*MSc in Environmental Monitoring, Modelling and Management Thesis*

# Spatial data quality assessment in the World Database on Protected Areas for Spain and Ecuador

*by*

*María Feria Aguaded*

*(Student number: 1256079)*

*Submitted in 2014*

This dissertation is submitted as part of a MSc degree in Environmental Monitoring, Modelling and Management at King's College London.

## Dissertation Declaration

**KING'S COLLEGE LONDON**

**UNIVERSITY OF LONDON**

**DEPARTMENT OF GEOGRAPHY**

**MSc DISSERTATION**

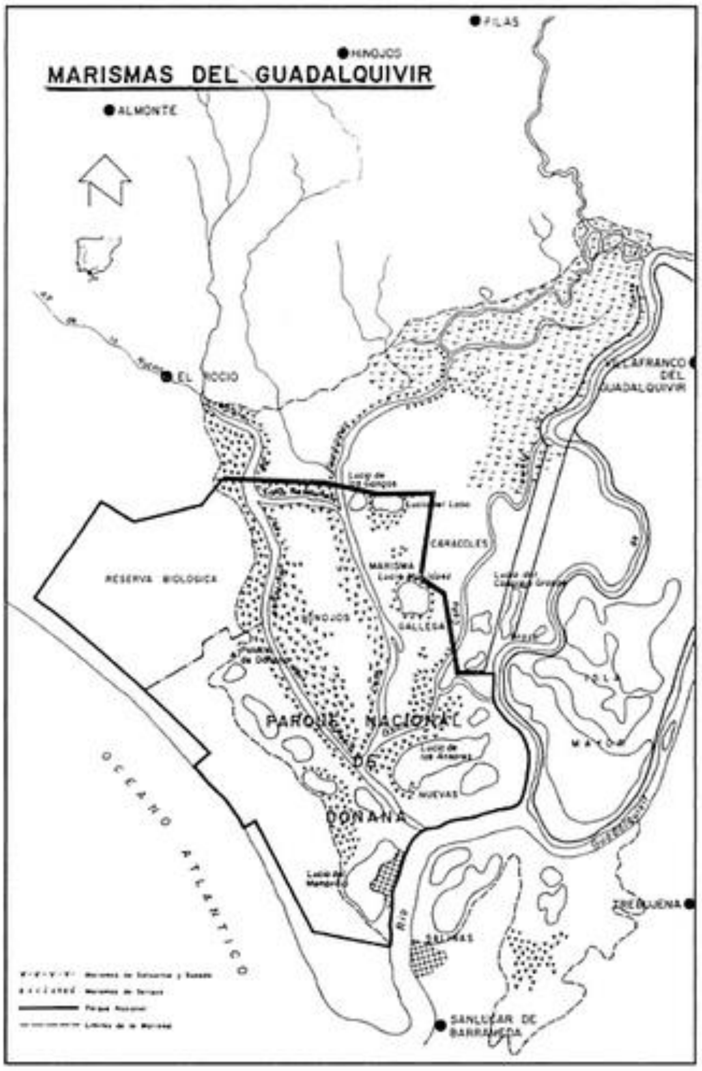
I, María Feria Aguaded hereby declare (a) that this Dissertation is my own original work and that all source material used is acknowledged therein; (b) that it has been specially prepared for a degree of the University of London; and (c) that it does not contain any material that has been or will be submitted to the Examiners of this or any other university, or any material that has been or will be submitted for any other examination.

This Dissertation is 9688 words.

Signed:

Date: 29 August 2014

## Dedication




“Adventure is the enterprise that implies risk, requires constancy and seeks reward in its own vital drive, three conditions that were met.”

José Antonio Valverde in La aventura de Doñana: cómo crear una reserva (p.15), my translation.

Much in the same way as José Antonio Valverde felt about the foundation of the Doñana National Park, I have also had a humbling experience, immersing myself in the study of this master’s thesis. I want to dedicate this study to my grandmother Rosario because her steps, together with many others, have guided me in carrying out this piece of work. I know she would have felt proud.

Many thanks.

## Abstract of Study

The World Database on Protected Areas (WDPA) is the geodatabase used as frame of reference on protected areas at global scale and provides a source for assessing the commitment of global conservation targets. Measuring data quality is important to evaluate the uncertainty on the estimation of conservation indicators. Two national datasets for Ecuador and Spain are analysed in terms of spatial data uncertainty. Lineage and logical consistency are analysed based on missing value analysis. Positional accuracy for polygon features is measured through the buffer overlay method (Goodchild and Hunter, 1997) in a sample of 136 protected areas. The results of these analyses are stochastically simulated to get a prognostic of propagation of the distance bias. Finally, a model is used to obtain a random set of realisations for the protected area features. The output features are compared with the reference geometry and a value of bias (in area units) is generated.

## List of Contents

## List of Tables

## List of Figures

## List of abbreviations or glossary, and special symbols used

AMP. *Áreas Marinas Protegidas, Spain*

CDDA. *Common Databse on Designed Areas*

CEN. *European Committee for Standardization*

CORINE. *Coordination of Information on the Environment*

DCW. *Digital Chart of the World*

DUE. *Data Uncertainty Engine software*

EEA. *European Environment Agency*

Eionet. *European Environment Information and Observation Network*

ENP. *Espacios Naturales Protegidos, Spain*

ETCBD. *European Topic Centre on Biological Diversity*

*EUROPARC*, *Federation of Nature and National Parks of Europe*

FAO. *Food and Agriculture Organization*

FGDC. *Federal Geographic Data Committee*

GEO. *Group on Earth Observations*

GIS. *Geographic Information System*

GPS. *Global Positioning System*

IBAS. *Important Bird Area*

ICA. *International Cartographic Association*

ID. *Identification/Identity/Identifier*

ISO. *International Organization for Standardization*

IUCN. I*nternational Union for Conservation Nature*

LIC. *Lugar de Importancia Comunitaria, Spain*

LMC. *Linear model of coregionalisation*

MAB. *Man and the Biosphere Programme*

*MAGRAMA*. *Ministerio de Agricultura, Alimentación y Medio Ambiente, Spain*, *Ministerio de Agricultura, Alimentación y Medio Ambiente, Spain*

mpdfs. *Marginalprobability distribution functions*

OSPAR. *Network of Marine Protected Areas for the North-East Atlantic*

PA. *Protected Area*

PANE. *Patrimonio de Areas Naturales del Estado, Ecuador*

PFE. *Patrimonio Forestal del Estado, Ecuador*

PN. *Parques Nacionales, Spain*

QA4EO. *Quality Assurance for Earth Observation*

QI. *Quality Indicator*

RAMSAR. *The Convention on Wetlands of International Importance*

RM. *Reservas Marinas, Spain*

RMSD. *Root Squared Mean Deviation*

SAC. *Special Areas of Conservation*

SEO. *Sociedad Española de Ornitologia*

SUIA. *Unique System of Environmental Information, Ecuador*

UNEP. *Environmental Programme of United Nations*

UTM. *Universal Transverse Mercator coordinate system*

WCMC. *World Centre on Monitoring Conservation*

WDPA. *World Database on Protected Areas*

ZEPA. *Zona de Especial Protección para la Aves, Spain*

ZEPIM. *Zonas Especialmente Protegidas de Importancia para el Mar Mediterráneo (Barcelona Convention)*

ZIE. *Zonas Intangibles del Ecuador*

## Acknowledgments

At the start of the dissertation process I did not expect the subject matter to be so intellectually stimulating. Analysing data allocated in the WDPA has been particularly challenging due to the myriad of possibilities and questions that emerge. Limiting the scope of the work has been the most difficult task.

Most of all, I owe my gratitude to my supervisor for have been patient and giving me the opportunity to work on this interesting field.

I would also like to thank my friend Yoanna for her time and comments on the stochastic simulation.

Special thanks to Lorena, Anita, Patty, Majo, Juanita and Eva for their moral support.

I will always be grateful to my sponsor, Andalusia Government, for believing in me and financing the MSc degree and thus funding this project.

Many thanks to all my professors at King's for making this experience so rich. Their stimulation to critical thinking has been a priceless contribution to my professional (and personal) life.

Finally, I am indebted to my family and especially my parents for their unconditional support.

# 1. INTRODUCTION

## *1.1. Importance of quality data in the WDPA*

The World Database on Protected Areas (WDPA) (IUCN and UNEP, 2014) is the geodatabase used as the frame of reference for protected areas on the global scale. The WDPA has been managed by the UNEP-WCMC since 1981 (Fish *et al.*, 2005; *IUCN and UNEP*, 2014). The data contained in the WDPA provides a source for assessing the commitment of global conservation targets and the number of protected areas and their coverage is an indicator of achievement for global targets and environmental assessments (Chape *et al.*, 2005).

## *1.2. Spatial data quality concepts and gaps*

There are different kinds of uncertainties in the spatial data. The president of the International Cartographic Association (ICA) pointed out the multi-dimensional aspect of the spatial data quality analysis beyond positional accuracy (Guptill and Morrison, 1995).

Inaccuracy can arise from three main categories according to (Aalders, 1996) in Burrough and McDonnell (1998): thematic, positional and temporal accuracy. Errors detected in spatial data can originate from either the observation or presentation stages (Burrough and McDonnell, 1998). The process for a spatial data quality analysis is often set up in three steps: definition, measuring and publication of the results for proper use.

Regarding to normalisation, metadata standards have experienced a significant development in the last two decades through initiatives from the Federal Geographic Data Committee (FGDC), the European Committee for Standardization (CEN) and the International Organization for Standardization (ISO). Nevertheless the documentation regarding data quality specifications is frequently incomplete.

In addition, in spite of an increasing number of papers that have defined data quality concepts and identified user needs (Devillers *et al.*, 2007), actual methods to determine and evaluate data quality in feature objects (Tveite and Langaas, 1995; Kiiveri, 1997; Tveite, 1999; Zhang and Kirby, 2000; Caprioli *et al.*, 2003; Leung *et al.*, 2004a; Leung *et al.*, 2004b; Leung *et al.*, 2004c; Bley and Haller, 2006; Heo *et al.*, 2009; Hernando Gallego *et al.*, 2012; Tong *et al.*, 2013; Zheng and Zheng, 2014) have not been sufficiently addressed in a consensual manner. In the below subchapter, a focused review of relevant methodologies for linear and area features is undertaken.

### *1.2.1. Definition: accuracy, uncertainty and statistical error*

To establish the frame of this research project, some key definitions are well worth exploring. First, accuracy is a qualitative value which can be defined as the closeness of the agreement between the observed value and the true value. Uncertainty refers to the margin of doubt that exists about any measurement (Bell, 1999). The extent of security that exits about the true value within the interval is called the confidence level. In order to express uncertainty, both the width of the margin and the confidence level are needed.

Error is the difference between the observation and the (unobservable) true value of the observed object (Bell, 1999; Yang *et al.*, 2013). Statistical error according to Yang *et al.* (2013) is "the amount by which an observation differs from its (unobservable) true value". So errors can be statistically quantified using residuals—which are the amounts by which observations differ from some estimators of their true values

### *1.2.2. Components of spatial data accuracy*

Positional accuracy, attribute accuracy, completeness, logical consistency, temporal accuracy are all known in the field of spatial data accuracy as the 'famous five' quality indicators (Yang *et al.*, 2013). The lineage is another component in which all the other components rely on. So lineage, more than a dimension of data quality, "is a prerequisite for assessing data quality based on deductive estimation methods" (Veregin and Hargitai, 1995).

#### 1.2.2.1. Lineage

Lineage refers to information about the origin of the data, including any method of derivation or transformation carried out to produce the final data allocated in the database. This information is in the WDPA mainly within a table which relates to every PA throughout a metadata ID. This information enables data traceability. The metadata in the WDPA follows the standards described by the International Organization for Standardization (ISO). The provision of metadata is mandatory when data is submitted to the WDPA and details at attribute level may also be recorded.

#### 1.2.2.2. Positional Accuracy

The coordinates *X*, *Y* and *Z* are measured in their degree of compliance or closeness to the reality. It should include all the transformations as projections that are been performed on data. Estimations on positional error and its propagation are developed in the following sections.

#### 1.2.2.3. Attribute Accuracy

Attributes in GIS are particularly relevant when the observational phenomenon is represented by features, as is the case in the features of the WDPA.

#### 1.2.2.4. Completeness

Presence or gaps of data in any of the three forms: geographic, alphanumeric or relationships. Two means definitions are expanded: data completeness and model completeness. The former is defined by the data gaps and the latter by the model fitness of use. Missing values may lead to less precision of calculated statistics.

#### 1.2.2.5. Logical Consistency

It is an aspect of how well the objects and attributes are related and properly encoded in the geodatabase. It deals with the representation of spatial data and the relationships among objects in a dataset (Kainz, 1995).

#### 1.2.2.6. Semantic Accuracy

The quality with which the alphanumeric elements of the geodatabase have been recorded to represent the geographic observed phenomena (Salgé, 1995).

#### 1.2.2.7. Temporal Accuracy

This component includes the time at which a change occurs, the observation date and the time at which the observed object was recorded in the database (Guptill and Morrison, 1995).

### *1.2.3. Standards*

The standardisation of data collected from multiple sources is a relevant stage concerning the maintenance of data integrity. Spatial accuracy of geospatial data in the WDPA is supposed to be highly varied in relation with the variation in the sources and types of information (points, polygons, buffers).

The communication of data quality has been recognized as key piece of information in order for users to assess the degree the data suits their purpose (Devillers *et al.*, 2007; Yang *et al.*, 2013). There have been several initiatives so far trying to fulfil this need. The Quality Assurance Framework for Earth Observation (QA4EO), which is coordinated by the Group on Earth Observations (GEO), establishes that all data and derived product must be accompanied by a Quality Indicator (QI). The QA4EO was set to develop a GEO data quality assurance strategy with space-based earth observations and potentially with field observations (Fox, 2010)

Standards for data quality are applicable to both data producers and data users. The former assesses the characteristics of its product and the latter determines how well the geodata suit their purpose.

The ISO 19157:2013 (ISO, 2013) has been recently released and describes the principles for the quality of geographic data from components, data quality measures methods, content structure for data quality measures, procedures for evaluation and reporting data quality.

The ISO 19157:2013 supersedes ISO 19113:2002, ISO 19114:2003 and ISO/TS 19138:2006. The former of these three standards, defined already, quality principles for geographic data. The ISO 19114:2003 provided a framework for determining, valuating and reporting data quality results in the form of metadata or an evaluation report. ISO/TS 19138:2006 described quality measures methods. Other standards are found in the Guide to the Uncertainty in Measurement series (*GUM*, 2014).

### *1.2.4. Positional accuracy methods*

To measure the positional accuracy of vector features the available methods vary according to the feature shape. The distance between the reference object and the tested object is a measure of accuracy for point features. Statistics such as the Root Squared Mean Deviation (RMSD), the 90$^{th}$ or 95$^{th}$ percentile are used broadly. However, the complexity increases in the case of linear or area features (Chrisman and Yandell, 1988; Ramirez and Ali, 2003; Goodchild, 2004; Leung *et al.*, 2004c; Goodchild, 2008b; Goodchild, 2008a; Heo *et al.*, 2009; Radoux and Bogaert, 2014). In the table I, a list and description of relevant methodologies for linear and area features is displayed.

Table I. Positional uncertainty in liner and area features methods review.

| Method | Author | Description |
|---|---|---|
| Epsilon band error | Perkal (1966), Chrisman (1982) | It takes the smaller buffer distance around the reference object that completely contains the observed object. It is very sensitive to outliers. |
| Buffer overlay | Goodchild and Hunter (1997) | It consists of a set of buffer widths around the reference object which split the tested object. A probability distribution is generated. |
| BOS method | Tveite (1999) | This method is based on the comparison between the buffers generated from both the tested and the reference object. |
| Error grids | Hunter and Goodchild (1996) | It creates a set of error grids which are overlaid with the vector data to create new possible versions of the original data. |
| Map perturbation functions | Kiiveri (1997) | It takes the observed location and a vector of distortion by error function of the location. |
| Positional errors of the vertices | Chrisman and Yandell (1988), Leung and Yan (1998), Shi and Liu (2000) | Modelling positional uncertainty in the vertices, considering that linear and area features are comprised of vertices and straight lines. This method establishes joint probability distribution functions considering random standard deviations of the positional uncertainty. Deformable objects can be modelled when the uncertainties of vertices are considered independent. |

### *1.3. Spatial data quality in the WDPA*

The WDPA data have been collected at different time periods, by different organizations, using various acquisition technologies, standards and specifications. The WDPA data may be acquired from field measurements using GPS or from maps through digitisation process. Both techniques can be influenced by different sources of error such as measuring instrument bias (e.g. GPS signal), objects being measuring (e.g. map scale), operator skills or sampling issues. The WDPA platform is used by both expert users using specialised software and users using web or mobile mapping technologies.

