information on the accuracy of the map. See MEGIS (2006) for an example of detailed metadata on a landcover map. On a positive note, even if current metadata for land use maps tend to be poor if they exist at all, the situation may be reversed soon, at least in Europe. The recently issued INSPIRE Directive (EU, 2007) specifies that “metadata shall include information on the quality and validity of spatial data sets”. This unprecedented legal requirement will in all likelihood increase concern about metadata among European map producers.

In the absence of detailed metadata, the overall accuracy estimate (derived from a confusion matrix) of a land use map could in general be used when propagating uncertainty in e.g. a hydrogeological model that uses this map as one of the input layers to estimate evapotranspiration (ET). But in order for this propagation to make sense, the model must just give an overall ET estimate for the whole study area, i.e. it should be a non-distributed model, unless we are ready to assume that accuracy is randomly distributed throughout the map. A better alternative for propagating uncertainty due to land use data in a distributed model would be a spatially explicit error model, which unfortunately is not provided by mapmakers in current compilations. Nevertheless, if in the previous example we had the confusion matrix of the map, we could, having class-specific ET estimates, propagate uncertainty in the distributed model. However, in doing this we would again be relying on an unrealistic assumption, that is, that errors are randomly distributed in space.

Finally, in the sadly common case of a land use map for which accuracy information is lacking, a possible solution

would be, using the threshold proposed by Anderson et al. (1976), to grant the map an 85% overall accuracy, on the assumption that the agency that entrusted the map uses high standards that in turn were followed by the contractor, and apply it to all the cells within the model. If that solution can be regarded as reasonable by both managers and stakeholders, then the outcome of the uncertainty propagation exercise may well be wrong, but at least it will be *legitimate* (see Hofmann and Mitchell, 1998). In short, the key question when assessing uncertainty in land use maps is to what degree the map allows managers (or models) to make decisions (or computations) that do not differ significantly from those that they would have made had they a direct knowledge (or perfect map) of the mapped region. Current practice does not provide a full answer to this question since the reliability of a map can only be fully appreciated through metadata. Hence, a final recommendation for users is to compel map producers to compile well documented standard metadata files, on the grounds that no map is acceptable as input to a numerical model without them.

*Acknowledgements.* The present work was carried out within the Project “Harmonised Techniques and Representative River Basin Data for Assessment and Use of Uncertainty Information in Integrated Water Management (HarmoniRiB)”, which is partly funded by the EC Energy, Environment and Sustainable Development programme (Contract EVK1-CT2002-00109). This manuscript has also been generously supported by a University of Calgary Start-up Grant, an Alberta Ingenuity New Faculty Grant, and an NSERC Discovery grant to G. J. Hay. The opinions expressed here are those of the authors, and do not necessarily reflect the views of their funding agencies. Finally, the authors also wish to thank two anonymous reviewers for helpful comments.

