

**ERROR AND UNCERTAINTY IN THE ACCURACY ASSESSMENT
OF LAND COVER MAPS**

by

Pedro Alexandre Reis Sarmiento

Dissertation submitted to the NOVA Information Management School in
partial fulfillment of the requirements for the Degree of

Doutor em Gestão de Informação – Sistemas de Informação Geográfica

(Doctor of Philosophy in Information Management – Geographic
Information Systems)

NOVA Information Management School

2015

**ERROR AND UNCERTAINTY IN THE ACCURACY ASSESSMENT
OF LAND COVER MAPS**

by

Pedro Alexandre Reis Sarmiento

Dissertation supervised by

Professor Mário Caetano

Professor Cidália Fonte

Professor Stephen V. Stehman

NOVA Information Management School

2015

DEDICATION

To my parents and grandmother.

ACKNOWLEDGEMENTS

To my parents for their unconditional support.

To Xana for all their love, care, motivation and patience during the period of my PhD studies and elaboration of this dissertation.

To Professor Mário Caetano for their friendship, ideas and lines of investigation that resulted in the work presented in this dissertation and for being my advisor in my graduation, master degree and PhD.

To Professor Cidália Fonte for all the ideas, support in the work developed during this PhD and also for their hospitality when I worked with her at University of Coimbra.

To Professor Stephen Stehman despite the distance, for all their guidance, advices and reviews, in the submission of this dissertation and the papers elaborated during this PhD.

To Artur Gil and my ex-colleagues of the extinguished Remote Sensing Unit of the Portuguese Geographic Institute, namely Hugo Carrão, António Nunes, António Araújo, Vasco Nunes, Maria da Conceição Pereira and Joel Dinis for all the friendship, motivation and work, that also contributed to this dissertation.

To my professors of the NOVA Information Management School, by everything I learned with them in classes.

To the directors that passed by the direction of the extinguished Portuguese Geographic Institute (now days *Direção-Geral do Território*) for providing me, all the conditions to develop my PhD studies.

The research presented in this dissertation was funded and/or partially supported by the following institutions:

- NOVA Information Management School (NOVA IMS)
- *Fundação para a Ciência e Tecnologia* (FCT) – SFRH/BD/61900/2009;
- *Direção-Geral do Território* (DGT);
- *Faculdade de Ciências e Tecnologia da Universidade de Coimbra* (FCTUC);
- *Instituto de Engenharia de Sistemas e Computadores de Coimbra* (INESC Coimbra).



RESUMO

Tradicionalmente a avaliação da exatidão temática de mapas de ocupação do solo é realizada através da comparação destes mapas com uma base de dados de referência, na qual se pretende representar a “verdadeira” ocupação do solo, sendo esta comparação reportada com os índices de exatidão temática através de matrizes de confusão. No entanto estas bases de dados de referência são também elas próprias uma representação da realidade, contendo erros devidos à incerteza humana na atribuição da classe de ocupação do solo que melhor caracteriza uma determinada área. Isto pode levar ao enviesamento dos índices de exatidão temática que são reportados aos utilizadores finais destes mapas. O principal objetivo desta dissertação é o desenvolvimento de uma metodologia que permita a integração da incerteza humana associada à criação de bases de dados de referência na avaliação da exatidão temática de mapas de ocupação do solo e analisar os impactos que esta incerteza pode ter nas medidas de exatidão temática que são reportados aos utilizadores finais de mapas de ocupação do solo. Investigámos a utilidade da inclusão da incerteza humana na avaliação da exatidão temática de mapas de ocupação do solo. Especificamente estudamos a utilidade da teoria fuzzy, mais concretamente da aritmética fuzzy, para uma melhor compreensão da incerteza humana associada à elaboração de bases de dados de referência e os seus impactos nos índices de exatidão temática que são derivados de matrizes de confusão. Para este propósito foram utilizadas escalas linguísticas transformadas em intervalos fuzzy que incorporam a incerteza na elaboração de bases de dados de referência, para elaborar matrizes de confusão fuzzy. A metodologia proposta é ilustrada num caso de estudo em que é realizada a avaliação da exatidão temática de um mapa de ocupação do solo para Portugal Continental derivado de imagens do sensor Medium Resolution Imaging Spectrometer (MERIS). Os resultados obtidos demonstram que a inclusão da incerteza humana nas bases de dados de referência proporciona muito mais informação acerca da qualidade de mapas de ocupação do solo, quando comparada com a abordagem tradicional de avaliação da exatidão temática de mapas de ocupação do solo.

Palavras-chave: ocupação do solo, avaliação da exatidão temática, bases de dados de referência, incerteza humana, matrizes de confusão, intervalos fuzzy, índices de exatidão temática.

ABSTRACT

Traditionally the accuracy assessment of land cover maps is performed through the comparison of these maps with a reference database, which is intended to represent the “real” land cover, being this comparison reported with the thematic accuracy measures through confusion matrixes. Although, these reference databases are also a representation of reality, containing errors due to the human uncertainty in the assignment of the land cover class that best characterizes a certain area, causing bias in the thematic accuracy measures that are reported to the end users of these maps. The main goal of this dissertation is to develop a methodology that allows the integration of human uncertainty present in reference databases in the accuracy assessment of land cover maps, and analyse the impacts that uncertainty may have in the thematic accuracy measures reported to the end users of land cover maps. The utility of the inclusion of human uncertainty in the accuracy assessment of land cover maps is investigated. Specifically we studied the utility of fuzzy sets theory, more precisely of fuzzy arithmetic, for a better understanding of human uncertainty associated to the elaboration of reference databases, and their impacts in the thematic accuracy measures that are derived from confusion matrixes. For this purpose linguistic values transformed in fuzzy intervals that address the uncertainty in the elaboration of reference databases were used to compute fuzzy confusion matrixes. The proposed methodology is illustrated using a case study in which the accuracy assessment of a land cover map for Continental Portugal derived from Medium Resolution Imaging Spectrometer (MERIS) is made. The obtained results demonstrate that the inclusion of human uncertainty in reference databases provides much more information about the quality of land cover maps, when compared with the traditional approach of accuracy assessment of land cover maps.

Keywords: land cover, accuracy assessment, reference databases, human uncertainty, fuzzy intervals, thematic accuracy measures.

LIST OF ABBREVIATIONS AND ACRONYMS

<i>C</i>	Centroid
CEOS/WGCV	Committee on Earth Observation Satellites/Working Group on Calibration & Validation
ESA	European Space Agency
<i>fom</i>	First of maxima
GOFC/GOLD	Global Observation for Forest and Land Cover Dynamics
<i>lom</i>	Last of maxima
MERIS	MEdium Resolution Imaging Spectrometer
NASA	National Aeronautics and Space Administration
<i>psu</i>	Primary sampling unit
RSU/IGP	Remote Sensing Unit of the Portuguese Geographic Institute
<i>ssu</i>	Secondary sampling unit

TABLE OF CONTENTS

DEDICATION.....	i
AKNOWLEDGEMENTS	iii
RESUMO.....	vii
ABSTRACT	ix
LIST OF ABBREVIATIONS AND ACRONYMS	xi
TABLE OF CONTENTS	xiii
LIST OF TABLES.....	xvi
LIST OF FIGURES	xviii
1 INTRODUCTION.....	1
1.1 Motivation.....	1
1.2 Overall research objectives	4
1.3 Dissertation organization	6
2 ACCURACY ASSESSMENT OF LAND COVER MAPS	7
2.1 Brief historical review on the accuracy assessment of remotely sensed data....	7
2.2 Basic components of the accuracy assessment of land cover maps.....	8
2.2.1 Sampling design	8
2.2.1.1 Sample frame.....	9
2.2.1.2 Sampling unit	10
2.2.1.2.1 Sampling units as points.....	12
2.2.1.2.2 Sampling units as areas	13
2.2.1.3 Sampling protocol	16
2.2.1.3.1 Simple random sampling.....	16
2.2.1.3.2 Systematic sampling.....	18
2.2.1.3.3 Stratified random sampling.....	21
2.2.1.3.4 Cluster sampling.....	23
2.2.1.4 Sample size.....	27
2.2.2 Response design	29
2.2.2.1 Evaluation protocol	30
2.2.2.2 Labeling protocol	31
2.2.3 Analysis and estimation protocol	33
2.2.3.1 Accuracy measures.....	35

2.2.3.1.1	Overall accuracy.....	35
2.2.3.1.2	User's accuracy.....	36
2.2.3.1.3	Producer's accuracy.....	36
2.3	Uncertainty of reference databases in the accuracy assessment of land cover maps	37
2.4	Challenges and future directions on the accuracy assessment of land cover maps	44
3	FUZZY SETS THEORY	47
3.1	Fuzzy sets.....	47
3.1.1	Core and support.....	48
3.1.2	Alpha-levels.....	48
3.1.3	Fuzzy quantities.....	48
3.2	Fuzzy arithmetic.....	49
3.3	Defuzzification.....	51
4	ADDRESSING REFERENCE DATA UNCERTAINTY IN THE ACCURACY ASSESSMENT OF LAND COVER MAPS THROUGH FUZZY SETS	52
4.1	Linguistic scales.....	53
4.2	Conversion of linguistic scales into fuzzy intervals	54
4.2.1	Ideal case	54
4.2.2	Interpreter-derived	57
4.2.2.1	Selection of a set of control sites.....	57
4.2.2.2	Assigning reference information to the control sites	57
4.2.2.3	Modeling the linguistic values with fuzzy intervals.....	58
4.3	Fuzzy confusion matrices	62
4.4	Fuzzy accuracy measures.....	62
5	ADDRESSING REFERENCE DATA UNCERTAINTY IN THE ACCURACY ASSESSMENT OF A MAP OF CONTINENTAL PORTUGAL	69
5.1	Data, nomenclature and map classification	69
5.2	Definition of the linguistic scales	71
5.3	Conversion of linguistic scales into fuzzy intervals	73
5.3.1	Selection of a set of control sites.....	73
5.3.2	Assigning reference information to the control sites.....	73
5.3.3	Modeling the linguistic values with fuzzy intervals.....	75
5.4	Elaboration of the reference databases	77

5.5	Accuracy assessment with the ideal and interpreter-derived fuzzy intervals ..	79
5.6	Comparison of the fuzzy accuracy measures with traditional accuracy measures.....	86
5.7	Comparison of the fuzzy accuracy measures between photo-interpreters.....	90
6	CONCLUSION	97
	REFERENCES	103

LIST OF TABLES

Table 2.1 - Useful concepts and results to guide selection of an accuracy assessment spatial unit (Source: Stehman and Wickham 2011).	11
Table 2.2 - Characteristics of the sampling designs used in the accuracy assessment of land cover maps (Source: Stehman (1999)).	26
Table 5.1 - Generalization of the LANDEO nomenclature.....	71
Table 5.2 - Linguistic values and respective order. Land cover coverage defined for each linguistic value of the linguistic scales with five and seven values.	72
Table 5.3 - Example of the linguistic values collected in a control site by the photo-interpreter and the land cover percent collected by the chief photo-interpreter. U: "Understandable but wrong"; G: "Good"; AW: "Absolutely wrong"; W: "Wrong".....	75
Table 5.4 - Land cover proportion intervals obtained for the core and support of the ideal case and interpreter-derived fuzzy intervals with five linguistic values.....	75
Table 5.5 - Land cover proportion intervals obtained for the core and support of the ideal case and interpreter-derived fuzzy intervals with seven linguistic values.....	76
Table 5.6 - Extract of the reference database with eight sample sites (n=8); five land cover classes, namely urban areas (UA), agriculture (AG), natural vegetation (NV), forest (F) and water and wetlands (WW); and five linguistic values: "Wrong" (W), "Understandable but wrong" (U), "Reasonable or acceptable" (A), "Good" (G), "Right" (R).....	78
Table 5.7 - Fuzzy confusion matrix for the photo-interpreter elaborated with the fuzzy intervals for the ideal case with five linguistic values.....	80
Table 5.8 - Fuzzy confusion matrix for the photo-interpreter elaborated with the interpreter-derived fuzzy intervals with five linguistic values.	81
Table 5.9 - Fuzzy confusion matrix for the photo-interpreter elaborated with the fuzzy intervals for the ideal case with seven linguistic values.....	82
Table 5.10 - Fuzzy confusion matrix for the photo-interpreter elaborated with the interpreter-derived fuzzy intervals with seven linguistic values.	83
Table 5.11 - Support and core intervals of the fuzzy thematic accuracy measures obtained with the ideal and empirical fuzzy intervals using five linguistic values and seven linguistic values. Thematic accuracy measures obtained with the centroid (C)...	84

Table 5.12 - Defuzzified user's accuracy in percentage, obtained with first of maxima (<i>fom</i>), centroid, last of maxima (<i>lom</i>) and traditional user's accuracy obtained with MAX and RIGHT.....	87
Table 5.13 - Defuzzified producer's accuracy in percentage, obtained with first of maxima (<i>fom</i>), centroid, last of maxima (<i>lom</i>) and traditional user's accuracy obtained with MAX and RIGHT.....	88
Table 5.14 - Defuzzified overall accuracy in percentage, obtained with first of maxima (<i>fom</i>), centroid, last of maxima (<i>lom</i>) and traditional user's accuracy obtained with MAX and RIGHT.....	89
Table 5.15 - Land cover proportion intervals obtained for the core and support of photo-interpreter 1 and photo-interpreter 2 with the interpreter-derived fuzzy intervals.....	90
Table 5.16 - Fuzzy confusion matrix and fuzzy accuracy measures elaborated with the interpreter-derived fuzzy intervals of photo-interpreter 1.....	92
Table 5.17 - Fuzzy confusion matrix and fuzzy accuracy measures elaborated with the interpreter-derived fuzzy intervals of photo-interpreter 2.....	93
Table 5.18 - Support and core intervals of the fuzzy thematic accuracy measures obtained with the interpreter-derived fuzzy intervals using seven linguistic values for each photo-interpreter. Thematic accuracy measures obtained with the centroid (<i>C</i>), MAX and RIGHT.....	94

LIST OF FIGURES

Figure 2.1 - Example of the distribution of the sampling units over a study area using a simple random sampling (Source: Congalton and Green 1999).....	17
Figure 2.2 - Example of the distribution of the sampling units over a study area using a systematic sampling (Source: Congalton and Green 1999).....	18
Figure 2.3 - Example of the distribution of the sampling units over a study area using a stratified random sampling (Source: Congalton and Green 1999).....	21
Figure 2.4 - Example of the distribution of the sampling units over a study area using cluster sampling (Source: Congalton and Green 1999).....	24
Figure 2.5. Representation of a confusion matrix. (Adapted from Congalton and Green (2009)).	34
Figure 3.1 - (a) A trapezoidal fuzzy interval (the core of the fuzzy interval is the interval [2,3] and the support is defined by the interval]1,4[); (b) A triangular fuzzy number (the core of the fuzzy interval contains only one element of R, the value 6, and the support is defined by the interval]5,7[).	49
Figure 4.1 - Example of the elaboration of five fuzzy intervals that correspond to an ideal case with five linguistic values.	56
Figure 4.2 - Example of the elaboration of the five fuzzy intervals that correspond to the ideal case with seven linguistic values. The linguistic values “Absolutely wrong” and “Absolutely right” are represented by the black dots corresponding respectively to the crisp numbers 0 and 1.	56
Figure 4.3 - Example of the elaboration of the five fuzzy intervals when intervals overlap. Intervals (a), (b) and (c) correspond to the intervals obtained respectively for "Understandable but wrong", "Reasonable or acceptable" and "Good", corresponding to intervals [0.11, 0.44] and [0.33, 0.67] and [0.56, 0.88].	60
Figure 4.4 - Example of the elaboration of the five fuzzy intervals when intervals do not overlap. Intervals (a), (b) and (c) correspond to the intervals obtained respectively for "Understandable but wrong", "Reasonable or acceptable" and "Good", corresponding to intervals [0.22, 0.33], [0.44, 0.56] and [0.68, 0.78].	60
Figure 4.5 - Fuzzy interval that intersects another fuzzy interval that is not adjacent. The fuzzy interval “Reasonable or acceptable” intersects the fuzzy intervals “Understandable but wrong” and “Good” and also the fuzzy interval “Wrong”.	61

Figure 4.6 - Core overlap of two fuzzy intervals. The core of the fuzzy interval “Reasonable or acceptable” (dashed line) is partially overlapped with the core of the fuzzy interval “Understandable but wrong” (bold solid line).....	61
Figure 4.7 - Flowchart of the proposed methodology of accuracy assessment of land cover maps.....	68
Figure 5.1 - Map for Continental Portugal with 5 land cover classes.	70
Figure 5.2 - Control site overlaid with 100 points systematically distributed.....	74
Figure 5.3 - Interpreter-derived fuzzy intervals obtained using a linguistic scale with five linguistic values (dashed line) and fuzzy intervals for an ideal case with five linguistic values (solid line).....	75
Figure 5.4 - Interpreter-derived fuzzy intervals obtained using a linguistic scale with seven linguistic values (dashed line) and fuzzy intervals for an ideal case with seven linguistic values (solid line).....	76
Figure 5.5 - Sample site of the reference database overlaid with the aerial image.	78
Figure 5.6 - Interpreter derived fuzzy intervals obtained for photo-interpreter 1 (solid line) and photo-interpreter 2 (dashed line), using a linguistic scale with seven linguistic values. The linguistic values "Absolutely wrong" and "Absolutely right" are represented by the black dots corresponding respectively to the crisp numbers 0 and 1.	90

1 INTRODUCTION

1.1 Motivation

Land cover is a fundamental variable for the comprehension of the interaction and impacts that human activity has in the environment. At no other time in history of human civilization, man printed so deep changes in land cover, and therefore it is necessary to understand the influence those changes can have in our life style, as well as in our survival as humanity. Accordingly to Lesschen et al. (2005), land cover change is the result of the interaction between social economic, institutional and environmental factors. Land cover change is referred to as the most important variable in the alteration of the ecological system at global level (Vitousek 1994); as well as a determinant factor in the alteration of the climatic system (Brovkin et al. 2004); since it is the variable with major impact in biodiversity in the next 100 years (Chapin et al. 2000).

Although, to monitor land cover at global scale, in short time and with costs that enterprises can support, it is necessary to resort to remote sensing satellite images. Indeed, Foody (2002) refers that satellite images are an attractive source of data for the production of land cover maps, that allow a spatially continuous representation of the Earth's surface at several spatial and temporal scales. For these reasons, remote sensing data are nowadays a growing source of information to the production of land cover maps (Caetano et al. 2006).

Despite the need for the production of land cover maps, it is also important to verify how these maps represent reality. This is an aspect that is many times ignored, but it can influence the decisions taken by the end users of these maps. If decisions are taken based on this information, its better or worst quality will also influence the quality of

Introduction

the decisions taken. For these facts, before land cover maps are used in scientific research and decision making processes, they should be subject to a rigorous assessment of their accuracy (Stehman and Czaplewski 1998).

Traditionally, the accuracy assessment of land cover maps is held through the comparison of the produced maps with a reference database. This comparison is made for a set of points or areas and is represented in a confusion matrix (Story and Congalton 1986). Generally the reference data is represented in the matrix columns, which are compared with the map data, represented in the matrix rows. The values in the matrix diagonal represent the agreement between the two data sets. The confusion matrix, for allowing an individual analysis of thematic accuracy by land cover class, is considered a very valid method to represent the accuracy of land cover maps (Congalton and Green 1999). It has been recommended by many researchers to represent accuracy and should be adopted as the standard report convention (Congalton, 1991).

Although, in the traditional methods of accuracy assessment of land cover maps several aspects are assumed that limit the quality of the accuracy assessment. Accordingly to Gopal and Woodcock (1994) and Woodcock and Gopal (2000), it is assumed that only one land cover class could be assigned to each sample point or sample area, being assumed that this assignment is made with full certainty. For the elaboration of reference databases the technician has to, in field visits or photo-interpretation of aerial images, choose a land cover class for each location, even though more than one land cover class may exist in that same location (Gopal and Woodcock 1994). This aspect occurs mainly in areas of transition between land cover classes and because land cover presents a *continuum* natural, and therefore rarely presents abrupt transitions between classes. For these reasons, land cover maps and reference databases have uncertainty and it is not likely that two persons achieve two equal classifications (Goodchild 2003).

Introduction

To avoid this problem, when the reference database is elaborated, more than one land cover class can be collected at each sample observation (Woodcock et al. 1996; Edwards et al. 1998; Zhu et al. 2000; Wickham et al. 2004).

Besides the referred methods to include uncertainty in the elaboration of reference databases, Gopal and Woodcock (1994) developed another approach based on fuzzy sets that allow the inclusion of uncertainty of reference databases in the accuracy assessment of land cover maps. Gopal and Woodcock (1994), developed a linguistic scale for the technicians to describe their perception about land cover classes at each sample observation of the reference database through field visits. With this approach the technician is not limited to just one right or wrong answer at each sample observation, contrary to the traditional methods of accuracy assessment of land cover maps. Some examples of the application of this approach are present in Gill et al. (2000); Ma et al. (2001); Gilmore et al. (2008); Laba et al. (2008).

The research described in this dissertation takes in consideration the lines of investigation defined as priority by the European Space Agency (ESA), through initiatives like the Global Observation for Forest and Land Cover Dynamics (GOFD/GOLD), in which is also promoted the collaboration and cooperation with the Committee on Earth Observation Satellites-Working Group on Calibration & Validation (CEOS/WGCV) inserted in the National Aeronautics and Space Administration (NASA). Some of the priority activities of both initiatives are to promote the international cooperation in the development of calibration and validation methodologies of Earth observation data and develop an operational program of validation of global land cover maps. Concerns about Land Cover Map validation are also referred in the manual "Global Land Cover Validation: Recommendations for Evaluation and Accuracy Assessment of Global Land Cover Maps" (Strahler et al.

2006), elaborated by the (CEOS/WGCV), which indicates areas in which research needs to be developed.

1.2 Overall research objectives

The research developed aims to improve the methodologies of accuracy assessment of land cover maps, continuing the research conducted in the extinguished Remote Sensing Unit of the Portuguese Geographic Institute (RSU/IGP) (Sarmiento et al. 2008a; Sarmiento et al. 2008b; Sarmiento et al. 2009). The main objective is to develop a methodology that allows the integration of human uncertainty present in reference databases in the accuracy assessment of land cover maps, and analyse the impacts that uncertainty may have in the thematic accuracy measures reported to the end users of land cover maps.

The methodology developed in this research is based on three steps to assess the applicability of fuzzy sets to address human uncertainty of reference databases used for the accuracy assessment of land cover maps. The methodology was applied to a case study where an accuracy assessment of a land cover map for Continental Portugal was performed. The main problems that human uncertainty of reference databases can cause in the accuracy assessment of land cover maps are reviewed and well as their impacts in thematic accuracy measures. The potential of the use of fuzzy numbers and intervals, together with fuzzy arithmetic, to address uncertainty of reference databases in the accuracy assessment of land cover maps is investigated.

In the first step a methodology was developed to deal with the uncertainty of reference databases providing a fuzzy accuracy assessment. The methodology consists in the use of a five value linguistic scale. This linguistic scale was then converted in fuzzy intervals that translate what is supposed to be the ideal perception of a photo-interpreter

Introduction

about land cover coverage at each sample observation of a reference database to collect the reference data, and through fuzzy arithmetic was possible to elaborate fuzzy confusion matrices, where fuzzy accuracy measures similar to the traditional user's, producer's and overall accuracy were derived from it. These fuzzy accuracy measures were then defuzzified, allowing a direct comparison with the traditional crisp accuracy measures.

Although, this ideal perception could not correspond to the real behavior of a photo - interpreter relative to land cover coverage at each sample observation of a reference database. To analyze if this was the case, a second step was carried out where a methodology was developed to elaborate interpreter-derived fuzzy intervals based on linguistic scales with five and seven values. These interpreter-derived fuzzy intervals were then introduced in fuzzy confusion matrices, and fuzzy accuracy measures were derived from it based on the methodology developed in the first step. A comparison between the ideal and interpreter-derived fuzzy accuracy measures was made.

To test the differences in the perception of land cover coverage at each sample observation and the impacts that these differences have in the fuzzy accuracy measures, a third step was carried out. In this study interpreter-derived fuzzy intervals from two photo-interpreters were elaborated following the methodology developed in the second step. These interpreter-derived fuzzy intervals for the two photo-interpreters were then introduced in fuzzy confusion matrices, and fuzzy accuracy measures were computed, allowing a comparison between the accuracy results achieved by the two photo-interpreters. The purpose is to understand the impacts that human uncertainty have in accuracy assessment of land cover maps and whether this methodology provides more useful information about map errors, than the traditional approach of accuracy assessment of land cover maps.

1.3 Dissertation organization

This dissertation is divided in six chapters. In Chapter 1 an introduction is made to the problem to be analyzed and the motivation to develop this dissertation is provided. In Chapter 2 the essential steps to perform a traditional accuracy assessment are described. The fundamental notions and definitions of fuzzy sets theory used in the work are presented in Chapter 3. The methodological strategy for addressing the uncertainty of reference databases in the accuracy assessment of land cover maps is presented in Chapter 4. In Chapter 5 the application of the proposed methodology to a land cover map for Continental Portugal is presented, and finally in Chapter 6 the conclusions of the studies described in this dissertation are presented.

2 ACCURACY ASSESSMENT OF LAND COVER MAPS

Traditionally the accuracy assessment of land cover maps is made through the comparison of the produced maps with a reference database which is supposed to represent the 'true' land cover. This comparison is made for a set of sampling units and represented in a confusion matrix (Story and Congalton 1986). In this chapter a description is made of the several steps that should be taken into account to perform a traditional accuracy assessment of land cover maps.

2.1 Brief historical review on the accuracy assessment of remotely sensed data

The accuracy assessment of remotely sensed data was not always considered an important requisite to accompany the remotely sensed data provided to the end user's. Indeed, the accuracy assessment of remotely sensed data just begun since 1975 and before this date the maps derived from remotely sensed data were rarely subjected to any kind of quantitative accuracy assessment (Congalton 2004). Just when photo-interpretation began to be used as reference data to compare with the produced maps, the concerns about the quality of maps derived from remotely sensed data arise. Congalton (2004) divided the history of accuracy assessment into four developmental ages:

- 1st age: in this period no real accuracy assessment was performed, instead a "it looks good" mentality about the maps quality prevailed;

- 2nd age or age of non site-specific assessment: at this period the overall areas were compared between ground estimates and the map without regard for location (Meyer et al. 1975);
- 3rd age or age of the site-specific assessment: at this period the overall accuracy was computed, comparing the actual places on the ground to the same place in the map;
- 4th age or age of the confusion matrix: this period correspond to the present age and the confusion matrix is at their core. In this epoch various analysis techniques were developed based on the confusion matrix that has been accepted as the standard descriptive reporting tool for accuracy assessment of remotely sensed data (Congalton 2001).

2.2 Basic components of the accuracy assessment of land cover maps

According to Stehman and Czaplewski (1998), the three basic components of an accuracy assessment are: 1) the sampling design used to select the reference sample; 2) the response design used to obtain the reference land cover classification for each sampling unit; and 3) the estimation and analysis procedures. In the following sections a description is made of the three basic components of an accuracy assessment.