Initially the WDPA was an aspatial database holding current information of protected areas, where no lineage of spatial information was recorded. Maps were digitalised systematically at WCMC, around the late 1980s with a gross scale around 1:5,000,000 (Fish *et al.*, 2005). Later, these maps were upgraded by matching them with ESRI's Digital Chart of the World 1:1,000,000 (DCW). Currently the scale of digital spatial information is usually much higher and challenges face all the information available. The high number of sources and the common lack of information about the quality of data makes the identification of the most accurate available data a difficult task.

Ultimately, the data from the WDPA come from sources that are heterogeneous spatially, temporally and more importantly semantically. That is why the assessment on the quality of data becomes even more complex as Devillers *et al.* (2007) highlight.

### *1.4. Research scope and objectives*

This study will analyse statistically the different kinds of uncertainty within the WDPA (see figure 1). What is the accuracy related to the boundaries of protected areas? What are the implications of having some missing information or data gaps in the WDPA from a quality data perspective? What is the positional accuracy of terrestrial protected area polygons? Or what is the proportion of area assigned to terrestrial protected area that we can be sure is actually covered by this figure of protection? How reliable are the reported areas? Is there discrepancy between the reported area and the current protected area? Those protected areas which do not have polygon features are registered as geographic point. How are geographic boundary missed values affecting the areal statistical accounting?

Specifically, the goal of this project is to provide an assessment and comparison of the level of uncertainty within the WDPA in two selected countries. The selected countries are Spain and Ecuador, and are both suitable for comparison in terms of their protected area coverage and availability of regional and authority data.

The scope of this master's dissertation does not include the possible errors of the instruments used to measure the area or delimit the boundaries, and it does not include the possible changes in the area itself due physical changes, or bias that may have occurred during the measuring process by the operator taking the measurement.

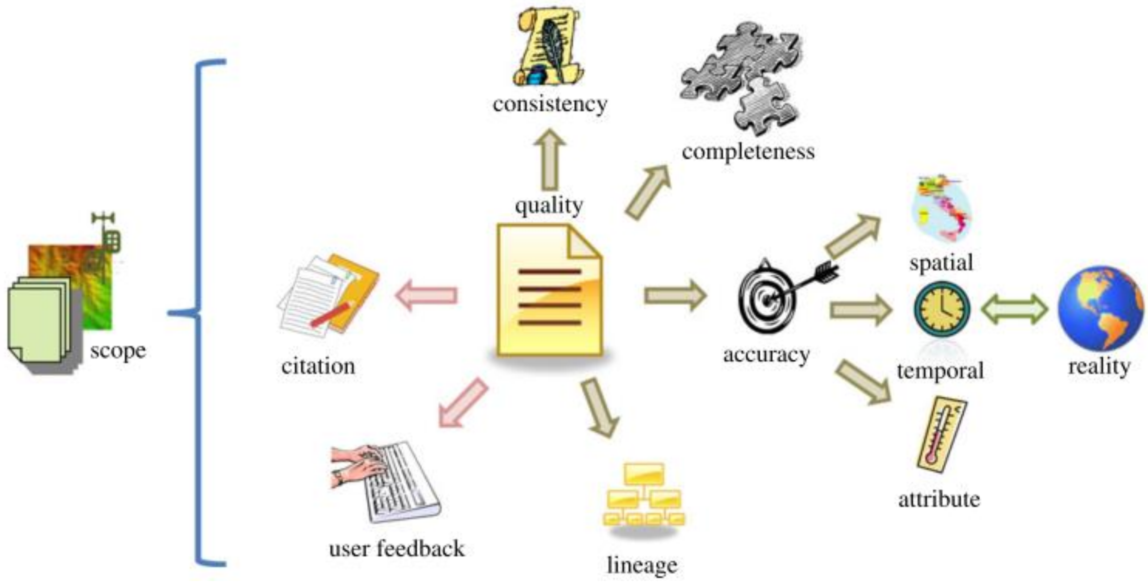



Figure 1. Scope of data quality. From Yang *et al.* (2013)

The above part of this dissertation attempted to define the concepts related with the spatial data quality in the WDPA and stated the objectives of the research project. In addition, established the reasons for evaluating the data quality in the WDPA. The below part of the dissertation focuses on explaining the methods to measure uncertainties on the data. Then a discussion of the results obtained and of how this results answer the questions that have been set above is undertaken.

.

# 2. METHODS

## 2.1. Datasets

### 2.1.1. WDPA data

The spatial data used in this study is openly available on the Protected Planet platform (*IUCN and UNEP*, 2014). This was downloaded in an ESRI geodatabase structure for the whole Earth in May 2014. The geodatabase contained a point feature class, a polygon feature class and a metadata table. The geographical information is recorded as polygon when the PA has known boundary or as point when only the location is known. The WDPA is constantly changing and being updated. The whole database at the time of being downloaded contained 24,639 points without associated boundary information and 180,013 polygons with known boundaries. The metadata table contained 766 rows.

Every geometric object in the WDPA has associated a set of attributes to the name, designation, IUCN management category or the reported area. A complete list of attributes and their descriptions can be seen in Annexe II.

### 2.1.2. Study area

The WDPA original dataset is queried to subtract all the spatial entities located in the Ecuadorian and Spanish territories. The logical query used was:

'ISO3' LIKE '%ECU%'
'ISO3' LIKE '%ESP%'

These queries selected all the entities that are registered in the WDPA as part of the national territory of Spain and Ecuador. It also included designed protected areas that share extensions with neighbouring nations (e.g. Pyrenees – Mont Perdu, WDPA ID 145590 in France and Spain).

The extension of the Spanish dataset is restricted to the peninsula and the Baleares Islands. The Canary Islands have not been included in the analysis for simplification reasons. The UTM grid zone lying on the location of the Canary Islands is distant from the peninsula and the Baleares UTM zone, so it would have required the use and interchange of different coordinate systems for better accuracy. Therefore, the Spanish dataset was queried to extract the protected areas located in the Canary Islands.

### *2.1.3. Reference sources*

The reference sources have been collected from different providers with an average resolution of 1:50000. A detailed list of the dataset used is shown in tables II and III.

#### 2.1.3.1. Ecuadorian reference dataset

A layer with all the protected areas in Ecuador has been courteously provided by the Ecuadorian Government (table II) (*SUIA*, 2014).

A dataset containing the protected areas within the 'Patrimonio de Areas Naturales del Estado' (PANE) is sorted by categories nationally declared by the Ministry of Environment of Ecuador (PANE, 2012). The dataset is comprised of 47 natural areas belonging to the PANE and the Biosphere Reserve of Madagascar.

The other information resource was a layer with protective or productive forests which is covered with protective vegetation. These areas belong to the so called Patrimonio Forestal del Estado (Ministry of Environment of Ecuador, 2014).

#### 2.1.3.2. Spanish reference dataset

The Spanish reference dataset on Protected Areas has been compiled using various national and regional datasets provided mainly by Spanish agencies ADDIN EN.CITE (*EUROPARC*. There are many official agencies that provide the same information. It has been observed that the different datasets vary in date, scale, and data structure and category labels.

The Spanish regulation (Ley 42/2007, de 13 de Diciembre, del Patrimonio Natural y de la Biodiversidad, 2007) establishes five protected figures which are: 'Parques', 'Reservas Naturales', 'Áreas Marinas Protegidas', 'Monumentos Naturales', 'Paisajes Protegidos'. There are also regional legal statuses concerning nature conservations which increase the denominations of protected areas to more than 40 types.

**Table II. List of Ecuadorian dataset and its associated information.**

| *Title* | Included | *Identifier* | *Creator* | *Date* | *Status* | *Scale* |
|---|---|---|---|---|---|---|
| Patrimonio de Areas Naturales del Estado | Yes | PANE | Ministry of Environment, Republic of Ecuador | 2012 | Displays the protected areas within the PANE, sorted by categories nationally declared by the Ministry of Environment of Ecuador (PANE, 2012). The PANE consists of 48 natural areas. | 250000 |
| Zonas Intangibles del Ecuador | No | ZIE | Ministry of Environment, Republic of Ecuador | 2010 | Protected areas of outstanding cultural and biological importance in which no extractive activities can be performed because of their environmental value. It recognises collective rights and domestic use of natural resources by indigenous peoples, which means ensuring the survival of the area and enhancing the development of their social, economic and cultural systems. | 50000 |
| Biosphere Reserves | Yes | MAB | Ministry of Environment, Republic of Ecuador | | Biosphere reserves are areas that represent marine / coastal and terrestrial ecosystems which are internationally recognized by UNESCO, under the jurisdiction of the countries in which they are located. These areas are selected for their scientific interest, based on a set of criteria that determine if a space is included in the program. | 250000 |
| Patrimonio Forestal del Estado | Yes | PFE | Ministry of Environment, Republic of Ecuador | 2012 | Belonging to the Ecuadorian government forest area. Public lands suitable for forestry according to agrological classification which are covered by protective forests or producers; and are covered with protective vegetation. | 250000 |

Publisher: Ministry of Environment, Republic of Ecuador.
All the maps were published at a resolution of 1:50000 and provided at the UTM ETRS89 Zone 30N reference system.

Table III. List of Spanish dataset and its associated information.

| | Title | Included | Identifier | Creator | Date | Status | CDDA Designation |
|---|---|---|---|---|---|---|---|
| Protected Areas | Áreas Marinas Protegidas | Yes | AMP | Banco de Datos de la Naturaleza | November 2011 | Definitive | Marine Protected Area |
| | Reservas Marinas. Estatal | Yes | RM | Banco de Datos de la Naturaleza | December 2012 | Updated up to December 2012 | Marine Reserve |
| | Espacios Naturales Protegidos | Yes | ENP | Banco de Datos de la Naturaleza | December 2013 | Updated up to December 2013 | - |
| | Parques Nacionales | Yes | PN | Banco de Datos de la Naturaleza | May 2012 | - | National Park |
| Nature 2000 | Lugar de Importancia Comunitaria | Yes | LIC | Banco de Datos de la Naturaleza | December 2012 | Definitive | Special Areas of Conservation (SAC) |
| | Zona de Especial Protección para la Aves | Yes | ZEPA | Banco de Datos de la Naturaleza | December 2013 | Definitive | Special Areas of Conservation (SAC) |
| International Conventions | Man and the Biosphere Programme | Yes | MaB | Banco de Datos de la Naturaleza | 2013 | Definitive | Biosphere Reserve |
| | Zonas Especialmente Protegidas de Importancia para el Mar Mediterráneo (Barcelona Convention) | Yes | ZEPIM | Banco de Datos de la Naturaleza | 2003 | Definitive | - |
| | The Convention on Wetlands of International Importance (Ramsar) | Yes | RAMSAR | Banco de Datos de la Naturaleza | May 2014 | Updated | Protected Wetland |
| | Network of Marine Protected Areas (OSPAR) for the North-East Atlantic | Yes | OSPAR | Banco de Datos de la Naturaleza | May 2010 | Definitive | Marine Protected Area |
| Others | Áreas Importantes para las Aves marinas en España | Not included | IBAS | SEO BirdLife | 2009 | Definitive | - |
| | Áreas Importantes para las Aves en España | Not included | IBAS | SEO BirdLife | 2008 | Definitive | - |

Publisher: Ministry of Agriculture, food and Environment, Spanish government
All the maps were provided at the WCS 84 16S reference system.

Spain reports the information annually to the Common Database on Designed Areas (CDDA) maintained by the European Environment Agency (EEA) as part of the Eionet inventory with support from the European Topic Centre on Biological Diversity (ETCBD).

The CDDA started in 1995 under the CORINE programme and is now part of the European Environment Information and Observation Network (Eionet) inventory. The information is related to any PA limit amendment or reclassification as well as any modification in denomination or protection figure type. The EEA publishes the dataset (*Eionet*, 2014) and makes it available to UNEP-WCMC in a long term collaboration partnership(Fish *et al.*, 2005).

## *2.2. Missing values analysis*

To perform an attribute accuracy evaluation, an analysis of missing records has been carried out in SPSS software. Missing value analysis address incomplete data uncerrtainty analysis.

The attributes in the dataset are recorded mainly in Latin characters except for the 'Original Name' which supports multiple language characters. Also 'Country' and Sub-National Location are documented under ISO Code. While 'English Designation' is in English alphabet (*IUCN and UNEP*, 2014).

It considered attributes allocated in both the metadata table and the spatial attribute table. The relationship between these tables is one to many under terminology of relational database. A metadata row can have many protected areas associated. This situation often happens when a layer with many protected areas is provided at the same time. The information associated with these protected areas is often homogeneous and it may be recorded in a single row on a table. The tables are related by the field 'Metadata ID'.

The data are recoded to new variables from the original fields. String fields are recoded as follows:

```
MISSING --> SYSMIS
'Not Reported' --> 99
ELSE --> 1
```

or

```
MISSING --> SYSMIS
'Not Reported' --> 99
'0' --> 0
ELSE --> 1
```

Numeric fields are recoded in a similar way:

```
SYSMIS --> SYSMIS
MISSING --> SYSMIS
ELSE --> 1
```

or

```
SYSMIS --> SYSMIS
MISSING --> SYSMIS
0 --> 0
ELSE --> 1
```

Essentially, the records with 'Not reported' value are recoded as 99 and blanks are recoded as 0. All other values are recoded as 1.

## *2.3. Logical consistency and coverage analysis*

An approach to analysing the uncertainties in the logical component consists of subsequent overlay analysis. The polygons representing protected areas are basic elements itself. These polygons were treated as discrete geographic entities in order to perform a spatial overlay analysis. As stated before, in order to analyse the logical consistency and positional accuracy of the protected area, a reference source of higher accuracy was selected (see tables II and III). Then a dissolving process was computed in each dataset to avoid area double counting.

Once the objects to be compared were selected, the dataset were dissolved and clipped with a terrestrial mask to differ marine areas from terrestrial ones. Then an overlay analysis was performed following logical or boolean operations. The boolean operations can be described as follows (see figure 2) (Burrough and McDonnell, 1998) where the objects from the WDPA are set as *A* and the reference objects are set as *B*:

- Areas inside *A* and inside *B*: *A* AND *B* (intersection).
- Areas outside *A* and outside *B*: *A* XOR *B* (symmetric difference)
- Areas inside *A* but outside *B*: *A* NOT *B* (difference)
- Areas outside *A* and inside *B*: *B* NOT A (difference)

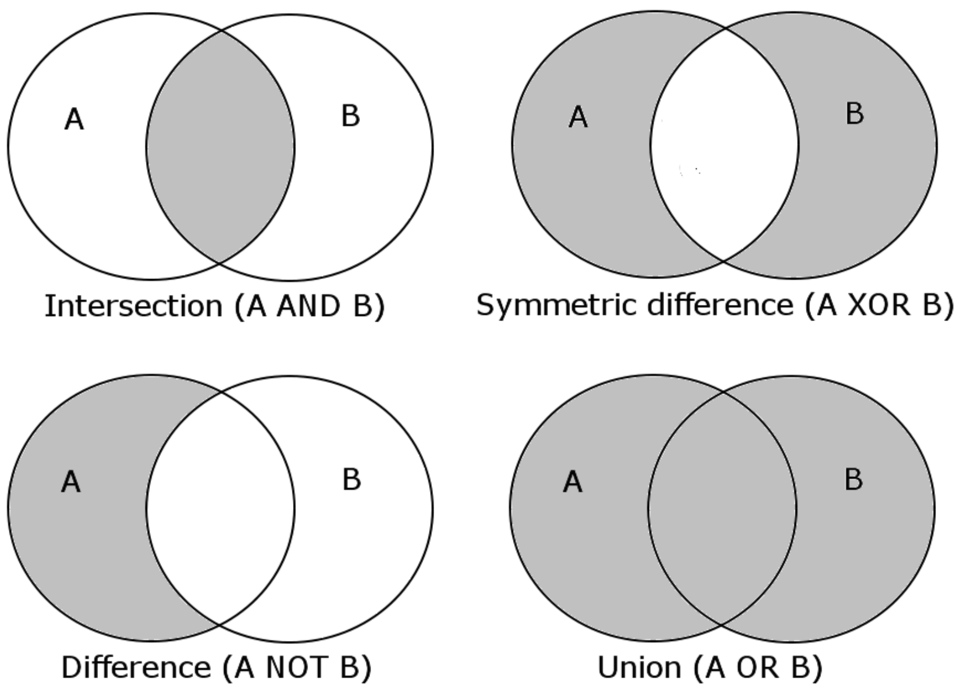



**Figure 2. Logical operations illustrated by set diagrams that have been used in the spatial analysis.**

Following this, statistics as total area, number of polygons, total perimeter, perimeter and area ratio (Tveite, 1999) for each polygon, were calculated.