Edited by: J. Freer

## References

- Anderson, J., Hardy, E., Roach, J., and Witmer, R.: A Land Use and Land Cover classification system for use with remote sensor data, U.S. Geological Survey, Professional Paper 964, Washington D.C., 1976.
- Bie, S. W. and Beckett, D. H. T.: Comparison of four independent soil surveys by air-photo interpretation, Paphos area (Cyprus), *Photogrammetria*, 29, 189–202, 1973.
- Blakemore, M.: Generalization and error in spatial databases, *Cartographica*, 21, 131–139, 1984.
- Brown, J. D., Heuvelink, G. B. M., and Refsgaard, J. C.: An integrated framework for assessing uncertainties in environmental data, *Water Sci. Technol.*, 52(6), 153–160, 2005.
- Burrough, P. A., Wilson, J. P., van Gaans, P. F. M., and Hansen, A. J.: Fuzzy k-means classification of topo-climatic data as an aid to forest mapping in the Greater Yellowstone area, USA, *Landscape Ecology*, 16, 523–546, 2001.
- Burnett, C. and Blaschke, T.: A multi-scale segmentation/object relationship modelling methodology for landscape analysis, *Ecol. Model.*, 168(3), 233–249, 2003.
- Buttner, G., Hajos, T., and Korandi, T.: Improvements to the effectiveness of supervised training procedures, *Int. J. Remote Sens.*, 10(6), 1005–1013, 1989.
- Castilla, G.: Object-oriented analysis of remote sensing images for land cover mapping: conceptual foundations and a segmentation method to derive a baseline partition for classification, Ph.D. Thesis, Polytechnic University of Madrid, [http://oa.upm.es/133/01/07200302\\_castilla\\_castellano.pdf](http://oa.upm.es/133/01/07200302_castilla_castellano.pdf), 2003.
- Castilla, G. and Hay, G. J.: Image objects and geographics objects, in: *Object-Based Image Analysis – Spatial concepts for knowledge-driven remote sensing applications*, edited by: Blaschke, T., Lang, S., and Hay, G. J., Springer-Verlag, Berlin, in press, 2007.
- Castilla, G., Hay, G. J., and Ruiz-Gallardo, J. R.: Size-constrained region merging (SCRM): an automated delineation tool for assisted photointerpretation, *Photogramm. Eng. Rem. S.*, in press, 2007.
- Cohn, A. G. and Gotts, N. M.: The “Egg-Yolk” Representation of Regions with Indeterminate Boundaries, in: *Geographic Objects with Indeterminate Boundaries*, edited by: Burrough, P. A. and Frank, A. U., 171–188, Taylor and Francis, 1996.
- Chrisman, N. R.: The error component in spatial data, in: *Geographical Information Systems: Principles and Applications*, edited by: Maguire, D. J., Goodchild, M. F., and Rhind, D. W., Longman Scientific and Technical, 1991.
- Cressie, N. A. C.: Change of support and the modifiable areal unit problem, *Geogr. Systems*, 3, 159–180, 1996.
- Davis, F. W. and Simonett, D. S.: GIS and remote sensing, in: *Geographic Information Systems: Principles and Applications*, edited by: Maguire, D. J., Goodchild, M. F., and Rhind, D., Longman, Essex, England, 191–210, 1991.
- Eckhardt, K., Breuer, L., and Frede, H. G.: Parameter uncertainty and the significance of simulated land use change effects, *J. Hydrol.*, 273, 164–176, 2003.
- European Union: Directive 2007/2/EC of the European Parliament and of the Council establishing an Infrastructure for Spatial Information in the European Community (INSPIRE), *Official Journal of the European Union* 50, 25 April 2007.
- Finke, P.A., Wladis, D., Kros, J., Pebesma, E. J., and Reinds, G. J.: Quantification and simulation of errors in categorical data for uncertainty analysis of soil acidification modelling, *Geoderma*, 93, 177–194, 1999.
- FAO: Africover land cover classification. Environment and Natural Resources Service, FAO, Rome, 1997.
- Frank, A. U.: *Ontology for Spatio-temporal Databases*, Lecture Notes in Computer Science, 2520, 9–77, 2003.
- Fukunaga, K.: *Introduction to Statistical Pattern Recognition*. Academic Press, New York, 1972.
- Foody, G. M.: Local characterization of thematic classification accuracy through spatially constrained confusion matrices, *Int. J. Remote Sens.*, 26(6), 1217–1228, 2005.
- Foody, G. M.: Status of land cover classification accuracy assessment, *Remote Sens. Environ.*, 80, 185–201, 2002.
- Gopal, S. and Woodcock, C.: Theory and methods for accuracy assessment of thematic maps using fuzzy sets, *Photogramm. Eng. Rem. S.*, 60(2), 181–188, 1994.
- Goodchild, M. F., Sun, G., and Yang, S.: Development and test of an error model for categorical data, *Int. J. Geogr. Inf. Syst.*, 6(2), 87–104, 1992.