2.2.1 Sampling design

The sampling design is the protocol by which the reference sample units are selected (Stehman and Czaplewski 1998). The selection of a sample design requires a consideration of the specific objectives of the accuracy assessment and a prioritized list of desirable design criteria, being the most critical recommendation that the sampling design should be a probability sampling design (Olofsson et al. 2014). Choosing a probability sampling design also requires defining a sample frame, along with the

sampling unit which form the basis of an accuracy assessment. Stehman (1999) proposed seven major criteria that should be used as guidance to choose a sampling design for accuracy assessment: 1) the sampling protocol satisfies the requirements of a probability sampling design; 2) the sampling design must be practical; 3) the design must be cost effective; 4) the sample is spatially well distributed; 5) the sampling variability of the accuracy estimates should be small; 6) sampling variability or precision of the accuracy estimators should be estimated without undue reliance on approximations other than those related to sample size; 7) ability to accommodate a change in sample size at any step in the implementation of the design. Besides the criteria to choose the sampling design for the accuracy assessment of a given application, Stehman (2009) defined three questions that should be answered to determine the sampling design: 1) are pixels selected as individual entities or are they grouped into clusters and selected in these clusters?; 2) are the pixels or clusters grouped into strata?; 3) is the sample selection protocol applied to the pixels or clusters (possibly grouped into strata) simple random or systematic? In the following subsections the fundamental concepts to define the sampling design for the accuracy assessment of land cover maps are presented.

2.2.1.1 Sample frame

In statistics a sampling frame is the source material or device from which a sample is drawn (Särndal et al. 1992). There are two types of sample frames: 1) list frames; and 2) area frames. A list frame consists of a list of all sampling units, for example either pixels or mapped polygons, in the target region (Stehman and Czaplewski 1998) and the sample is selected directly from this list of sampling units. The sampling protocol used with an area frame is based on first selecting a sample of spatial locations, followed by associating a sampling unit with each sample location (Stehman and Czaplewski 1998).

The sets of elements drawn with the aid of an area frame are often called clusters and in a secondary selection step, the selected clusters may be sub sampled (Särndal et al. 1992). An area frame is preferable to a list frame when a systematic design is planned and if the area frame is a map of all pixels, converting the map to a one-dimensional list frame of pixels, it would lose much of the spatial structure important for systematic sampling (Stehman and Czaplewski 1998).

2.2.1.2 Sampling unit

The sampling unit in the accuracy assessment of land cover maps, could be defined as the comparison unit between the reference data and the map data. It will be through the comparison between the map classification and the reference data classification that will be obtained the thematic accuracy of the produced maps. The sampling unit is the fundamental unit in which the accuracy assessment is based i.e., is the link between the spatial location in the map and the correspondent location in the terrain (Stehman and Czaplewski 1998).

Accordingly with Stehman and Czaplewski (1998), there are two types of sampling units that could be implemented in the accuracy assessment of land cover maps: 1) sampling units under the form of points, and 2) sampling units under the form of areas, which may present various extensions and forms. The sampling units implemented as areas could have, for example, the same dimension of an image pixel, or be defined as polygons containing several pixels (e.g. block of pixels, or polygons).

The sampling unit should be defined before the conception of the sampling protocol. The several sampling protocols that exist, have different characteristics, and should be used taking into account the sampling units that have been adopted (Stehman and Czaplewski, 1998). Indeed Stehman and Wickham (2011) stated that the choice of the

appropriate sampling unit should be considered in terms of the ramifications of the three basic components of the accuracy assessment of land cover maps (i.e. the sampling design, the response design and the analysis and estimation protocol), providing also useful concepts that should be considered to guide the selection of the appropriate sampling unit. The referred concepts are presented in Table 2.1.

Table 2.1 - Useful concepts and results to guide selection of an accuracy assessment spatial unit (Source: Stehman and Wickham 2011).

<p>Defining the population and accuracy parameters</p>	<ul style="list-style-type: none"> • The population that determines the values of the accuracy parameters targeted by the assessment is conceptualized as resulting from overlaying a complete coverage reference classification (i.e. a reference map) and the target map to be evaluated. The population may be viewed as a difference map resulting from this overlay showing where the map and reference classifications agree and where they disagree. The corresponding areas of class-specific agreement and class-specific disagreement may be obtained from the difference map. The difference map will depend on the partition of the region of interest (ROI) created from the spatial unit chosen for the assessment. Different populations will yield different values of the accuracy parameters (e.g. different overall accuracy values) • For an area-based accuracy assessment, the per-class area of agreement and area of disagreement are summarized by a population error matrix where the cells of the error matrix represent the proportion of area of agreement for each class and the proportion of area misclassified for each type of error. The population and accuracy parameter values obtained from the population error matrix will differ depending on the choice of the spatial unit. Changing the spatial unit changes the population and consequently the results of the accuracy assessment. • Pixels, blocks of pixels and polygons are all arbitrary spatial units. The validity of an accuracy assessment does not depend on whether the spatial assessment unit is a real surface feature of the earth but instead depends on the area representation of agreement and disagreement resulting from use of the spatial unit to partition the ROI. • The population and values of the accuracy parameters are not affected by the choice of the sampling design.
<p>Response design - determining the reference classification and definition of agreement</p>	<ul style="list-style-type: none"> • Heterogeneity within the spatial unit requires specification of how heterogeneity will be accommodated in the labeling protocol and definition of agreement. These decisions will have a substantial impact on the population and values of accuracy parameters. As a general rule, the likelihood of heterogeneity within a spatial unit will increase with the size of the unit. • Map polygons do not necessarily represent real earth surface features and use of map polygons does not support reporting of accuracy results for different levels of class aggregation. The potentially large variation in polygon size creates practical challenges to the response design protocol and also motivates consideration of stratifying the sampling design by polygon size. Little research has been conducted exploring how variation in polygon size affects reference data collection, sampling and analysis. • Polygons defined by the reference classification ("reference polygons") are appealing because they represent real objects of interest but present practical challenges when constructing the response and sampling designs. Methodological developments are needed before reference polygons can be considered a viable option.
<p>Sampling design and estimation</p>	<ul style="list-style-type: none"> • The sampling design requires specifying the universe of all spatial units forming a partition of the ROI. The spatial units making up the universe must be non-overlapping and spatially exhaustive. • For a stratified design, each spatial unit must be assigned to one and only stratum. Block assessment units may be internally heterogeneous and this complicates the protocol for assigning each block to a stratum. • Polygons are less amenable than pixel or block assessment units for the purpose of cluster sampling because polygons have no natural grouping or nested structure to form clusters.

	<ul style="list-style-type: none"> • It is possible to construct an unbiased or a consistent estimator for any accuracy parameter of interest when a probability sampling design is implemented and working within the design-based inference framework. Consequently, it is pointless to consider comparing sampling designs on the basis of bias or consistency of estimators derived for different designs.
	<ul style="list-style-type: none"> • Different sampling designs result in different precision (i.e. variance) for the estimator of a given accuracy parameter, so it is meaningful to compare sampling designs on the basis of variance of the accuracy estimators.
Location or registration error (map and reference data are not spatially aligned)	<ul style="list-style-type: none"> • Registration errors affect all spatial units. Spatial misregistration can be conceptualized as a "halo" around the assessment unit.
	<ul style="list-style-type: none"> • The ultimate impact of location error is observed in the change in the population values of the accuracy parameters relative to the parameter values of a population determined by overlaying perfectly spatially aligned reference and target maps (i.e. location error is absent).
	<ul style="list-style-type: none"> • Pixel-based assessments are generally more sensitive to location error than are block and polygon-based assessments as evidenced by larger changes in the values of the accuracy parameters between spatially aligned and spatially unaligned reference and target maps. However location error, can still have a considerable effect on accuracy results when block or polygon units are used.

2.2.1.2.1 Sampling units as points

Statistically, the main distinction between the sampling units as points and areas, is that the first one is viewed as a continuous population and the second one as a discrete population. A continuous population avoids the difficulty in the interpretation of a sampling reference database based on pixels (Moisen et al. 1994). When points are used as the sampling units, the response design may evaluate a spatial extent larger than just the point location to obtain the reference classification at that point, but the comparison between the map and reference classification remains on a per point basis (Stehman and Czaplewski, 1998). Points could also be used to select a sample when it is impractical to construct a list frame. Stehman and Wickham (2011), provide an example where the spatial unit is a polygon defined by the reference classification where a list frame of those polygons is not available. To obtain a sample of reference polygons using point sampling, the polygons that contain the sample point locations are selected. Only these reference polygons would need to be identified, so it is no longer necessary to create a list of all reference polygons comprising the population (Stehman and Wickham 2011).

2.2.1.2.2 Sampling units as areas

A sampling unit defined as an area may be of three types: pixels, polygons or clusters. Any of these three types of sampling units divide the population in a finite number of discrete units (Stehman and Czaplewski, 1998). Both pixels and polygons, correspond to structures used in Geographic Information Systems (GIS) to represent a land cover class, while clusters do not necessarily possess that correspondence, i.e. may represent more than one land cover class.

The classification contained in pixels is derived from the automatic classification of satellite images, and are units homogeneous in size and shape. In fact, this structure, also designated as matrix model, is widely used in GIS, namely in mathematical operations that evolve several layers of information, in which each pixel contains just one value. The larger or smaller number of pixels is usually related with the spatial resolution of satellite images. The satellites with high spatial resolution (e.g. IKONOS, QUICKBIRD), possess spatial resolutions in the order of meters, i.e. represent small areas, but divide the population in a finite and high number of sampling units, being for this reason more related with sampling with points (Stehman and Czaplewski, 1998). On the other hand, satellites with medium spatial resolution (e.g. MERIS, MODIS), possess spatial resolutions in the order of hundreds of meters, i.e. divide the population in larger areas, but in a more reduced number of sampling units, being for this reason more related with sampling with areas.

Congalton and Green (2009) indicate three reasons why a single pixel is a poor choice for the sampling unit: 1) a pixel is an arbitrary rectangular delineation of the landscape that have little relation to the actual delineation of land cover; 2) it is almost impossible to exact align one pixel on the map to the same exact area in the reference data: and 3)

few classifications specify the pixel as the minimum mapping unit, and in this sense if the minimum mapping unit is larger than a single pixel, then a single pixel is inappropriate as the sampling unit.

However Stehman and Wickham (2011) stated that one option to cope with the different mapping unit between the map classification and the final map is to divide the study area into square blocks, where the area of each block is equivalent to the minimum mapping unit (e.g. a 9 pixel minimum mapping unit would translate to a 3x3 pixel block unit). A problem related with the choice of a block of pixels as the sampling unit is that generally it will contain a mix of land cover classes and this heterogeneity will affect the response design, sampling design and analysis (Stehman and Wickham 2011).

Despite a block of pixels is also an arbitrary representation of landscape like the pixel (Congalton and Green 2009), Stehman and Wickham (2011) stated that the results of an accuracy assessment does not depend on whether the sampling unit is a meaningful entity, but instead depends on whether the area representation portrayed by the population error matrix is meaningful. In an accuracy assessment perspective the problem is not related on how the sampling unit represents landscape but instead if the sampling unit preserves the areas of agreement and disagreement between the map data and reference data, and in this sense a pixel or a block of pixels could be used as the sampling unit. Unlike the pixel, a block of pixels minimizes registration problems, because it is easier to locate on the reference data or in the field (Congalton and Green 2009).

Polygons as sampling units are different from pixels, since generally they are different in shape and size. Each polygon represents an homogeneous area of a land cover class in the classified image (polygon), or may be identified through aerial images (photo-

interpreted polygon) (Stehman and Czaplewski, 1998). This type of sampling unit is related with the vector model. Polygons should be used as sampling unit if the map to be assessed is a polygon map (Congalton and Green 2009).

According to Stehman and Wickham (2011), polygons are viewed as more natural on the basis that they represent real features of the landscape. However this vision is suitable if the reference polygons are defined by the reference classification but not if polygons are defined by the map classification (Stehman and Wickham 2011). This is an important aspect because generally the polygons chosen for the accuracy assessment are based on the map classification and polygons could be added, deleted or modified prior the final version of the map. Indeed this aspect can cause confusion if the accuracy assessment polygons are collected during the initial training, data/calibration, field work (Congalton and Green 2009). In this sense if the polygon assessment unit is different from the polygon that is been assessed on the map, such a polygon will not be relevant to the objective of assessing accuracy of the final map (Stehman and Wickham 2011). However it is difficult to implement an accuracy assessment based on reference polygons as the sampling unit, because it is not feasible to construct a complete list frame of the universe of reference polygons because a census of the reference classification would be required (Stehman and Wickham 2011). Clusters are different from the two types of sampling units mentioned above, due to not being necessarily related to an homogeneous area of a land cover class classified automatically or a photo-interpreted digitized polygon. Generally, clusters possess a regular shape and area. There are several examples of the use of this kind of sampling units in the accuracy assessment of land cover maps (Wickham et al. 2004; Stehman et al. 2003; Wulder et al. 2006). One of the main advantages of clusters is that can reduce costs dramatically because travel time and/or setup time is decreased (Congalton and Green 2009). Unlike

the block of pixels, the polygons contained by a cluster of polygons could be considered a single sampling unit because polygons by definition separate map class types that have more between than within variation (Congalton and Green 2009).

2.2.1.3 Sampling protocol

Define which sampling method to use, is one of the most complex and important steps in the accuracy assessment of land cover maps (Wulder et al. 2006, Congalton and Green, 1999). Sampling is the protocol in which sampling units are selected (Stehman and Czaplewski, 1998). Accordingly with Stehman (2001), the sampling protocol and the analysis of their components are directly related with statistic inference, and for that reason motivate the proposed criteria for statistics rigor. The same author refers that a sampling protocol is the one in which the inclusion of probabilities are known for all the elements in the sample, and the probability is different from zero for all the elements of the population. A probabilistic sampling requires that all included probabilities are higher than zero and that all probabilities should be known for all the sampling units over the study area. Without this inclusion and knowledge of the probability it is not possible to obtain a statistically valid accuracy assessment of land cover maps (Stehman and Czaplewski, 1998).

There are five sampling designs widely used that could be implemented for the collection of reference data: simple random sampling, stratified random sampling, cluster sampling, systematic sampling and unaligned systematic sampling (Stehman and Czaplewski, 1998; Congalton and Green, 1999).

2.2.1.3.1 Simple random sampling

In a simple random sampling (Figure 2.1) each sampling unit over the study area has the same probability of being selected (Congalton and Green, 1999).



Figure 2.1 - Example of the distribution of the sampling units over a study area using a simple random sampling (Source: Congalton and Green 1999).

Stehman (1999), refers that this method is very easy to implement because the statistical estimators are much less complex when compared with other sampling designs. This sampling method presents also a high flexibility, adapting easily to the need of increasing or decreasing the number of sampling units, due to some over or under estimation of the costs of elaboration of the sampling (Stehman 1999). Other advantage of this method is the possibility the sampling units have to be collected simultaneously (Congalton and Green, 1999).

Foody (2002) refers that the simple random sampling could be appropriate for the accuracy assessment of land cover maps if the sampling size is high enough to guarantee that all land cover classes are appropriately represented. The adoption of a simple random sampling could also be useful to respond to the needs of a wide group of users (Stehman and Czaplewski 1998), despite the objectives of all users could not be anticipated (Stehman et al. 2000).

Although, Congalton and Green (1999) refer that simple random sampling have the disadvantage of underestimate the land cover classes less representative, that could be also important, unless there is a significantly increase of the number of sampling units.

Sampling a map with an adequate number of sampling units for each land cover class is essential to determine the overall accuracy of the maps (Rosenfield et al. 1982). Stehman (1999) makes reference to the fact that simple random sampling is not well spatially distributed, and if this is one of the criteria to perform an accuracy assessment should choose a systematic sampling or a stratified random sampling.

2.2.1.3.2 Systematic sampling

The simplicity and convenience of systematic sampling are extremely appealing to the end users of land cover maps (Stehman 1992). Accordingly with Freund and Williams (1972) the systematic sampling distributes the sampling units equitably for all study area and so could be treated as random. Figure 2.2 show an example of a systematic sampling.

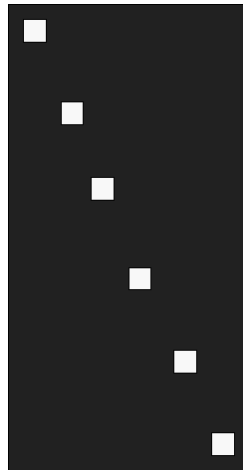


Figure 2.2 - Example of the distribution of the sampling units over a study area using a systematic sampling (Source: Congalton and Green 1999).

However, there's some controversy relatively to systematic sampling properties. According to Stehman (1992), this controversy is generally due to the non existence of an unbiased estimator to the variance computation. This aspect leads to an overestimation of the variance (Stehman and Czaplewski 1998). For that reason Congalton (1988) states that systematic sampling could not be considered random

because all the land cover classes over the study area (and consequently the less represented land cover classes) don't have the same probability of being chosen. In fact the non randomness of systematic sampling could be considered, but only when the first selected sampling unit is not chosen randomly. However, in remote sensing the previous mentioned aspect doesn't apply, because as long as it is applied to a systematic sampling where the first sampling unit is randomly selected, we can say that systematic sampling is random (Stehman 1992).

To conduct a systematic sampling, a first pixel with equal probability assigned to all pixels of the population should be selected. A square sampling grid of pixels is obtained sampling every pixel in horizontal and vertical directions relatively to the first random pixel (Stehman 1999). The constant intervals are equal to the side size of the sampling units, establishing a link between all the sampling units. This aspect facilitates field visits, because when the first sampling unit is randomly selected, the remaining sampling units are placed in a fix distance and direction relatively to the first sampling unit (Stehman 1999).

As the systematic sampling distributes evenly the sampling units over the study area, Stehman and Czaplewski (1998) claim that this method generally produces better results when compared with simple random sampling. The facility of implementation is another advantage of the systematic sampling, being for this reason one of the methods more used by the scientific community (Stehman 1999). Stehman (1999) refers also, that this method allows a simpler analysis of the data, and the post-stratification could be combined with systematic sampling with the aim of obtaining a more efficient sampling scheme.

In systematic sampling, a higher or lower value of variance depends on how the error is spatially distributed (Stehman 1999). Accordingly with Cochran (1977), if the errors are located in particular areas of the population, the systematic sampling will have a minor variance when compared with simple random sampling, due to the superior variability of simple random sampling. The linearity of error is an important criteria before opt by systematic sampling, because in the presence of errors distributed evenly over the study area, the implementation of this method is not desirable, due to the high values of the variance when compared with simple random sampling (Stehman 1999).

In the attempt of minimize the effects of error periodicity (e.g. valley topography, alignment of agricultural fields), could be applied a variation of systematic sampling, called unaligned systematic sampling (Cochran 1977). To implement an unaligned systematic sampling, the area of interest is divided into smaller, regularly spaced regions, being a sample unit chosen randomly inside of each one of these regions. With this method the sample units are also evenly dispersed by the study area but not so regularly positioned like in systematic sampling. This method can be viewed as a compromise between systematic sampling and simple random sampling, in other words, if the linearity of error is present, the unaligned systematic sampling is less susceptible to error (Stehman 1999). However, Stehman (1992) claim that the unaligned systematic sampling decreases the advantage of systematic sampling when the sampling units interval and the spatial distribution of error favors systematic sampling over simple random sampling.

Due to the nonexistence of an unbiased estimator for the variance of systematic sampling, Cochran (1977) claims that should be used the simple random sampling estimators and perform an approximation to the systematic sampling estimators. The systematic sampling is in this way a viable option if the approximation of the variance

computation be acceptable, because obtaining a reduced variance through simple random sampling is generally more important than a biased variance estimation (Stehman 1999).

2.2.1.3.3 Stratified random sampling

The stratified random sampling is very similar to simple random sampling. The difference relatively to the first one is due to the fact that is being used a previous knowledge to divide the study area in strata, where each one is sampled after (Congalton and Green 1999). Figure 2.3 shows an example of a stratified random sampling.



Figure 2.3 - Example of the distribution of the sampling units over a study area using a stratified random sampling (Source: Congalton and Green 1999).

There are two primary purposes to implement stratification in the accuracy assessment:

1) when the strata are of interest for reporting results (e.g. accuracy and area are reported by land cover class or by geographic subregion); and 2) when there's the need to improve the precision of the accuracy and area estimates (Olofsson et al. 2014). In the accuracy assessment of land cover maps, usually each stratum represents a land cover class, so stratified sampling allows for increasing the sample size from rare land cover classes, which in turn decreases the standard errors of the accuracy estimates for these rare classes (Olofsson et al. 2012). Accordingly with Stehman (1999), the

stratified random sampling could be used to ensure that the size of the sample in each stratum is suitable to satisfy the precision requirements in the accuracy assessment of each land cover class. This characteristic is the main advantage of this sample method since each land cover class is always sampled (Congalton and Green 1999). Besides the stratification by land cover class, the region of interest could be stratified by geography in order to control sample allocation (e.g. when the intent is to increase the sample size within one or more relatively small geographic areas), or to construct a spatially balanced sample (Stehman 2009b).

Comparatively with simple random sampling, Cochran (1977) states that the stratified random sampling shows a decrease in standard deviation if the proportions are very different between strata. In fact the need to satisfy several objectives through a multi stratification, is a hard task, due to the creation of many strata. If not many reference information is available it might be problematic to consider many regions and also many strata. However the stratification by geographic regions does not result in a precision gain, when compared with simple random sampling (Cochran 1977). The geographic stratification could be also used to ensure a good spatial distribution of the sample over the study area. The advantages of this approach are similar to the aligned systematic sampling and unaligned systematic sampling (Stehman 1999). Although the stratified random sampling is less susceptible to precision losses due to the periodicity of systematic sampling, which results in a variance decrease (Stehman 1999). The stratified random sampling allows the option to use different sampling designs in different strata, as sample selection in each stratum is implemented independently of other strata, providing this way flexibility to tailor the design to address different requirements in different strata (Stehman 2009b).

The stratified random sampling presents some drawback. Stehman and Czaplewski (1998) stated that the stratified random sampling only allows the accuracy assessment of the map that origins the strata, as also requires that the map in which this method is to be implemented to be available *à priori*. A stratified random sampling with an optimal or equal allocation of all sampling units by stratum, generally leads to different probabilities of inclusion of the sampling units in the different strata (e.g. larger polygons have higher probabilities of being sampled than smaller polygons) (Stehman and Czaplewski 1998). The same author stated that the different probabilities of certain strata being sampled could not be a problem, as long as probabilities are known and taken into account in the estimators computation. However, Stehman (1999) regards that the estimators formulas of the stratified random sampling are more complex than for simple random sampling or the systematic sampling, which difficult the probabilities inclusion in the computation of the estimators.

2.2.1.3.4 Cluster sampling

Cluster sampling has been frequently used in the accuracy assessment of land cover maps for allowing a quick collection of reference data in many sample units (Congalton and Green 1999). Accordingly with Sarndal et al. (1992), in cluster sampling the population is grouped in sub-populations. In cluster sampling the sampling units are grouped in sets, designated by clusters (Caetano et al. 2006). Figure 2.4 shows an example of cluster sampling.

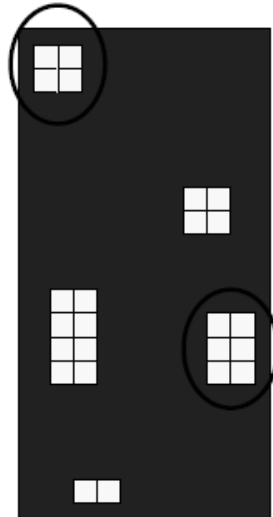


Figure 2.4 - Example of the distribution of the sampling units over a study area using cluster sampling (Source: Congalton and Green 1999).

In cluster sampling, two sizes of sampling units are employed: 1) the primary sampling unit (*psu*) which is the cluster itself, and 2) the secondary sampling unit (*ssu*) which is the sampling unit within the *psu* (Stehman 1999).

Accordingly with Caetano et al. (2006), a cluster is randomly selected and all *ssu* are inspected. Alternatively, the *ssu* could also be randomly selected. These sampling designs are designated respectively by one-stage cluster sampling and two-stage cluster sampling (Caetano et al. 2006). Both methods of cluster sampling have been implemented in the accuracy assessment of land cover maps (Strahler et al. 2006).

In this sampling method the *psu* used, is many times the area defined by an aerial image limits, since it allows a reduction of the costs inherent to travel (Stehman 2001), or else could be used blocks of 3x3 or 5x5 pixels (Stehman 1999). In fact, according to Strahler et al. (2006), the use of cluster sampling is due to its minor cost when compared with other sampling designs used in the accuracy assessment of land cover maps. Grouping the reference pixels reduces the acquisition cost of reference data, the time travel in the cases where field visits are necessary and reduces the processing time and quantity of

aerial images or satellite images used in sampling protocol (Strahler et al. 2006). It is less expensive sampling all the 9 pixels defined as the *psu* (i.e. a cluster of 3x3 pixels), than sampling 9 pixels randomly distributed over the study area (Stehman and Czaplewski 1998). The spatial proximity of the *ssu* sampled in each *psu* allows the collection of reference information in the sampling units with reduced costs, when compared with simple random sampling or systematic sampling (Stehman 1999). Although if we consider a one-stage cluster sampling (considering large clusters), reference data must be collected for all pixels within each sampled cluster, and in this sense, cluster size and number of sample clusters will be strongly affected by cost (Stehman 2009b; Stehman et al. 2009). Cluster sampling is also well suited to provide a sample of assessment units of different sizes for a multiple-objective accuracy assessment (Stehman 2009b). In his work Stehman (2009) provides an example where the objectives of the accuracy assessment stipulates a traditional error matrix assessment at the pixel level (e.g. 30 m pixel), and an assessment of land cover composition accuracy for both a 3 km by 3 km support and a 12 km by 12 km support. A three-stage cluster sampling design using 12 km by 12 km blocks as the *psu*, 3 km by 3 km blocks as the *ssu* nested within the sampled *psu*'s, and pixels as the final stage sampling unit nested within the sampled *ssu*'s would provide the data required. However this sampling method also possesses some disadvantages. Despite the relation cost-benefit of cluster sampling be its major advantage, this aspect implies a decrease in statistic consistency, due to spatial correlation of the error that exists between the sampling units in the *psu*, causing an increase on the value of the estimation error variance (Cochran 1977), as also an increase of the complexity of the formula to compute the standard deviation (Stehman and Czaplewski 1998). However, accordingly with Stehman (1999), when is used a two-stage cluster sampling, exists a compromise between the strong spatial

Accuracy assessment of land cover maps

relation of the *ssu* in one-stage cluster sampling and the weak spatial relation of simple random sampling and systematic sampling. Other negative aspect of the application of a two-stage cluster sampling is the complexity of the formulas to compute the variance as also are less known by the end users when compared with more simpler sampling designs (Stehman 1999). Table 2.2 show the main characteristics of the sampling designs used in the accuracy assessment of land cover maps.

Table 2.2 - Characteristics of the sampling designs used in the accuracy assessment of land cover maps
(Source: Stehman (1999)).

Method	Analytic simplicity	Variance	Variance estimation	Spatial distribution	Spatial control
Simple random	High	High, unless when combined with more complex analytic methods	Simple	Poor	None
Systematic	High ^a	Low, unless sampling in phase with periodicity	Simple but biased	Very good	None
Stratified	<i>By land cover (equal allocation)</i> Moderate ^b	Low, for stratum-specific estimates, moderate to high for overall estimates	Moderate	Poor ^c	None
	<i>By geography</i> Moderate ^b	Moderate	Moderate	Very good	Low ^d
Cluster	<i>One-stage</i> High	High, unless spatial correlation is weak	Difficult	Poor to moderate ^e	Moderate
	<i>Two-stage</i> High ^f	Moderate, even if spatial correlation is weak	Difficult	Moderate to good ^g	Very high

^a Systematic sampling of polygons results in unequal inclusion probabilities and a more complex analysis.