## *2.4. Positional accuracy*

To compare the current world protected area database a reference source is needed. Positional accuracy measurements are based on the comparison between the observed object to be tested and an observed reference object that is considered to have a higher positional accuracy. This reference object is treated as a true location and the difference between them may be ignored when the accuracy of the reference object is high enough (Drummond, 1995; Goodchild and Hunter, 1997).

### *2.4.1. Perimeter positional accuracy: an adapted technique for accurate assessment of area features through a simple buffer method*

The methodology applied has been proposed by Goodchild and Hunter (1997) for linear objects and adapted to polygon objects in this work. The model provides a set of buffers of gradually increasing width ($x_1$,… $x_n$) around the reference object. Each buffer intersects with the protected area perimeter being assessed. Then, the ratio between the fragmented perimeter and the total perimeter of the protected area is computed:

$$\mathrm{p(x)} = \frac{\text{Length of perimeter segments within a buffer}}{\text{Total perimeter of the protected area being assessed}} \quad (1)$$

where x is the buffer width matching the percentiles of a cumulative probability function $p(x)=y$, with $p(0)=0$ and $p(\infty)=1$. The initial $x_i$ value is based on a measure of data accuracy. This is established in this case according to the highest source scale found (1:20,000). If a measurement of 1 mm is assumed as the maximum likely error then 20 m can be set as our $x_i$ value.

Once we have set the $x_i$ value, the buffer width can be increasingly risen to reach a target percentile of $y$ (Heo *et al.*, 2009). More complex than a gradually increasing method, Goodchild and Hunter (1997) proposed a more elegant solution as:

$$\mathrm{x_{i+1}} = \frac{(\mathrm{y}-\mathrm{p_{i-1}})(\mathrm{x_i}-\mathrm{x_{i-1}})}{\mathrm{p}-\mathrm{p_{i-1}}} + \mathrm{x_{i-1}} \quad (2)$$

Setting i = i + 1, this equation is repeated as many times as is required until $|p_i - y| < 0.001$ or $y$ reaches the target percentile.

The polygons allocated in the WDPA are the tested objects and they are examined by taking into consideration two cases. Those protected areas with unknown boundaries are allocated in the global database as points. Since the WDPA is comprised of both polygons and points, these objects are analysed separately. A register of the area is compulsory and a buffer around these points is computed in order to perform area indicators (*UNEP-WCMC*, 2012). Therefore, this thesis takes into account these two scenarios when measuring the positional accuracy. In summary, there are two scenarios considered, one for the known boundaries objects and another for the unknown boundaries objects. The analysis for these two scenarios are performed in each country. In total, 136 polygons were analysed with the buffer overlay method. A detailed count of the sample size by country and scenario is depicted in table IV.

**Table IV. Sample size of protected area allocated in the WDPA for the overlay method (Goodchild and Hunter, 1997).**

| Country | Known boundaries (scenario 1) | Unknown boundaries (scenario 2) | Total |
|---|---|---|---|
| **Ecuador** | 30 | 30 | 60 |
| **Spain** | 50 | 26 | 76 |

136

To implement this methodology in each sample (N=136), a python code was written to automatically provide the operations. The script admits polygon features as inputs and loops over the buffer widths. The buffer widths are already implemented inside the code and can be easily changed for user convenience. The code runs the arcpy module and can be incorporated as a geoprocessing Arctoolbox in ArcGIS software (see annexe III).

### *2.4.2. Monte Carlo simulation of the 90th percentile buffer width bias*

Error analysis can be modelled using stochastic model equations as Monte Carlo Simulation (Zhang and Goodchild, 2002; Refsgaard *et al.*, 2007). The use of this method was applied by using normal distribution density function and the random function in Microsoft Excel. The structure of the distribution sample of the 90th percentile buffer width is analysed.

Due to the fact that the sample distribution is positively skweded a transformation was performed to bring the data to a parametric distribution. This was reached by transforming the data into a logarithmic function and subsequently modelling it by its normal distribution.

Then the stochastic simulation was computed in excel. The main functions used have been RAND() and NORMINV(rand(), mu , sigma). The RAND() function generates a number that is equally likely to take any value between 0 and 1. The NORMINV(rand(), mu , sigma) generates a modelled value of a parametric random variable having a mean mu and standard deviation sigma. Having a random value between 0 and 1, the formula NORMINV(rand(), mu , sigma) generates the percentile of a parametric distribution with a mean mu and a standard deviation sigma.

The simulation was made up of a number of 4000 times and the result is transformed again by reversing the initial transformation.

### *2.4.3. Polygonal Area Error Model*

Finally, an error model based on parameters from the outputs of the Goodchild method is performed in two cases: a protected area in Ecuador and another in Spain. This model allows us to obtain a value of bias in terms of area, not distance as does the Goodchild method.

The model takes into consideration that the true coordinates of points being measured are subject to observational error. The point measurement errors are transferred to the area calculated from these coordinate points (Chrisman and Yandell, 1988). A statistical model based on Chrisman and Yandell (1988) is applied which depicts areas confined by straight lines between the simulated points. The model input data are the ($x_i$, $y_i$) coordinates of n points or vertices recorded in the WDPA for the particular PA. The model assumes independence in error values at each point. Also it assumes these error values have an identical distribution.

The observational errors are the variables $X$ and $Y$ with marginal probability distribution functions (mpdfs) $F_x$ and $F_y$ respectively:

$$\mathbf{F_x(x) = P(X \leq x)} \text{ and } \mathbf{F_y(y) = P(Y \leq y)} \quad \textbf{(3)}$$

The continuous numerical variables $X$ and $Y$ that vary in space are defined by their cumulative joint probability distribution functions:

$$\mathbf{F_{XX_1Y_1 \ldots X_nY_n}(x_1, y_1, \ldots x_n, y_n) = P(X_1 \leq x_1, Y_1 \leq y_1, \ldots, X_n \leq x_n, Y_n \leq y_n)} \quad \textbf{(4)}$$

The above error model was performed in the Data Uncertainty Engine version 3.0 (DUE, 2006). The protected area was considered as being a deformable object, by which is understood that each point can be relocated independently from the other points. In other words the uncertainty is modelled independently for each of the vertices that constitutes the area shape. The model consider a normal shape function for the coordinates.

Two input parameters are needed to perform a normal distribution. The average position of each point can be fixed as the measured value of the coordinate point. That assumes zero mean positional bias in the coordinate axes ($\mu_X = \mu_Y = 0$) for the tested points.

The spread parameter represents the uncertainty of the measured point (Brown and Heuvelink, 2007). These values are established according to the spread of the value obtained throughout the buffer overlay method simulation. The spread of the likely maximum error is represented by the spread of the buffer distance comprising each protected area in the national level simulated 4000 times. So this value is set up differently in the case of Doñana National Park (WDPA ID = 61611) in Spain and Cotacachi Cayapas Ecological Reserve (WDPA ID = 184) in Ecuador (see Fig X).

In the model, the correlations are required for the relationship between each coordinate point (cross-correlation between *X* and *Y*), each abscissa (autocorrelation in *X*) and each ordinate (autocorrelation in *Y*). It is assumed that these correlation functions are defined for uncertainties that are joint normally distributed. The correlation is set up as 0.8, assuming statistical dependence. We assume that the procedure used to gather or register the positional data has consistent positional errors.

Next, in order to assign correlation coefficients, a stationary function is performed. This stationary function simulates direction and Euclidian distance between points. The direct and cross variograms of the correlation matrix is built following a linear model of coregionalisation (LMC) (Goovaerts, 1997; Pebesma, 2004). This model requires an identical function shape for the auto-correlations and the cross-correlation. We set up an exponential function as is illustrated by Wackernagel (1996) and cited in Zhang (2007).

Finally the DUE software performs simulations by factorising the covariance matrix Σ:

$$\sum = \mathbf{L}\mathbf{L}^{\mathbf{T}} \quad (5)$$

where *T* is the transpose. Then, a random number is drawn and simulations are obtained from the standard normal distribution *N*(0*, I*) (*DUE*, 2006; Brown and Heuvelink, 2007; de Bruin *et al.*, 2008). The covariance matrix is equal to the identity matrix *I*. Sampling for the joint normal distribution then follows the next equation:

$$\mathbf{x} = \boldsymbol{\mu} + \mathbf{L}\,\mathbf{z} \quad (6)$$

where z is a random sample from *N*(0, *I*) and x is a random sample from the distribution *N*(μ, Σ).

A number of 500 realisations were performed using DUE software package (Brown and Heuvelink, 2007) for the described error model. The 500 realisations produced by DUE were post-processed to eliminate sliver polygons due to auto intersection of protected area boundaries (De Bruin *et al.*, 2007). Then the areas of the simulated polygons were geoprocessed in ArcGIS 10.1 and compared to the reference area shape. The difference in area between the reference source and the simulations were analysed.

## 3. RESULTS AND DISCUSSION

### 3.1. *WDPA dataset versus reference datasets: number of sites and nodes*

The dataset of Ecuador contains 23 polygons and 98 points at the time the database was provided, May 2014 (see table V). The number of PA in Spain is considerably higher than the PA in Ecuador with 3,445 polygons and 185 points. The nodes density has been displayed for each dataset that has been used in the analysis. In comparison the PANE dataset has the highest density of nodes followed by the OSPAR and ZEPA datasets. Notice that a higher density of nodes does not imply necessarily a better accuracy of the dataset. However, it can be considered as an indicator with other pointers as the visualization of each feature with a high resolution aerial ortophoto.

**Table V. Summary of number of sites and nodes per dataset.**

| | Source | Feature shape | Protected Area (PA) | Nodes | Node density by PA |
|---|---|---|---|---|---|
| Ecuador | WDPA (2014) | Polygons | 23 | 74,250 | 3,228 |
| | | Points | 98 | 98 | 1 |
| | PANE (2012) | Polygons | 48 | 494,239 | 10,297 |
| | Bosques y vegetacion protectores (2012) | Polygons | 171 | 123,399 | 722 |
| Spain | WDPA (2014) | Polygons | 3,445 | 5,095,306 | 1,479 |
| | | Points | 185 | 185 | 1 |
| | ENP (2013) | Polygons | 1,405 | 3,321,394 | 2,364 |
| | Ramsar (2014) | Polygons | 147 | 273,227 | 1,859 |
| | Mab (2013) | Polygons | 659 | 520,431 | 790 |
| | ZEPIM (2003) | Polygons | 344 | 222,197 | 646 |
| | OSPAR (2010) | Polygons | 5 | 23,687 | 4,737 |
| | LIC (2013) | Polygons | 1,272 | 4,730,015 | 3,719 |
| | ZEPA (2013) | Polygons | 555 | 2,169,481 | 3,909 |

In the case of Spain, the study area included only the peninsula and Baleares Islands. Canary Islands were not included.

### *3.2. Missing values analysis*

This first subcharter of results and discussion describe the pattern of missing data from the attributes in the table associated to the geographic entinties and the metadata table. In other words, this part describes aspects of lineage, attribute accuracy and completeness.

The table VI and VII show the number of records by attribute with missing values separated by country. In the analysis blanks and 'Not reported' records are treated equally. The number of nonmissing values for each attribute is shown in the column N. The number and percent of missing values appear in the missing columns. This compares the extent of missing data among variables.

The column 'Requirement' show how the attributes in the WDPA are classified according to their requirement to be reported: minimum attributes, core attributes and enhanced attributes. Minimum attributes are mandatory to be reported but core and enhanced attributes are optional. Providers are encouraged by the UNEP-WCMC to report these two last types, mainly the core attributes.

The missing values are located differently in Spain and Ecuador. Spain has 72 protected areas without the 'Reported Area' value associated, being a 2% of the whole national protected area pool in the WDPA. Ecuador only has a record without the 'Reported Area' value; however, this acounts almost 1% of the Ecuadorian dataset. Another minimum attribute with missing values is the 'Status Year' where Spain has around a 4% and Ecuador around a 15%. The missing value analysis shows that the minimum required attributes are fully complete in both countries except for 'Reported Area' and 'Status Year'. Those protected areas with these missing values are not being included in some of the analysis. For instance, these missing data may have implications in the annual statistics of coverage indicator by country and by the known year of establishment produced by the WCMC-UNEP.

In the case of Spain, 'Management' and 'Management Plan' have the greatest number of records with missing values, with 99.1% and 99.6% respectively. In the case of Ecuador, the 'Governance Type' and the 'Management' have the highest number of missing values with 96.9% and 97.7% respectively.

On the other hand, although the attributes 'Original Name' and 'English Designation' are no compulsory they present a full completeness in both countries.

Table VI. Extent of missing data among spatial attributes in Ecuador and Spain.

| Field Name | Full Name | Requirement | Spain[1] | | | Ecuador | | |
|---|---|---|---|---|---|---|---|---|
| | | | N | Missing | | N | Missing | |
| | | | | Count | Percent | | Count | Percent |
| WDPAID | WDPA ID | Minimum [2] | 3,621 | 0 | 0 | 128 | 0 | 0 |
| WDPA_PID | WDPA Parent ID | Minimum [2] | 3,621 | 0 | 0 | 128 | 0 | 0 |
| NAME | Name | Minimum | 3,621 | 0 | 0 | 128 | 0 | 0 |
| COUNTRY | Country | Minimum | 3,621 | 0 | 0 | 128 | 0 | 0 |
| DESIG | Designation | Minimum | 3,621 | 0 | 0 | 128 | 0 | 0 |
| DESIG_TYPE | Designation Type | Minimum | 3,621 | 0 | 0 | 128 | 0 | 0 |
| MARINE | Marine | Minimum | 3,621 | 0 | 0 | 128 | 0 | 0 |
| REP_M_AREA | Reported Marine Area (km2) | Minimum | 3,621 | 0 | 0 | 128 | 0 | 0 |
| REP_AREA | Reported Area (km2) | Minimum | 3,549 | 72 | 2.0 | 127 | 1 | 0.8 |
| STATUS | Status | Minimum | 3,621 | 0 | 0 | 128 | 0 | 0 |
| STATUS_YR | Status Year | Minimum | 3,488 | 133 | 3.7 | 109 | 19 | 14.8 |
| ORIG_NAME | Original Name | Core | 3,621 | 0 | 0 | 128 | 0 | 0 |
| SUB_LOC | Sub-national Location | Core | 423 | 3,198 | 88.3 | 123 | 5 | 3.9 |
| DESIG_ENG | English Designation | Core | 3,621 | 0 | 0 | 128 | 0 | 0 |
| IUCN_CAT | IUCN Category | Core | 969 | 2,652 | 73.2 | 26 | 102 | 79.7 |
| GOV_TYPE | Governance Type | Enhanced | 1,674 | 1,947 | 53.8 | 4 | 124 | 96.9 |
| MANG_AUTH | Management | Enhanced | 32 | 3,589 | 99.1 | 3 | 125 | 97.7 |
| MANG_PLAN | Management Plan | Enhanced | 16 | 3,605 | 99.6 | 14 | 114 | 89.1 |
| NO_TAKE | No Take | Enhanced | 3,481 | 140 | 3.9 | 115 | 13 | 10.2 |
| NO_TK_AREA | No Take Area | Enhanced | 3,621 | 0 | 0 | 128 | 0 | 0 |
| METADATAID | Metadata ID | UNEP-WCMC Assigned | 3,621 | 0 | 0 | 128 | 0 | 0 |
| GIS_AREA | GIS Area (km2) | UNEP-WCMC Assigned | 3,445 | 176 | 4.9 | 32 | 96 | 75.0 |
| GIS_M_AREA (if applicable) | GIS Marine Area (km2) | UNEP-WCMC Assigned | 3,621 | 0 | 0 | 128 | 0 | 0 |
| INT_CRIT | Assigned International Criteria | UNEP-WCMC Assigned | 1,681 | 1,940 | 53.6 | 106 | 22 | 17.2 |

[1] Canary Islands have not been included in the analysis.
[2] WDPA and WDPA ID are assigned by the UNEP-WCMC although are responsibility of the data provider.
See a detailed field description in Annexe II.

Another attribute that is worth comparing is the 'GIS Area'. This has around 5% of miising values in Spain and 75% in Ecuador. These percentages correspond to the protected areas with not known boundaries. They are registered as points in the WDPA.