- Goodchild, M. F.: Integrating GIS and remote sensing for vegetation analysis and modelling: methodological issues, *J. Veg. Sci.*, 5, 615–626, 1994.
- Green, D. R. and Hartley, S.: Integrating photointerpretation and GIS for vegetation mapping: some issues of error, in: *Vegetation Mapping: From Patch to Planet*, edited by: Millington, A. and Alexander, R., Wiley, 103–134, 2000.
- Hay, G. J. and Castilla, G.: Geographic Object-Based Image Analysis (GEOBIA): a new name for a new discipline, in: *Object-Based Image Analysis – Spatial concepts for knowledge-driven remote sensing applications*, edited by: Blaschke, T., Lang, S., and Hay, G. J., Springer-Verlag, Berlin, in press, 2007.
- Hofmann, N. and Mitchell, B.: The RESPECT model: evolving decision-making approaches in water management, *Water Policy*, 1, 341–355, 1998.
- Honeycutt, D. M.: Epsilon bands based on probability, *Proceedings of Autocarto-8*, 1987.
- Horttanainen, P. and Virrantaus, K.: Uncertainty evaluation of military terrain analysis results by simulation and visualization, *Proc. 12th Int. Conf. on Geoinformatics*, 474–480, Gävle, Sweden, 2004.
- Landgrebe, D.: Information extraction principles and methods for multispectral and hyperspectral image data, in: *Information Processing for Remote Sensing*, edited by: Chen, C. H., The World Scientific Publishing Co., New Jersey, 1999.
- Lillesand, T. M., Kiefer, R. W., and Chipman, J. W.: *Remote Sensing and Image Interpretation*, 5th Edition, John Wiley and Sons, New York, 2003.
- Lins, K.: Land use and land cover mapping in the United States: An overview and history of the concept, in: *Gap Analysis: A landscape approach to biodiversity planning*, edited by: Scott, M., Tear, T., and Davis, F., ASPRS, Bethesda, Maryland, 57–66, 1996.
- Marceau, D. J.: The scale issue in social and natural sciences, *Can. J. Remote. Sens.*, 25, 347–356, 1999.
- Markham, B. L. and Townshend, J. R. G.: Land cover classification accuracy as a function of sensor spatial resolution, *Proc. 15th Int. Symp. Remote Sens. Environ.*, Ann Arbor, 1075–1090, 1981.
- Mausel, P. W., Kramber, W. J., and Lee, J. K.: Optimum Band Selection for Supervised Classification of Multispectral Data, *Photogramm. Eng. Rem. S.*, 56(1), 55–60, 1990.
- MEGIS: Maine land cover map metadata, <http://megisims.state.me.us/metadata/melcd.htm>, 2006.
- Meinel, G. and Hennemersdorf, J.: Classifications systems of land cover and land use and their challenges for picture processing of Remote Sensing data – Status of international discussion and programs, *Proc. 3rd Int. Symp. Remote Sens. of urban areas*, Istanbul, 472–479, 2002.
- Næsset, E.: Use of the weighted Kappa coefficient in classification error assessment of thematic maps, *Int. J. Geog. Info. Sci.* 10, 591–604, 1996.
- Openshaw, S. and Taylor, P.J.: The modifiable areal unit problem, in: *Quantitative Geography: A British View*, edited by: Wrigley, N. and Bennett, R., Routledge, London, 60–69, 1981.
- Orzanco, M. G., Lowell, K., and Fortin, M. J.: Assessing the Spatial Uncertainty of Boundaries on Forest Maps Using an Ecological Similarity Index. 6th Int. Symp. Spatial Accuracy Assessment. Portland, Maine, USA 28 June– 1 July, 2004.
- Plourde, L. and Congalton, R.: Sampling method and sample placement: How do they affect the accuracy of remotely sensed maps? *Photogramm. Eng. Rem. S.*, 69(3), 289–297, 2003.
- Richards, J. A. and Jia, X.: *Remote Sensing Digital Image Analysis: An Introduction*, Springer, Berlin, 181–238, 1999.
- Schowengerdt, R. A.: *Remote Sensing, models and methods for image processing*, Second edition, Academic Press, London, 1997.
- Smith, J. H., Stehman, S. V., Wickham, J. D., and Yang, L.: Effects of landscape characteristics on land-cover class accuracy, *Remote Sens. Environ.*, 84, 342–349, 2003.
- Smits, P. C., Dellepiane, S. G., and Schowengerdt, R. A.: Quality assessment of image classification algorithms for landcover mapping: a review and a proposal for a cost-based approach, *Int. J. Remote Sens.*, 20(8), 1461–1486, 1999.
- Stehman, S. V. and Czaplewski, R. L.: Design and analysis for thematic map accuracy assessment: Fundamental principles, *Remote Sens. Environ.*, 64(3), 331–344, 1998.
- Stehman, S. V., Wickham, J. D., Smith, J. H., and Yang, L.: Thematic accuracy of the 1992 National Land-Cover Data (NLCD) for the eastern United States: statistical methodology and regional results, *Remote Sens. Environ.*, 86, 500–516, 2003.
- Story, M. and Congalton, R. G.: Accuracy assessment: a user's perspective, *Photogramm. Eng. Rem. S.*, 52, 397–399, 1986.
- Toutin, T.: Geometric processing of remote sensing images: models, algorithms and methods, *Int. J. Remote Sens.*, 25(10), 1893–1924, 2004.
- Woodcock, C. E. and Strahler, A. H.: The factor of scale in remote sensing, *Remote Sens. Environ.*, 21, 311–332, 1987.
- Wulder, M. A., Franklin, S. E., White, J. C., Linke, J., and Magness, S.: An accuracy assessment framework for large area land cover classification products derived from medium resolution satellite data, *Int. J. Remote Sens.*, 27(4), 663–683, 2006a.
- Wulder, M. A., White, J. C., Luther, J. E., Strickland, G., Rempel, T.K. and Mitchell, S. W.: Use of vector polygons for the accuracy assessment of pixel-based land cover maps, *Can. J. Remote Sens.*, 32(3), 268–279, 2006b.
- Zweig, M. H. and Campbell, G.: Receiver-Operating Characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine, *Clin. Chem.*, 39, 561–577, 1993.