^b High simplicity if proportional allocation is used.

^c Good if the land cover classes are strongly associated with geographic regions.

^d High if unequal allocation is used to focus sampling effort in only a few strata.

^e Poor if *psus* are selected via simple random sampling, moderate if *psus* are selected via systematic sampling or by geographic stratification.

^f Difficult if unequal inclusion probabilities at second stage.

^g Moderate if *psus* are selected via simple random sampling, good if *psus* are selected via systematic sampling or by geographic stratification.

2.2.1.4 Sample size

For a statistically valid accuracy assessment of land cover maps, the collection of an adequate number of sampling units for each land cover class is necessary (Congalton and Green 1999). Indeed Foody (2009) stated that sample size will impact on the precision with which accuracy is estimated as also impact on the interpretation of differences in accuracy. Congalton and Green (1999), refer that the minimum number of sampling units that should be collected, taking into account the practical constrains, is of 50 sampling units for each land cover class. Although if the used land cover nomenclature possess more than 12 land cover classes, the minimum number of sampling units that should be collected, should be between 75 and 100 (Congalton and Green 1999). According to Stehman (2001), a sample with 100 sampling units per land cover class, ensures that the accuracy could be estimated with a standard deviation not superior than 0,05. Despite this aspect, the minimum value of sampling units per land cover class could vary, depending of the importance of the evaluation of certain land cover classes. It could be more useful to increase the number of sampling units relative to a certain land cover class, and decrease the number of sampling units relative to a land cover class less important (Congalton and Green 1999; Stehman 2001).

To derive an estimate of the required sample size, basic sampling theory may be used (Foody 2009). The objective is to estimate accuracy expressed as the proportion of correctly allocated cases, to a certain degree of precision. The sample size could be determined for the case of simple random sampling through Equation (2.1), where P is an estimate of the actual population value, $z_{\alpha/2}^2$ is the critical value of the normal distribution for the two-tailed significance level α and h is the half width of the desired confidence interval.

$$n = \frac{z_{\alpha/2}^2 P(1-P)}{h^2} \quad (2.1)$$

Although Foody (2009) stated that there are two problems with the use of Equation (2.1). One of the problems is that it may not be appropriate usable, since the terms of the equation are not always easy to define. To overcome with this problem the terms of Equation (2.1) could be defined using prior knowledge or conservative values. Without the knowledge of P could be used a conservative estimate where $P = 0.5$, maximizing the term $P(1-P)$ despite the value defined for h is not necessarily clear. The other problem with the use of Equation (2.1) is that it simply gives the sample size required to estimate proportion at a given level of precision, which is only a part of the common accuracy assessment scenario. Indeed in many applications there's the need to evaluate the derived proportion with a target value, to determine if the estimated proportion satisfies the end user's needs. For example if a target accuracy of 0.70 (70%) is defined, the analyst could use the testing set using Equation (2.1) to estimate the proportion of corrected allocated cases, and compare the lower value of the confidence interval with the target accuracy (Foody 2009). If the lower value of the confidence interval exceeds the target accuracy then the user's requirements in terms of accuracy are achieved. Although a very narrow confidence interval will be needed if the classification accuracy is only just above the target accuracy, being necessary the use of a very large testing sample. In this sense, Foody (2009) stated that it is preferable to design the sample more fully around the factors that influence the proportions comparison, providing a manner of determining the sample size accordingly with project goals.

The definition of the project goals is extremely important in the definition of the sample size, since the project goals will influence the choice of the equations to be used in the computation of the sample size. As an example, Stehman (2012) for determining the

impacts of sample size allocation in land cover land change maps under stratified random sampling, stated that if the objective is to estimate overall accuracy Neyman optimal allocation should be used to determine the sample size in the change and no-change strata. If the objective is to determine user's accuracy, equal allocation should be used to determine the sample size in the change and no-change strata.

According to Dicks and Lo (1990), the higher the sample size, the higher will be the confidence in the accuracy estimation made with that sample. Although, it is necessary to find a compromise between the necessary number of sampling units and the cost of obtaining reference data for those sampling units. Cochran (1977) refers that a very high sample size implies a waste of resources and a very reduced sample size, diminish the utility of the results. In this sense it is necessary to find a relation cost-benefit to determine the ideal sample size.

To determine the sample size researchers have used, an equation based on a binomial distribution or a normal distribution approximate to binomial distribution (Congalton and Green 1999). The same authors also claim that the choice of the sample size should be dependent of two aspects: 1) the acceptable level of accuracy; and 2) the confidence level of the estimation. Although when considering an error matrix is not a matter of correct or incorrect (the binomial distribution). Instead it is a matter of which error or which categories are being confused and for that the multinomial distribution is more adequate (Congalton and Green 2009).

2.2.2 Response design

The response design defines the protocol for determining the ground condition (i.e. the reference classification) at the selected sample sites (Olofsson et al. 2012). A critical feature of the response design protocol is that the spatially explicit character of the

accuracy assessment should be retained, and practitioners should aim to have reference data with an equal or finer level of detail than the data used to create the map (Olofsson et al. 2014). Conceptually is useful to separate the response design into two components, the evaluation protocol, which consists of the procedures used to collect information contributing to the reference classification determination, and the labeling protocol, which assigns a land cover classification to the sampling unit based on the information obtained from the evaluation protocol (Stehman and Czaplewski 1998). The quality of the reference classification plays a key role in the quality of the assessment, since it is at this stage that the land cover information that serves as the "ground truth", is collected to be afterwards compared with the map classification. Although collecting reference data with high accuracy is not an easy task, namely due to: variation in classification and delineation of the reference data due to inconsistencies in human interpretation of heterogeneous vegetation; registration differences between the reference data and the remotely sensed map classification; error in interpretation and delineation of the reference data; delineation error encountered when the sites chosen for accuracy assessment are digitized; data entry error when the reference data is entered into the accuracy assessment database; changes in land cover between the date of the remotely sensed data and the date of the reference data; errors in the remotely sensed map classification; and errors in the remotely sensed map delineation (Congalton and Green 1993).

2.2.2.1 Evaluation protocol

To develop the response design, first it is necessary to choose the spatial support region on which the reference land cover evaluation will be based (Stehman and Czaplewski 1998). The spatial support region is an area surrounding a sampling unit that is used as context for determining the reference classification of the sampling unit (Wulder et al.

2006). Specifying the area and shape of the support region is dependent on the type of land cover. For example, linear features such as stream riparian zones may be evaluated differently from forest stands or agricultural fields (Stehman and Czaplewski 1998). An intelligent choice of the support region can decrease some of the difficulties in collecting the reference data, namely the spatial registration error between the map data and the sampling unit, and the spatial heterogeneity of land cover, using for example a block of 3x3 pixels as the support region, where the main land cover could be the mode of land cover at that support region (Wulder et al. 2006). The spatial support region could be also sub sampled (e.g. using line transects, point samples or cluster plots) in order to estimate quantitative characteristics that are useful to determine the land cover classification of the sampling unit or collect information about land cover heterogeneity within the sampling unit (Stehman and Czaplewski 1998). The selection of the appropriate spatial support region for an accuracy assessment is extremely important because the spatial unit has implications on the sampling design (Stehman and Wickham 2011).

2.2.2.2 Labeling protocol

The labeling protocol assigns the reference classification (or classifications) to the sampling unit based on the information obtained from the evaluation protocol (Stehman and Czaplewski 1998). Generally the assignment of reference land cover information within a sampling unit, is restricted to just one land cover class. Although, other possibilities exist, namely the collection of a primary and a secondary reference label when in the support region of the sampling unit the land cover consists in more than one land cover class or represent a transition or mixed class not easily identified as a single land cover type. In addition an interpreter confidence rating that represents the interpreter's perception of uncertainty in the reference classification for a sampling unit

could be also collected (Olofsson et al. 2014). This issue will be further discussed in Section 2.3.

The reference land cover information could be collected using several methods namely through the photo-interpretation with higher resolution imagery, qualitative observations by a field crew, or sub-sampling and physical measurements by a field crew (Czaplewski 2003), and using several data sources like field plots, aerial photography, forest inventory data, airborne video, lidar and satellite imagery (Olofsson et al. 2014). The use of the methods mentioned above are dependent on the size of the support region of the sampling unit. For example if the support region is large, inexpensive methods are more appropriate, such as the photo-interpretation of high resolution imagery. On the other hand, if the support region is small, the reference land cover information collected by a field crew, is more valuable. According with Congalton (2001) the reference data should be collected with the pixel size and/or the minimum mapping unit of the map in mind and additionally, the same exact classification scheme using the same exact rules must be used to label the reference data and to generate the map. However, Czaplewski (2003) refers that could exist advantages with a more detailed classification system for the reference data, giving an example where the reference data separates "shrubby wetlands" from "forested wetlands", but the map groups both types of wetland into a single category called "wooded wetlands". The reference data statistically estimate the proportion of wooded wetlands that are actually shrubby and forested woodlands, even though these detailed categories are not separated in the map. This aspect is emphasized by Olofsson et al. (2014) where is recommended that the practitioners should aim to have reference data with an equal or finer level of detail than the data used to create the map. The same authors refer also the need to specify a minimum mapping unit for the reference classification because minimum

mapping unit can have important implications for accuracy assessment and area estimation.

2.2.3 Analysis and estimation protocol

The analysis and estimation protocols applied to the reference sample data constitute the third main component of an accuracy assessment (Stehman and Czaplewski 1998). The main goal of an accuracy assessment is the estimation of the area for each land cover class on the map that is correctly classified, as also the confidence intervals of the accuracy measures for those land cover classes (Carrão 2006; Olofsson et al. 2014). The more used method to perform an accuracy assessment is the confusion matrix that plays a very important role in accomplishing the objectives of an accuracy assessment (Foody 2013; Stehman 2013). Generally in a confusion matrix, the reference data is represented in the matrix columns, and is compared with the map data, generally represented in the matrix rows. The values in the diagonal of the matrix indicate the agreement between the two datasets i.e. between the classified map and the reference data. For allowing an individual analysis of the accuracy per land cover class, the confusion matrix is considered a good method to represent the accuracy of land cover maps (Congalton and Green 1999).

According to Congalton and Green (2009), assume that n samples are distributed into k^2 cells, here each sample is assigned to one of k categories in the map (usually the rows), and independently to one of the same k categories in the reference data set (usually the columns). Let n_{ij} denote the number of samples classified into category i ($i = 1, 2, \dots, k$) in the map and category j ($j = 1, 2, \dots, k$) in the reference dataset (Figure 2.5).

		j = columns (reference)			Row total n_{i+}
		1	2	k	
i = rows (classification)	1	n_{11}	n_{12}	n_{1k}	n_{1+}
	2	n_{21}	n_{22}	n_{2k}	n_{2+}
	k	n_{k1}	n_{k2}	n_{kk}	n_{k+}
Column total n_{+j}		n_{+1}	n_{+2}	n_{+k}	n

Figure 2.5. Representation of a confusion matrix. (Adapted from Congalton and Green (2009)).

Several approaches of area estimation could be derived from the confusion matrix. Stehman (2013) presented four approaches for area estimation under simple random and stratified random sampling, namely: 1) calculating area from the map; 2) direct estimation of area from the reference classification; 3) estimating area by adjusting for map classification error; and 4) model-assisted estimation area. From the three sample based approaches (i.e. approaches 2); 3); and 4)) the stratified random sampling yield the same stratified estimator, while for simple random sampling the three approaches yield different estimators (Stehman 2013). Other examples of approaches for area estimation are recursive restrictive estimation (Czaplewski 2010) and the incorporation of other auxiliary variables in addition to the map information (Stehman 2009a).

One important feature of the estimation protocol is that the specific estimators for accuracy, area and the variances of these estimators depend on the sampling design implemented, and in this sense is essential that only unbiased or consistent estimators should be used in the accuracy assessment (Olofsson et al. 2014).

Once correctly generated, the confusion matrix can be used as a starting point for a series of descriptive and analytical statistical techniques (Congalton 2001). In the

following section the accuracy estimators for simple random sampling will be further discussed.

2.2.3.1 Accuracy measures

In this section some accuracy measures that could be derived from a confusion matrix will be presented. These techniques show how powerful is the confusion matrix and why this matrix should be included in any accuracy assessment of land cover maps (Congalton and Green 1999). However, Stehman (1997) states that any accuracy measure is universally adequate in the accuracy assessment and that different accuracy measures could lead to conflicting conclusions, because accuracy measures do not represent accuracy in the same manner. The choice of the appropriate accuracy measures that allows to elucidate the end users of land cover maps if their accuracy objectives were accomplished, is a fundamental aspect in the accuracy assessment of land cover maps. The more common accuracy measures in the accuracy assessment of land cover maps are the overall accuracy, user's accuracy and the producer's accuracy.

2.2.3.1.1 Overall accuracy

The overall accuracy is computed through the division of the sum of the sampling units in the diagonal of the matrix by the sample size (Story and Congalton 1986). This is the more common accuracy measure and reflects the proportion of area that is correctly classified (Stehman 1997), where i represents the matrix rows, n_{ii} the sampling units that are correctly classified for each land cover class k , and n the sample size, and is given by Equation (2.2).

$$\text{Overall accuracy} = \sum_{i=1}^k \frac{n_{ii}}{n} \quad (2.2)$$

2.2.3.1.2 User's accuracy

When the number of sampling units correctly classified (n_{kk}) in one class is divided by the total sampling units of that class in the map (n_{k+}), this value indicates the probability of a certain land cover class in the map represent that same class in reality. In this case the commission errors are being measured, which are defined as the inclusion of an area on the map in a land cover class in which that area should not be included (Congalton and Green 1999). This value is defined as users accuracy, and indicates how a map represents reality. The users accuracy for a land cover class i , indicates the probability of an area classified as i on the produced map, be classified as i in the reference data (Stehman 1997), and is given by Equation (2.3).

$$\text{User's accuracy} = \frac{n_{ii}}{n_{i+}} \quad (2.3)$$

2.2.3.1.3 Producer's accuracy

Story and Congalton (1986) refer another computation method of accuracy, in which the number of sampling units correctly classified (n_{kk}) in one class is divided by the total number of sampling units in the reference data (column total), i.e. n_{+k} for that class. This proportion represents the probability of the sampling units in the reference data to be well classified, i.e. the omission errors are being measured. The omission errors arise when an area is excluded on the map from the land cover class it should belong to (Congalton and Green 1999). This value is designated as producer's accuracy where for a land cover j , indicates which is the probability of a land cover class j in the reference data being classified as j , in the produced map (Stehman 1997) and is given by Equation (2.4).

$$\text{Producer's accuracy} = \frac{n_{jj}}{n_{+j}} \quad (2.4)$$

2.3 Uncertainty of reference databases in the accuracy assessment of land cover maps

Reference databases are a key factor for the accuracy assessment of land cover maps. The reference classifications determined from the reference database are intended to represent the 'real' land cover data and are usually assumed to be free of errors. According to Olofsson et al. (2014) there are two sources of uncertainty in reference databases: 1) the uncertainty associated with spatial co-registration of the map and reference location (Pontius 2000); and 2) the uncertainty associated with the interpretation of the reference data (Pontius and Lippitt 2006). Geolocation error is defined as a mismatch between the location of the spatial assessment unit identified from the map and the location identified from the reference data (Olofsson et al. 2014). The interpreter uncertainty can be divided in two parts: 1) interpreter bias is defined as an error in the assignment of the reference class to the spatial unit; and 2) interpreter variability is a difference between the reference class assigned to the same spatial unit by different interpreters (Olofsson et al. 2014).

Only one land cover class is traditionally assigned at each sample location chosen for the accuracy assessment, and this information is obtained through field visits and/or photo interpretation of higher resolution images. However, this procedure presents limitations that could lead to an erroneous evaluation of the map's accuracy. Reference databases are also classifications and consequently are imperfect, and the assumption that they are free of errors is not correct (Congalton and Green 1999), leading to large biases of the estimators of classification accuracy and area (Foody 2010; Foody 2013).

The assignment of only one class to each sample location requires the choice of the class that best describes the land cover at that location (Gopal and Woodcock 1994; Woodcock and Gopal 2000). Frequently, due to landscape heterogeneity and the fact that land cover rarely presents abrupt transitions between classes, the choice of the correct land cover class is not obvious, and this leads to uncertainty in the assignment of the correct land cover class to the sample pixels.

Several approaches to deal with uncertainty in the reference classification have been adopted. One approach is to assign a primary and a secondary reference label (Woodcock et al. 1996; Edwards et al. 1998; Zhu et al. 2000; Wickham et al. 2004). In addition, a qualitative interpretation confidence rating may be assigned to each sample location to represent the confidence in the reference classification at that location (Zhu et al. 2000; Wickham et al. 2004; Sarmiento et al. 2009). A variety of analyses are available from this information. Accuracy estimates based on all sample observations compared to estimates based on just the high confidence sample observations or just sample observations where land cover is homogeneous, provide insight into the impact reference data uncertainty has on the accuracy of the land cover maps (Yang et al. 2000; Stehman et al. 2003). Although, if just the sample observations where the land cover is homogeneous are used or just the sample observations where the photo-interpreter assigned with high confidence a certain land cover class, the results of the accuracy assessment may be biased, since it is not taken into account the uncertainty derived from landscape fragmentation, providing an optimistic perspective of accuracy. Another disadvantage of estimating accuracy for various subsets of the sample is that this approach does not provide a unique thematic accuracy measure that accommodates the uncertainty present in a reference database, providing instead a unique thematic accuracy measure for each subset of the reference database. Another important issue is

that the reference classifications used for these analyses are crisp and consequently the representation of land cover as a continuum is not accommodated. The technician can only assign one or two land cover classes at each sample observation, considering that all the area of the sample observation is occupied by just one land cover class even if more than one land cover class exists in the area occupied by the sample observation.

Mayaux et al. (2006) developed another approach where a similarity matrix that accounts for the relationship between land cover classes was used to derive thematic accuracy measures that incorporates the similarity between land cover classes. The rationale of this approach is based on the treatment of errors in a unequal way. For example in this approach an error between deciduous needle forest and cropland will contribute more to error than an error between deciduous needle forest and deciduous broadleaf forest since there is more relation between both these forest land cover classes than between each of them and cropland.

Other authors proposed to specify the proportion of area that each land cover class occupies within a sampling unit (Foody et al. 1992; Lewis and Brown 2001).

Fuzzy approaches that account for partial membership of the land cover classes to the sampling unit have been also developed (Foody, 1996; Foody and Arora, 1996), where the agreement between the fuzzy classification and the fuzzy ground data is made through simple distance measures. To deal with fuzzy classifications and fuzzy reference data, Binaghi et al. (1999) developed a fuzzy error matrix to extend the applicability of the error matrix to the evaluation of soft classifiers, and more recently Silván-Cardenas and Wang (2008) introduced the sub-pixel confusion-uncertainty matrix for sub-pixel accuracy assessment and Pontius and Cheuk (2006) a cross-tabulation matrix to compare soft classified maps at multiple spatial resolutions.

Gopal and Woodcock (1994) proposed another approach for the accuracy assessment of land cover maps using fuzzy sets, which uses a linguistic scale to translate the degree of match or mismatch between what is observed in the sample locations and the classes. This linguistic scale is composed by five values: (1) absolutely wrong; (2) understandable but wrong; (3) reasonable or acceptable answer; (4) good answer; (5) absolutely right, which are assumed to translate degrees of membership of the sample sites to the classes. Based on this linguistic scale, the authors developed several operators (denoted by capital letters) to evaluate the frequency of matches and mismatches (MAX and RIGHT); the magnitude of errors (DIFFERENCE); the source of errors (MEMBERSHIP) and the nature of errors (CONFUSION and AMBIGUITY). Examples of use of this approach include Muller et al. (1998); Mickelson et al. (1998); Woodcock and Gopal (2000); Townsend (2000); Falzarano and Thomas (2004); and Laba et al. (2002). Even though the proposed methodology enables the inclusion of the uncertainty in the reference database in the accuracy assessment, further refinement of the methods can yield improved insights into the accuracy of a land cover map. For example, a simple discrete value between zero and one is used to express the fuzzy membership function associated to each value of the linguistic scale but this does not make full use of the potential of fuzzy sets theory to model uncertainty. An analysis that incorporates uncertainty in a more unified manner is also desirable.

The concept of fuzzy confusion matrix to incorporate the uncertainty of reference data has been developed by other authors. Congalton and Green (2009) presented three methods to incorporate uncertainty in the accuracy assessment of land cover maps, namely: 1) expanding the major diagonal of the error matrix; 2) measuring map class variability; and 3) using a fuzzy error matrix approach.

Expanding the major diagonal of the confusion matrix is the simplest and most straightforward method to incorporate fuzziness in the accuracy assessment process (Congalton and Green 2009). In the traditional approach the confusion matrix diagonal represent the agreement between the map data and reference data. With this approach the map classification could be expanded to accept as correct plus or minus one adjacent class of the actual class (Congalton and Green 2009). In this sense the major diagonal is no longer just a cell, but instead a wider group of cells. Accordingly with Congalton and Green (2009) this method works well for continuous classifications (e.g. tree size class or forest crown closure), but if the classification scheme is discrete this method cannot be used. With this method the accuracy increases but this could be a problem since accepting one or minus one class cannot be adequately justified or does not meet the map user's requirements (Congalton and Green 2009).

The rationale for measuring map class variability, is that it is difficult to control variation in human interpretation, but is possible to measure this variation, compensating this way the differences between the map data and reference data that are caused by variation in interpretation and not caused by map error (Congalton and Green 2009). Congalton and Green (2009) defined two options to control the variation in human interpretation: 1) measure each reference site with great precision to minimize the variance in the reference sites labels; and 2) measure the variance and use these measurements to compensate for non error differences between the map data and reference data. The first option could be prohibitively expensive because requires extensive field sampling and detailed measurements. Measure the variance requires having multiple analysts assess each reference site (Congalton and Green 2009), collecting the reference information through field visits or photo-interpretation. Although the assessment of the reference sites by several analysts could be prohibitively

expensive and therefore is not usually a valid component of most remotely sensed data projects (Congalton and Green 2009).

Green and Congalton (2004) introduced the fuzzy error matrix approach, which allows the analyst to compensate for situations in which the classification scheme breaks represent artificial distinctions along a continuum of land cover and/or where observer variability is difficult to control (Congalton and Green 2009). With this approach, reference data incorporates the likelihood of each land cover class at each sample site, through a linguistic scale ("best", "good", "acceptable" and "poor"). In the traditional error matrix, the sample sites where the land cover classes is labelled as "best" and match the map classification, correspond to the diagonal of the matrix. The nondiagonal cells of the matrix contain two tallies, which can be used to distinguish land cover classes that are uncertain, where the first number represents the sample sites where the map label matched a "good" or "acceptable" reference label in the fuzzy assessment (Green and Congalton 2004). The second number represents the sites where the map label was considered "poor" (i.e. an error). The fuzzy overall accuracy is estimated as the percentage of sites where the "best", "good", or "acceptable" reference labels matched the map label (Green and Congalton 2004). Despite this method can be used for any accuracy assessment, Congalton and Green (2009) defined three criteria when it works best: 1) there are issues in collecting good reference data because of limitations in the reference data collection methods; 2) when interpreter variability cannot be controlled; or 3) when the ecosystem being mapped is highly heterogeneous. The accuracy measures derived from this approach are similar to the RIGHT operator (Gopal and Woodcock 1994) and in recent studies the RIGHT operator demonstrates providing optimistic perspectives of accuracy (Sarmiento et al. 2013).

The methods developed in this dissertation incorporate the uncertainty of the reference database in the accuracy assessment of land cover maps. Gopal and Woodcock's (1994) approach provides the foundation of the proposed analysis. To each sample location and each land cover class is assigned one value of a linguistic scale. The proposed extension of the Gopal and Woodcock (1994) approach is that these linguistic values are then converted into fuzzy intervals that express the percentage of the sample pixel that is typically occupied by the class (when a particular value of the linguistic scale is used) as well as the uncertainty that is associated with the possible variability of that percentage. Since computations can be done with fuzzy intervals using fuzzy arithmetic, it is possible to incorporate this information to generate a fuzzy confusion matrix analogous to the traditional approach used to construct a confusion matrix based on a crisp classification of the reference database. Using this fuzzy confusion matrix, fuzzy overall, user's and producer's accuracy measures may be computed. These accuracy measures are fuzzy intervals that incorporate information on the percentage of the pixel covered by the class as well the uncertainty existing in this value. To obtain from the fuzzy accuracy measures traditional crisp measures, defuzzification methods can be applied to the fuzzy confusion matrix. The importance of providing additional thematic accuracy measures that address the uncertainty of reference databases is highlighted by several authors (Stehman 1997; Arbia et al. 1998; Muller et al. 1998).

The research presented in this dissertation gives continuity to the research previously developed in the RSU/IGP where to accommodate the difficulty of identifying a single 'true' or 'reference' land cover class, the reference data protocol of an accuracy assessment included the identification of a primary and a secondary reference label along with a rating of the interpreter's confidence. This additional reference information was used to construct one nominal variable (called CONF) in which the categories

represent the 'confidence' in the correctness of the map land cover classification at a given location. One accuracy measure that incorporate uncertainty in the reference classification was then derived by assigning partial credit weights to each CONF categories (Sarmiento et al. 2009).

2.4 Challenges and future directions on the accuracy assessment of land cover maps

Despite the methods for the accuracy assessment of land cover maps based on the confusion matrix be widely adopted by the scientific community, many further developments are needed to improve the quality of the accuracy assessment of land cover maps. One of the developments that are needed is referred by Stehman (2004) where a compromise between a good precision and low costs is necessary. For one side a good precision could be obtained using a stratified random sampling while the low costs could be obtained using a cluster sampling. Both sampling designs referred are widely adopted in many accuracy assessments, but individually. Implementing a design with both characteristics is difficult, but combining the characteristics of stratification and clusters would bring more efficiency to the accuracy assessment. In this sense, develop methods to integrate both sampling designs in the accuracy assessment of land cover maps are needed.

Besides the accuracy assessment of land cover maps is nowadays considered an important aspect of map production projects, there isn't an accepted standard method for the report of the accuracy assessment (Foody 2002). However, the confusion matrix lies in the core of the major part of map production projects and research projects due to their simplicity and arrange of several accuracy metrics. Nevertheless the assumptions made on the accuracy assessment of land cover maps based on the confusion matrix

must be carefully made. Indeed there are many assumptions that could lead to misinterpretation of the accuracy of a land cover map. In this sense develop methods to deal with the problem of mixed pixels and misregistration between the map and reference data are essential to cope with the limitations on the interpretation of the confusion matrix and the accuracy metrics.

The assumption that the reference data is free of errors is another issue in the accuracy assessment of land cover maps. Strahler et al. (2006) refers the need to integrate reference data error in the overall estimates of the maps through more practical techniques. Another issue related with reference data is that this data is many times difficult to acquire (Foody and Boyd 2013), since it is needed to collect a great quantity of reference data to meet the requirements of the sampling designs most commonly applied in the accuracy assessment of land cover maps. To cope with this situation develop methods to validate land cover maps without reference data are in the agenda of the remote sensing scientific community.

One of the topics receiving attention lately is area estimation based on data for the accuracy assessment (Stehman 2013; Olofsson et al. 2014) as also the impacts that imperfect ground truth data have in area estimation (Foody 2013). Although there's a lack of research for area estimation using fuzzy data sets and further developments on this topic are needed.