**Table VII.Extent of missing data among metadata attributes in Ecuador and Spain.**

| Field Name | Full Name | Requirement | Spain[1] | | | Ecuador | | |
|---|---|---|---|---|---|---|---|---|
| | | | N | Missing | | N | Missing | |
| | | | | Count | Percent | | Count | Percent |
| METADATAID | Metadata ID | Minimum | 3,621 | 0 | 0 | 128 | 0 | 0 |
| DATA_TITLE | Data Set Title | Minimum | 3,621 | 0 | 0 | 128 | 0 | 0 |
| RESP_PARTY | Responsible Party | Minimum | 3,608 | 13 | 0.4 | 126 | 2 | 1.6 |
| RESP_EMAIL | Responsible Party Contact E-mails | Minimum | 3,328 | 293 | 8 | 18 | 110 | 85.9 |
| YEAR | Date | Minimum | 3,620 | 1 | 0 | 125 | 3 | 2.3 |
| LANGUAGE | Dataset Language | Minimum | 3,621 | 0 | 0 | 86 | 42 | 32.8 |
| CHAR_SET | Dataset Character Set | Minimum | 3,328 | 293 | 8.1 | 18 | 110 | 85.9 |
| REF_SYSTEM | Coordinate System | Minimum | 3,328 | 293 | 8.1 | 20 | 108 | 84.4 |
| SCALE | Scale | Optional | 1,458 | 2,163 | 59.7 | 4 | 124 | 96.9 |
| LINEAGE | Lineage | Optional | 3,579 | 42 | 1 | 111 | 17 | 13.3 |
| CITATION | Citation | Optional | 3,392 | 229 | 6 | 0 | 128 | 100 |
| DISCLAIMER | Disclaimer | Optional | 3,240 | 381 | 11 | 0 | 128 | 100 |

See a detailed field description in Annexe I

[1] Canary Islands have not been included in the Analysis.

Respect to the metadata missing value analysis, it is remakable that the 'Scale' has around 60% of missing values in Spain and around 97% in Ecuador. The 'Reference System' is missed in the 8% of the records in Spain and almost 85% in Ecuador. Notice the importance of these fields as lineage information to determine the positional accuracy of the feature associated. The pair 'Citation' and 'Disclaimer' have values missing in the entire dataset of Ecuador.

The tabulated patterns tables VIII-XI show whether the data tend to be missing for multiple attributes at each protected area (row level). They describe the jointly missing data. Table VIII shows only attributes with missing values patters with more than 1 % of the protected areas. Those attributes with no missing data are not displayed.

There are several patterns of jointly missing data that occur in more than 1% of the cases and they are accounted by each row in tables VIII-XI. In the case of Spain and for spatial attributes (table VIII), more than 50% of the protected areas show a pattern of jointly missing data of the 'GOV_TYPE', 'INT_CRIT', 'IUCN_CAT', 'SUB_LOC', 'MANG_PLAN' and 'MANG_AUTH'. In the case of Ecuador for the spatial attributes (table IX), the most frequent missing pattern is among the attributes 'PARENT_ISO', 'GIS_AREA', 'IUCN_CAT', 'GOV_TYPE', 'MANG_AUTH' and 'MANG_PLAN'. It may be observed that for both countries the number of cases with the list of attributes is fully recorded is equal to cero.

**Table VIII. Missing patterns for spatial attributes in the WDPA subset for Spain**

| Number of Cases | REP_AREA | STATUS_YR | NO_TAKE | GIS_AREA | GOV_TYPE | INT_CRIT | IUCN_CAT | SUB_LOC | MANG_PLAN | MANG_AUTH | Complete |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | | | | | | | | | | | 0 |
| 131 | | | | | | | | | X | X | 131 |
| 89 | | | | X | | | | | X | X | 220 |
| 708 | | | | | | | | X | X | X | 840 |
| 613 | | | | | | | X | X | X | X | 1465 |
| 66 | X | X | | | | | X | | X | X | 214 |
| 1710 | | | | | X | X | X | X | X | X | 3208 |
| 55 | | X | | | X | X | X | X | X | X | 3266 |
| 76 | | | X | | X | X | X | X | X | X | 3315 |

Patterns with less than 1% cases (36 or fewer) are not displayed.
Number of complete cases if variables missing in that pattern (marked with X) are not used.

**Table IX. Missing patterns of spatial attributes in the WDPA subset for Ecuador**

| Number of Cases | SUB_LOC | NO_TAKE | INT_CRIT | STATUS_YR | PARENT_ISO | GIS_AREA | IUCN_CAT | GOV_TYPE | MANG_AUTH | MANG_PLAN | Complete |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | | | | | | | | | | | 0 |
| 19 | | | | | | | | X | X | X | 20 |
| 2 | | X | | | | | | X | X | X | 23 |
| 2 | | | X | | | | X | X | X | X | 23 |
| 2 | | X | X | | | | X | X | X | X | 30 |
| 3 | | X | X | | X | X | X | X | X | | 10 |
| 6 | | | X | | X | X | X | X | X | | 6 |
| 3 | | | X | | X | X | X | X | X | X | 92 |
| 58 | | | | | X | X | X | X | X | X | 81 |
| 2 | | | | | X | X | | X | X | X | 22 |
| 18 | | | | X | X | X | X | X | X | X | 99 |
| 4 | X | | X | | X | X | X | X | X | | 10 |

Patterns with less than 1% cases (1 or fewer) are not displayed.
Number of complete cases if variables missing in that pattern (marked with X) are not used.

The tabulated patterns for metadata attributes (table X and XI) show significant differences between Ecuador and Spain. In the case of Spain, the metadata information is fully complete for 1,412 designed protected areas. Even more, if the 'Scale' is not used as missing pattern the number of complete protected areas increases up to 3,240. In the case of Ecuador for metadata attributes (table XI), the more repeated missing pattern is composed by the attributes 'REF_SYSTEM', 'CHART_SET', 'RESP_EMAIL', 'SCALE', 'CITATION' and 'DISCLAIMER' with 64 cases.

**Table X. Missing patterns of metadata attributes in the WDPA subset for Spain**

| Number of Cases | LINEAGE | CITATION | DISCLAIMER | RESP_EMAIL | CHART_SET | REF_SYSTEM | SCALE | Complete |
|---|---|---|---|---|---|---|---|---|
| 1412 | | | | | | | | 1412 |
| 1828 | | | | | | | X | 3240 |
| 45 | | X | X | | | | | 1457 |
| 41 | X | X | X | | | | X | 3328 |
| 140 | | X | X | X | X | X | X | 3567 |
| 140 | | | X | X | X | X | X | 3380 |

Patterns with less than 1% cases (36 or fewer) are not displayed.
Number of complete cases if variables missing in that pattern (marked with X) are not used.

**Table XI. Missing patterns of metadata attributes in the WDPA subset for Ecuador**

| Number of Cases | RESP_PARTY | YEAR | LINEAGE | LANGUAGE | REF_SYSTEM | CHART_SET | RESP_EMAIL | SCALE | CITATION | DISCLAIMER | Complete |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | | | | | | | | | | | 0 |
| 3 | | | | | | | | | X | X | 3 |
| 14 | | | X | | | | | X | X | X | 18 |
| 2 | | | | | | X | X | X | X | X | 6 |
| 64 | | | | | X | X | X | X | X | X | 70 |
| 40 | | | | X | X | X | X | X | X | X | 110 |
| 2 | X | X | X | X | X | X | X | X | X | X | 128 |

Patterns with less than 1% cases (1 or fewer) are not displayed.
Number of complete cases if variables missing in that pattern (marked with X) are not used.

## *3.3. Logical consistency and coverage analysis*

Some protected areas have several designations and appears repetitively in their geographic component whether this is different or not. This may not meet the requirements of second normal form (Codd, 1972) upon the geographical component which establishes that a single entity should not appear repeated in a database. Instead other field or other relational table should be created. The field 'WDPA Parent ID' is an attempt to address this issue. This field groups all the real protected areas as one single entity in a more similar approach as it is in reality. So the representation of a protected area in the WDPA may result more consistent semantically. However, currently the 'WDPA Parent ID' is not fully operative. And most of the protected areas show a 'WDPA Parent ID' identical to its 'WDPA ID'.

For instance, the protected area Manglares Cayapa has several inconsistencies in its data structure. The same physical reality appears with three different geometric representations or objects (ID numbers: 97517, 900907 and 301976). One of them is a polygon feature and two are registered as points due to a lack of boundary information.

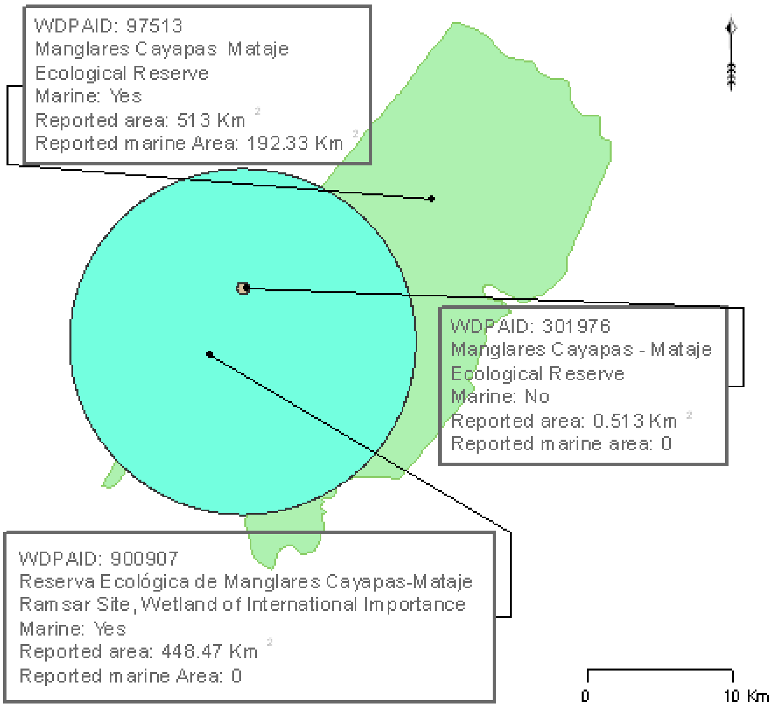



**Figure 3. Manglares Cayapas Mataje Case in Ecuador.**

As a result there may be several inconsistencies in the analysis of accounting coverage area. A land object should be unique and not overlap when an analysis of area and other geometric characteristics are performed. To solve this overlapping of features a dissolve process is performed before accounting any coverage area indicator. However, in the case illustrated in figure 3 notice that when a dissolve technique is performed (e.g. to avoid the double counting of surface) the resultant area is overestimating the real surface covered by the PA.

To address this issue two scenarios are analysed separately. The results are split according to these two scenarios. The scenario one considers only those protected areas with known boundaries. The scenario two considers both protected areas (with known boundaries and not known boundaries).

The results of overlaying a reference source (*B*) with the WDPA dataset (*A*) for Spain and Ecuador are summarised in the table XII and XIII. The tables show the figures of area in km$^2$ that were the result of the different Boolean operations described above in the methodology chapter.

Table XII. Overlay analysis for all the protected areas of Ecuador allocated in the WDPA in $Km^2$. *A* represents the WDPA dataset and *B* represents the reference dataset. The figures highlighted in grey shadow are a bias measure.

**Scenario 1: Only considered known boundaries objects**

| | Terrestrial PA ($Km^2$) | Marine PA ($Km^2$) | Totals ($Km^2$) | No match ($Km^2$) (*B* NOT *A*) |
|---|---|---|---|---|
| **Intersection with reference source ($Km^2$) (*A* AND *B*)** | **33,426.2** | **136,613.7** | **170,040.0** | **30,966.7** |
| Percentage of Protected Area | 17.7% | 72.3% | 89.9% | 16.4% |
| Percentage of national terrestrial area* | 12.9% | 52.9% | 65.9% | 12.0% |
| **No match with reference source ($Km^2$) (*A* NOT *B*)** | **3,502.8** | **15,501.9** | **19,004.7** | |
| Percentage of Protected Area | 1.8% | 8.2% | 10.0% | |
| Percentage of national terrestrial area and territorial waters* | 1.0% | 4.5% | 5.5% | |
| **Totals** | **36,929.0** | **152,115.7** | **189,044.7** | |
| Percentage of national terrestrial area and territorial waters* | 10.6% | 43.8% | 54.5% | |

**Scenario 2: Including points**

| | Terrestrial PA | Marine PA | Totals | No match (*B* NOT *A*) |
|---|---|---|---|---|
| **Intersection with reference source ($Km^2$) (*A* AND *B*)** | **43,248.2** | **136,957.5** | **180,205.7** | **20,800.9** |
| Percentage of Protected Area | 19.5% | 61.7% | 81.2% | 9.4% |
| Percentage of national terrestrial area* | 16.7% | 53.1% | 69.8% | 6.0% |
| **No match with reference source ($Km^2$) (*A* NOT *B*)** | **25,814.4** | **15,884.7** | **41,699.1** | |
| Percentage of Protected Area | 11.6% | 7.2% | 18.8% | |
| Percentage of national terrestrial area and territorial waters* | 7.4% | 4.6% | 12.0% | |
| **Totals ($Km^2$)** | **69,062.6** | **152,842.3** | **221,904.8** | |
| Percentage of national terrestrial area and territorial waters* | 19.9% | 44.0% | 63.9% | |

*Surface area is a country's total area, including areas under inland bodies of water and some coastal waterways (according to Food and Agriculture Organization).

Table XIII. Overlay analysis for all the protected areas of Spain (except Canary Islands) allocated in the WDPA in Km². *A* represents the WDPA dataset and *B* represents the reference dataset. The figures highlighted in grey shadow are a bias measure.

**Scenario 1: Only considered known boundaries objects**

| | Terrestrial PA ($Km^2$) | Marine PA ($Km^2$) | Totals ($Km^2$) | No match ($Km^2$) (*B* NOT *A*) |
|---|---|---|---|---|
| **Intersection with reference source ($Km^2$) (*A* AND *B*)** | **134,441.2** | **8,458.9** | **142,900.2** | **26,221.4** |
| Percentage of Protected Area | 90.7% | 5.7% | 96.4% | 17.7% |
| Percentage of national terrestrial area* | 26.5% | 1.7% | 28.2% | 5.2% |
| **No match with reference source ($Km^2$) (*A* NOT *B*)** | **4,857.7** | **412.9** | **5,270.7** | |
| Percentage of Protected Area | 3.3% | 0.3% | 3.6% | |
| Percentage of national terrestrial area and territorial waters* | 0.8% | 0.1% | 0.8% | |
| **Totals ($Km^2$)** | **139,299.0** | **8,871.9** | **148,170.9** | |
| Percentage of national terrestrial area and territorial waters* | 22.3% | 1.4% | 23.7% | |

**Scenario 2: Including points**

| | Terrestrial PA ($Km^2$) | Marine PA ($Km^2$) | Totals ($Km^2$) | No match (*B* NOT *A*) |
|---|---|---|---|---|
| **Intersection with reference source (*A* AND *B*)** | **140,402.5** | **8,642.5** | **149,045.0** | **20,054.5** |
| Percentage of Protected Area | 78.2% | 4.1% | 83.0% | 11.2% |
| Percentage of national terrestrial area* | 27.7% | 1.7% | 29.4% | 4.0% |
| **No match with reference source (*A* NOT *B*)** | **22,736.6** | **7,766.0** | **30,502.63** | |
| Percentage of Protected Area | 12.7% | 4.3% | 17.0% | |
| Percentage of national terrestrial area and territorial waters* | 3.6% | 1.2% | 4.9% | |
| **Totals** | **163,139.1** | **16,408.5** | **179,547.6** | |
| Percentage of national terrestrial area and territorial waters* | 26.1% | 2.6% | 28.7% | |

*Surface area is a country's total area, including areas under inland bodies of water and some coastal waterways (according to FAO)

Estimations of bias appear on the shadow grey cells. They reflect differences between the coverage of protected areas in the WDPA and the reference source. A summary of these figures is shown in table XIV.

In the case of Ecuador, for the scenario one there are about 10% of protected areas which do not match with the reference source. The figure increase up to almost 19% when protected areas without boundaries information are incorporated to the analysis. On the other hand, the coverage of protected areas which is not represented in the WDPA is around 16% and 9% for the scenario one and two, respectively. Scenario two has a surface of 22,694 $Km^2$ more than scenario one that is not matching the reference source. Scenario two agrees more with the reference source than the scenario one with 10,165 $Km^2$. In total, the difference is of 12,528 $Km^2$ which is being overestimated in scenario two respect to scenario one. Thus, it could be argued that the scenario one may be more precise than the scenario two.

In the case of Spain, the overestimation of scenario two with respect the scenario one is about 25,231 $Km^2$. In addition, the better concordance of protected area which is not being cover in the WDPA is 6,166 $Km^2$ smaller in scenario two in comparison to scenario one. In total, scenario two overestimates in approximately 19,065 $Km^2$ the area being measured. Therefore, for the Ecuadorian dataset the accuracy seems to be higher in the scenario one than in the scenario two. This agrees with the case exposure for Spain.

**Table XIV. Summary of coverage analysis bias**

| | | | **Scenario 1** | **Scenario 2** | **Difference** | **Total overestimation Sc 2 respet Sc 1** |
|---|---|---|---|---|---|---|
| Spain | B NOT A ($Km^2$) | Underestimation | 26,221.4 | 20,054.5 | 6,167.0 | 19,065.0 |
| | A NOT B ($km^2$) | Overestimation | 5,270.7 | 30,502.6 | 25,231.9 | |
| Ecuador | B NOT A ($Km^2$) | Underestimation | 30,966.7 | 20,800.9 | 10,165.8 | 12,528.6 |
| | A NOT B ($km^2$) | Overestimation | 19,004.7 | 41,699.1 | 22,694.3 | |

## 3.4. Positional accuracy

### 3.4.1. Perimeter positional accuracy: an adapted technique for accurate assessment of area features through the simple buffer method

A protected area is selected to illustrate the results of applying the overlay method (Goodchild and Hunter, 1997). This PA is called Cotachi Cayapas and is identified by the number 184 within the WDPA (see figure 4).

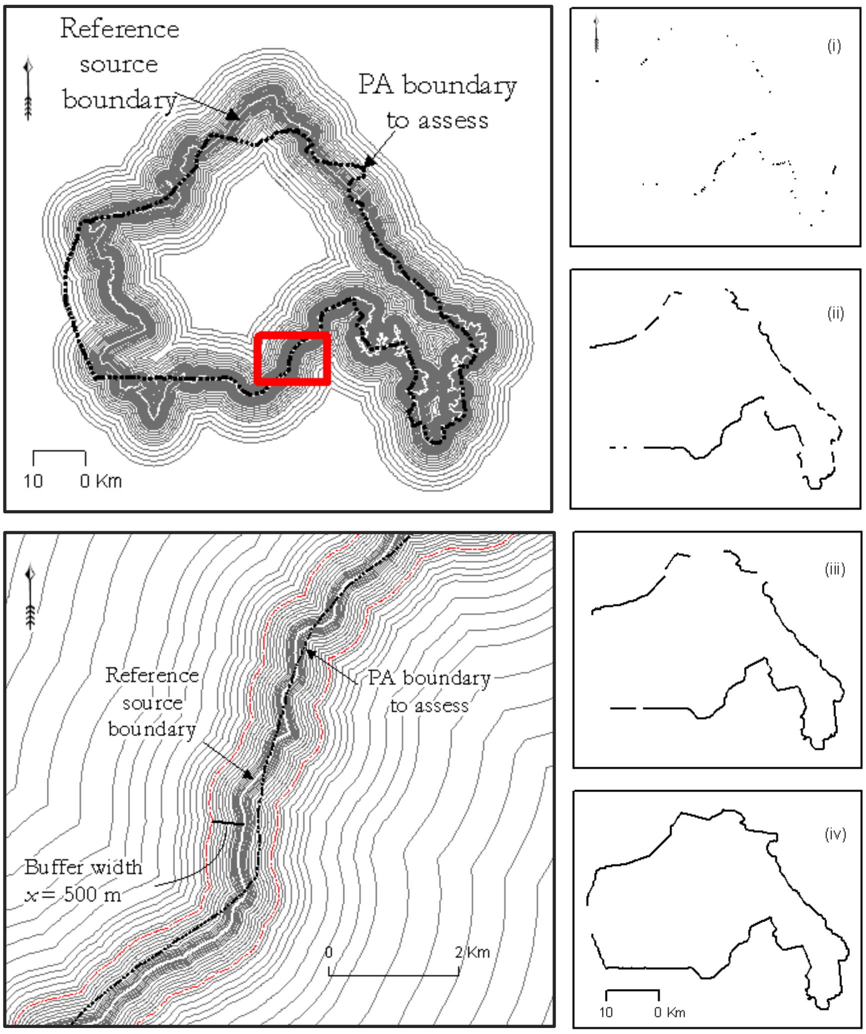



**Figure 4. Illustrations showing the application of the buffer overlay method for Cotachi Cayapas protected area (ID = 184) in Ecuador. On the upper left corner, the protected area's shape is depicted in a black dotted line and a set of buffers around the reference object are depicted in consecutive grey lines. The fringe close to the reference object perimeter appears as a continuous grey buffer as result of the high density of buffer lines. On the bottom left corner, a detailed view from the above picture is shown. On the right side of the image, the results of the intersection between the protected area perimeter and the set of buffers with widths of 20 (i), 1000 (ii), 2000 (iii) and 8000 m (iv) are depicted.**

The example PA is compared with a higher resolution shape which has been provided by a national online platform, the Unique System of Environmental Information (*SUIA*, 2014) from the Ministry of Environment of Ecuador. The online platform supplies a layer with 49 protected areas belonging to the 'Patrimonio de Areas Naturales del Estado' (see table II). The scale of this layer is 1:25,000. The tested PA has 536 nodes and the reference PA has 3403 nodes. A verification to assess the higher accuracy of the reference source is made by photointerpretation. The polygon from the WDPA was transformed to the local projected coordinate system WGS 84 17S. Therefore, both sources were treated under the same coordinate system.

The figure 5 shows the results of plotting buffer width and the percentage of PA perimeter lying on the buffer. Table XV shows these values and the segment length that actually lies on the corresponding buffer. For example, only 2.2% of the perimeter length is enclosed within 20m of the true position. The 90$^{th}$ percentile for the Cotacachi Cayapas ecological reserve is close to 4090 m. This number is obtained applying a linear trend joining the closest higher and lower values to the 90$^{th}$ percentile.

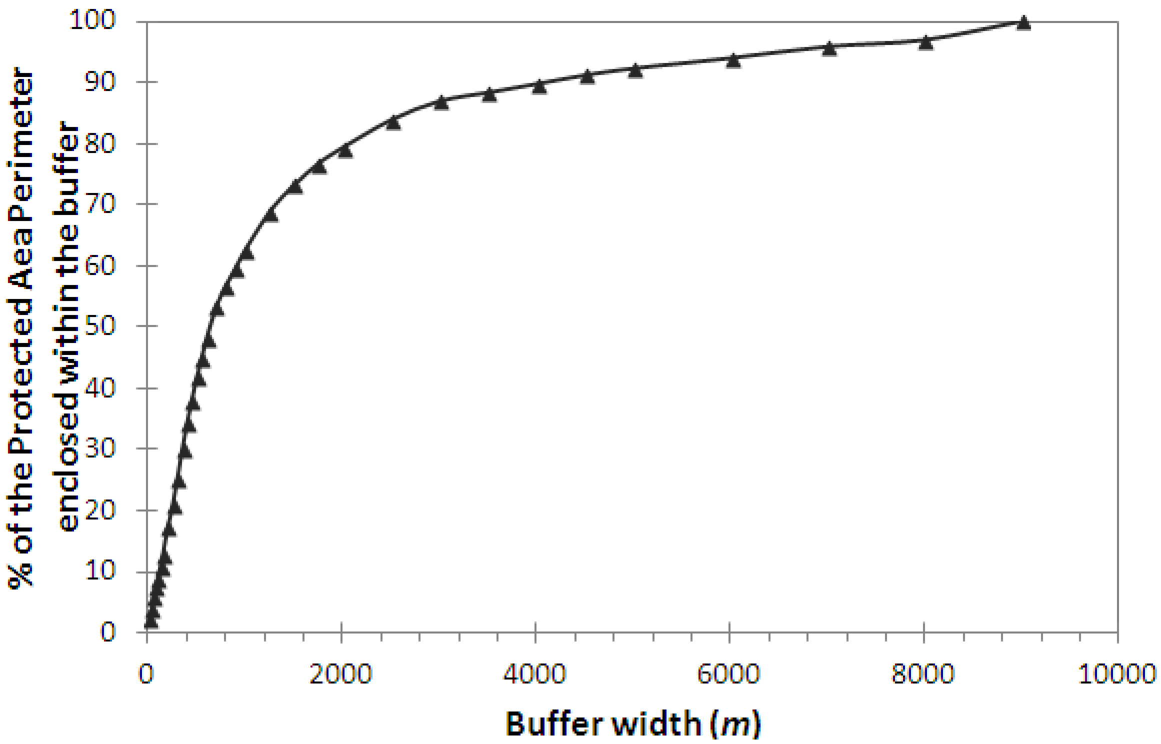



**Figure 5. Percentage of the protected area enclosed within a width buffer (m). This chart corresponds to the protected area perimeter analysis of Cotacachi Cayapas ecological reserve (WDPA ID = 184) in Ecuador.**

Table XV. Length and percentage of the protected area perimeter lying within the buffer of different widths.

| Buffer width (m) | Length of protected area perimeter within buffer (Km) | Percentage of protected area perimeter within buffer (%) |
|---|---|---|
| 20 | 6.2 | 2.2 |
| 40 | 11.4 | 4.1 |
| 60 | 16.3 | 5.8 |
| 80 | 21.0 | 7.5 |
| 100 | 25.3 | 9.0 |
| 125 | 30.6 | 10.9 |
| 150 | 35.9 | 12.8 |
| 200 | 48.4 | 17.2 |
| 250 | 58.5 | 20.8 |
| 300 | 71.0 | 25.3 |
| 350 | 84.2 | 30.0 |
| 400 | 96.6 | 34.4 |
| 450 | 107.2 | 38.1 |
| 500 | 117.5 | 41.8 |
| 550 | 125.9 | 44.8 |
| 600 | 135.0 | 48.0 |
| 700 | 149.6 | 53.2 |
| 800 | 159.4 | 56.7 |
| 900 | 167.5 | 59.6 |
| 1000 | 175.9 | 62.6 |
| 1250 | 193.4 | 68.8 |
| 1500 | 205.8 | 73.2 |
| 1750 | 215.7 | 76.7 |
| 2000 | 223.0 | 79.3 |
| 2500 | 235.8 | 83.9 |
| 3000 | 244.5 | 86.9 |
| 3500 | 248.4 | 88.3 |
| 4000 | 252.3 | 89.7 |
| 4500 | 256.4 | 91.2 |
| 5000 | 259.5 | 92.3 |
| 6000 | 264.2 | 94.0 |
| 7000 | 269.6 | 95.9 |
| 8000 | 272.6 | 96.9 |
| 9000 | 281.2 | 100.0 |

The procedure illustrated in the Cotacachi Cayapas ecological reserve case is repeated for the whole sample of 136 polygons (see table IV). The results of the analysis of the 90$^{th}$ percentile width buffer for both known boundary and not known boundary scenarios are shown in the figure 6 and table XVI. As expected, the not known boundary areas show a substantial higher bias than the known boundary PAs in both countries. A strong skew in data can be interpreted from the statistics in table XVI.

There is a particularly significant difference between the mean and the median which means that data is not homogeneously distributed. In fact, the data shows a positive skew. For instance, the median may be a more convenient indicator of bias than the mean which is not representative of this type of sample.

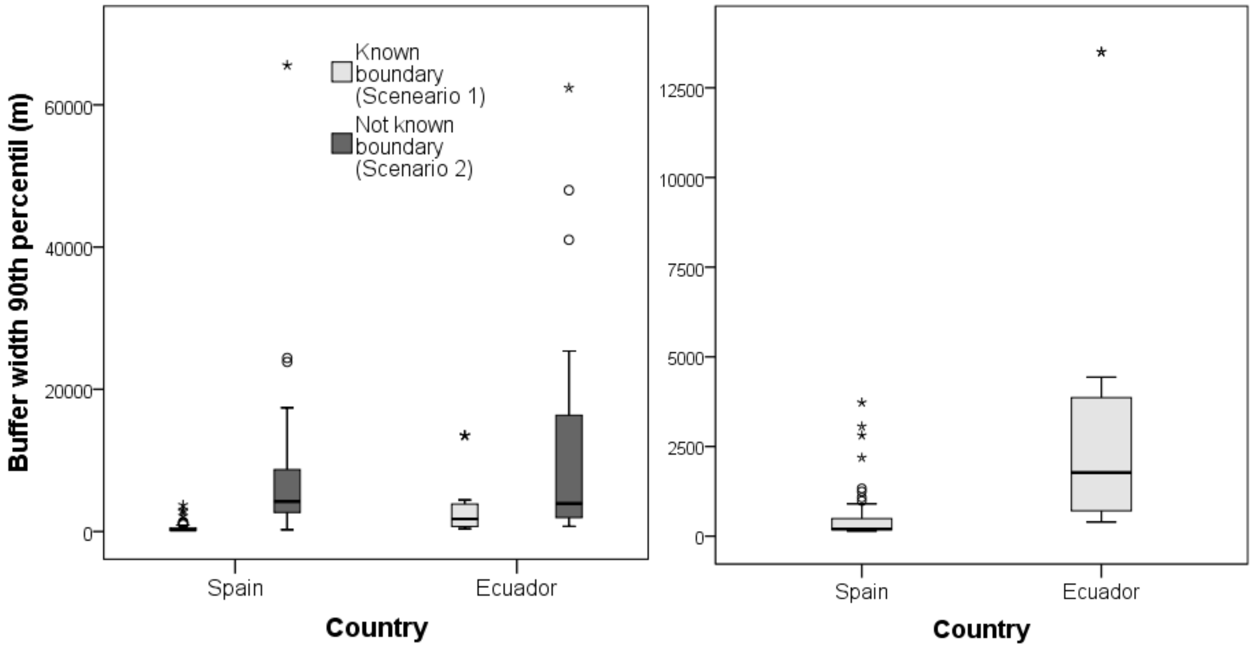



**Figure 6. Boxplot for positional bias measured with the overlay method (Goodchild and Hunter, 1997). The left chart depicts the two scenarios: known boundary polygons (light grey) and polygons built from an area value (dark grey). The right chart shows only the first scenario for known boundary polygons for a friendlier visualisation. Outliers are defined by values more than 1.5 times the spread, box length, from the upper or lower edge of the box. The box length is the interquartile range.**

**Table XVI. Statistic summary for the 90$^{th}$ percentile buffer width (N=136).**

| Country | Scenario | N | Mean (m) | Median (m) | Standard Deviation (m) | Minimum (m) | Maximum (m) |
|---|---|---|---|---|---|---|---|
| Spain | 1 | 50 | 565.9 | **199.4** | 784.2 | 140.2 | 3,725.9 |
| | 2 | 26 | 8,837.4 | **4,214.5** | 13,495.8 | 239.6 | 65,563.2 |
| Ecuador | 1 | 30 | 3,028.6 | **1,771.1** | 3,777.1 | 395.4 | 13,500.0 |
| | 2 | 30 | 12,203.9 | **3,930.9** | 15,314.8 | 727.7 | 62,384.8 |

As Goodchild and Hunter (1997) point out, the method has not been confirmed in practice for area features. To verify the robustness of the method to be applied in area features, an analysis of the actual area and perimeter length of each tested protected area is correlated with the 90th percentile of buffer width (see figure 7).