Another potential source of reference data that have been explored by the scientific community is the volunteer data. Indeed the recent availability of volunteer data in a great quantity lead the scientific community to test the potential of this data to be used as reference data. Although many issues related with volunteer information remain to be solved, namely their variability, unknown quality as also ethical and legal concerns with

its collection and use (Goodchild and Glennon 2010). Recent work have been developed on this topic, namely to study the potential of Geo-Wiki project to improve global land cover maps (Fritz et al. 2009; Fritz et al. 2012), the accuracy of volunteered geographic information (Foody et al. 2013; Foody et al. 2014) and produce estimates of land available for biofuel production (Fritz et al. 2013).

Another line of research that has received very little attention is the accuracy assessment of several maps in an integrated and cohesive manner (Olofsson et al. 2012). In this sense Olofsson et al. (2012) and Stehman et al. (2012) developed a conceptual framework and specific methodological details for constructing a global validation database that could be used to assess the accuracy of multiple land cover maps.

3 FUZZY SETS THEORY

In this chapter a brief description of how and why fuzzy sets theory started to be developed, as well the fundamental notions of fuzzy sets theory necessary for the work developed in this thesis are presented.

3.1 Fuzzy sets

Since the beginning of civilization, the boolean logic has been the basis of science. In fact, this line of thought is the basis of most of the mathematic and computational science that we deal with nowadays. Although, frequently human mind can't deal with exact definitions of objects or classes (Burrough and Frank 1996). Even if we can define classes in an exact way we may not assign correctly the individuals to the defined classes, due to ambiguity of the rules or because we cannot do it with sufficient accuracy (Burrough and McDonnell 1998). In this sense methods that allow dealing with imprecise information are needed.

Fuzzy sets theory was introduced by Zadeh (1965) in order to characterize the ability of the human brain to deal with vague relations. Unlike boolean logic that just allows binary membership functions (an individual belongs to a determined set or not), fuzzy sets theory admits the possibility of partial belonging of an individual to a determined set, being a generalization of boolean sets in situations where the classes boundaries cannot be well defined (Burrough and Frank 1996). Land cover is vague and uncertain, and his definition is difficult. In this context fuzzy sets theory allowed the development of techniques to deal with uncertainty.

In classical sets theory, a characteristic function that can only take values 0 and 1 is

assigned to each element x of a universal set X to indicate whether that element belongs to a certain crisp set A , with zero corresponding to non-membership and 1 to membership. In fuzzy set theory the characteristic function is generalized into the membership function $\mu_A(x)$, allowing the assignment to each element x of the universal set a value in a range of possible values, which correspond to grades of membership of element x to set A . In these conditions set A is called a fuzzy set, and is characterized by the membership function (Klir and Yuan 1995). The range of the membership function is usually considered to correspond to the unit interval, that is, $\mu_A(x): X \rightarrow [0,1]$ (Dubois et al. 2000a).

3.1.1 Core and support

The support of a fuzzy set A , $S(A)$, is the set of all elements of X that have positive degrees of membership to A . The core of a fuzzy set is the set of all elements that satisfy $\mu_A(x)=1$, that is, the elements that have full membership to A . A fuzzy set is called normal if there is at least one element such that $\mu_A(x)=1$ and subnormal when, for all elements x of X , $\mu_A(x) < 1$.

3.1.2 Alpha-levels

One of the most important concepts associated with fuzzy sets is the alpha level cut (Klir and Yuan 1995). A fuzzy quantity is represented in a unique way by its alpha levels. For a fuzzy set A defined in set X and any number $\alpha \in [0,1]$, the alpha level of A is the set of elements that have a degree of membership to A equal or larger than α , that is,

$${}^\alpha A = \{x : \mu_A(x) \geq \alpha\} \quad (3.1)$$

3.1.3 Fuzzy quantities

The term fuzzy quantity is used to express any normal fuzzy set defined on the set of

real numbers \mathbb{R} (Dubois et al. 2000b). Fuzzy intervals and fuzzy numbers are particular cases of fuzzy quantities. A fuzzy interval is a convex fuzzy subset of the real line. A fuzzy number is an upper semi continuous fuzzy interval with limited support, whose core only contains one element of \mathbb{R} (Klir and Yuan 1995). The alpha levels of fuzzy intervals and numbers are closed intervals of real numbers. A fuzzy interval is called trapezoidal if its membership function has a trapezoidal shape and a fuzzy number is called triangular if the membership function has a triangular shape (Figure 3.1).

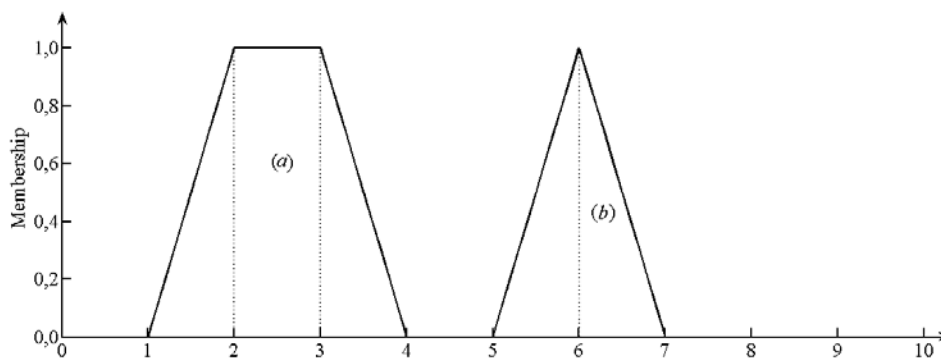


Figure 3.1 - (a) A trapezoidal fuzzy interval (the core of the fuzzy interval is the interval $[2,3]$ and the support is defined by the interval $]1,4[$); (b) A triangular fuzzy number (the core of the fuzzy interval contains only one element of \mathbb{R} , the value 6, and the support is defined by the interval $]5,7[$).

3.2 Fuzzy arithmetic

The notion of fuzzy quantity encouraged the development of fuzzy arithmetic, which enables the execution of arithmetic operations with fuzzy numbers and intervals (Kaufmann and Gupta 1991; Klir and Yuan 1995). One of the methods available to perform operations with fuzzy quantities uses interval arithmetic (Moore 1966) and is supported by the following two characteristics of fuzzy intervals and numbers: 1) a fuzzy quantity is represented in a unique way by its alpha levels; 2) the alpha levels of fuzzy intervals and numbers are closed intervals of real numbers.

The second characteristic enables the use of interval arithmetic and the first one enables

the arithmetic operations on fuzzy quantities to be performed applying the corresponding interval arithmetic operations to all alpha levels of the involved quantities. Therefore, if A and B are two fuzzy intervals or numbers and * is one of the four basic arithmetic operations (addition, subtraction, multiplication or division), the fuzzy interval or number $A * B$, with alpha values ${}^\alpha(A * B)$, is obtained considering

$${}^\alpha(A * B) = {}^\alpha A * {}^\alpha B \quad (3.2)$$

When * represents division, it is necessary to satisfy the condition $0 \notin {}^\alpha B$ for all $\alpha \in]0,1]$. The four basic arithmetic operations on intervals are performed as indicated in Equations (3.3), (3.4), (3.5) and (3.6).

$$[a,b] + [c,d] = [a+c, b+d] \quad (3.3)$$

$$[a,b] - [c,d] = [a-d, b-c] \quad (3.4)$$

$$[a,b] \times [c,d] = [\min(a \times c, a \times d, b \times c, b \times d), \max(a \times c, a \times d, b \times c, b \times d)] \quad (3.5)$$

$$\begin{aligned} [a,b] / [c,d] &= [a,b] \times [1/d, 1/c] \\ &= [\min(a/c, a/d, b/c, b/d), \max(a/c, a/d, b/c, b/d)] \end{aligned} \quad (3.6)$$

Although, fuzzy arithmetic shows a serious drawback which consists in getting different results for the same problem depending on the form of solution procedure applied if independency restrictions are not taken into consideration (Hanss 2002). As an example consider the following expression

$$(p1+p2)/p1 \quad (3.7)$$

to be evaluated for $p1=[4, 5]$ and $p2=[2, 3]$, the application of interval arithmetic directly to the form of Equation (3.7) leads to

$$([4, 5]+[2, 3])/[4, 5]=[2, 4] \quad (3.8)$$

The expression in Equation (3.7), however, can also be rewritten in the form

$$1+(p2/p1) \quad (3.9)$$

for which the interval arithmetical evaluation then results in

$$1+([2, 3]/[4, 5])=[0.4, 0.75] \quad (3.10)$$

3.3 Defuzzification

Fuzzy quantities may be converted into crisp numbers using several defuzzification methods described in the literature (Ross 1995; Klir and Yuan 1995). Within this dissertation three defuzzification methods are used: 1) for a fuzzy quantity A defined on a universal set X, the first of maxima (*fom*) defuzzification method is given by Equation (3.11) (where inf stands for infimum and sup stands for supremum) and returns the smallest value of the core of the fuzzy quantity; 2) the last of maxima (*lom*) is given by Equation (3.12) and returns the largest value of the core of fuzzy quantity A; and 3) the centroid method is given by Equation (3.13) and returns the center of gravity of a fuzzy quantity (Ross 1995).

$$fom_A = \inf \left\{ x \in X : \mu_A(x) = \sup_{x \in A} (\mu_A(x)) \right\} \quad (3.11)$$

$$lom_A = \sup \left\{ x \in X : \mu_A(x) = \sup_{x \in A} (\mu_A(x)) \right\} \quad (3.12)$$

$$centroid_A = \frac{\int_{-\infty}^{+\infty} \mu_A(x) x dx}{\int_{-\infty}^{+\infty} \mu_A(x) dx} \quad (3.13)$$

4 ADDRESSING REFERENCE DATA UNCERTAINTY IN THE ACCURACY ASSESSMENT OF LAND COVER MAPS THROUGH FUZZY SETS

In the previous Chapter were presented the fundamental aspects of fuzzy sets theory and the base knowledge necessary to understand the proposed methodology for addressing the uncertainty of reference data in the accuracy assessment of land cover maps through fuzzy sets, that is presented in the present Chapter. The proposed methodology for addressing the uncertainty of reference data in the accuracy assessment of land cover maps through fuzzy sets is performed in 6 steps: (1) Definition of the linguistic scales; (2) Selection of a set of control sites; (3) Assigning reference information to the control sites; (4) Modeling the linguistic values with fuzzy intervals; (5) Elaboration of fuzzy confusion matrices; and (6) Computation of the fuzzy accuracy measures.

The proposed methodology for addressing the uncertainty of reference data in the accuracy assessment of land cover maps through fuzzy sets presented in this Chapter, originated the submission of two papers to International Journal of Remote Sensing (Sarmiento et al. 2013; Sarmiento et al. 2015).

4.1 Linguistic scales

When fuzzy numbers or fuzzy intervals represent linguistic concepts, such as "very small", "small", "medium", etc., we face a linguistic scale. The linguistic scale is expressed by linguistic values expressing each one a specific fuzzy number or interval defined in terms of a base variable (e.g. temperature) which are real numbers within a specific range. In this study the base variable is land cover proportion and the linguistic values of the linguistic scales used, provide information about the area proportion occupied by each land cover class at each sample site.

As the elaboration of reference databases with a crisp approach presents difficulties, the use of a linguistic scale that translates the proportion of area occupied by each land cover class at each sample site also presents some challenges, mainly due to two main problems: (1) difficulty in defining the land cover coverage threshold that separates two land cover classes (e.g. a sample observation to be classified as forest needs to have at least 40% of tree cover, else is considered shrubland or grassland); and (2) difficulty in the perception of the proportion of land cover coverage that each land cover class occupies within each sample site.

In this study is assumed that the reference database is created by photo-interpretation of aerial images. To decrease the difficulties to assign the linguistic values guidelines are provided to the photo-interpreter about the intervals of land cover proportion that ideally should correspond to each one of the linguistic values. If the uncertainty is derived from the presence of more than one class in the sample sites (mixed pixels), each linguistic value will express the degree to which the sample site is occupied by the land cover class. The number of values of the linguistic scale should take into account two aspects. On one hand the number of linguistic values should not be too high in

order to decrease the uncertainty in the assignment of the linguistic values by the photo-interpreter. On the other hand, the number of linguistic values should not be too small to prevent that each linguistic value is used to express a large range of possible proportions, which would result in high values of uncertainty about the real land cover proportion.

The assignment of the linguistic values in each sample site enables the creation of a traditional reference database by choosing at each sample site the land cover class that was assigned with the highest linguistic value, or if two alternative classes are considered, the classes corresponding to the highest and second highest linguistic values.

4.2 Conversion of linguistic scales into fuzzy intervals

Two approaches may be considered to build the fuzzy interval corresponding to each linguistic value. One considers that the photo-interpreters use the linguistic scale according to their ideal definition, and the other has into consideration the effective use of the linguistic scale by each photo-interpreter, which may vary with the interpreter, the nomenclature, the image resolution or quality of the reference data, and the characteristics of the terrain in the study area. In the case of mixed pixels, the conversion of linguistic scales into fuzzy intervals will translate the variability of the percentage of the sample site area occupied by the class when a particular linguistic value is assigned by an interpreter, and the difficulty in choosing the most adequate value of the linguistic variable.

4.2.1 Ideal case

Two possibilities were considered in this research for the ideal response of a photo-interpreter, one considering five linguistic values and the other seven. In the first case

the linguistic values used are the ones proposed by Gopal and Woodcock (1994), and the fuzzy intervals are built based on the following assumptions of an interpreter's label assignments: 1) the two extremes of the linguistic scale (i.e. "Right" and "Wrong") would only be used in cases where almost no uncertainty exists. Therefore, fuzzy numbers were used with these linguistic values with the core of "Wrong" being 0% and the core of "Right" being 100% of the pixel area. The possibility that some assignments may not correspond to this ideal case was considered by admitting the assignment of these linguistic values to situations where the pixel area covered by the class presents a variation relative to the core value of up to 10% of the pixel area, with the degree of membership to the linguistic variable decreasing linearly with the increasing distance to the core; 2) the remaining linguistic values (i.e. "Understandable but wrong"; "Reasonable or acceptable" and "Good") were considered to have the same core and support amplitudes. Since only five values of the linguistic scale are considered, the cores of these linguistic values must include a relatively large range of possible pixel area proportions. For this reason, trapezoidal membership functions were used for these three linguistic values. A core amplitude of 20% was chosen for each one of them, and partial memberships were considered for amplitudes of 10% for each side of the core, decreasing linearly with the increasing distance to the core. Since there is uncertainty regarding which linguistic value should be chosen in the transition from one linguistic value to the contiguous ones, contiguous fuzzy intervals overlap for the values of the pixel area percentage that do not correspond to full membership to the linguistic values. Figure 4.1 shows the five fuzzy intervals that correspond to the considered ideal case for the linguistic scale with five linguistic values.

Addressing reference data uncertainty in the accuracy assessment of land cover maps through fuzzy sets

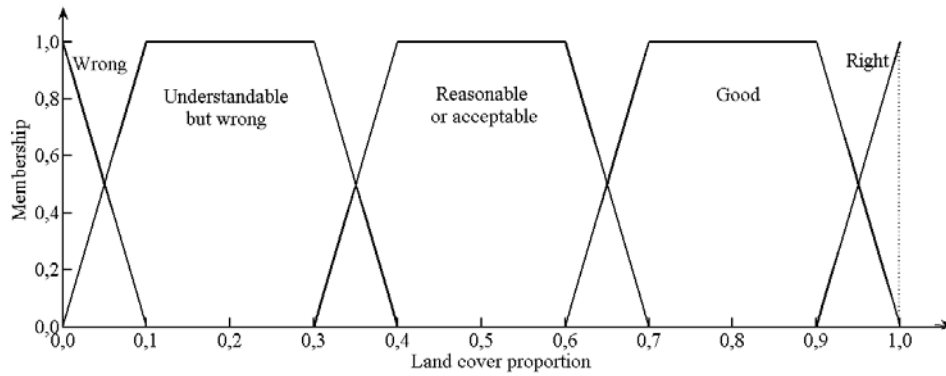


Figure 4.1 - Example of the elaboration of five fuzzy intervals that correspond to an ideal case with five linguistic values.

The other case used in this research considers the linguistic scale with seven linguistic values, where two additional values were considered in both ends of the scale which should be used when there is no uncertainty, designated respectively by "Absolutely wrong" and "Absolutely right", which correspond respectively to the crisp numbers 0 and 1. The fuzzy intervals corresponding to the other linguistic values are defined considering equal amplitudes for the core and overlap between contiguous fuzzy intervals, which to nine intervals of 0.11 proportion values. Figure 4.2 shows an example of five fuzzy intervals that correspond to an ideal case with seven linguistic values.

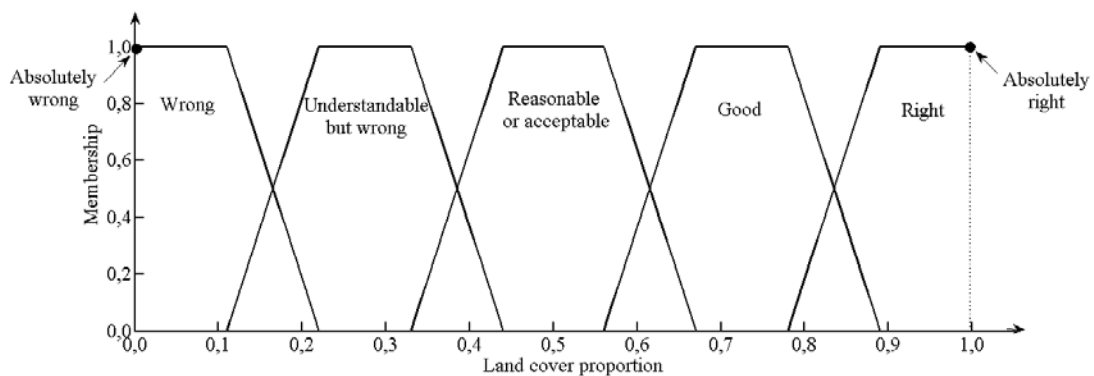


Figure 4.2 - Example of the elaboration of the five fuzzy intervals that correspond to the ideal case with seven linguistic values. The linguistic values "Absolutely wrong" and "Absolutely right" are represented by the black dots corresponding respectively to the crisp numbers 0 and 1.

4.2.2 Interpreter-derived

The fuzzy intervals defined for the ideal case represent what should be the ideal response of a photo-interpreter relative to land cover coverage of each land cover class at each sample site. Although this hypothetical response doesn't reflect necessarily the real behavior of the photo-interpreter relative to land cover coverage at each sample site. Therefore, to understand the real behavior of a photo-interpreter relative to land cover coverage, the linguistic scales may be converted in interpreter-derived fuzzy intervals. To convert the linguistic scales in interpreter-derived fuzzy intervals three steps were considered: 1) selection of a set of control sites; 2) assigning reference data to the control sites; and 3) modeling the linguistic values with interpreter-derived fuzzy intervals.

4.2.2.1 Selection of a set of control sites

To analyze the human perception and use of the linguistic scales to assess the land cover proportion, a set of control sites are selected. These sample sites are selected deterministically, taking into account that they should represent, as much as possible, a significant variety of situations in terms of land cover classes and land cover proportion, to make sure that each linguistic value is sufficiently represented.

4.2.2.2 Assigning reference information to the control sites

Since the control sites are used to control the photo-interpreter responses when using a linguistic scale, the proportion of area covered at each one by each class needs to be assessed with higher accuracy. Several approaches may be used for this aim, such as 1) using sample sites located in regions where higher spatial resolution products are available that enable the assessment of the area occupied by each land cover; 2) use a chief photo-interpreter to assign the real proportion of each land cover at each control

site.

For this assignment two fundamental aspects should be taken into account: 1) minimize the thematic uncertainty of the photo-interpreter in the interpretation of the land cover classes of the legend, and 2) assign the linguistic values considering the ideal proportion that should be associated to each linguistic value.

The photo-interpreter then assigns, for the control sample, a linguistic value to each land cover class of the legend. The comparison of the results obtained by the photo-interpreter and the chief photo-interpreter enables a preliminary control of the photo-interpreter's perception of the linguistic values in relation to the land cover proportion, and well as the detection of possible thematic errors, permitting the inclusion of additional rules to identify the land cover classes, if necessary. This step, besides enabling a calibration of the criteria used in the classification by the photo-interpreters, also enables the construction of fuzzy intervals for each photo-interpreter based on the information collected using the control sites.

4.2.2.3 Modeling the linguistic values with fuzzy intervals

The next step is the transformation of each linguistic value into a fuzzy interval, reflecting the intervals of land cover proportions that correspond to the linguistic value for the photo-interpreter. Knowing the land cover proportion assigned in the control sample to each linguistic value, it is possible to assess its variability and build a fuzzy interval that expresses that information (Sarmiento et al. 2012). The rationale behind the proposed rules is to allow partial memberships to contiguous fuzzy intervals of the proportion area values that were, for different sites, allocated to different linguistic values. This enables the representation of the uncertainty associated with the photo-interpreter estimated area proportions. The fuzzy numbers cores for each linguistic

Addressing reference data uncertainty in the accuracy assessment of land cover maps through fuzzy sets

value represent the interval of values of land cover proportion that were always assigned by the photo-interpreter to the same linguistic value, and therefore there is no uncertainty in their assignment to a linguistic value.

The general approach used for the construction of membership functions of the fuzzy intervals corresponding to each value of the linguistic scale depends on whether the intervals obtained for contiguous linguistic values in the control sample do or do not overlap.

If the intervals overlap, these intervals define the support of the fuzzy interval corresponding to each linguistic value; if they do not overlap they define the core of the fuzzy intervals. Even though the intervals generally overlap, the opposite may also occur, since the intervals depend on the characteristics of the control sites and the photo interpreters outputs.

When the intervals overlap and the fuzzy intervals support are defined first, the cores are defined for a linguistic value considering that the minimum value of the core of a fuzzy interval corresponds to the maximum value of the support of the previous fuzzy interval (e.g. the minimum value of the core for the linguistic value "Good" will match the maximum value of the support for the linguistic value "Reasonable or acceptable"). On the other hand, the maximum value of the core corresponds to the minimum value of the support of the next fuzzy interval (e.g. the maximum value of the core for the linguistic value "Understandable but wrong" matches the minimum value of the support for the linguistic value "Reasonable or acceptable"). When the intervals do not overlap, the fuzzy intervals cores are defined first and the support limits are defined as the core limit of the adjacent linguistic values. Figure 4.3 and Figure 4.4 show respectively the example of a fuzzy interval built with the prior definition of the supports and cores.

Addressing reference data uncertainty in the accuracy assessment of land cover maps through fuzzy sets

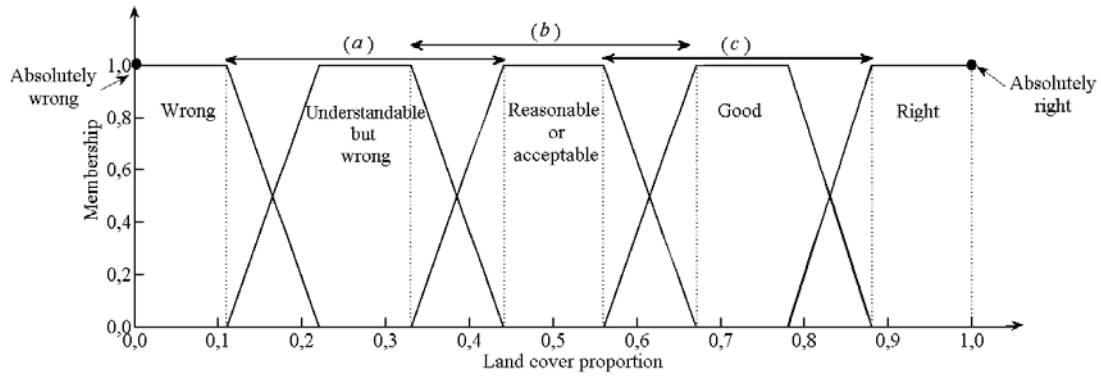


Figure 4.3 - Example of the elaboration of the five fuzzy intervals when intervals overlap. Intervals (a), (b) and (c) correspond to the intervals obtained respectively for "Understandable but wrong", "Reasonable or acceptable" and "Good", corresponding to intervals [0.11, 0.44] and [0.33, 0.67] and [0.56, 0.88].

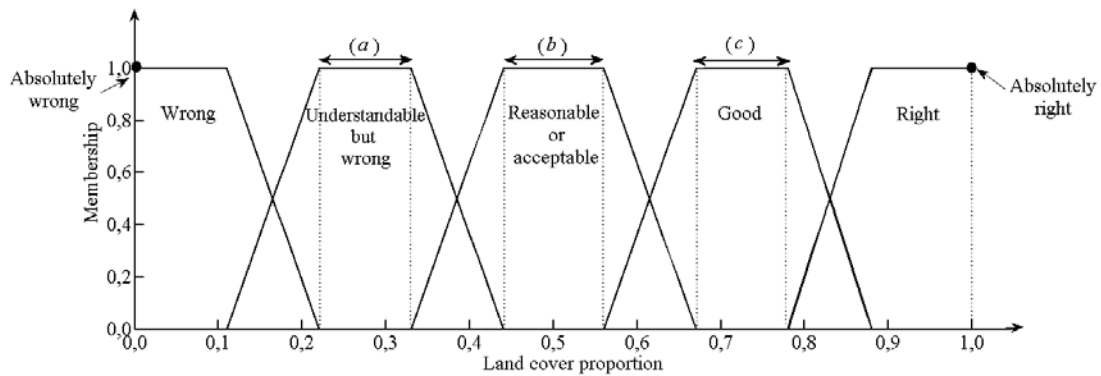


Figure 4.4 - Example of the elaboration of the five fuzzy intervals when intervals do not overlap. Intervals (a), (b) and (c) correspond to the intervals obtained respectively for "Understandable but wrong", "Reasonable or acceptable" and "Good", corresponding to intervals [0.22, 0.33], [0.44, 0.56] and [0.68, 0.78].

The rules mentioned above are applied to all linguistic values, except for the first and last values of the linguistic scale, where additional rules are necessary, since respectively no previous or next values exist. These rules are, for the situation when these first and last values are considered to have uncertainty: 1) the lowest value of the support and core of the first linguistic value are both equal to 0; 2) the largest value of the support and core of the last linguistic value are both equal to 1. When the initial and last value are considered without uncertainty, they correspond respectively to the crisp numbers zero and one.

Addressing reference data uncertainty in the accuracy assessment of land cover maps through fuzzy sets

Since some thematic interpretation errors may still exist in the assignment of the linguistic values to the control sites, due to particularly difficult cases. If the fuzzy intervals use all information provided by the control sites, only one erroneous value may result in meaningless fuzzy intervals, such as the examples shown in Figure 4.5, for a five value linguistic scale where fuzzy intervals corresponding to linguistic values that are not adjacent intersect, and Figure 4.6 where the core of two fuzzy intervals that are adjacent overlap. Therefore, it is convenient to remove from the analysis the outliers and extreme values of land cover proportion corresponding to each linguistic value. This can be done considering the 10th percentile and 90th percentile of land cover proportion for each linguistic value.

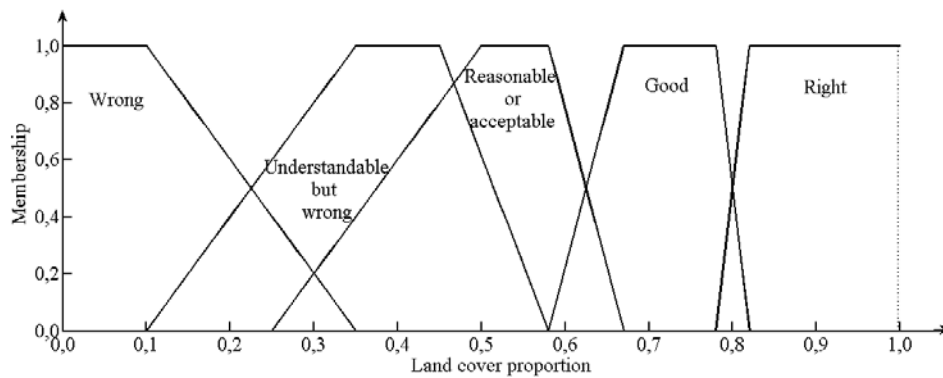


Figure 4.5 - Fuzzy interval that intersects another fuzzy interval that is not adjacent. The fuzzy interval “Reasonable or acceptable” intersects the fuzzy intervals “Understandable but wrong” and “Good” and also the fuzzy interval “Wrong”.

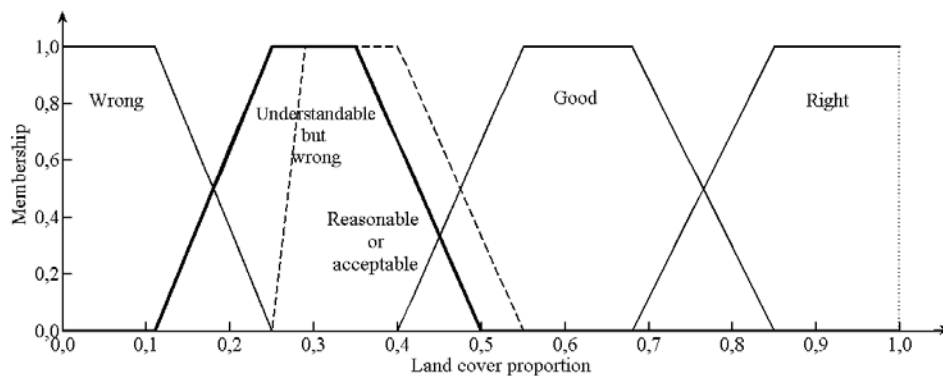


Figure 4.6 - Core overlap of two fuzzy intervals. The core of the fuzzy interval “Reasonable or acceptable” (dashed line) is partially overlapped with the core of the fuzzy interval “Understandable but wrong” (bold solid line).

4.3 Fuzzy confusion matrices

The confusion matrix has been the standard method to represent the accuracy assessment of remotely sensed data (Story and Congalton 1986). In the traditional approach, each cell of the confusion matrix indicates the number of sample pixels that were jointly assigned to the class indicated by the map (the matrix rows) and the class indicated by the reference data (the matrix columns). With the proposed approach, a sample pixel will have a value of the linguistic scale for each of the classes in the legend, and consequently each sample pixel may contribute to more than one column of the confusion matrix. For example, if a sample pixel has a linguistic value “Good” assigned to class 1 and a linguistic value “Reasonable or acceptable” assigned to class 2, this sample pixel will contribute to the columns of the confusion matrix corresponding to classes 1 and 2 (i.e. a “partial count” will apply to both columns).

For each sample pixel, the fuzzy intervals corresponding to each class in the reference database are placed in the appropriate row and column and added with all fuzzy intervals placed in that same cell. Since the addition of fuzzy intervals generates fuzzy intervals, each cell in the confusion matrix will also be a fuzzy interval. This interval will correspond to the sum of the percentage of area occupied by the class at each sample pixel in the reference database, and the uncertainty associated with this value.

4.4 Fuzzy accuracy measures

The confusion matrix proposed in this study allows the computation of overall, user's and producer's accuracy using fuzzy intervals instead of crisp numbers and traditional arithmetic like in the approach proposed by Sarmiento et al. (2013). To compute the fuzzy accuracy measures analogous to the traditional overall, user's and producer's

accuracy presented in Section 2.2.3.1, some modifications in the formulas are necessary because of the nature of fuzzy arithmetic.

When applying the traditional formulas with fuzzy intervals and arithmetic to compute the accuracy indices, the fuzzy interval in the main diagonal appears in the numerator and the denominator of the ratios for user's and producer's accuracy. Since the calculations made with fuzzy arithmetic consider the combination of all possible values of all fuzzy intervals involved, the fuzzy interval of the main diagonal may take, for example, its minimum value in the numerator and its maximum value in the denominator (as well as all combinations of their possible values), which may result in obtaining a wide interval, including values larger than one. This however makes no sense for an accuracy measure because even though the true value of the main diagonal is not known (and is therefore represented by a fuzzy interval), it will not take different values independently when used in the numerator and the denominator of the ratios. A simple example illustrates the issue. When dividing a fuzzy number \tilde{a} by itself, for each alpha level the resulting interval will be obtained from Equation (3.6). Suppose that the 0.5 alpha level of \tilde{a} is the interval [1,2]. Then, the resulting interval will be

$$\begin{aligned} [1,2]/[1,2] &= [\min(1/1,1/2,2/1,2/2), \max(1/1,1/2,2/1,2/2)] \\ &= [\min(1,0.5,2,1), \max(1,0.5,2,1)] \\ &= [0.5,2] \end{aligned}$$

The lower limit of the interval is obtained when the 0.5 alpha level of \tilde{a} takes the value 1 when in the numerator and the value 2 when in the denominator and the higher limit of the interval is obtained when the 0.5 alpha level of \tilde{a} takes the value 2 when in the numerator and the value 1 when in the denominator. But this only makes sense if the 0.5 alpha level of \tilde{a} may take different values independently when in the numerator and the

denominator. However, if they represent the same value, the division will always be equal to one, independently of the real value of \tilde{a} . That is, these values cannot be considered non-interactive in the calculations, since this assumption introduces additional uncertainty in the obtained fuzzy intervals. To avoid performing these calculations with linked variables instead of non-interactive ones (Dubois et al. 2000c), a transformation is made in the formulas so that each fuzzy quantity appears no more than once in the accuracy formula. To achieve this aim, the formulas are inverted twice and the denominator is transformed into an addition of ratios, where one of the ratios is the division of the fuzzy interval in the main diagonal by itself, and therefore is considered to be equal to one. For the same reason, the calculation of the overall accuracy is also computed using in the denominator the crisp number of samples used in the confusion matrix, since that is a quantity known without uncertainty. Equations (4.1), (4.2) and (4.3) are the equations obtained after the application of these transformations, where f_{ij} is the fuzzy quantity in row i and column j of the fuzzy confusion matrix, m is the number of land cover classes and N the sample size (Sarmiento et al. 2013).

Addressing reference data uncertainty in the accuracy assessment of land cover maps through fuzzy sets

$$\text{Overall accuracy} = \frac{\sum_{i=1}^m f_{ii}}{N} \quad (4.1)$$

$$\text{User's accuracy} = \frac{1}{1 + \frac{\sum_{\substack{j=1 \\ j \neq i}}^m f_{ij}}{f_{ii}}} \quad (4.2)$$

$$\text{Producer's accuracy} = \frac{1}{1 + \frac{\sum_{\substack{i=1 \\ i \neq j}}^m f_{ij}}{f_{jj}}} \quad (4.3)$$

However, the user's and producer's accuracy may also be computed adopting additional modifications to the methodology proposed above (Sarmiento et al. 2013), so that their computation uses all information available that may decrease the degree of uncertainty. The main modification consists in considering that the sum of class proportions in each row of the confusion matrix cannot exceed the number of pixels in the reference data located in each map land cover class, since the classes are considered mutually exclusive, and therefore the sum of the area occupied by all land cover classes can't exceed the total area of the sample pixels considered in the reference database. Therefore the fuzzy user's accuracy is computed for each class dividing the fuzzy interval in the main diagonal by the crisp number of pixels considered in the map class. The same approach cannot be considered for the producer's accuracy, since the exact number of pixels assigned to each class in the reference database is not known without uncertainty. However, it is known that the total number of pixels assigned to all classes in the reference database cannot exceed the total number of pixels used for the accuracy assessment, since the classes are considered as mutually exclusive. Therefore, the computation of the producer's accuracy is performed considering the following steps: 1) add the fuzzy intervals for each map class of the confusion matrix (matrix rows); 2)

Addressing reference data uncertainty in the accuracy assessment of land cover maps through fuzzy sets

defuzzify the sum of the fuzzy intervals for each row using the defuzzification methods indicated in Section 3.3, and select the method that provides a value closer to the number of pixels used for each map class; 3) add the fuzzy intervals of each land cover class in the reference database (matrix columns) and defuzzify the sums with the defuzzification method found in the previous step; 4) compute the producer's accuracy per class dividing the fuzzy interval in the main diagonal of each class by the defuzzified values for each class obtained in step 3).

The user's, producer's and overall accuracy are expressed as fuzzy intervals, representing in this case the percentage of the correctly classified regions of the sample sites. Therefore, the uncertainty associated to these values may be assessed considering several levels of confidence, corresponding to the several alpha levels. The values more likely to be correct correspond to the core of the fuzzy numbers. Thus, the amplitude of the core (*lom-fom*) provides the degree of uncertainty of the values with the highest levels of confidence. The support corresponds to the range of values that are possible, even though some are less likely to be the real value, and so the amplitude of the support indicates the uncertainty associated to the values considering the lowest values of confidence. Pessimistic perspectives may be obtained considering the lower value of the alpha cuts of the desired level of confidence, and optimistic perspectives considering the largest value of each of these intervals.

Defuzzification approaches can then be used to transform the obtained fuzzy accuracy measures into crisp values. These defuzzified accuracy values provide a simpler characterization in terms of a single number for each accuracy measure and these defuzzified values can be easily compared to traditional accuracy measures. In this dissertation three methods are considered, namely the first of maxima (*fom*), the last of

Addressing reference data uncertainty in the accuracy assessment of land cover maps through fuzzy sets

maxima (*lom*) and the centroid method indicated in Section 3.3. In Figure 4.7 is presented a flowchart of the proposed methodology of accuracy assessment of land cover maps.

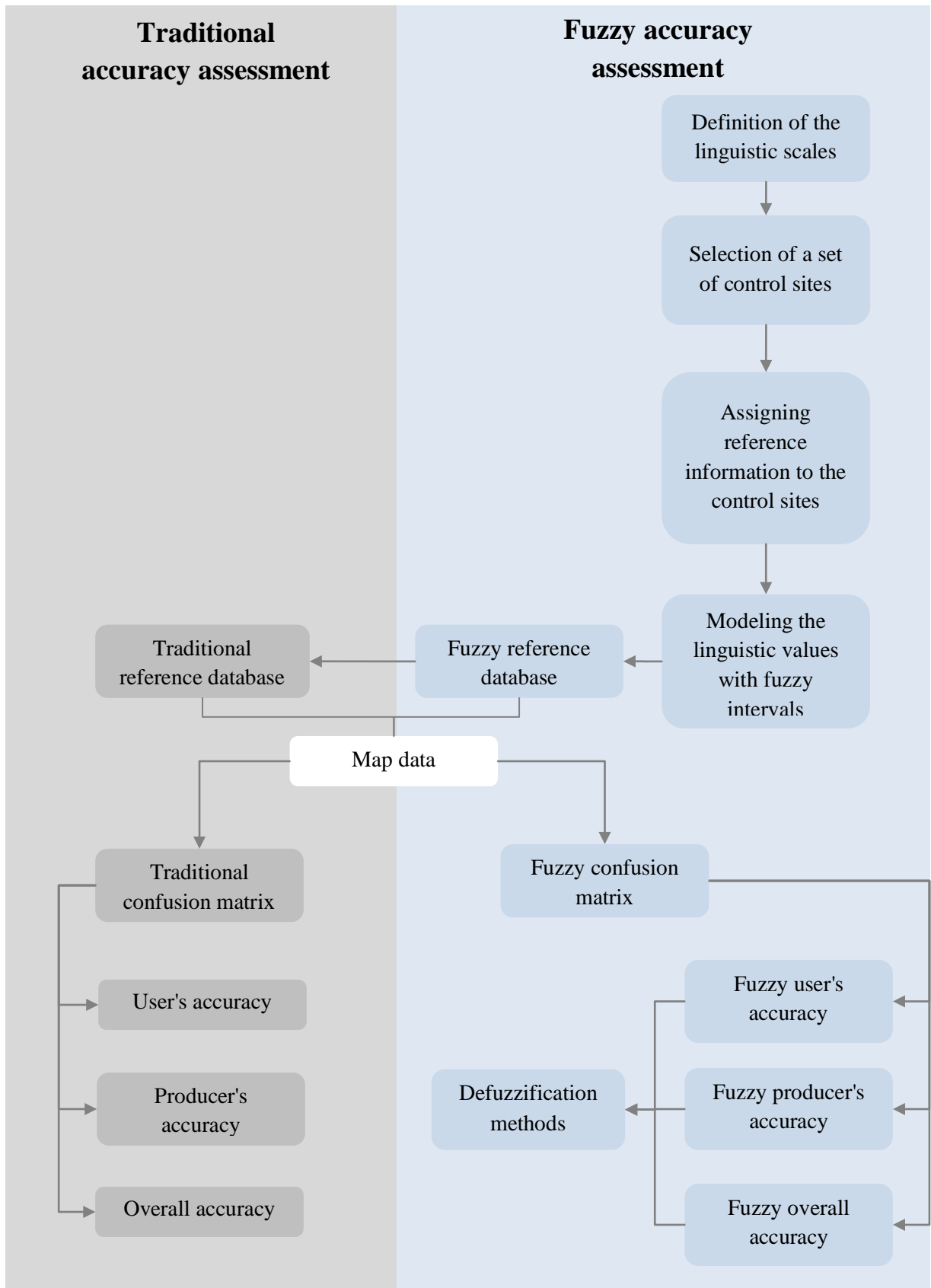


Figure 4.7 - Flowchart of the proposed methodology of accuracy assessment of land cover maps.

5 ADDRESSING REFERENCE DATA UNCERTAINTY IN THE ACCURACY ASSESSMENT OF A MAP OF CONTINENTAL PORTUGAL

In this Chapter the proposed approach for addressing the uncertainty of reference databases in the accuracy assessment of land cover maps presented in the previous Chapter, is applied to a map of Continental Portugal. A comparison between the accuracy results achieved with the ideal and interpreter-derived fuzzy intervals for five and seven values linguistic scales is made. The obtained fuzzy accuracy measures are also compared with traditional accuracy measures and a comparison is performed between the accuracy results obtained by two photo-interpreters.

5.1 Data, nomenclature and map classification

A map of Continental Portugal derived from a set of six bimonthly image composites of 2005 from MERIS (Medium Resolution Imaging Spectrometer) was used to demonstrate the proposed accuracy assessment methodology. These images have a 300 m spatial resolution and 13 spectral bands. The map was produced with two different techniques: 1) a linear discriminant classifier was used to classify the intra-

Addressing reference data uncertainty in the accuracy assessment of a map of Continental Portugal

annual stationary land cover classes; 2) a vegetation index differentiating to classify intra-annual transient land cover classes (Carrão et al. 2010). This map was originally produced with 16 land cover classes, based on LANDEO nomenclature (Carrão et al. 2010). To decrease the number of classes used in this case study a generalization of the initial classification was done, considering a nomenclature with five land cover classes, namely urban areas (UA), agriculture (AG), natural vegetation (NV), forest (F) and water and wetlands (WW). Figure 5.1 shows the resulting generalized land cover map.

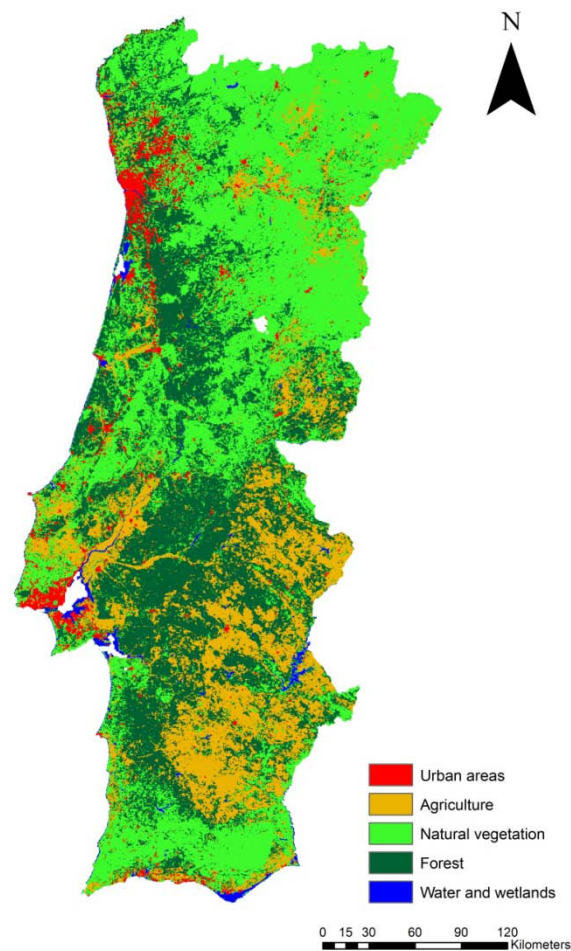


Figure 5.1 - Map for Continental Portugal with 5 land cover classes.

Table 5.1 shows the procedure used to perform the generalization of the original map and the description of each land cover class of the LANDEO nomenclature.

*Addressing reference data uncertainty in the accuracy assessment of a map of
Continental Portugal*

Table 5.1 - Generalization of the LANDEO nomenclature.

LANDEO generalization	LANDEO land cover class	Description
Urban areas	Continuous artificial areas	The land cover consists of artificial areas (e.g. buildings, roads). At least 80% of the total surface must be impermeable.
	Discontinuous artificial areas	The land cover consists of artificial areas (e.g. buildings, roads (including slopes and berm of the roads). Includes gardens that are adjacent to houses. Between 30-80% of the total surface must be impermeable.
Agriculture	Non-irrigated herbaceous crops	The land cover consists of rainfed herbaceous crops. These crops are annually harvested and followed by a bare soil period.
	Irrigated herbaceous crops	The land cover consists of irrigated herbaceous crops. These crops are annually harvested and followed by a bare soil period.
	Rice crops	The land cover consists of rice crops. These crops are annually harvested and followed by a bare soil period.
	Vineyards	The land cover consists of permanent deciduous crops, namely vineyards.
Natural vegetation	Agro-forestry areas	The land cover consists of broadleaf trees with at least 5 m height and with a crown cover between 15-40% and with understory agricultural systems.
	Bare to sparsely vegetated areas	The land cover consists of natural areas with less than 15% of vegetation cover during all time of year. It includes areas like bare rock and sands.
	Shrubland	The land cover consists of woody vegetation (shrubs) with more than 15% cover and with less than 5 m height. Tree cover is less than 40%.
	Grassland	The land cover consists of natural and cultivated herbaceous vegetation for livestock feeding with more than 15% cover. Tree and shrub cover is less than 40%.
Forest	Burnt areas and clear cuts	The main layer consists of closed to open trees or shrubs affected by forest fires or clear cuts.
	Broadleaf forest	The land cover consists of broadleaf trees with at least 5 m height and with a crown cover of more than 40%.
	Needleleaf forest	The land cover consists of needleleaf trees with at least 5 m height and with a crown cover of more than 40%.
Water & wetlands	Mixed forest	The land cover consists of a mixture or mosaic of broadleaf and needleleaf trees with at least 5 m height and with a crown cover of more than 40%. None of the tree types exceeds 80% of the mixture or mosaic.
	Water	The land cover consists of natural/artificial water bodies. Can be either fresh or salt-water bodies.
	Wetlands	The land cover consists of a permanent mixture of water and vegetation. The vegetation can be present in salt, brackish, or fresh water.

5.2 Definition of the linguistic scales

In this study two linguistic scales were used: 1) a five value linguistic scale, where all values were considered to include uncertainty, which propagates to the accuracy results; 2) a linguistic scale with seven linguistic values, that does not account for uncertainty for the first and last value, allowing the consideration of situations in which pixels are a

*Addressing reference data uncertainty in the accuracy assessment of a map of
Continental Portugal*

perfect match ("Absolutely right") or mismatch ("Absolutely wrong") to one class. This aspect enables the consideration of no uncertainty, when in fact there is none. The use of a linguistic scale with seven values is not more complex for the photo-interpreter because all that is required is identifying the perfect matches and mismatches. The two linguistic scales used in this study are presented in Table 5.2. The linguistic values "Absolutely wrong" and "Absolutely right" for the linguistic scale with seven linguistic values, were considered to have no uncertainty.

An order was associated to the values of the linguistic scales, from 1 to 5 for the linguistic scale with five linguistic values, and from 1 to 7 for the linguistic scale with seven linguistic values. The sum of values of the order of the linguistic values in each sample site must be close to 9 when the linguistic scale with five linguistic values is used, and close to 11 when the linguistic scale with seven linguistic values is used. The sample sites for which major deviations from these values were detected, were inspected and their classification corrected.

Table 5.2 - Linguistic values and respective order. Land cover coverage defined for each linguistic value of the linguistic scales with five and seven values.

Linguistic scales	Order	Linguistic values	Land cover coverage (%)
Five linguistic values	1	Wrong (W)	0 - 5
	2	Understandable but wrong (U)	5 - 35
	3	Reasonable or acceptable (A)	35 - 65
	4	Good (G)	65 - 95
	5	Right (R)	95 - 100
Seven linguistic values	1	Absolutely wrong (AW)	0
	2	Wrong (W)	0 - 16
	3	Understandable but wrong (U)	16 - 38
	4	Reasonable or acceptable (A)	38 - 61
	5	Good (G)	61 - 83
	6	Right (R)	83 - 100
	7	Absolutely right (AR)	100

5.3 Conversion of linguistic scales into fuzzy intervals

5.3.1 Selection of a set of control sites

To analyze the human perception and the use of the linguistic scales to assess the land cover proportion at the control sites, a set of 32 control sites were deterministically selected. For each one of these sites, the photo-interpreter had to assign a linguistic value to each one of the land cover classes, taking into consideration the description of the land cover classes (Table 5.1) and the description of the linguistic values shown in Table 5.2, which include information about land cover proportion intervals that are supposed to be associated to each linguistic value. These intervals correspond to the values of land cover proportion where there is no overlap between the fuzzy intervals defined for each linguistic value for the ideal case.

5.3.2 Assigning reference information to the control sites

In this study a chief photo-interpreter determined the land cover proportion of each class at each control site considering a set of 100 points systematically overlaid to the control sites, assigning a land cover coverage to each point (Figure 5.2). The decision to use 100 points per control site was based on the fact that if the true proportion of area of a land-cover class for the site is p , the standard error for the estimated proportion from the sample would be the square root of $p(1-p)/100$. Consequently, for examples of p of 0.5, 0.25, and 0.10, the standard errors would be 0.05, 0.043, and 0.03.

The information about land cover was collected through the observation of aerial images. These aerial images have a spatial resolution of 0.5 m and four spectral bands, corresponding to the years 2004, 2005 and 2006.

*Addressing reference data uncertainty in the accuracy assessment of a map of
Continental Portugal*

This allows a link to be established between the linguistic values assigned to each land cover class at the control sites and the area proportion occupied by the land cover classes. Since this is done for all control sites, it enables the assessment of the variability of the area proportion that the photo-interpreter assigns to each linguistic value.

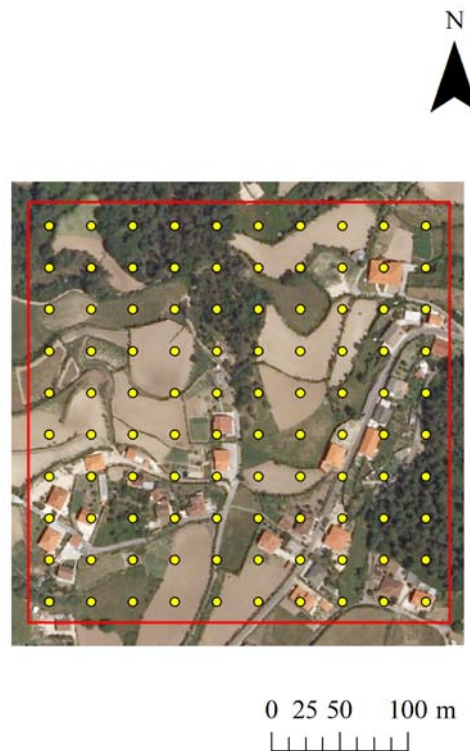


Figure 5.2 - Control site overlaid with 100 points systematically distributed.

Table 5.3 shows an example of the information collected by the photo-interpreter and the chief photo-interpreter in the control sites classifying the 100 systematically distributed points represented in Figure 5.2. In this case, for example for urban areas, the photo-interpreter assigned the linguistic value "Understandable but wrong" (U) and the chief photo-interpreter considered that 21 points belonged to this class, out of the 100.

Addressing reference data uncertainty in the accuracy assessment of a map of Continental Portugal

Table 5.3 - Example of the linguistic values collected in a control site by the photo-interpreter and the land cover percent collected by the chief photo-interpreter. U: "Understandable but wrong"; G: "Good"; AW: "Absolutely wrong"; W: "Wrong".

	Land cover classes				
	<i>UA</i>	<i>AG</i>	<i>NV</i>	<i>F</i>	<i>WW</i>
Linguistic values	U	G	AW	W	AW
Land cover (%)	21	63	0	16	0

5.3.3 Modeling the linguistic values with fuzzy intervals

The fuzzy intervals corresponding to each of the linguistic values were built using the methodology described in Section 4.2.2.3 for the linguistic scales with five (Figure 5.3 and Table 5.4) and seven values (Figure 5.4 and Table 5.5).

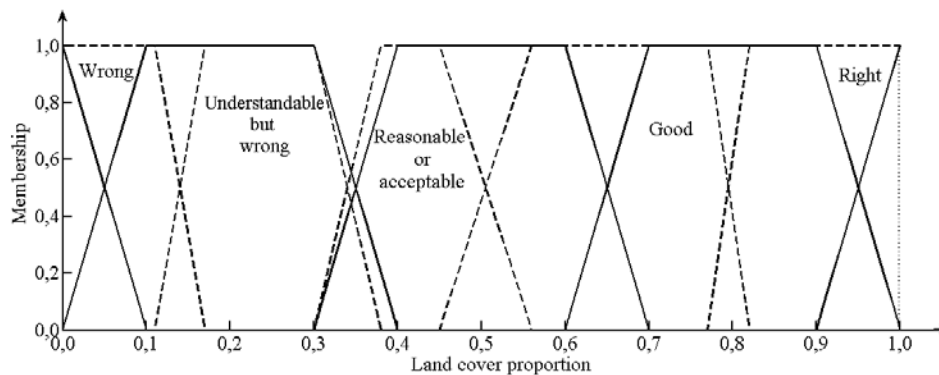


Figure 5.3 - Interpreter-derived fuzzy intervals obtained using a linguistic scale with five linguistic values (dashed line) and fuzzy intervals for an ideal case with five linguistic values (solid line).

Table 5.4 - Land cover proportion intervals obtained for the core and support of the ideal case and interpreter-derived fuzzy intervals with five linguistic values.