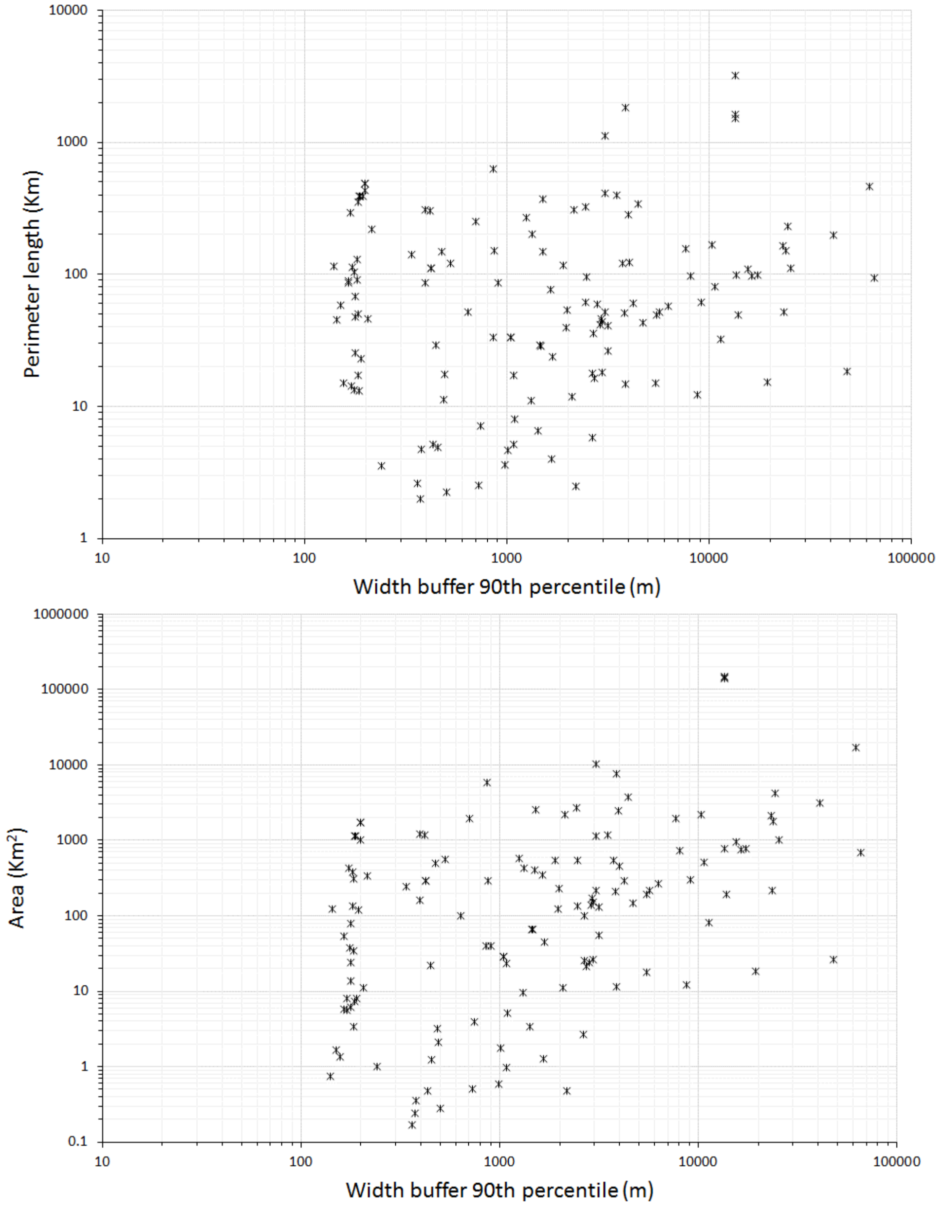



**Figure 7. Scattergrams of the 90th percentile width buffer bias (m) versus its perimeter length in km (above chart) and its polygon area in km² (below chart) calculated by geoprocessing. Axes are shown in logarithmic scale in base ten. The Pearson correlation coefficient had a value of 0.2 (p=0.06) for the perimeter length and a value of 0.4 (p<0.01) for the area. The sample size is N=136.**

The variables are non-parametric distributed with a positive skew. To compute bivariate correlation, the variables are first transformed with a logarithmic function to bring them to a normal distribution. The results of the correlation has a Pearson coefficient of 0.2 (p=0.06) and 0.4 (p<0.01) for both perimeter and area respectively. Therefore, it can be said that there is no evidence that the method systematically depends on the area or the length of the feature tested.

### *3.4.2. Monte Carlo simulation of the 90th percentile buffer width bias*

The stochastic simulation of the 90th percentile was performed 4000 times. The 90th percentile distribution was obtained by the buffer overlay method for a sample of 136 protected areas. This bias value is modelled by a distribution normal after its normalisation as it is described in the methodology chapter.

The result show mean values of around 339 m and 1730 m for Spain and Ecuador respectively. The distance bias is five times bigger in Ecuador than in Spain. The interval of confidence are between approximately 349 and 330 for Spain and between 1791 and 1673 for Ecuador with p=0.95. It is important to note that these values come from the original non-parametric distribution. The retransformation to the original values bring statistics that have been moved from a normal to a non-normal function. However, the simulated figures give a mean representative bias of Euclidian distance in meters around a protected area.

**Table XVII. Summary Statistics for stochastic simulation of the Width 90th percentile (m) with a sample of 4000 realisations.**

| ***Spain*** | | | | |
|---|---|---|---|---|
| ***Central Tendency (Location)*** | | | | |
| **Mean:** | 339.3 | **Median:** | 342.7 | |
| StErr: | 0.0 | | | |
| ***95% Confidence Interval for Mean*** | | | | |
| **Upper Bound** | 348.8 | **Lower Bound** | 330.0 | |
| ***Spread*** | | | | |
| **StDev:** | 2.5 | | | |
| Max: | 8522.2 | Q(.75): | 620.4 | |
| Min: | 13.5 | Q(.25): | 188.3 | |
| **Range:** | 629.3 | **IQ Range:** | 432.0 | |
| ***Quartiles, Percentiles, Intervals*** | | | | |
| | **90% Interval** | | **95% Interval** | |
| Q(.05): | 76.8 | Q(.025): | 55.5 | |
| Q(.95): | 1478.0 | Q(.975): | 1904.6 | |
| **Alpha (a):** | 0.1 | Q(a/2): | 55.5 | |
| % Interval: | 1.0 | Q(1a/2): | 1904.6 | |

| ***Ecuador*** | | | |
|---|---|---|---|
| ***Central Tendency (Location)*** | | | |
| **Mean:** | 1730.9 | **Median:** | 1753.6 |
| StErr: | 0.0 | | |
| ***95% Confidence Interval for Mean*** | | | |
| **Upper Bound** | 1791.0 | **Lower Bound** | 1672.8 |
| ***Spread*** | | | |
| **StDev:** | 3.0 | | |
| Max: | 62101.6 | Q(.75): | 3615.7 |
| Min: | 52.1 | Q(.25): | 820.4 |
| **Range:** | 1191.5 | **IQ Range:** | 2795.2 |
| ***Quartiles, Percentiles, Intervals*** | | | |
| | **90% Interval** | | **95% Interval** |
| Q(.05): | 265.8 | Q(.025): | 201.9 |
| Q(.95): | 10160.0 | Q(.975): | 14756.7 |
| **Alpha (a):** | 0.1 | Q(a/2): | 201.9 |
| % Interval: | 1.0 | Q(1a/2): | 14756.7 |

### *3.4.3. Polygonal area error model*

The simulations of the error model carried out 500 realisations in shapefile structures for two study sites: the Cotacachi Cayapas Ecological Reserve (WDPA ID = 184) in Ecuador and the Doñana National Park (WDPA ID = 61611) in Spain (see figure 8). These features representing possible protected area shapes were compared with the reference source and then the difference in area were computed.

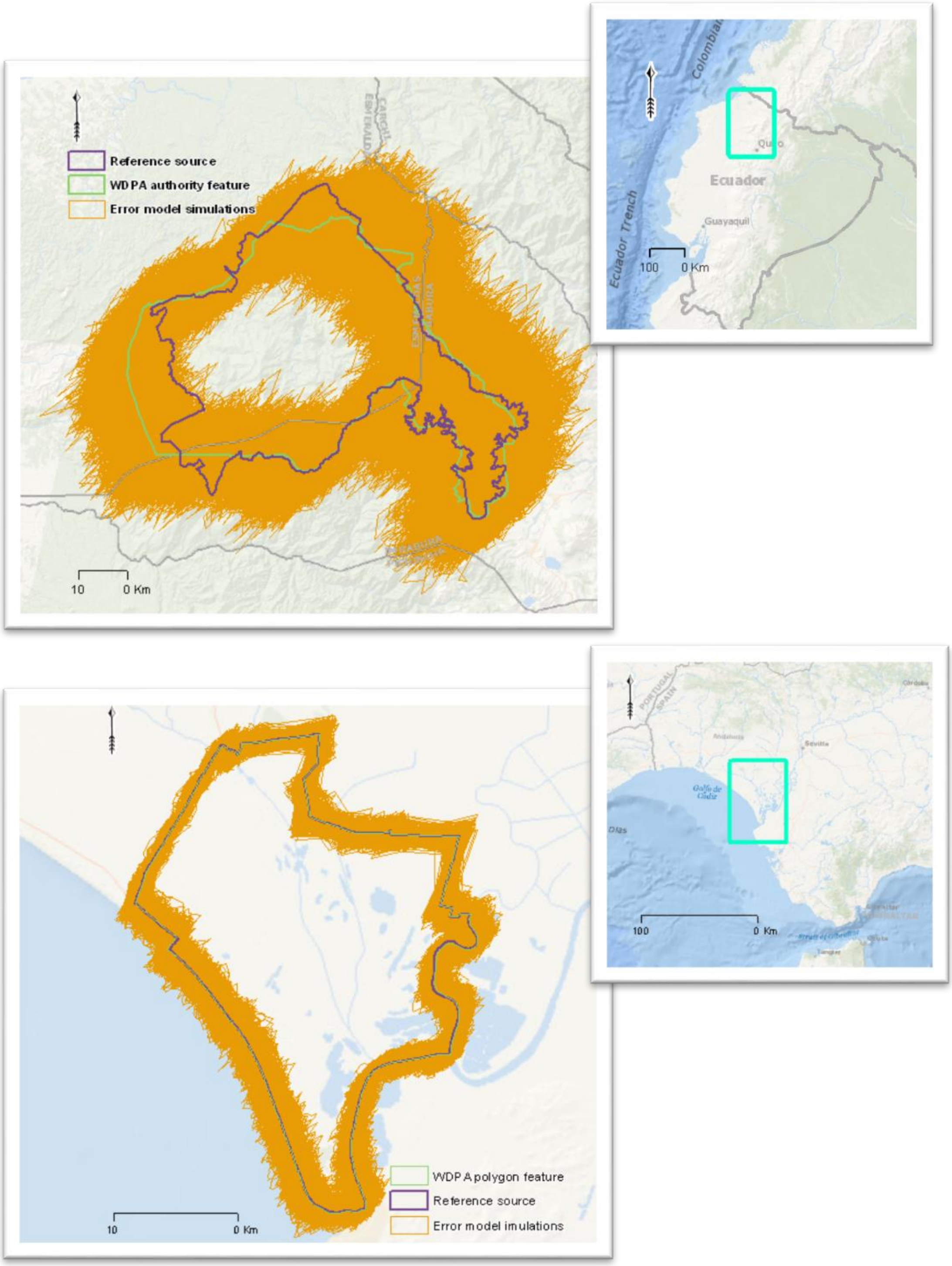



**Figure 8. Maps illustrating the simulations that were performed in DUE software for the Cotacachi Cayapas Ecological Reserve (WDPA ID = 184) in Ecuador (above map) and the Doñana National Park (WDPA ID = 61611) in Spain (below map).**

The histograms of the difference in area are depicted in figure 9. These values are computed taking into account gain and loss in horizontal surface. Negative values represent that the PA has less surface than the reference source. On the contrary, positive values mean that the PA has more surface compared to the reference source.

The WDPA sample data show an average bias of approximately 166 and 2 km$^2$ for the Cotacachi Cayapas Ecological Reserve (WDPA ID = 184) in Ecuador and the Doñana National Park (WDPA ID = 61611) in Spain, respectively. In both counties the averages have positive values which means that the surface simulated is on average higher than the reference source.

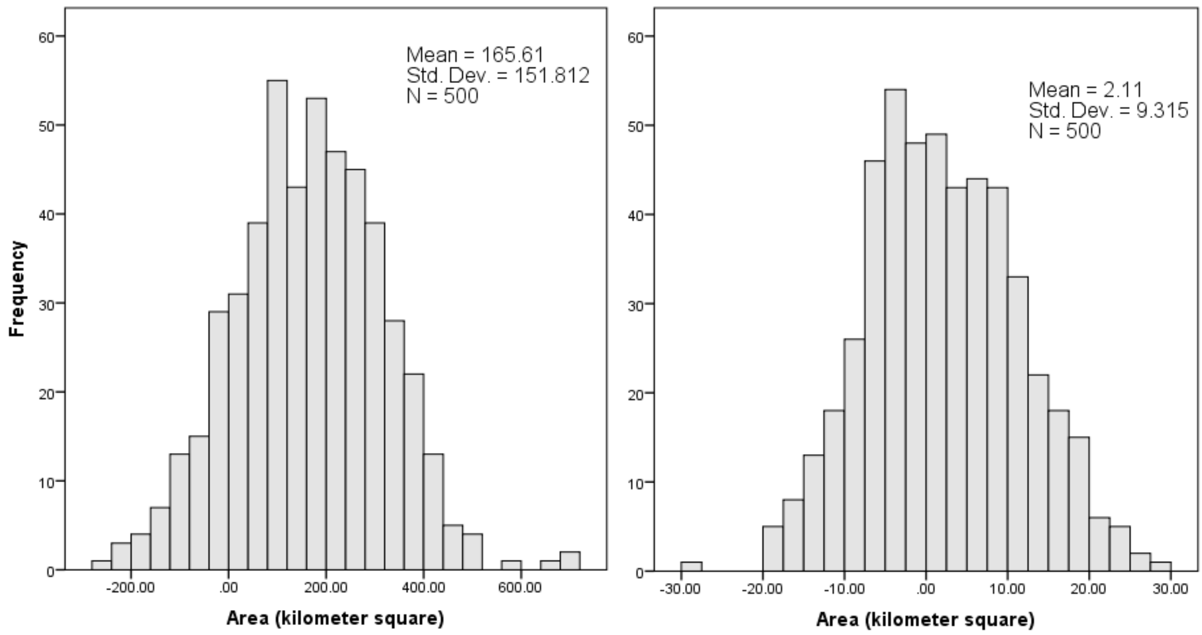



**Figure 9. Histograms with the frequencies of the error model simulation outputs. On the left hand, the Cotacachi Cayapas Ecological Reserve (WDPA ID = 184) in Ecuador. On the left hand, Doñana National Park (WDPA ID = 61611) in Spain. The horizontal axes shows the difference in area between the simulated polygon and the reference source of higher accuracy.**

## 4. CONCLUSIONS

Regarding to the logical consistency, two minimum required attributes show missing values. The 'Status Year' field is reported 4% and 15% of the time in Spain and Ecuador, respectively. In Spain, 2% of protected areas have missing values in the 'Reported Area' attribute while Ecuador has 1% missing. These missing values may have implications in the accounting of annual statistics reports produced by the WCMC-UNEP. For instance, the protected area layers created by the known year of establishment are not considering these entities with missing values.

In the case of Ecuador, when comparing coverage analysis in protected areas represented by polygons against those being represented by points (buffers), the bias is approximately 12,528 $km^2$ higher when point features are included in the analysis. This number increases up to about 19,065 $km^2$ in the case of Spain. Therefore it can be concluded that the inclusion of points in a coverage analysis implies less accuracy.

The overlay buffer method (Goodchild and Hunter, 1997), originally formulated for linear features, has been put in practice for area features. It has been demonstrated that the robustness of the method does not depend on the length or area of the tested polygons.

The distance bias obtained by this method and stochastically simulated by a normalized distribution gives mean values around 339 m and 1730 m for Spain and Ecuador respectively. The distance bias is significantly higher in Ecuador than in Spain according to the analysis made in this essay. Further analysis could address the cause of these discrepancies. Accordingly to these results, the area simulation performed in DUE software for two protected areas showed average bias of approximately 1662 $km^2$ and 2 $km^2$ for Cotacachi Cayapas Ecological Reserve in Ecuador and Doñana National Park in Spain respectively.

In conclusion, the WDPA datasets for Spain and Ecuador show significant differences in terms of spatial data accuracy.  A geographical dependence may be related with national mechanism of collection, gathering and reporting geographical information to the WCMC-UNEP.