Fuzzy intervals	Linguistic values	Support	Core
Ideal case	Wrong (W)	[0, 0.1]	[0, 0]
	Understandable but wrong (U)	[0, 0.4]	[0.1, 0.3]
	Reasonable or acceptable (A)	[0.3, 0.7]	[0.4, 0.6]
	Good (G)	[0.6, 1]	[0.7, 0.9]
	Right (R)	[0.9, 1]	[1, 1]
Interpreter derived	Wrong (W)	[0, 0.17]	[0, 0.11]
	Understandable but wrong (U)	[0.11, 0.38]	[0.17, 0.30]
	Reasonable or acceptable (A)	[0.30, 0.56]	[0.38, 0.45]
	Good (G)	[0.45, 0.82]	[0.56, 0.77]
	Right (R)	[0.77, 1]	[0.82, 1]

Addressing reference data uncertainty in the accuracy assessment of a map of Continental Portugal

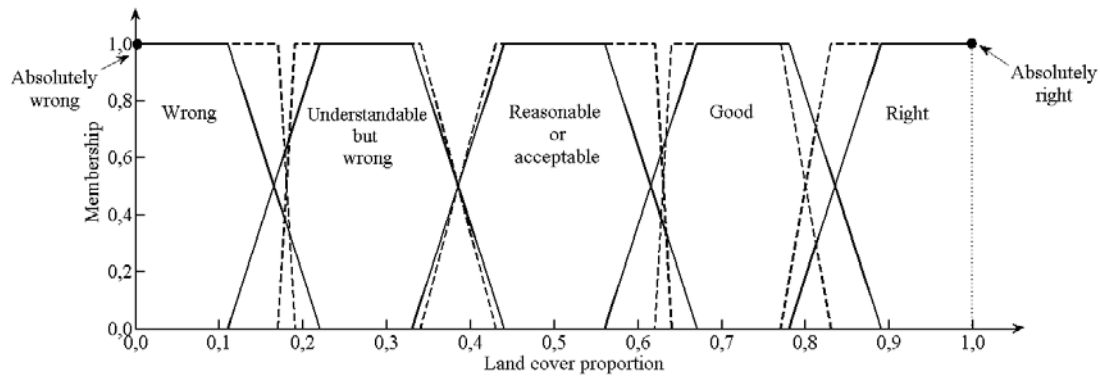


Figure 5.4 - Interpreter-derived fuzzy intervals obtained using a linguistic scale with seven linguistic values (dashed line) and fuzzy intervals for an ideal case with seven linguistic values (solid line).

Table 5.5 - Land cover proportion intervals obtained for the core and support of the ideal case and interpreter-derived fuzzy intervals with seven linguistic values.

Fuzzy intervals	Linguistic values	Support	Core
Ideal case	Absolutely wrong (AW)	[0, 0]	[0, 0]
	Wrong (W)	[0, 0.22]	[0, 0.11]
	Understandable but wrong (U)	[0.11, 0.44]	[0.22, 0.33]
	Reasonable or acceptable (A)	[0.33, 0.67]	[0.44, 0.56]
	Good (G)	[0.56, 0.89]	[0.67, 0.78]
	Right (R)	[0.78, 1]	[0.89, 1]
	Absolutely right (AR)	[1, 1]	[1, 1]
Interpreter derived	Absolutely wrong (AW)	[0, 0]	[0, 0]
	Wrong (W)	[0, 0.19]	[0, 0.17]
	Understandable but wrong (U)	[0.17, 0.43]	[0.19, 0.34]
	Reasonable or acceptable (A)	[0.34, 0.64]	[0.43, 0.62]
	Good (G)	[0.62, 0.83]	[0.64, 0.77]
	Right (R)	[0.77, 1]	[0.83, 1]
	Absolutely right (AR)	[1, 1]	[1, 1]

An analysis of the fuzzy intervals obtained with the five values linguistic scale shows that there are substantial differences between the ideal fuzzy intervals and the interpreter derived fuzzy intervals (Figure 5.3). The major differences are in the extremes of the linguistic scale, namely "Wrong" and "Right". For example, in the case of "Wrong", for the ideal case the core is 0, while for the interpreter-derived fuzzy interval the core is the interval [0, 0.11]. For "Right" the core for the ideal case is 1, while for the interpreter-derived fuzzy interval the core is the interval [0.82, 1]. This reflects that the photo-interpreter usually used these extreme values for more area values than the ones

expected in the situation considered as ideal. As a consequence, for the intermediate linguistic values the cores of the interpreter-derived fuzzy intervals are smaller than for the ideal case. For example, the core of the linguistic value "Reasonable or acceptable", for the interpreter-derived fuzzy interval has an amplitude of 0.07, while for the ideal case has an amplitude of 0.2. Another important aspect is the core of linguistic value "Good" that for the interpreter-derived fuzzy interval overlaps part of the core of the linguistic value "Reasonable or acceptable" for the ideal case.

On the contrary, the interpreter-derived fuzzy intervals obtained for the seven values linguistic scale (Figure 5.4) are very similar to the ideal ones. The main differences are in the linguistic values "Wrong" and "Right". For "Wrong" the core of the interpreter-derived fuzzy interval is the interval $[0, 0.17]$ while for the ideal case is the interval $[0, 0.11]$, and for "Right" the core of the interpreter-derived fuzzy interval is the interval $[0.83, 1]$ while for the ideal case is the interval $[0.89, 1]$, corresponding in both cases to a difference of only 0.06. It can also be seen that the region of transition between fuzzy intervals is smaller for the interpreter-derived fuzzy intervals than for the fuzzy intervals for the ideal case.

5.4 Elaboration of the reference databases

After the analysis of the control sites, a simple random sample of 250 sites was used to build two reference datasets, to be used in the accuracy assessment, one, based on five and the other on seven linguistic values. Each sample site has the same dimension of a MERIS pixel (300 x 300 m) and the information about land cover was collected through the observation of the same aerial images used for the assignment of the reference information to the control sites. At each sample site of the two reference databases, the photo-interpreter assigned one of the five and one of the seven linguistic values for each

*Addressing reference data uncertainty in the accuracy assessment of a map of
Continental Portugal*

one of the land cover classes of the legend. Figure 5.5 shows one sample site overlaid with the aerial image, where there is a clear difficulty in assigning one class to the pixel, and even to linguistic values. Table 5.6 shows an example of the information collected for eight sample sites.



Figure 5.5 - Sample site of the reference database overlaid with the aerial image.

Table 5.6 - Extract of the reference database with eight sample sites (n=8); five land cover classes, namely urban areas (UA), agriculture (AG), natural vegetation (NV), forest (F) and water and wetlands (WW); and five linguistic values: "Wrong" (W), "Understandable but wrong" (U), "Reasonable or acceptable" (A), "Good" (G), "Right" (R).

Site	Map class	Reference class				
		UA	AG	NV	F	WW
1	F	U	A	W	G	W
2	UA	G	U	W	U	W
3	UA	G	A	W	W	W
4	UA	G	U	A	W	W
5	NV	W	R	W	U	W
6	AG	W	R	U	W	W
7	NV	W	W	R	W	W
8	WW	W	W	W	W	R

5.5 Accuracy assessment with the ideal and interpreter-derived fuzzy intervals

Fuzzy confusion matrices were built with the reference data for the 250 sample observations using five and seven linguistic values and both with the interpreter-derived fuzzy intervals or the ones considered ideal, to assess if there are considerable differences in the accuracy results when using one or the other.

Table 5.7 and Table 5.8 show the fuzzy confusion matrices obtained for the considered photo-interpreter with the ideal and the interpreter-derived fuzzy intervals using five linguistic values, and Table 5.9 and Table 5.10 show the fuzzy confusion matrices obtained for the seven linguistic values.

Table 5.7 - Fuzzy confusion matrix for the photo-interpreter elaborated with the fuzzy intervals for the ideal case with five linguistic values.

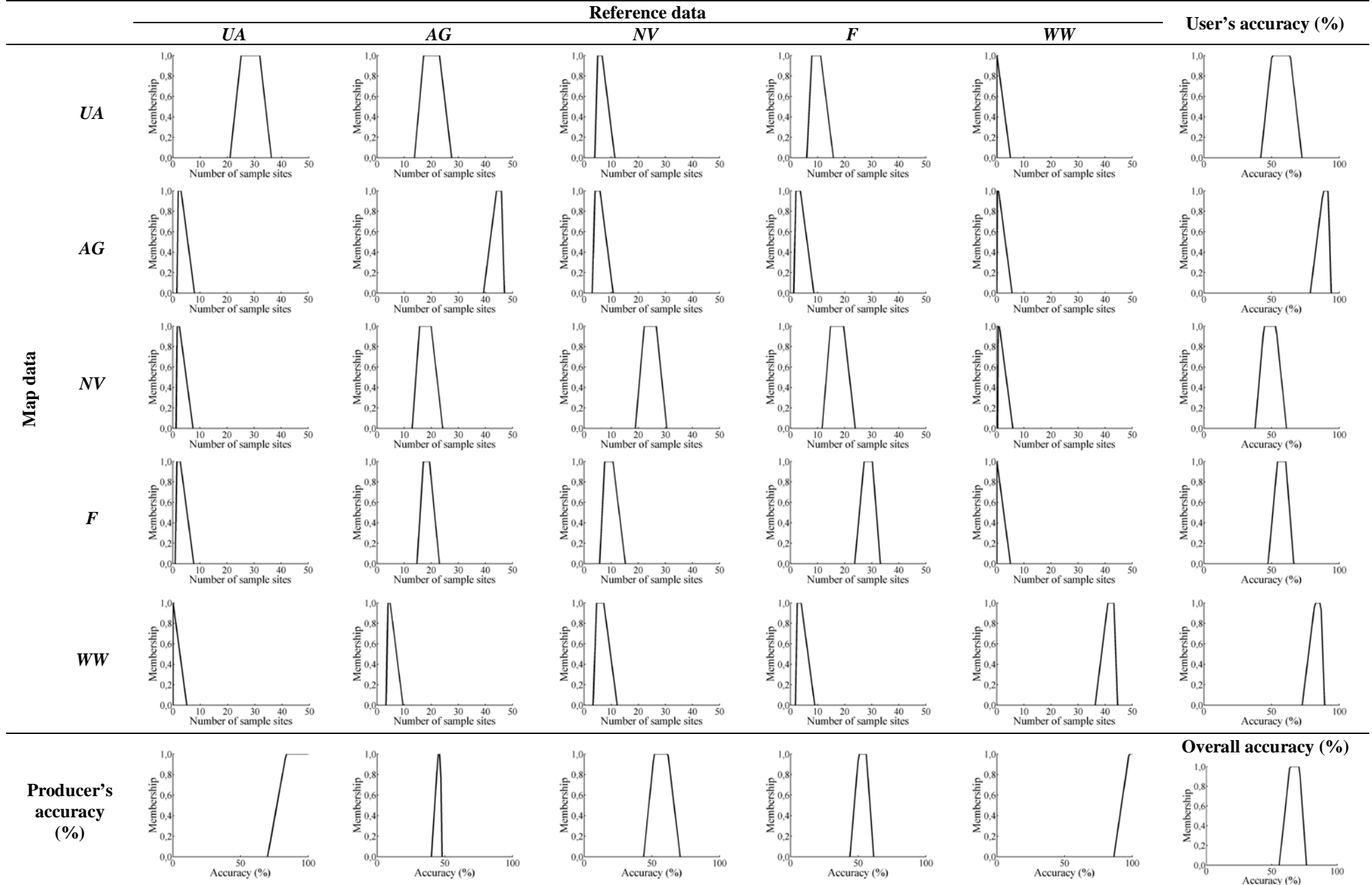


Table 5.8 - Fuzzy confusion matrix for the photo-interpreter elaborated with the interpreter-derived fuzzy intervals with five linguistic values.

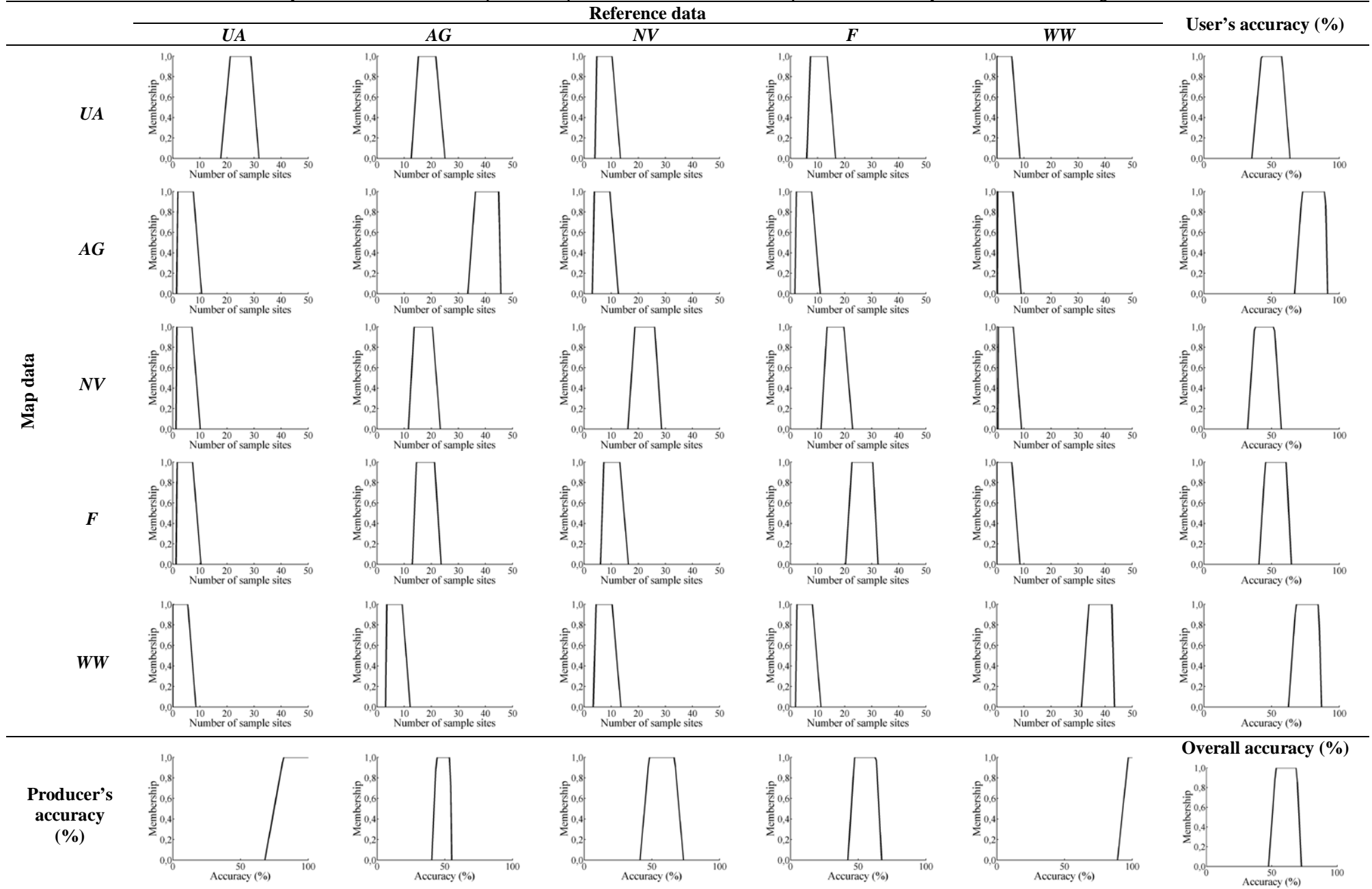


Table 5.9 - Fuzzy confusion matrix for the photo-interpreter elaborated with the fuzzy intervals for the ideal case with seven linguistic values.

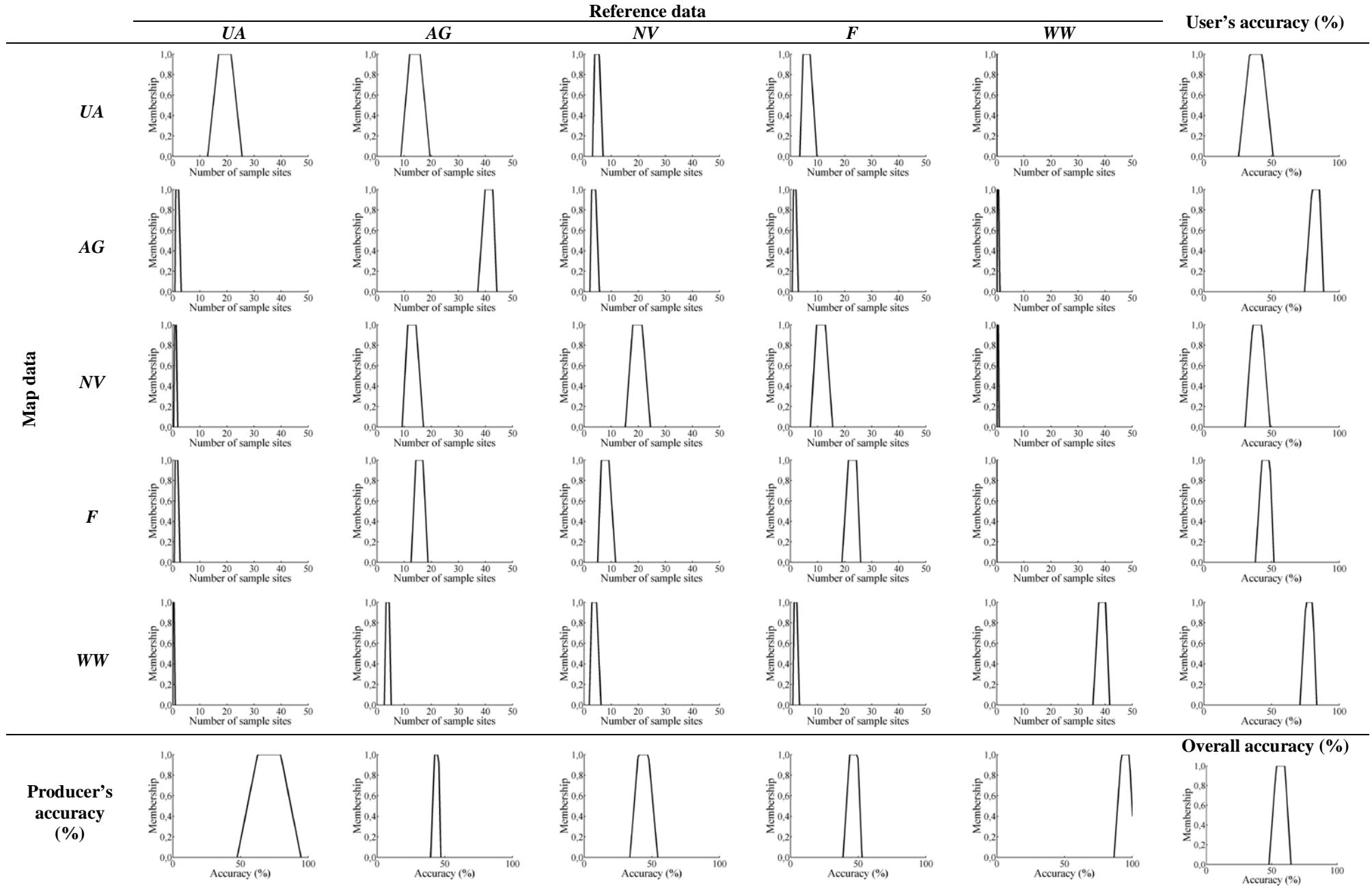
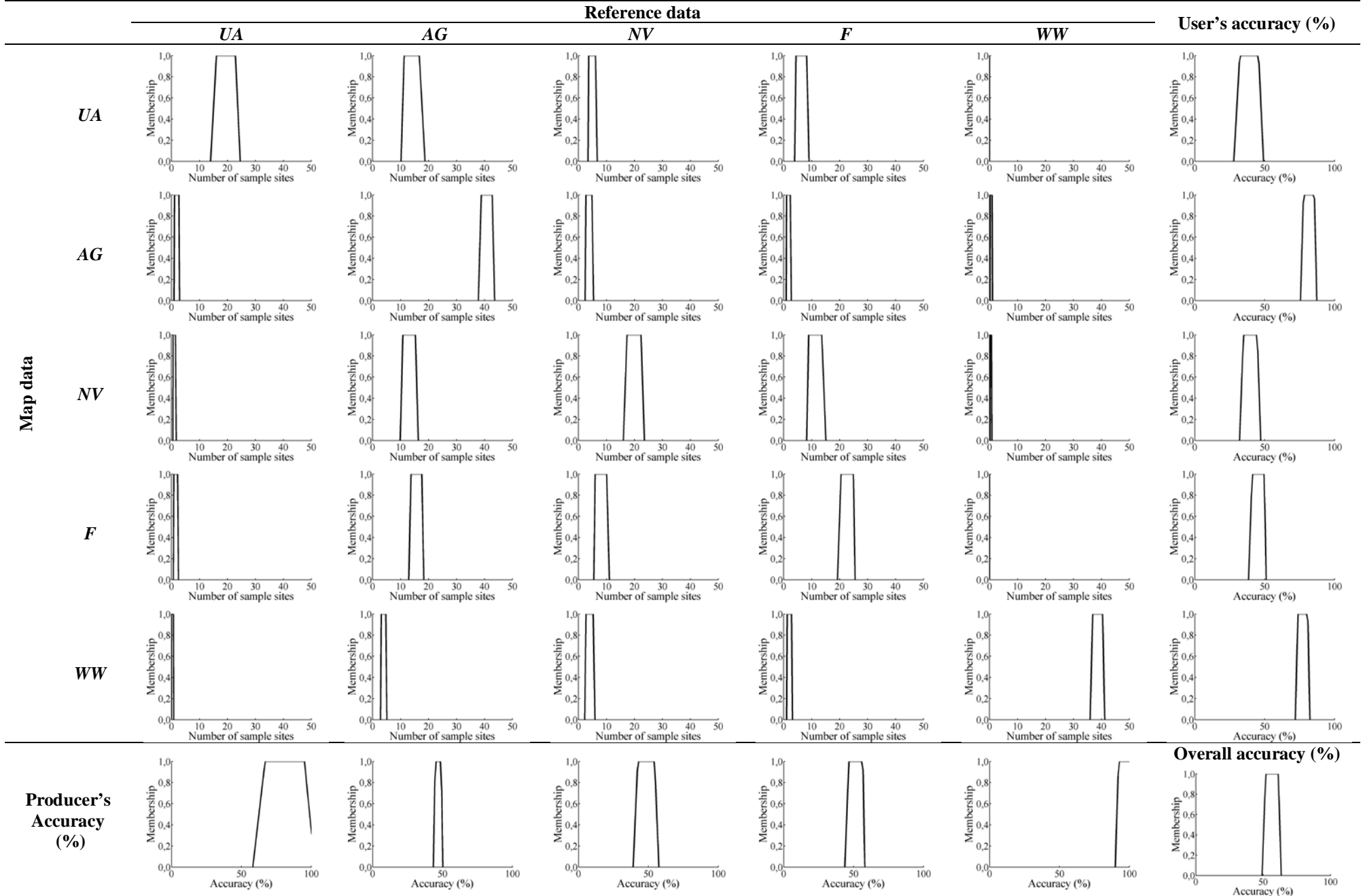


Table 5.10 - Fuzzy confusion matrix for the photo-interpreter elaborated with the interpreter-derived fuzzy intervals with seven linguistic values.



*Addressing reference data uncertainty in the accuracy assessment of a map of
Continental Portugal*

Table 5.11 - Support and core intervals of the fuzzy thematic accuracy measures obtained with the ideal and empirical fuzzy intervals using five linguistic values and seven linguistic values. Thematic accuracy measures obtained with the centroid (*C*).

Fuzzy intervals	Land cover classes	Five linguistic values			Seven linguistic values			
		<i>Support</i>	<i>Core</i>	<i>C</i>	<i>Support</i>	<i>Core</i>	<i>C</i>	
Ideal case	User's accuracy (%)	<i>UA</i>]42, 72[[51, 63]	57]26, 51[[35, 43]	38
		<i>AG</i>]79, 94[[89, 92]	87]74, 88[[80, 85]	82
		<i>NV</i>]38, 61[[45, 53]	49]30, 49[[36, 42]	40
		<i>F</i>]47, 66[[54, 60]	57]38, 52[[43, 48]	45
		<i>WW</i>]73, 89[[83, 86]	82]71, 83[[76, 80]	77
	Producer's accuracy (%)	<i>UA</i>]70, 100[[84, 100]	95]48, 95[[64, 80]	71
		<i>AG</i>]40, 47[[45, 46]	45]40, 47[[43, 45]	44
		<i>NV</i>]44, 71[[52, 61]	57]34, 54[[41, 47]	44
		<i>F</i>]44, 62[[51, 56]	53]39, 53[[44, 49]	46
		<i>WW</i>]89, 100[[100, 100]	100]87, 100[[93, 98]	94
Overall accuracy]56, 77[[65, 71]	67]48, 65[[54, 60]	57	
Interpreter derived	User's accuracy (%)	<i>UA</i>]35, 64[[43, 57]	50]28, 49[[33, 45]	39
		<i>AG</i>]67, 91[[73, 89]	80]76, 87[[79, 85]	82
		<i>NV</i>]32, 57[[38, 51]	45]32, 47[[35, 44]	40
		<i>F</i>]41, 65[[46, 61]	53]39, 51[[42, 50]	45
		<i>WW</i>]62, 87[[68, 84]	76]72, 82[[74, 80]	77
	Producer's accuracy (%)	<i>UA</i>]68, 100[[82, 100]	96]58, 100[[67, 95]	80
		<i>AG</i>]40, 55[[44, 53]	48]43, 50[[45, 48]	46
		<i>NV</i>]41, 73[[48, 66]	57]39, 57[[43, 54]	48
		<i>F</i>]42, 67[[47, 62]	55]44, 58[[47, 56]	51
		<i>WW</i>]89, 100[[97, 100]	100]90, 100[[93, 100]	97
Overall accuracy (%)]48, 73[[54, 69]	61]49, 63[[52, 61]	57	

In a general overview of the fuzzy confusion matrices obtained for the five value linguistic scale, we can state that the fuzzy intervals inside the matrix elaborated with the fuzzy intervals for the ideal case have core amplitudes smaller when compared with the ones obtained with the interpreter-derived fuzzy intervals. Indeed this reflects the difference between the fuzzy interval cores that correspond to the extreme values of the linguistic scale with five linguistic values, namely "Wrong" and "Right", since for the ideal case the core is a real number and for the interpreter-derived fuzzy interval the core is an interval (Figure 5.3).

These differences in the core and support amplitudes are reflected in some fuzzy thematic accuracy measures obtained, as shown in Table 5.11. The major differences in the fuzzy user's accuracy were obtained for the land cover classes AG and WW. For

*Addressing reference data uncertainty in the accuracy assessment of a map of
Continental Portugal*

example, the support and core of the fuzzy user's accuracy of AG obtained with the fuzzy intervals for the ideal case, are respectively the intervals [79, 94] and [89, 92], while with the interpreter-derived fuzzy intervals the values are respectively [67, 91] and [73, 89], which correspond to lower values of accuracy and larger uncertainty. The same occurs for the class WW.

Relative to fuzzy producer's accuracy the major differences of the support and core between the results obtained with the interpreter-derived and the ideal case fuzzy intervals were for the classes of AG and F. For AG, the fuzzy producer's accuracy with the fuzzy intervals for the ideal case has a support interval between [40, 47] (amplitude of 7%) and a core interval of [45, 46] (amplitude of 1%), while with the interpreter-derived fuzzy intervals, AG presents a support interval of [40, 55] (amplitude of 15%), and a core interval of [44, 53] (amplitude of 9%), corresponding the values obtained with the user-derived fuzzy intervals to larger uncertainty.