# 5. APPENDICES

# Appendix I. Metadata attributes

| Field Name | Full Name | Requirement | Attribute Definition |
|---|---|---|---|
| **METADATAID** | Metadata ID | Minimum | A unique identification number assigned by UNEP-WCMC. |
| **DATA_TITLE** | Data Set Title | Minimum | The title of the dataset being provided as an update to the WDPA. Several PA features can have the same Metadata ID associated. |
| **RESP_PARTY** | Responsible Party | Minimum | The organisation, consultancy, national government, private company or other entity that claims ownership/authorship of the data or that is providing the data on behalf of the ownership/authorship entity. |
| **RESP_EMAIL** | Responsible Party Contact E-mails | Minimum | Contact e-mails of person(s) and organisation(s) associated with the resource. |
| **YEAR** | Date | Minimum | The reference date, as a four digit year, indicating when the dataset was last updated or created prior to inclusion in the WDPA. |
| **LANGUAGE** | Dataset Language | Minimum | Language(s) used within the dataset (before translation into English or transliteration into Latin characters). |
| **CHAR_SET** | Dataset Character Set | Minimum | Full name of the character coding standard used for the dataset. |
| **REF_SYSTEM** | Coordinate System | Minimum | Name and parameters of the coordinate system of the original dataset including where applicable, datum, ellipsoid or projection. The WDPA is based on Geographic Coordinate System: World Geodetic Survey (WGS) 1984. |
| **SCALE** | Scale | Optional | The scale of the source data given as the denominator of the representative fraction. For example, on a scale of 1:150000, the denominator would be 150000. |
| **LINEAGE** | Lineage | Optional | Information about an event, change or transformation in the life of a dataset including the process used to create and maintain the dataset including dates associated with each event. |
| **CITATION** | Citation | Optional | Recommended text to be used referencing for the dataset on www.protectedplanet.net. |
| **DISCLAIMER** | Disclaimer | Optional | Warnings/exceptions to use of the data, displayed on www.protectedplanet.net. |

# *Appendix II. Spatial data attributes*

| Field Name | Full Name | Requirement | Attribute Definition |
|---|---|---|---|
| **WDPAID** | WDPA ID | Minimum [1] | A unique identification number assigned by UNEP-WCMC. If a protected area has already been assigned a WDPA ID it should be included in update submissions. If a protected area is not already listed in the WDPA, UNEP-WCMC will assign new IDs and report those back to the data provider. |
| **WDPA_PID** | WDPA Parent ID | Minimum [1] | Parent ID is assigned by UNEP-WCMC to legal zones of a protected area, therefore only records representing zones in the WDPA will have WDPA_PIDs. The WDPA ID of the overarching or 'parent' protected area becomes the WDPA Parent ID of the zone. |
| **NAME** | Name | Minimum | The name of the protected area provided in Latin characters (including accents). Numeric strings, addresses, acronyms and abbreviations are not accepted. |
| **COUNTRY** | Country | Minimum | The country, territory or other administrative unit of geographical interest that a protected area jurisdictionally resides within, as given by its ISO 3166-1 alpha-3 code. |
| **DESIG** | Designation | Minimum | The type of protected area as legally/officially established or recognised (e.g. Parque Nacional, World Heritage Site, etc.) supported by UTF 8 encoding. |
| **DESIG_TYPE** | Designation Type | Minimum | Describes whether a protected area is 'National' or 'International' by designation. International applies to protected areas designated under a convention, commission or regional agreement such as ASEAN Heritage, Barcelona, OSPAR, HELCOM, Natura2000, RAMSAR, UNESCO World Heritage or Man and Biosphere Programme. |
| **MARINE** | Marine | Minimum | Marine protected areas, as defined for the WDPA, encompass any portion of the marine environment in whole or in part according to a protected area's geographic location and management strategy. Either '1' for True or '0' for False. Mixed marine and terrestrial protected areas should be reported as '1'. |
| **REP_M_AREA** | Reported Marine Area ($km^2$) | Minimum | If Marine is '1' a reported marine area must be given as the total marine extent of the protected area in square kilometres. |
| **REP_AREA** | Reported Area ($km^2$) | Minimum | Total protected area extent, including both marine (if applicable) and terrestrial areas in square kilometres. |
| **STATUS** | Status | Minimum | The current legal or 'official' standing of the protected area. Either 'Proposed' or 'Designated. |
| **STATUS_YR** | Status Year | Minimum | The year in which the current status was officially decreed. |
| **ORIG_NAME** | Original Name | Core | The name of the protected area in any language supported by UTF 8 encoding. |
| **SUB_LOC** | Sub-national Location | Core | The principle subdivision that a protected area geographically resides within, given by an ISO 3166-2 sub-national code, e.g. autonomous region, overseas territory, dependency, possession, etc. as long as it does not already have an ISO 3166-1 alpha-3 character code in which case it should be reported as a country. |
| **DESIG_ENG** | English Designation | Core | The type of protected area as legally/officially established or recognized translated into English. |
| **IUCN_CAT** | IUCN Category | Core | The classification of IUCN Management Category (Ia, Ib, II, III, IV, V or VI) adopted for national protected areas. For reporting on international protected areas the option of listing 'Not Applicable' is accepted. For national protected areas where an IUCN category has not been adopted 'Not Reported' can be listed. |
| **GOV_TYPE** | Governance Type | Enhanced | A description of the governance structure of a protected area, written as one of the 11 governance sub-types, described in the IUCN Guidelines on protected areas. |
| **MANG_AUTH** | Management | Enhanced | Authority The organisation(s) or agency (ies) responsible for management of the protected area. |
| **MANG_PLAN** | Management Plan | Enhanced | Reference to an official management plan for the protected area as a link to the document(s) on-line or a full bibliographic reference. |
| **NO_TAKE** | No Take | Enhanced | Listed when part or all of a marine protected area is no take, meaning that the taking of fish or living resources is strictly prohibited in the no take area. |
| **NO_TK_AREA** | No Take Area | Enhanced | The total size of the no take area in square kilometres. |
| **METADATAID** | Metadata ID | UNEP-WCMC Assigned | An ID assigned by UNEP-WCMC and is used to link source tables to WDPA shapefiles. |
| **GIS_AREA** | GIS Area ($km^2$) | UNEP-WCMC Assigned | The GIS area calculated by UNEP-WCMC using the Mollweide projection. The reliability of the GIS area depends on the accuracy of the polygon provided, and hence can differ substantially from the Reported Area 'REP_AREA'. |
| **GIS_M_AREA (if applicable)** | GIS Marine Area ($km^2$) | UNEP-WCMC Assigned | The reliability of a calculated marine area is dependent on accuracy of the polygon provided, and hence can differ substantially from the Reported Marine Area 'Rep_M_Area'. For mixed protected areas that are only marine the boundaries of the marine area may be submitted, if available, so partially that the GIS marine area can be calculated and included as 'GIS_M_Area'. The marine area is not stored as a separate record in the WDPA unless it is legitimately zoned. |
| **INT_CRIT** | Assigned International Criteria | UNEP-WCMC Assigned | Applicable only to World Heritage Sites where a set of criteria is used to define the protected area. For national protected areas this field is 'Not Applicable'. |

[1] WDPA and WDPA ID are assigned by the UNEP-WCMC although are responsibility of the data provider.

## *Appendix III. Python script to calculate the polygon positional accuracy based on Goodchild (1997)*

```
# ---------------------------------------------------------------------------
# Goodchildv13.py
# Created on: 2014-06-19
# Generated by ArcGIS/ModelBuilder and modified on PythonWin
# Description: Procedure to calculate the proportion of the Tested Area Perimeter enclosed within the buffer around the Reference Polygon Source based on Goodchild (1997)
# Author: Maria Feria Aguaded
# Master Dissertation Title: Spatial data quality assessment in the World Database on Protected Areas for Spain and Ecuador
# This dissertation is part of a Msc degree in Environmental Monitoring, Modelling and Management at King's College London
# ---------------------------------------------------------------------------


# Import arcpy module
import arcpy

# Local variables
tested = arcpy.GetParameterAsText(0)  #Location of Tested Source --> recorded in the database. To introduce in an arctoolbox window
reference = arcpy.GetParameterAsText(1)  #Location of Reference Source--> determined with higher accuracy. To introduce in an arctoolbox window
output = arcpy.GetParameterAsText(2) #Location of output with proportion of the Tested Perimeter wihin the buffer
arcpy.env.scratchWorkspace = arcpy.GetParameterAsText(3) #Path where of temporal scratch folder will be created
scratchFolder = arcpy.env.scratchFolder #Create scratch folder
outputBuff = arcpy.CreateScratchName("buff","","Shapefile",scratchFolder) #Buffer output
perimTested = arcpy.CreateScratchName("perimT","","Shapefile",scratchFolder) #Tested protected area as a boundary or line
perimReference = arcpy.CreateScratchName("perimR","","Shapefile",scratchFolder) #Reference protected area as a boundary or line
```

```
lineWithinBuff = arcpy.CreateScratchName("line","","Shapefile",scratchFolder) #Proportion of Protected Area Perimeter within the buffer


#Buffer width(m)
buff20 = 20
buff40 = 40
buff60 = 60
buff80 = 80
buff100 = 100
buff125 = 125
buff150 = 150
buff200 = 200
buff250= 250
buff300= 300
buff350 = 350
buff400 = 400
buff450 = 450
buff500 = 500
buff550 = 550
buff600 = 600
buff700 = 700
buff800 = 800
buff900 = 900
buff1000 = 1000
buff1250 = 1250
buff1500 = 1500
buff1750 = 1750
buff2000 = 2000
buff2500 = 2500
buff3000 = 3000
buff3500 = 3500
buff4000 = 4000
buff4500 = 4500
buff5000 = 5000
buff6000 = 6000
```

```
buff7000 = 7000
buff8000 = 8000
buff9000 = 9000
buff10000 = 10000
buff12000 = 12000
buff15000 = 15000
buff20000 = 20000
buff30000 = 30000
buff50000 = 50000
buff75000 = 75000

# Process: Polygon To line
arcpy.PolygonToLine_management(tested,perimTested,"IGNORE_NEIGHBORS")
arcpy.PolygonToLine_management(reference,perimReference,"IGNORE_NEIGHBORS")

# Process: Create a buffer of width equal to xi around the Reference Perimeter
# Create a list of buffering distances
buffList = [buff20, buff40, buff60, buff80, buff100, buff125, buff150, buff200, buff250, buff300,
buff350, buff400, buff450, buff500, buff550, buff600, buff700, buff800, buff900, buff1000, buff1250,
buff1500, buff1750, buff2000, buff2500, buff3000, buff3500, buff4000, buff4500, buff5000, buff6000,
buff7000, buff8000, buff9000, buff10000, buff12000, buff15000, buff20000, buff30000, buff50000,
buff75000]

#Loop through each distance and create a buffer
for buff in buffList:
   # Process: Buffer
   arcpy.Buffer_analysis(perimReference,outputBuff[:-4] + str(buff),
buff,"FULL","ROUND","NONE","#")
   # Process: Indicate Buffer Width
   arcpy.AddField_management(outputBuff[:-4] + str(buff)+
".shp","BWIDTH_M","INTEGER","#","#","#","#","NULLABLE","NON_REQUIRED","#")
   arcpy.CalculateField_management(outputBuff[:-4] + str(buff)+ ".shp","BWIDTH_M",
buff,"PYTHON_9.3","#")
   # Process: Intersect
   arcpy.Intersect_analysis([outputBuff[:-4] + str(buff) + ".shp" , perimTested], lineWithinBuff[:-4] +
str(buff), "NO_FID", "#", "INPUT")
```

```
    #Merge outputs from all intersects
    if buffList.index(buff) != 0:
        #temp arcpy.Merge_management(lineWithinBuff[:-4] + str(buff) + ".shp", lineWithinBuffAll, "#")
        arcpy.Append_management(lineWithinBuff[:-4] + str(buff) + ".shp",lineWithinBuff[:-4] + str(buffList[0]) + ".shp", "TEST","#","#")

lineWithinBuffAll = lineWithinBuff[:-4] + str(buffList[0]) + ".shp"

# Add Field
arcpy.AddField_management(perimTested,"PERIM_M","DOUBLE","#","#","#","#","NULLABLE","NON_REQUIRED","#")
arcpy.AddField_management(lineWithinBuffAll,"LENGTH_M","DOUBLE","#","#","#","#","NULLABLE","NON_REQUIRED","#")
arcpy.AddField_management(lineWithinBuffAll,"LINEWIB","DOUBLE","#","#","#","#","NULLABLE","NON_REQUIRED","#")

# Process: Calculate Perimeter Length
perimField = "PERIM_M"
lengthField = "LENGTH_M"
arcpy.CalculateField_management(perimTested,perimField,'float(!shape.length@meters!)',"PYTHON_9.3","#")
arcpy.CalculateField_management(lineWithinBuffAll,lengthField,'float(!shape.length@meters!)',"PYTHON_9.3","#")

# Call PERIMETER_M rows
perimField = "PERIM_M"
perimAverage = 0
perimTotal = 0
perimCount = 0

with arcpy.da.SearchCursor(perimTested, (perimField,)) as cursor:
    for perimRow in cursor:
        perimTotal += perimRow[0]
        perimCount += 1

perimAverage = perimTotal / perimCount
```

```
# Process: Calculate the proportion of the Tested Perimeter enclosed within the buffer
lineWIBuffField = "LINEWIB"
arcpy.CalculateField_management(lineWithinBuffAll,lineWIBuffField,'!LENGTH_M! * 100 / %f' % (perimAverage),"PYTHON_9.3","#")

#Copy result feauture
arcpy.CopyFeatures_management(lineWithinBuffAll, output)
```

# *Appendix IV. Geographic Risk Assessment Form*



## Geography Individual Research and Fieldwork

**(This form is available electronically via http://www.kcl.ac.uk/geography then "For Current Students")**

## 1 GENERAL GUIDANCE ON RISK ASSESSMENT: MANAGING HEALTH AND SAFETY

In most countries, national legislation provides the legal framework for health and safety management. For example, in the UK, universities abide by the *Health and Safety at Work Act 1974* (http://www.hse.gov.uk/legislation/hswa.htm) and associated regulations. The Act places a duty upon employers to take steps to ensure, so far as reasonably practicable, the health and safety of their employees and any other people affected by their activities including, in the case of universities, all students and members of the public.

Additional guidelines may also exist at the local or national scale. For example, there is a British standard concerning *Specification for the provision of visits, fieldwork, expeditions, and adventurous activities, outside the UK* (BS8848). These laws and guidelines typically require that 'risk assessments' are undertaken to identify what should be done in order to manage safety. Assessments of risk are usually focused around 'risk' and 'hazard'. Put simply:

- **Hazards** result from working in potentially dangerous environments and refer to environmental conditions, agents (including strangers) or substances that can cause harm.
- **Risk** is the chance that a person (you or someone else) might be harmed by the hazard.

During individual research (e.g., fieldwork), hazards and risks can change rapidly, for example, as a result of changing weather conditions or political actions, and should be continually reassessed.

A risk assessment should be completed for:

- All individual research and fieldwork taking place OUTSIDE the Department of Geography conducted as part of your degree,
- All laboratory work INSIDE the College premises
- All student dissertation work, whether human or physical, and whether undergraduate, postgraduate taught or postgraduate research.

The extent of your involvement in actually assessing the risks will vary according to the way in which the individual research and fieldwork is organised, but you have a responsibility to follow any precautions or safety measures laid down in the risk assessment.

This form is meant to be a formative process of considering risk — not just a form-filling 'bureaucratic exercise' — and for you to actively consider risks involved in your research/project that might be greater than everyday life and your normal activities. You should consider a five-step approach to risk assessment.

1. Identify the hazards: during the field work, what could cause harm, for example, slippery ground, high altitude, weather conditions, civil or political unrest.
2. Identify who might be harmed and how: this includes all the field workers and members of the public.
3. Evaluate and minimise the risks: once suitable precautions have been taken, for example, wearing appropriate protective clothing, how likely is it that someone will be harmed?
4. Record the findings: write down the identified hazards and precautions to be taken.
5. Review the assessment periodically: situations change and so do the potential hazards and risks, so risk assessment is not an isolated task but an ongoing evaluation to be updated and revised as often as necessary.

## 2 COMPLETING THE DEPARTMENT OF GEOGRAPHY RISK ASSESSMENT FORM

In order to participate in fieldwork activity and for all student dissertations, you must complete this Departmental Risk Assessment form. This form is designed to make the process intelligible and manageable, and you will need to demonstrate that you have considered *each* listed risk by indicating with either 'yes' or 'no' whether the risk applies to your fieldwork. ALL risks must be answered 'yes' or 'no' even if they do not apply to you. If the answer is to any particular risk is 'yes' then you must also indicate the degree of risk from 1–5 (1=low, 5=high) and complete the relevant comments. .At the end of the form you should add any further relevant risks for your fieldwork that have not been listed.

**THE MAJOR CATEGORIES OF RISK IN THIS FORM ARE AS FOLLOWS:**

| | |
|---|---|
| 1 Lone / out of hours working in the field. | 11 Health: Accidents and illness |
| 2 Environment: Hazardous Weather/Climate | 12 Health: Allergies |
| 3 Environment: Hazardous Terrain | 13 Health: Food and Drink |
| 4 Environment: Location | 14 Equipment: Using Equipment and Equipment Failure |
| 5 Environment: Animals | 15 Equipment: Checking Equipment |
| 6 Environment: Pollution | 16 People: Unexpected Behaviour |
| 7 Manual Handling: Loading/Unloading | 17 War or Military Zones |
| 8 Manual Handling: Moving Equipment to Site | 18 Additional Risks Related to the Specific Project (USE THIS SECTION FOR ADDITIONAL RELEVANT RISKS) |
| 9 Working Near Water: General | |
| 10 Working Near Water: Boat handling | |

## 3 RISK ASSESSMENT FORM AND ASSOCIATED DOCUMENTATION

After reading through ALL risk categories, please select RISK TYPE A or B below.