Regarding the results obtained using the centroid defuzzification method, the user's accuracy obtained with the ideal fuzzy intervals are higher for all land cover classes than the ones obtained with the interpreter-derived fuzzy intervals. Moreover, the uncertainty of these values is also smaller, since the amplitude of most fuzzy intervals is smaller. However, the ideal fuzzy intervals do not reflect the reality of how linguistic values were assigned by the photo-interpreter.

The comparison of both fuzzy confusion matrices elaborated with seven linguistic values shows that they are very similar, as seen on Table 5.9, Table 5.10 and Table 5.11. Indeed this results from the similarity between the interpreter-derived and the ideal fuzzy intervals using the seven linguistic values (Figure 5.4). This similarity between the thematic accuracy measures when the ideal and interpreter-derived fuzzy intervals

are used is also expressed by the defuzzification with the centroid method, since it is practically the same in both cases, especially with the fuzzy user's accuracy where all values are equal except for UA land cover class with only 1% difference between the result obtained with the ideal case and the interpreter-derived fuzzy intervals.

Considering the fuzzy producer's accuracy, the major differences are for UA. UA for the ideal case presents a support interval of [48, 95] (amplitude of 47%) and a core interval of [64, 80] (amplitude of 16%) and for the interpreter-derived fuzzy intervals the support interval is [58, 100] (amplitude of 42%) and the core interval is [67, 95] (amplitude of 28%). The cores of the fuzzy thematic accuracy measures for the ideal case are smaller than the ones obtained with the interpreter-derived fuzzy intervals. However, this aspect for the support is exactly the opposite: the supports of the fuzzy thematic accuracy measures for the ideal case are wider than the supports obtained with the interpreter-derived fuzzy intervals.

5.6 Comparison of the fuzzy accuracy measures with traditional accuracy measures

To allow a comparison between the obtained fuzzy accuracy measures and traditional accuracy measures, the fuzzy accuracy measures were defuzzified with the three methods referred in Section 3.3. The traditional accuracy measures were computed using the operators MAX and RIGHT (Gopal and Woodcock 1994). MAX considers that the map class and the reference class match when the map class corresponds to the class that was assigned with the highest linguistic value, while the RIGHT operator considers matches when the map class corresponds to the highest or second highest linguistic value (i.e. "Right" and "Good" when the five values linguistic scale is used; and when the seven value linguistic scale is used, "Absolutely right" in cases where the

*Addressing reference data uncertainty in the accuracy assessment of a map of
Continental Portugal*

pixel is occupied by just one land cover class, or "Right" and "Good" in case of mixed pixels). MAX corresponds to the traditional approach of defining agreement based on the most likely reference class and RIGHT corresponds to a more optimistic assessment.

Table 5.12 show the defuzzified user's accuracy, along with the user's accuracy obtained with MAX and RIGHT.

Table 5.12 - Defuzzified user's accuracy in percentage, obtained with first of maxima (*fom*), centroid, last of maxima (*lom*) and traditional user's accuracy obtained with MAX and RIGHT.

Linguistic scales	Land cover classes	Defuzzification methods			Traditional methods		
		<i>fom</i>	<i>Centroid</i>	<i>lom</i>	<i>MAX</i>	<i>RIGHT</i>	
Five linguistic values	Ideal case	<i>UA</i>	51	57	63	52	82
		<i>AG</i>	89	87	92	88	96
		<i>NV</i>	45	49	53	46	64
		<i>F</i>	54	57	60	56	68
		<i>WW</i>	83	82	86	82	90
	Interpreter derived	<i>UA</i>	43	50	57	52	82
		<i>AG</i>	73	80	89	88	96
		<i>NV</i>	38	45	51	46	64
		<i>F</i>	46	53	61	56	68
		<i>WW</i>	68	76	84	82	90
Seven linguistic values	Ideal case	<i>UA</i>	35	38	43	52	82
		<i>AG</i>	80	82	85	88	96
		<i>NV</i>	36	40	42	50	66
		<i>F</i>	43	45	48	50	64
		<i>WW</i>	76	77	80	82	90
	Interpreter derived	<i>UA</i>	33	39	45	52	82
		<i>AG</i>	79	82	85	88	96
		<i>NV</i>	35	40	44	50	66
		<i>F</i>	42	45	50	50	64
		<i>WW</i>	74	77	80	82	90

Relative to the ideal case for the five linguistic values scale we can state that the user's accuracy obtained with MAX is between the values obtained with *fom* and centroid, while the values obtained with RIGHT are always larger than *lom*. For the interpreter-derived case with the five values linguistic scale MAX is in most cases between the values obtained with the centroid and *lom*. As in for the ideal case, RIGHT is always larger than *lom*.

*Addressing reference data uncertainty in the accuracy assessment of a map of
Continental Portugal*

For the ideal case and interpreter-derived with the seven values linguistic scale, MAX is always higher than *lom* for all land cover classes since it less accuracy was obtained with the seven linguistic scale than with the five value linguistic scale.

Table 5.13 show the defuzzified producer's accuracy, along with producer's accuracy obtained with MAX and RIGHT.

Table 5.13 - Defuzzified producer's accuracy in percentage, obtained with first of maxima (*fom*), centroid, last of maxima (*lom*) and traditional user's accuracy obtained with MAX and RIGHT.

Linguistic scales	Land cover classes	Defuzzification methods			Traditional methods		
		<i>fom</i>	<i>Centroid</i>	<i>lom</i>	<i>MAX</i>	<i>RIGHT</i>	
Five linguistic values	Ideal case	<i>UA</i>	84	95	100	87	93
		<i>AG</i>	45	45	46	48	62
		<i>NV</i>	52	57	61	62	80
		<i>F</i>	51	53	56	56	79
		<i>WW</i>	100	100	100	100	100
	Interpreter derived	<i>UA</i>	82	96	100	87	93
		<i>AG</i>	44	48	53	48	62
		<i>NV</i>	48	57	66	62	80
		<i>F</i>	47	55	62	56	79
		<i>WW</i>	97	100	100	100	100
Seven linguistic values	Ideal case	<i>UA</i>	64	71	80	90	95
		<i>AG</i>	43	44	45	47	59
		<i>NV</i>	41	44	47	61	80
		<i>F</i>	44	46	49	56	80
		<i>WW</i>	93	94	98	100	100
	Interpreter derived	<i>UA</i>	67	80	95	90	95
		<i>AG</i>	45	46	48	47	59
		<i>NV</i>	43	48	54	61	80
		<i>F</i>	47	51	56	56	80
		<i>WW</i>	93	97	100	100	100

Relative to the ideal case with the five values linguistic scale the accuracy obtained with MAX is identical or higher to the one obtained with *lom* for all land cover classes, excepting for UA where MAX has an accuracy of 87% and is between the accuracy obtained with *fom* and centroid. For the interpreter-derived with five linguistic values the accuracy obtained with MAX is between centroid and *lom* excepting for AG where MAX equals the accuracy obtained with centroid (48%) and WW where MAX equals

*Addressing reference data uncertainty in the accuracy assessment of a map of
Continental Portugal*

the accuracy obtained with both *lom* and the centroid (100%). Such as for the ideal case, the accuracy obtained with MAX for UA is between the accuracy obtained with *fom* and centroid.

For the ideal case with seven linguistic values the producer's accuracy obtained with MAX is higher than *lom* for all land cover classes. The same pattern is registered for the interpreter-derived with seven linguistic values where MAX exceeds or is very similar to *lom*, excepting for UA and AG, where MAX is between centroid and *lom*.

Table 5.14 show the defuzzified overall accuracy, along with overall accuracy obtained with MAX and RIGHT.

Table 5.14 - Defuzzified overall accuracy in percentage, obtained with first of maxima (*fom*), centroid, last of maxima (*lom*) and traditional user's accuracy obtained with MAX and RIGHT.

Linguistic scales		Defuzzification methods			Traditional methods	
		<i>fom</i>	<i>Centroid</i>	<i>lom</i>	MAX	RIGHT
Five linguistic values	Ideal case	65	67	71	65	80
	Interpreter derived	54	61	69		
Seven linguistic values	Ideal case	54	57	60	64	80
	Interpreter derived	52	57	61		

Relative to the overall accuracy obtained with MAX (65%) and comparing it with the ideal case for the five values linguistic scale, it can be stated that it is identical to *fom*, while for the interpreter-derived, MAX is between the centroid (61%) and *lom* (69%). For the seven linguistic scale with MAX an overall accuracy of 64% was obtained, exceeding *lom* both for the ideal case (60%) and for the interpreter-derived (61%). As for user's and producer's accuracy, RIGHT is always considerably higher than *lom* confirming that RIGHT is a very optimistic approach (Sarmiento et al. 2009).

5.7 Comparison of the fuzzy accuracy measures between photo- interpreters

To compare the differences in the accuracy assessment for two photo-interpreters, interpreter-derived fuzzy intervals for photo-interpreter 1 and photo-interpreter 2 where built according to the methodology described in Section 4.2.2 using a seven value linguistic scale. Figure 5.6 and Table 5.15 show the interpreter-derived fuzzy intervals for photo-interpreter 1 and photo-interpreter 2.

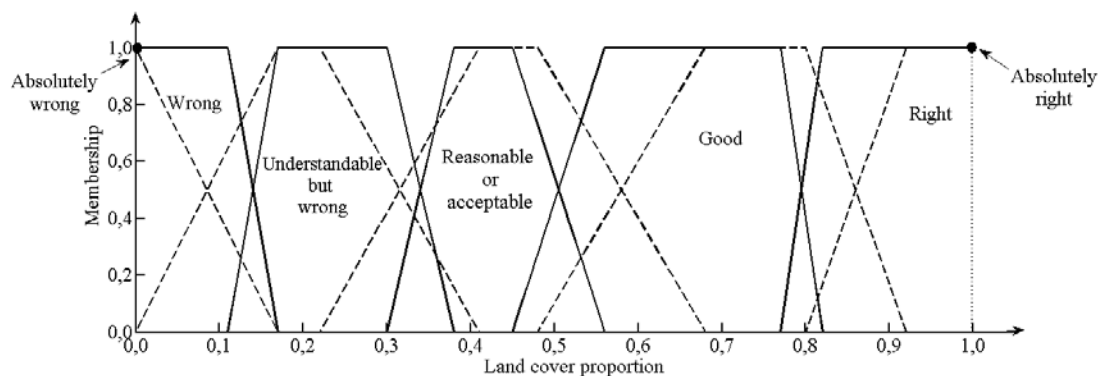


Figure 5.6 - Interpreter derived fuzzy intervals obtained for photo-interpreter 1 (solid line) and photo-interpreter 2 (dashed line), using a linguistic scale with seven linguistic values. The linguistic values "Absolutely wrong" and "Absolutely right" are represented by the black dots corresponding respectively to the crisp numbers 0 and 1.

Table 5.15 - Land cover proportion intervals obtained for the core and support of photo-interpreter 1 and photo-interpreter 2 with the interpreter-derived fuzzy intervals.

	Linguistic values	Support	Core
Photo-interpreter 1	Absolutely wrong (AW)	[0, 0]	[0, 0]
	Wrong (W)	[0, 0.17]	[0, 0.11]
	Understandable but wrong (U)	[0.11, 0.38]	[0.17, 0.30]
	Reasonable or acceptable (A)	[0.30, 0.56]	[0.38, 0.45]
	Good (G)	[0.45, 0.82]	[0.56, 0.77]
	Right (R)	[0.77, 1]	[0.82, 1]
	Absolutely right (AR)	[1, 1]	[1, 1]
Photo-interpreter 2	Absolutely wrong (AW)	[0, 0]	[0, 0]
	Wrong (W)	[0, 0.17]	[0, 0]
	Understandable but wrong (U)	[0, 0.41]	[0.17, 0.22]
	Reasonable or acceptable (A)	[0.22, 0.68]	[0.41, 0.48]
	Good (G)	[0.48, 0.92]	[0.68, 0.80]
	Right (R)	[0.80, 1]	[0.92, 1]
	Absolutely right (AR)	[1, 1]	[1, 1]

*Addressing reference data uncertainty in the accuracy assessment of a map of
Continental Portugal*

The interpreter-derived fuzzy intervals of photo-interpreter 1 have higher core amplitudes and smaller support amplitudes when compared with photo-interpreter 2. This indicates that the range of values of land cover proportion that belong to a certain linguistic value with full membership for photo-interpreter 2, are smaller than for photo-interpreter 1. The major difference of core amplitude between the two photo-interpreters is for the linguistic value "Good", which for photo-interpreter 1 has a core amplitude of 0.21 and for photo-interpreter 2 have a core amplitude of 0.12, with an overlap of 0.09.

Each photo-interpreter built a reference database with 250 sample observations, and fuzzy confusion matrices were assembled with the fuzzy intervals for each linguistic value showed in Table 5.15. Table 5.16 and Table 5.17 show the confusion matrices obtained for each photo-interpreter. Table 5.18 shows the support and core intervals of the fuzzy thematic accuracy measures obtained for each photo-interpreter.

Table 5.16 - Fuzzy confusion matrix and fuzzy accuracy measures elaborated with the interpreter-derived fuzzy intervals of photo-interpreter 1.

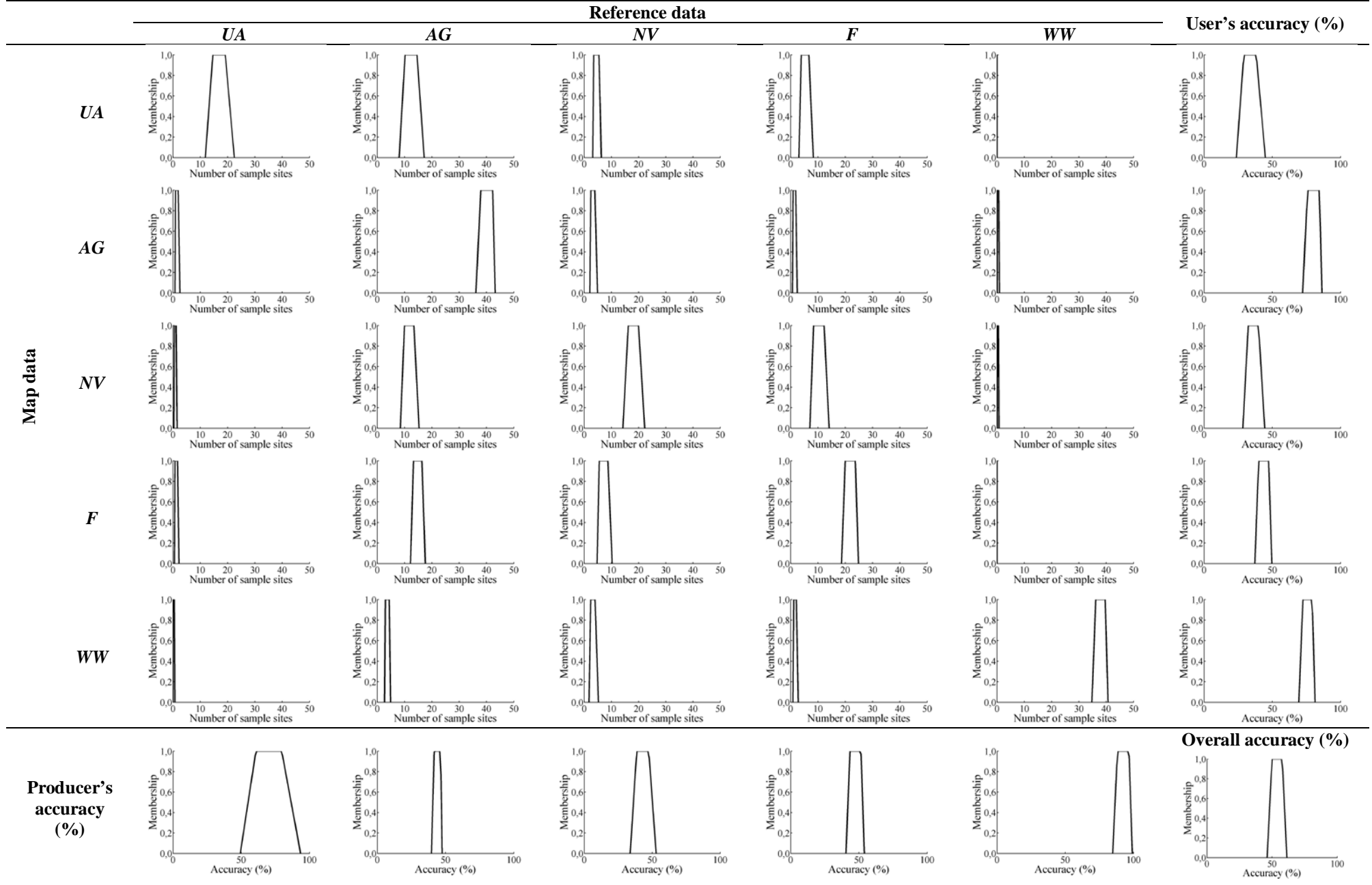
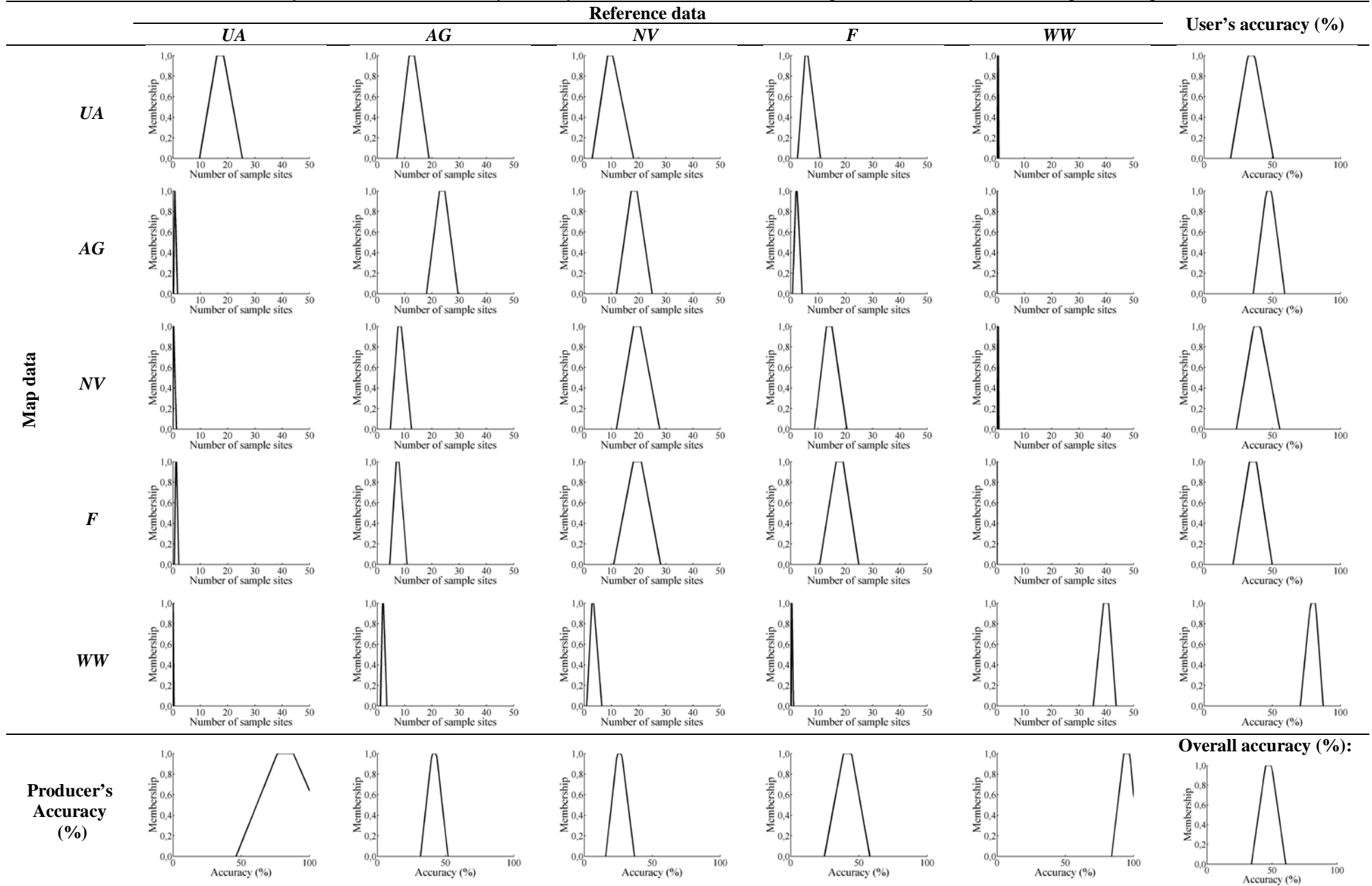


Table 5.17 - Fuzzy confusion matrix and fuzzy accuracy measures elaborated with the interpreter-derived fuzzy intervals of photo-interpreter 2.



*Addressing reference data uncertainty in the accuracy assessment of a map of
Continental Portugal*

Table 5.18 - Support and core intervals of the fuzzy thematic accuracy measures obtained with the interpreter-derived fuzzy intervals using seven linguistic values for each photo-interpreter. Thematic accuracy measures obtained with the centroid (*C*), *MAX* and *RIGHT*.

Fuzzy intervals		Land cover classes	Support	Core	<i>C</i>	<i>MAX</i>	<i>RIGHT</i>	
Photo-interpreter 1	User's accuracy (%)	<i>UA</i>]24, 45[[30, 38]	34	52	82	
		<i>AG</i>]72, 86[[76, 84]	80	88	96	
		<i>NV</i>]28, 44[[32, 39]	36	50	66	
		<i>F</i>]37, 50[[40, 47]	43	50	64	
		<i>WW</i>]69, 81[[72, 78]	75	82	90	
	Producer's accuracy (%)	<i>UA</i>]49, 93[[61, 79]	71	90	95	
		<i>AG</i>]40, 47[[43, 46]	44	47	59	
		<i>NV</i>]34, 53[[39, 47]	43	61	80	
		<i>F</i>]40, 54[[43, 50]	47	56	80	
		<i>WW</i>]85, 99[[89, 96]	92	100	100	
	Overall accuracy]46, 61[[50, 57]	54	64	80
	Photo-interpreter 2	User's accuracy (%)	<i>UA</i>]19, 51[[32, 36]	35	44	66
			<i>AG</i>]36, 59[[46, 49]	47	60	76
<i>NV</i>]24, 55[[37, 41]	39	44	82	
<i>F</i>]21, 50[[34, 38]	36	42	66	
<i>WW</i>]70, 87[[78, 81]	79	88	96	
Producer's accuracy (%)		<i>UA</i>]46, 100[[77, 88]	83	92	97	
		<i>AG</i>]32, 52[[41, 43]	42	42	58	
		<i>NV</i>]16, 37[[25, 27]	26	31	64	
		<i>F</i>]25, 58[[39, 44]	41	51	85	
		<i>WW</i>]84, 100[[93, 97]	94	100	100	
Overall accuracy]34, 60[[45, 49]	47	56	77	

The differences in the fuzzy intervals obtained for each photo-interpreter are also reflected in the fuzzy confusion matrices of each one, since the fuzzy intervals of the number of samples for photo-interpreter 1 (Table 5.16) have larger core amplitudes and smaller support amplitudes than for photo-interpreter 2 (Table 5.17). This difference is also reflected in the fuzzy thematic accuracy measures obtained for each photo-interpreter (Table 5.18).

Relative to overall accuracy we can state that photo-interpreter 1 achieved higher accuracy values than photo-interpreter 2. The core interval of the fuzzy overall accuracy for photo-interpreter 1 is [50, 57] (amplitude of 7%) and the support is the interval]46,

*Addressing reference data uncertainty in the accuracy assessment of a map of
Continental Portugal*

61] (amplitude of 15%) while for photo-interpreter 2 the core is the interval [45, 49] (amplitude of 4%) and the support interval is]34, 60[with an amplitude of 26%.

The fuzzy user's accuracy obtained for each photo-interpreter is quite similar taking into consideration the accuracy that resulted from the computation of the centroid method for all land cover classes, excepting for AG and F. The fuzzy user's accuracy for AG obtained with the centroid method for photo-interpreter 1 (80%) is 33% higher than for photo-interpreter 2 (47%). For photo-interpreter 1 the core interval of the fuzzy user's accuracy is [76, 84] (amplitude of 8%) and the support is the interval]72, 86[(amplitude of 14%) while for photo-interpreter 2 the core of the fuzzy user's accuracy is [46, 49] (amplitude of 3%) and the support is the interval]36, 59[(amplitude of 23%).

The major differences relative to fuzzy producer's accuracy for both photo-interpreters were for UA and NV. For photo-interpreter 1 the core interval of the fuzzy producer's accuracy for UA is [61, 79] (amplitude of 18%) and the support is the interval]49, 93[(amplitude of 44%), while for photo-interpreter 2 the core is the interval [77, 88] (amplitude of 11%) and the support is the interval]46, 100[(amplitude of 54%). Relative to fuzzy producer's accuracy of NV for photo-interpreter 1, the core interval is [39, 47] (amplitude of 8%) and the support is the interval]34, 53[(amplitude of 19%), while for photo-interpreter 2, the core is the interval [25, 27] (amplitude of 2%) and the support is the interval]16, 37[(amplitude of 21%).

MAX and RIGHT for photo-interpreter 1 exceeds the maximum value of the support while for photo-interpreter 2 MAX is between the maximum value of the core (*lom*) and the maximum value of the support. In this sense it can be considered that when is used a seven value linguistic scale, MAX and RIGHT provide an optimistic perspective of accuracy. In this case if there were no information about uncertainty a potential user of

*Addressing reference data uncertainty in the accuracy assessment of a map of
Continental Portugal*

the map would be misled about the accuracy of this land cover map, regarding just
MAX and RIGHT.

.

6 CONCLUSION

The main objective of this dissertation was to develop a methodology that allows the integration of human uncertainty present in reference databases in the accuracy assessment of land cover maps, and analyse the impacts that uncertainty may have in the thematic accuracy measures reported to the end users of land cover maps. The fuzzy interval approach proposed in this study to assess the accuracy of land cover maps enables propagating the uncertainty present in the reference database to the confusion matrix and the accuracy measures derived from it. Since the proportion of each class in each pixel is considered, the fuzzy confusion matrices and the accuracy measures obtained translate the accuracy of the classification having into consideration the proportion of the classes observed in the terrain for all pixels in the reference database, instead of evaluating only if the class assigned to each pixel is correct. That is, they estimate the accuracy of the classification independently of the pixel size, reflecting therefore a closer relation to reality.

This approach enables the computation of fuzzy accuracy measures, in particular fuzzy user's, producer's and overall accuracy, providing: 1) results with different levels of confidence (corresponding to different alpha-levels) as well as optimistic and pessimist values for each one of these levels, corresponding to the minimum and maximum of each alpha-level (enabling the choice of the value to consider according to the demands of each application); 2) information about the uncertainty of the accuracy for each level of confidence resulting from the uncertainty present in the reference database (corresponding to the amplitude of the alpha-levels); 3) defuzzified values similar to the ones usually obtained from confusion matrices, enabling a comparison with the

Conclusion

accuracy measures traditionally used and the use of one value that reflects all uncertainty present, given by the centroid of the fuzzy intervals.

To assess if the use of different fuzzy linguistic scales and fuzzy intervals for the same reference sample may lead to different accuracy results, fuzzy confusion matrices were built for the same reference sample using two linguistic scales, one with 5 and another with 7 linguistic values, and these linguistic values were converted into fuzzy intervals using two approaches. The first approach considers fuzzy intervals that correspond to the ideal response of the photo-interpreter for each linguistic value, and the other derived from the photo-interpreter assignment of the linguistic values for a control sample, where the “real” land cover classes proportions were assessed. The results showed that there are significant differences in the levels of uncertainty and accuracy values. The linguistic scale with five values, provided accuracy results with more uncertainty when compared with the results obtained with the linguistic scale with seven values. Indeed less linguistic values lead to wider fuzzy intervals and therefore more uncertainty will exist. This indicates that the use of a linguistic scale with seven linguistic values with crisp extremes is preferred, since it showed to have a more close relation to the photo-interpreter assignment of the linguistic values and no relevant additional effort is required to use the scale with seven values as opposed to the one with five values. Despite the use of very detailed rules for the elaboration of the reference database, in the case of the linguistic scale with five linguistic values, the interpreter-derived fuzzy intervals are very different when compared with the ideal case proposed by Sarmiento et al. (2013). However this is not the case when the linguistic scale with seven linguistic values is used. In this case, interpreter-derived fuzzy intervals are very similar to the ideal case, resulting in similar thematic accuracy measures. In this case, a close similarity was obtained between the fuzzy intervals for

Conclusion

the ideal case and the interpreter-derived ones when the linguistic scale with seven linguistic values was used. However, for the computation of the interpreter-derived fuzzy intervals is necessary to assess the photo-interpreter perception about land cover proportion and is useful to control his/her assignment of the linguistic values. This may be especially important when a reference database is elaborated by more than one photo-interpreter. The interpreter-derived fuzzy intervals can therefore translate the photo-interpreters behavior and should be the basis of an actual accuracy assessment, whereas the ideal fuzzy intervals are only hypothetical. Moreover, different photo-interpreters may have different perceptions about land cover proportion and the knowledge of these differences can allow, on one hand, the development of methodologies to adjust the criteria in the assignment of the linguistic values in the elaboration of the reference database, reducing the differences between them, and on the other hand, enabling the assessment of accuracy for each photo-interpreter using the human uncertainty present in the process.

The approach proposed has the advantage of incorporating uncertainty with little methodological changes from the traditional methodologies used to assess accuracy, providing an estimate closer to reality and additional information about the uncertainty of the obtained values. A disadvantage of this approach, that is also present for Gopal and Woodcock's (1994) approach, is that additional information has to be collected by the interpreter when constructing the reference database. Specifically, for each sample site a linguistic value must be provided for each class. For many classes the linguistic value to assign will be evident, but if the nomenclature includes many classes, a significant increase in effort may be required to obtain the data. The construction of interpreter-derived fuzzy intervals requires the use of control sites, where the area occupied by each land cover class is known. This is a difficulty of this approach, since,

Conclusion

if no data enabling this computation is available, requires a considerable additional effort.

To overcome this problem, the aim should be to find data that may provide the necessary information, because it is impractical to do all this process for every accuracy assessment. This data could be obtained through OpenStreetMap project, where in recent studies land use data was derived from the information collected by volunteers with good levels of accuracy (Arsanjani et al. 2013). Other advantages of using data collected by volunteers in OpenStreetMap is that doesn't requires the use of other remote sensing data sources, don't has financially costs and also the land use maps updating process is easy to achieve, accordingly with the latest information provided by volunteers. However there are several problems, one is OpenStreetMaps completeness and the other data quality. These issues are very relevant, since surely OpenStreetMaps cannot be used at present for many locations in the world due to lack of data.

An analysis of the results obtained in this study shows that the conclusions obtained from the confusion matrixes generated using the MAX and RIGHT operators can also be obtained from the fuzzy matrix with the addition of information about uncertainty, however with more optimistic results for the accuracy. MAX considers just the best class of the reference data (primary reference label) and when there is a match between the map data and reference data, MAX counts the entire pixel as correct. However, in most cases this is not true because of the existence of many mixed pixels, especially in images with medium and low spatial resolution. This aspect explains the lowest values of accuracy obtained with the proposed approach when compared with the traditional one. Indeed, the accuracy values obtained for MAX are slightly higher than the ones obtained with the optimistic values corresponding to the highest levels of confidence (*lom*). Although this is not verified when the five value linguistic scale is used. Indeed

Conclusion

the consideration of uncertainty in the extreme values of the scale (i.e. "Wrong" and "Right") contributed to an increase of the accuracy, being the accuracy obtained with MAX between the values obtained with the centroid and *lom*. RIGHT presents values always higher than the ones obtained in the most optimistic case (the maximum of the support of the fuzzy numbers). Since it considers the primary or secondary reference label as correct, counting the entire pixel as correct, even if the match is between the map data and the secondary reference label (which may occupy just a relatively small area of the pixel). Even though in the presented studies both the MAX and RIGHT operators provide relatively optimistic estimates for the user's, producer's and overall accuracy.

The studies developed in this dissertation provide several lines of research that can be explored in the future, namely: 1) adapt the traditional equation estimators of the several sampling designs normally used in the accuracy assessment of land cover maps (e.g. stratified random sampling, cluster sampling) with fuzzy arithmetic in order to develop their applicability within the proposed approach of accuracy assessment of land cover maps (although this line of research is far from being straight forward, since involves mixing probability theory and fuzzy sets theory); 2) use other types of membership functions to model the photo-interpreters behavior relative to land cover coverage and build the fuzzy intervals relative to each linguistic value (e.g. gaussian, generalized bell); 3) study the potential of volunteer geographic information data (e.g. data provided by OpenStreetMap) to collect the reference data in the control sample, necessary to model the interpreter-derived fuzzy intervals in order to decrease the effort in their collection, improving this way the applicability of the proposed approach; 4) test the potential of the proposed methodology in the accuracy assessment of land cover maps derived from very high spatial resolution satellite images; 5) develop a friendly user

Conclusion

application to extend the broad range of possible user's of the proposed methodology; 6)
derive area estimation from the proposed fuzzy confusion matrix.

REFERENCES

Arbia, G., D. Griffith, and R. Haining. 1998. "Error propagation modelling in raster GIS: overlay operations." *International Journal of Geographical Information Science* 12:145-167.

Aronoff, S. 1982. "Classification Accuracy: A User Approach." *Photogrammetric Engineering & Remote Sensing* 48:1299-1307.

Arsanjani, J. J., M. Helbich, M. Bakillah, J. Hagenauer, and A. Zipf. 2013. "Toward mapping land-use patterns from volunteered geographic information." *International Journal of Geographical Information Science* 27:2264-2278.

Binaghi, E., P. A. Brivio, P. Ghezzi, and A. Rampini. 1999. "A fuzzy set-based accuracy assessment of soft classifications." *Pattern Recognition Letters* 20:935-948.

Brovkin, V., S. Sitch, V. B. Werner, M. Claussen, E. Bauer, and W. Cramer. 2004. "Role of land cover changes for atmospheric CO₂ increase and climate change during the last 150 years." *Global Change Biology* 10:1253-1266.

Burrough, P. A. and A. U. Frank. 1996. *Geographic Objects with Indeterminate Boundaries*. London: Taylor & Francis.

Burrough, P. A. and R. A. McDonnell. 1998. *Principles of Geographical Information Systems*. New York: Oxford University Press Inc.

References

- Caetano, M., F. Mata, and S. Freire. 2006. "Accuracy assessment of the Portuguese CORINE Land Cover Map." In *Global Developments in Environmental Earth Observation from Space*, edited by A. Marçal, 459-467. Rotterdam: Millpress.
- Carrão, H. 2006. *Land Cover Cartography Accuracy Assessment: An approach in the framework of LANDEO project*. Lisbon: IGP.
- Carrão, H., A. Araújo, P. Gonçalves, and M. Caetano. 2010. "Multitemporal MERIS images for land cover mapping at national scale: the case study of Portugal." *International Journal of Remote Sensing* 31:2063-2082.
- Chapin, F. S. III, E. S. Zavaleta, V. T. Eviner, R. L. Naylor, P. M. Vitousek, H. L. Reynolds, D. U. Hooper, S. Lavorel, O. E. Sala, S. E. Hobbie, M. C. Mack, and S. Diaz. 2000. "Consequences of changing biodiversity." *Nature* 405:234-242.
- Cochran, W. G. 1977. *Sampling Techniques*. New York: Wiley.
- Congalton, R. G. 1988. "A comparison of sampling schemes used in generating error matrices for assessing the accuracy of maps generated from remotely sensed data." *Photogrammetric Engineer and Remote Sensing* 54:593-600.
- Congalton, R. G. 2004. "Putting the Map Back in Map Accuracy Assessment." In *Remote Sensing and GIS Accuracy Assessment*, edited by R. S. Lunetta and J. G. Lyon, 1-11. Boca Raton: CRC Press.
- Congalton, R. G., and K. Green. 1993. "A Practical Look at the Sources of Confusion in Error Matrix Generation." *Photogrammetric Engineer & Remote Sensing* 59:641-644.
- Congalton, R. G., and K. Green. 1999. *Assessing the accuracy of remotely sensed data: Principles and practices*. Boca Raton: Lewis Publishers.

References

- Congalton, R. G., and K. Green. 2009. *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*. Boca Raton: CRC Press.
- Czaplewski, R. L. 2003. "Accuracy assessment of maps of forest condition: statistical design and methodological considerations." In *Remote Sensing of Forest Environments: Concepts and Case Studies*, edited by M. A. Wulder, and S. E. Franklin, 115-140. Boston: Kluwer Academic Publishers.
- Czaplewski, R. L. 2010. *Recursive restriction estimation: an alternative to post-stratification in surveys of land and forest cover*. Fort Collins, CO: U. S. Department of Agriculture, Forest Service, Rocky Mountain Research Station.
- Dicks, S. E., and T. H. C. Lo. 1990. "Evaluation of Thematic Map Accuracy in a Land-Use and Land-Cover Mapping Program." *Photogrammetric Engineering & Remote Sensing* 56:1247-1252.
- Dubois, D., E. Kerre, R. Mesier, and H. Prade. 2000c. "Fuzzy Interval Analysis." In *Fundamentals of Fuzzy Sets. The Handbook of Fuzzy Sets Series*, edited by D. Dubois and H. Prade, 483-561. New York: Kluwer Academic Publishers.
- Dubois, D., H. Nguyen, and H. Prade. 2000b. "Possibility theory, probability and fuzzy sets." In *Fundamentals of Fuzzy Sets. The Handbook of Fuzzy Sets Series*, edited by D. Dubois and H. Prade, 343-438. New York: Kluwer Academic Publishers.
- Dubois, D., W. Ostasiewick, and H. Prade. 2000a. "Fuzzy Sets: History and Basic Notions." In *Fundamentals of Fuzzy Sets. The Handbook of Fuzzy Sets Series*, edited by D. Dubois and H. Prade, 21-124. New York: Kluwer Academic Publishers.
- Edwards, T. C. Jr., G. G. Moisen, and D. R. Cutler. 1998. "Assessing map accuracy in an ecoregion-scale cover-map." *Remote Sensing of Environment* 63:73-83.

References

- Falzarano, S. R., and K. A. Thomas. 2004. "Fuzzy Set and Spatial Analysis Techniques for Evaluating Thematic Accuracy of a Land-Cover Map." In *Remote Sensing and GIS Accuracy Assessment*, edited by R. S. Lunetta and J. G. Lyon, 189-207. Boca Raton: CRC Press.
- Foody, G. M. 1996. "Approaches for the production and evaluation of fuzzy land cover classification from remotely sensed data." *International Journal of Remote Sensing* 17:1317-1340.
- Foody, G. M. 2002. "Status of land cover classification accuracy assessment." *Remote Sensing of Environment* 80:185-201.
- Foody, G. M. 2009. "Sample size determination for image classification accuracy assessment and comparison." *International Journal of Remote Sensing* 30:5273-5291.
- Foody, G. M. 2010. "Assessing the accuracy of land cover change with imperfect ground reference data." *Remote Sensing of Environment* 114:2271-2285.
- Foody, G. M. 2013. "Ground reference data error and the mis-estimation of the area of land cover change as a function of its abundance." *Remote Sensing Letters* 4:783-792.
- Foody, G. M., and D. S. Boyd. 2013. "Using Volunteered Data in Land Cover Map Validation: Mapping West African Forests." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 6:1305-1312.
- Foody, G. M., and M. K. Arora. 1996. "Incorporating mixed pixels in the training, allocation and testing stages of supervised classifications." *Pattern Recognition Letters* 17:1389-1398.

References

- Foody, G. M., L. See, S. Fritz, M. Van der Velde, C. Perger, C. Schill, and D. S. Boyd. 2013. "Assessing the Accuracy of Volunteered Geographic Information arising from Multiple Contributors to an Internet Based Collaborative Project." *Transactions in GIS* 17:847-860.
- Foody, G. M., L. See, S. Fritz, M. Van der Velde, C. Perger, C. Schill, D. S. Boyd, and A. Comber. 2014. "Accurate Attribute Mapping from Volunteered Geographic Information: Issues of Volunteer Quantity and Quality." *The Cartographic Journal* 10.1179/1743277413Y.0000000070.
- Foody, G. M., N. A. Campbell, N. M. Trodd, and T. F. Wood. 1992. "Derivation and applications of probabilistic measures of class membership from the maximum-likelihood classification." *Photogrammetric Engineering & Remote Sensing* 58:1335-1341.
- Freund, J. E., and F. J. Williams. 1972. *Elementary business statistics*. New Jersey: Prentice-Hall.
- Fritz, S., I. McCallum, C. Schill, C. Perger, R. Grillmayer, F. Achard, F. Kraxner, and M. Obersteiner. 2009. "Geo-Wiki.Org: The Use of Crowdsourcing to Improve Global Land Cover." *Remote Sensing* 1:345-354.
- Fritz, S., I. McCallum, C. Schill, C. Perger, L. See, D. Schepaschenko, M. Van der Velde, F. Kraxner, and M. Obersteiner. 2012. "Geo-Wiki: An online platform for improving global land cover." *Environmental Modelling & Software* 31:110-123.

References

- Fritz, S., L. See, M. Van der Velde, R. A. Nalepa, C. Perger, C. Schill, I. McCallum, D. Schepaschenko, F. Kraxner, X. Cai, X. Zhang, S. Ortner, R. Hazarika, A. Cipriani, C. Di Bella, A. H. Rabia, A. Garcia, M. Vakolyuk, K. Singha, M. E. Beget, S. Erasmi, F. Albrecht, B. Shaw, and M. Obersteiner. 2013. "Downgrading Recent Estimates of Land Available for Biofuel Production." *Environmental Science & Technology* 47:1688-1694.
- Gill, S. J., J. Milliken, D. Beardsley, and R. Warbington. 2000. "Using a mensuration approach with FIA vegetation plot data to assess the accuracy of tree size and crown closure classes in a vegetation map of northeastern California." *Remote Sensing of Environment* 73:298-306.
- Gilmore, M. S., E. H. Wilson, N. Barrett, D. L. Civco, S. Prisloe, J. D. Hurd, and C. Chadwick. 2008. "Integrating multi-temporal spectral and structural information to map wetland vegetation in a lower Connecticut River tidal marsh." *Remote Sensing of Environment* 112:4048-4060.
- Goodchild, M. 2003. "Geographic information science and systems for environmental management." *Annual Review of Environment and Resources* 28:493-519.
- Goodchild, M., and J. A. Glennon. 2010. "Crowdsourcing geographic information for disaster response: a research frontier." *International Journal of Digital Earth* 3:231-241.
- Gopal, S., and C. Woodcock. 1994. "Theory and methods for accuracy assessment of thematic maps using fuzzy sets." *Photogrammetric Engineering & Remote Sensing* 60:181-188.
- Green, K., and R. G. Congalton. 2000. "An Error Matrix Approach to Fuzzy Accuracy Assessment: The NIMA Geocover Project." In *Remote Sensing and GIS Accuracy Assessment*, edited by R. S. Lunetta and J. G. Lyon, 163-172. Boca Raton: CRC Press.

References

- Hanss, M. 2002. "The transformation method for the simulation and analysis of systems with uncertain parameters." *Fuzzy Sets and Systems* 130:277-289.
- Kaufmann, A., and M. Gupta. 1991. *Introduction to Fuzzy Arithmetic: Theory and Applications*. New York: Van Nostrand.
- Klir, G., and B. Yuan. 1995. *Fuzzy Sets and Fuzzy Logic: Theory and Applications*. New Jersey: Prentice Hall PTR.
- Laba, M., R. Downs, S. Smith, S. Welsh, C. Neider, S. White, M. Richmond, W. Philpot, and P. Baveye. 2008. "Mapping invasive wetland plants in the Hudson River National Estuarine Research Reserve using quickbird satellite imagery." *Remote Sensing of Environment* 112:286-300.
- Laba, M., S. K. Gregory, J. Braden, D. Ogurcak, E. Hill, E. Fegraus, J. Fiore, and S. D. DeGloria. 2002. "Conventional and fuzzy accuracy assessment of the New York Gap Analysis Project Land cover map." *Remote Sensing of Environment* 81:443-455.
- Lesschen, J. P., P. H. Verburg, and S. J. Staal. 2005. *Statistical methods for analyzing the spatial dimension of changes in land use and farming systems*. Nairobi: ILRI.
- Lewis, H. G., and M. Brown. 2001. "A generalized confusion matrix for assessing area estimates from remotely sensed data." *International Journal of Remote Sensing* 22:3223-3235.
- Ma, Z., M. M. Hart, and R. L. Redmond. 2001. "Mapping vegetation across large geographic areas: integration of remote sensing and GIS to classify multisource data." *Photogrammetric Engineering & Remote Sensing* 67:295-307.

References

- Mickelson, J. G., D. L. Civco, and J. A. Silander. 1998. "Delineating forest canopy species in the Northeastern United States using multi-temporal TM imagery." *Photogrammetric Engineering & Remote Sensing* 64:891-904.
- Moisen G. G., T. C. Jr. Edwards, & D. R. Cutler. 1994. "Spatial sampling to assess classification accuracy of remotely sensed data." In *Environmental Information Management and Analysis: Ecosystem to Global Scales*, edited by W.K. Michener, J.W. Brunt and S.G. Stafford, 159-176. London: Taylor & Francis.
- Moore, R. E. 1966. *Interval Analysis*. New York: Prentice-Hall.
- Muller, S. V., D. A. Walker, F. E. Nelson, N. A. Auerbach, J. G. Bockheim, S. Guyer, and D. Sherba. 1998. "Accuracy Assessment of a Land-Cover Map of the Kuparuk River Basin, Alaska: Considerations for Remote Regions." *Photogrammetric Engineering & Remote Sensing* 64:619-628.
- Olofsson, P., G. M. Foody, M. Herold, S. V. Stehman, C. E. Woodcock, and M. A. Wulder. 2014. "Good practices for estimating area and assessing accuracy of land change." *Remote Sensing of Environment* 148:42-57.
- Olofsson, P., S. V. Stehman, C. E. Woodcock, D. Sulla-Menashe, A. M. Sibley, J. D. Newell, M. A. Friedl, and M. Herold. 2012. "A global land-cover validation data set, part I: fundamental design principles." *International Journal of Remote Sensing* 33:5768-5788.
- Pontius, R. G. 2000. "Quantification error versus location error in comparison of categorical maps." *Photogrammetric Engineering & Remote Sensing* 66:1011-1016.
- Pontius, R. G., and C. D. Lippitt. 2006. "Can error explain map differences over time?" *Cartography and Geographic Information Science* 33:159-171.

References

- Pontius, R. G., and M. L. Cheuk. 2006. "A generalized cross-tabulation matrix to compare soft-classified maps at multiple spatial resolutions." *International Journal of Geographical Information Science* 20:1-30.
- Rosenfield, G. H., K. Fitzpatrick-Lins, and H. S. Ling. 1982. "Sampling for Thematic Map Accuracy Testing." *Photogrammetric Engineering & Remote Sensing* 48:131-137.
- Ross, T. J. 1995. *Fuzzy logic with engineering applications*. USA: McGraw-Hill.
- Sarmiento, P., C. Fonte, J. Dinis, S. V. Stehman, and M. Caetano. 2015. "Assessing the impacts of human uncertainty in the accuracy assessment of land cover maps using linguistic scales and fuzzy intervals." Manuscript submitted for publication.
- Sarmiento, P., C. Fonte, M. Caetano, and S. V. Stehman. 2013. "Incorporating the uncertainty of linguistic-scale reference data to assess accuracy of land-cover maps using fuzzy intervals." *International Journal of Remote Sensing* 34:4008-4024.
- Sarmiento, P., H. Carrão and M. Caetano. 2008a. "Avaliação da exactidão temática de cartografias de ocupação do solo através de funções fuzzy: primeira abordagem." In *ESIG'08: Actas do X Encontro de Utilizadores de Informação Geográfica*, edited by J. G. Rocha, 771-789. Oeiras: Universidade do Minho.
- Sarmiento, P., H. Carrão and M. Caetano. 2008b. "A fuzzy synthetic evaluation approach for land cover cartography accuracy assessment." In *Spatial Uncertainty*, edited by J. Zang and M. Goodchild, 348-355. Liverpool: World Academic Union.
- Sarmiento, P., H. Carrão, M. Caetano, and S. V. Stehman. 2009. "Incorporating reference classification uncertainty into the analysis of land cover accuracy." *International Journal of Remote Sensing* 30:5309-5321.

References

- Sarmento, P., C. Fonte, J. Dinis, and M. Caetano. 2012. "Influence of human uncertainty in the elaboration of reference databases for the accuracy assessment of land cover maps." In *Proceedings of the 10th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*, edited by C. Vieira, V. Bogomy, and A. Aquino, 79-84. Florianopolis: AGNUS.
- Särndal, C. E., B. Swensson, and J. Wretman. 1992. *Model-Assisted Survey Sampling*. New York: Springer-Verlag.
- Silván-Cárdenas, J. L., and L. Wang. 2008. "Sub-pixel confusion-uncertainty matrix for assessing soft classifications." *Remote Sensing of Environment* 112:1081-1095.
- Stehman, S. V. 1992. "Comparison of Systematic and Random Sampling for Estimating the Accuracy of Maps Generated from Remotely Sensed Data." *Photogrammetric Engineering & Remote Sensing* 58:1343-1350.
- Stehman, S. V. 1997. ". "Selecting and interpreting measures of thematic classification accuracy." *Remote Sensing of Environment* 62:77-89.
- Stehman, S. V. 1999. "Basic probability sampling designs for thematic map accuracy assessment." *International Journal of Remote Sensing* 20:2423-2441.
- Stehman, S. V. 2001. "Statistical Rigor and Practical Utility in Thematic Map Accuracy Assessment." *Photogrammetric Engineering & Remote Sensing* 67:727-734.
- Stehman S. V. 2004. "Sampling Design for Accuracy Assessment of Large-Area, Land Cover Maps: Challenges and Future Directions." In *Remote Sensing and GIS Accuracy Assessment*, edited by R. S. Lunetta and J. G. Lyon, 13-29. Boca Raton: CRC Press.

References

- Stehman, S. V. 2009a. "Model-assisted estimation as a unifying framework for estimating the area of land cover and land-cover change from remote sensing." *Remote Sensing of Environment* 113:2455-2462.
- Stehman, S. V. 2009b. "Sampling designs for accuracy assessment of land cover." *International Journal of Remote Sensing* 30:5243-5272.
- Stehman, S. V. 2012. "Impact of sample size allocation when using stratified random sampling to estimate accuracy and area of land-cover change." *Remote Sensing Letters* 3:111-120.
- Stehman, S. V. 2013. "Estimating area from an accuracy assessment error matrix." *Remote Sensing of Environment* 132:202-211.
- Stehman, S. V., and J. D. Wickham. 2011. "Pixels, blocks of pixels, and polygons: Choosing a spatial unit for thematic accuracy assessment." *Remote Sensing of Environment* 115:3044-3055.
- Stehman, S. V., and R. L. Czaplewski. 1998. "Design and analysis for thematic map accuracy assessment: Fundamental principles." *Remote Sensing of Environment* 64:331-344.
- Stehman, S. V., J. D. Wickham, J. H. Smith, and L. Yang. 2003. "Thematic accuracy of the 1992 National Land-Cover Data for the eastern United States: Statistical methodology and regional results.", *Remote Sensing of Environment*, 86:500-516.
- Stehman, S. V., J. D. Wickham, L. Fattorini, T. D. Wade, F. Baffetta, and J. H. Smith. 2009. "Estimating accuracy of land-cover composition from two-stage cluster sampling." *Remote Sensing of Environment* 113:1236-1249.

References

Stehman, S. V., J. D. Wickham, L. Yang, and J. H. Smith. 2000. "Assessing the Accuracy of Large-Area Land Cover Maps: Experiences from the Multi-Resolution Land Cover Characteristics (MRLC) Project." In *Proceedings of the 4th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*, edited by G. Heuvelink and M. Lemmens, 601-608. Delft: Delft University Press

Stehman, S. V., P. Olofsson, C. E. Woodcock, M. Herold, and M. A. Friedl. 2012. "A global land-cover validation data set, II: augmenting a stratified sampling design to estimate accuracy by region and land-cover class." *International Journal of Remote Sensing* 33:6975-6993.

Story, M., and R. G. Congalton. 1986. "Accuracy assessment: A user's perspective." *Photogrammetric Engineering & Remote Sensing* 52:397-399.

Strahler, A. H., L. Boschetti, G. M. Foody, M. A. Friedl, M. C. Hansen, M. Herold, P. Mayaux, J. T. Morissette, S. V. Stehman, and C. E. Woodcock. 2006. *Global Land Cover Validation: Recommendations for Evaluation and Accuracy Assessment of Global Land Cover Maps*. EUR 22156EN-DG. Luxembourg: Office for Official Publications of the European Communities.

Townsend, P. A. 2000. "A Quantitative Fuzzy Approach to Assess Mapped Vegetation Classifications for Ecological Applications." *Remote Sensing of Environment* 72:253-267.

Vitousek, P. M. 1994. "Beyond global warming: ecology and global change." *Ecology* 75:1861-1876.

References

- Wickham, J. D., S. V. Stehman, J. H. Smith, and L. Yang. 2004. "Thematic accuracy of the 1992 National Land-Cover Data for the western United States." *Remote Sensing of Environment* 91:452-468.
- Woodcock, C., and S. Gopal. 2000. "Fuzzy set theory and thematic maps: accuracy assessment and area estimation." *International Journal of Geographical Information Science* 14:153-172.
- Woodcock, C., S. Gopal, and W. Albert. 1996. "Evaluation of the potential for providing secondary labels in vegetation maps." *Photogrammetric Engineering & Remote Sensing* 62:393-399.
- Wulder M. A., S. E. Franklin, J. C. White, J. Linke, and S. Magnussen. 2006. "An accuracy assessment framework for large-area land cover classification products derived from medium-resolution satellite data." *International Journal of Remote Sensing* 27:663-683.
- Yang, L., S. V. Stehman, J. Wickham, S. Jonathan, and N. J. VanDriel. 2000. "Thematic validation of land cover data of the Eastern United States using aerial photography: feasibility and challenges." In *Proceedings of the 4th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*, edited by G. Heuvelink and M. Lemmens, 747-754. Delft: Delft University Press
- Zadeh, L. A. 1965. "Fuzzy Sets." *Information and Control* 8:338-353.
- Zhu, Z., L. Yang, S. V. Stehman, and R. L. Czaplewski. 2000. "Accuracy assessment from the US Geological Survey regional land cover mapping program: New York and New Jersey region." *Photogrammetric Engineering & Remote Sensing* 66:1425-1435.