### RISK TYPE A

**You are only eligible for RISK TYPE A if all of the following are true:**

- **Your work take place within: college premises or home or within organizations/premises that have their own clear risk assessment in place.**
- **Your work involves ONLY library/archival data or existing on-line/other data.**
- **Your work WILL NOT expose you to risks greater than in everyday life.**

DECLARATION: I have considered ALL categories in this form and I declare that I am undertaking a student project/dissertation where: a) NONE of my research will be outside of college premises or home or organizations/premises that have their own clear risk assessment in place; and b) it does not involve ANY of the risks identified in ANY of the categories of this risk assessment form. Should my research project change, such that there are now risks involved, then it is my responsibility to resubmit this form after completing an assessment for Risk Type B.

| SIGNATURES OF PERSON FILLING IN A RISK ASSESSMENT AND COUNTERSIGNATURE. |
|---|
| A. Person filling in this risk assessment |
| Name (Typed or printed in BLOCK letters): MARÍA FERIA AGUADED |
| Signature: |
| Date: 26th June 2014 |
| |
| **B. Countersignature and date** (Students – Research Supervisor; Research Staff – Project Leader; Academic Staff – Head of Department) |
| Name (Typed or printed in BLOCK letters): MARK MULLIGAN |
| Signature: |
| Date: 26th June 2014 |

Print this page in triplicate; the three copies signed and countersigned, and lodged with:
(1) Your supervisor.
(2) The Department Office.
(3) One for retention by yourself.
For UGT and PGT students, this signatures page of your risk assessment must be included in Appendix 1 of your dissertation.

### RISK TYPE B

**Fill out THIS PAGE and ALL OTHER PAGES in this form.**

DECLARATION: I have considered ALL categories in this form and have indicated which risks apply to me that are greater than in everyday life and normal activities (writing yes/no for every section). Where I have answered 'yes' then I have also indicated the degree of risk from 1–5 (1=low, 5=high) and, where appropriate, added notes or comments relating to the level of risk. I have identified and added any additional risks not explicitly covered by this form in the final section.

| SIGNATURES OF PERSON FILLING IN A RISK ASSESSMENT AND COUNTERSIGNATURE. |
|---|
| **A. Person filling in this risk assessment** |
| Name (Typed or printed in BLOCK letters): |
| Signature: |
| Date: |
| |
| **B. Countersignature and date** (Students – Research Supervisor; Research Staff – Project Leader; Academic Staff – Head of Department) |
| Name (Typed or printed in BLOCK letters): |
| Signature: |
| Date: |

All pages in this form should be printed in triplicate; the three copies signed and countersigned, and lodged with:
(1) Your supervisor.
(2) The Department Office.
(3) One for retention by yourself before fieldwork commences.
For UGT and PGT students, this signatures page of your risk assessment must be included in Appendix 1 of your dissertation.

For work outside of the UK, please do not forget to obtain insurance in accordance with College regulations (application form https://internal.kcl.ac.uk/about/ps/finance/treasry/insure.aspx).

# Appendix V. Research Ethics Screening Form

## Department of Geography Research Ethics Screening Form
### King's College London

**Please Note: Filling out this *Geography Research Ethics Screening Form* does NOT constitute *College Ethics Approval.***

This *Geography Research Ethics Screening Form* will help you to determine if you must submit a *College Research Ethics Application* to the *College Research Ethics Committees* before starting your research, under the guidelines for working with human participants set out by the Social Sciences, Humanities & Law Research Ethics Sub-Committee (SSHL RESC), and the Geography, Gerontology and Social Care Workforce Research Unit Panel (GGS REP).

In order to complete this process, please

(a) Familiarise yourself with the professional research ethics guidelines of *The British Sociological Association*: http://www.britsoc.co.uk/equality/ (Statement of Ethical Practice)

(b) Read "Which kinds of research require ethical approval through the KCL Research Ethics Committees?" (p. 2 of this form).

(c) Answer the questions in Table 1 below, sign the form and also obtain the signature of your supervisor.

(d) Return the signed (by both you and your supervisor) *Geography Ethics Screening form* to the Geography Department office and KEEP A COPY which you will place in Appendix 1 of your IGS/dissertation.

(e) If ethics approval is needed (answering 'yes' to question 2 in Table 1), you must apply for college ethics approval through the appropriate *College Research Ethics* committee, and **not start ANY research (including preliminary 'trials')** until ethics approval has been granted in writing.

*Table 1. Department of Geography Research Ethics Screening Questions.*

| | |
|---|---|
| 1) Have you read and familiarised yourself with the professional research guidelines of *The British Sociological Association*? | ☒ Yes ☐ No |
| 2) Does your research "involve human participants" and/or "raise other ethical issues with potential social or environmental implications"? | ☐ Yes ☒ No |

If you answered 'No' to question two, you do not need to submit your research for ethical review. If you answered 'Yes' to question two, please use the flowchart on http://www.kcl.ac.uk/innovation/research/support/ethics/about/index.aspx to establish your risk level and where you need to apply (see Table 2).

*Table 2. Three levels of risk for project types, and how to obtain College Research Ethics clearance.*

| *Project type* | *How to submit?* |
|---|---|
| Low risk: | Can be reviewed using an on-line process. The process includes guidelines and prompts to help ensure your project is low-risk: http://www.kcl.ac.uk/innovation/research/support/ethics/applications/lowrisk/index.aspx |
| Moderate risk:<br>Uncertain risk: | Should be submitted to the Geography, Gerontology and Social Care Workforce Research Unit Panel (GGS REP): http://www.kcl.ac.uk/innovation/research/support/ethics/applications/apply.aspx |
| High risk: | Should be submitted to the Social Sciences, Humanities and Law Research Ethics Sub-Committee (SSHL RESC): http://www.kcl.ac.uk/innovation/research/support/ethics/committees/sshl/highrisk.aspx |

In all cases, even if 'no' risk, you **MUST** sign and return this *Geography Research Ethics Screening Form* to be kept on file with the Department Office, and if an Undergraduate or Masters student, submit a copy of this at the end (as part of Appendix 1) of your IGS or Dissertation. In cases where there is low, moderate or high ethics risk, **you MUST complete the *College Research Ethics Application* at least one month before you intend to start your research and obtain written approval from them BEFORE carrying out any research.**

**Carrying out research without ethical approval by the College Ethics Committee may result in a charge under misconduct regulations as "*action that deviates from accepted institutional, professional, academic or ethical standards will be regarded as misconduct and an infringement of these regulations*" "Academic regulations, Regulations concerning students & General regulations" B3 – 1.1, King's College London. You should note that your research will not be covered by the College's insurance until you have completed the College ethical review process. This means that unless you receive ethical approval for your research, if a participant makes a legal claim regarding the research, then you would be personally liable. It is your responsibility to submit your research for *College Ethical Review* in good time to carry out any research.**

Provisional IGS/dissertation title: Quality Assessment and Analysis: How Much Do Data Gaps in the WDPA Matter?

Student Name: María Feria Aguaded Student Card No: 1256079

Student Signature: Date: 29/April/2014

Supervisor Name: Mark Mulligan

Supervisor Signature: Date: 07/May/2014

## 6. REFERENCES CITED

Aalders, H. Quality metrics for GIS. 7th International Symposium on Spatial Data Handling, 1996.

Ministerio de Agricultura, Alimentación y Medio Ambiente (MAGRAMA). 2014. *Banco de Datos de la Naturaleza* [Online]. Available at: http://www.magrama.gob.es/es/biodiversidad/servicios/banco-datos-naturaleza/ [Accessed 7th August 2014].

Bell, S. 1999. *A beginner's guide to uncertainty of measurement*, National Physical Laboratory.

Bley, D.and Haller, R. Checking the spatial accuracy of class boundaries with a varying range of accuracy requirements. 7th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences, 2006.

Brown, J. D.and Heuvelink, G. B. M. 2007. The Data Uncertainty Engine (DUE): A software tool for assessing and simulating uncertain environmental variables. *Computers & Geosciences*, 33**,** 172-190.

Burrough, P. A.and McDonnell, R. 1998. *Principles of geographical information systems*, Oxford university Press.

Caprioli, M., Scognamiglio, A., Strisciuglio, G.and Tarantino, E. Rules and standards for spatial data quality in GIS environments. Proc. 21st Int. Cartographic Conf. Durban, South Africa, 2003. 10-16.

Eionet. 2014. *Central Data Repository*. *CDDA-1: Designated Areas* [Online]. Available at: http://cdr.eionet.europa.eu/es/eea/cdda1 [Accessed 7th August 2014.

Chape, S., Harrison, J., Spalding, M.and Lysenko, I. 2005. Measuring the extent and effectiveness of protected areas as an indicator for meeting global biodiversity targets. *Philosophical Transactions of the Royal Society B: Biological Sciences,* 360**,** 443-455.

Chrisman, N. R.and Yandell, B. S. 1988. Effects of point error on area calculations: A statistical model. *Surveying and Mapping,* 48**,** 241-246.

Codd, E. F. 1972. Further Normalization of the Data Base Relational Model, Data Base Systems, Courant Computer Science Symposia Series 6, R. Rustin. Prentice-Hall, Englewood Cliffs, New Jersey.

UNEP-WCMC (2012) *Data Standards for the World Database on Protected Areas* [Online]. Cambridge, UK: UNEP-WCMC, [Accessed.

De Bruin, S., Heuvelink, G.and Brown, J. Propagation of positional measurement errors to field operations. Proceedings of the Fifth International Symposium on Spatial Data Quality, Enschede, The Netherlands (available at http://www. itc. nl/ISSDQ2007/proceedings/session02. html), 2007.

de Bruin, S., Heuvelink, G. B. M.and Brown, J. D. 2008. Propagation of positional measurement errors to agricultural field boundaries and associated costs. *Computers and Electronics in Agriculture*, 63**,** 245-256.

Devillers, R., Bédard, Y., Jeansoulin, R.and Moulin, B. 2007. Towards spatial data quality information analysis tools for experts assessing the fitness for use of spatial data. *International Journal of Geographical Information Science*, 21**,** 261-282.

Drummond, J. 1995. Positional accuracy. *Elements of spatial data quality***,** 31-58.

Fish, L., Fox, P., May, I.and Blyth, S. Development of the world database on protected areas. 2005 ESRI International User Conference Proceedings, 2005.

QA4EO (2010) *A guide to establish a Quality Indicator on a satellite sensor derived data product* [Online]. QA4EO, Available at: http://qa4eo.org/docs/QA4EO-QAEO-GEN-DQK-001_v4.0.pdf [Accessed 31th July 2014].

Goodchild, M. F. 2004. A general framework for error analysis in measurement-based GIS. *Journal of Geographical Systems*, 6**,** 323-324.

Goodchild, M. F. 2008a. Foreword; Imprecision and Spatial Uncertainty; Spatial Data Analysis. *In:* SHEKHAR, S. & XIONG, H. (eds.) *Encyclopedia of GIS.* New York: Springer.

Goodchild, M. F. Spatial accuracy 2.0. Proceedings of the eighth international symposium on spatial accuracy assessment in natural resources and environmental sciences, 2008b. Citeseer, 1-7.

Goodchild, M. F.and Hunter, G. J. 1997. A simple positional accuracy measure for linear features. *International Journal of Geographical Information Science*, 11**,** 299-306.

Goovaerts, P. 1997. *Geostatistics for natural resources evaluation*, Oxford university press.

Joint Committee for Guides in Metrology (JCGM). 2014. *Guides to the expression of uncertainty in measurement (GUM series)* [Online]. Available at: http://www.iso.org/sites/JCGM/GUM-introduction.htm [Accessed 28th July 2014].

Guptill, S. C.and Morrison, J. L. 1995. *Elements of Spatial Data Quality*, Oxford, Elsevier Science.

Heo, J., Kim, J. H.and Kim, J. W. 2009. A new methodology for measuring coastline recession using buffering and non‐linear least squares estimation. *International Journal of Geographical Information Science*, 23**,** 1165-1177.

Hernando Gallego, A., Castilla, G., Zang, C., Mazumdar, D.and Macdermic, G. 2012. An integrated framework for assessing the accuracy of geobia landcover products.

ISO 2013. ISO 19157:2013 Geographic information: Data quality. International Organization for Standardization.

Kainz, W. 1995. Logical consistency. *In:* GUPTILL, S. C. & MORRISON, J. L. (eds.) *Elements of Spatial Data Quality.* Amsterdam: Pergamon.

Kiiveri, H. T. 1997. Assessing, representing and transmitting positional uncertainty in maps. *International Journal of Geographical Information Science,* 11**,** 33-52.

Leung, Y., Ma, J.-H.and Goodchild, M. F. 2004a. A general framework for error analysis in measurement-based GIS Part 2: The algebra-based probability model for point-in-polygon analysis. *Journal of Geographical Systems,* 6**,** 355-379.

Leung, Y., Ma, J.-H.and Goodchild, M. F. 2004b. A general framework for error analysis in measurement-based GIS Part 3: Error analysis in intersections and overlays. *Journal of Geographical Systems,* 6**,** 381-402.

Leung, Y., Ma, J.-H.and Goodchild, M. F. 2004c. A general framework for error analysis in measurement-based GIS Part 4: Error analysis in length and area measurements. *Journal of Geographical Systems,* 6**,** 403-428.

Ley 42/2007, de 13 de Diciembre, del Patrimonio Natural y de la Biodiversidad. 2007. *BOE num 299.* BOE: Boletin Oficial del Estado (BOE).

2011. *Mapa espacios naturales protegidos 2011* [Online]. Available at: http://www.redeuroparc.org/descargasmapas.jsp [Accessed 07th August 2014].

Pebesma, E. J. 2004. Multivariable geostatistics in S: the gstat package. *Computers & Geosciences,* 30**,** 683-691.

Radoux, J.and Bogaert, P. 2014. Accounting for the area of polygon sampling units for the prediction of primary accuracy assessment indices. *Remote Sensing of Environment,* 142**,** 9-19.

Ramirez, J. R.and Ali, T. Progress in metrics development to measure positional accuracy of spatial data. Proceedings of the 21st ICA Conference, 2003.

Refsgaard, J. C., van der Sluijs, J. P., Højberg, A. L.and Vanrolleghem, P. A. 2007. Uncertainty in the environmental modelling process–a framework and guidance. *Environmental modelling & software,* 22**,** 1543-1556.

Salgé, F. 1995. Semantic accuracy. *Elements of spatial data quality***,** 139-151.

Ministerio del Ambiente de Ecuador. 2014. *Sistema Unico de Información Ambiental* [Online]. Available at: http://suia.ambiente.gob.ec/ [Accessed 9th August 2014].

Tong, X., Sun, T., Fan, J., Goodchild, M. F.and Shi, W. 2013. A statistical simulation model for positional error of line features in Geographic Information Systems (GIS). *International Journal of Applied Earth Observation and Geoinformation,* 21**,** 136-148.

Tveite, H. 1999. An accuracy assessment method for geographical line data sets based on buffering. *International Journal of Geographical Information Science,* 13**,** 27-47.

Tveite, H.and Langaas, S. Accuracy assessments of geographical line data sets: The case of the Digital Chart of the World. Proceedings from the Fifth Scandinavian Research Conference on Geographical Information Systems, 1995.

(2006) *User's Manual* [Online]. The Netherlands, Available at: http://harmonirib.geus.info/due_download/index.html [Accessed 31th July 2014].

Veregin, H.and Hargitai, P. 1995. An evaluation matrix for geographical data quality. *Elements of spatial data quality*, 167-188.

Wackernagel, H. Multivariate geostatistics: an introduction with applications. International Journal of Rock Mechanics and Mining Sciences and Geomechanics Abstracts, 1996. Elsevier Science.

UNEP-WCMC. 2014. *The World Database on Protected Areas (WDPA)* [Online]. Available at: www.protectedplanet.net [Accessed 26th March 2014].

Yang, X., Blower, J., Bastin, L., Lush, V., Zabala, A., Masó, J., Cornford, D., Díaz, P.and Lumsden, J. 2013. An integrated view of data quality in Earth observation. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 371**,** 20120072.

Zhang, H. 2007. Maximum‐likelihood estimation for multivariate spatial linear coregionalization models. *Environmetrics*, 18**,** 125-139.

Zhang, J.and Goodchild, M. F. 2002. *Uncertainty in geographical information*, CRC press.

Zhang, J.and Kirby, R. P. 2000. A Geostatistical Approach to Modelling Positional Errors in Vector Data. *Transactions in GIS*, 4**,** 145-159.

Zheng, S.and Zheng, J. 2014. Assessing the Completeness and Positional Accuracy of OpenStreetMap in China. *In:* BANDROVA, T., KONECNY, M. & ZLATANOVA, S. (eds.) *Thematic Cartography for the Society.* Springer International Publishing.