

Out of line, out of mind?

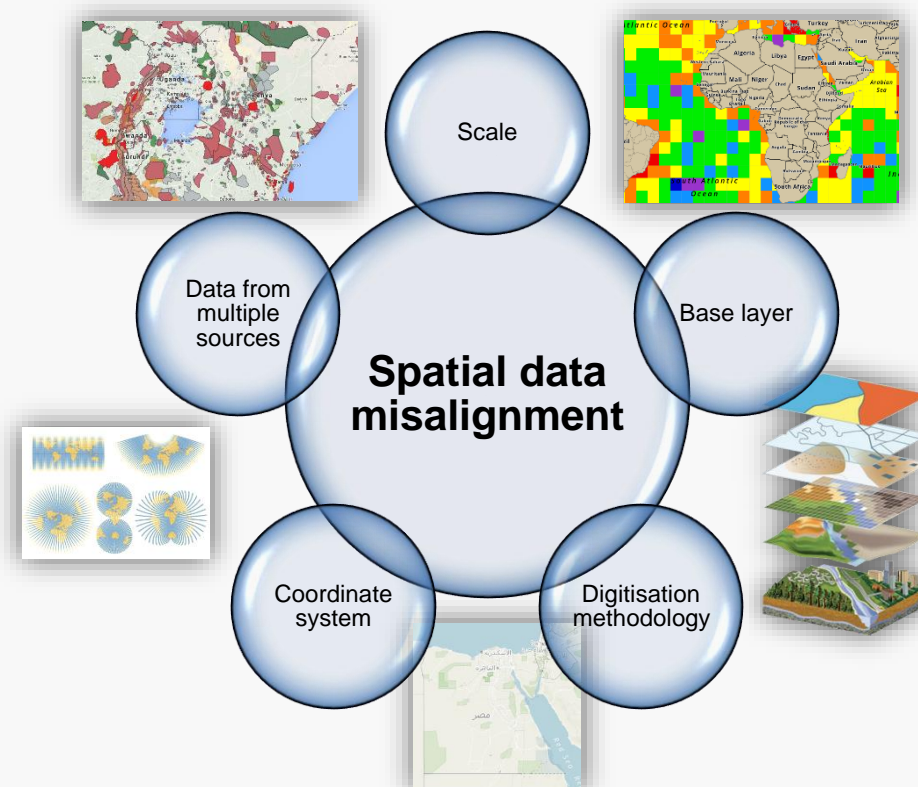
Working with data misalignments in global biodiversity spatial datasets

Key messages

Misalignments are a potential limitation if present in spatial datasets and may arise from differences in base layers, scale, digitisation methodology, coordinate system or when using datasets from multiple sources. However, they are often overlooked and 'out of mind'. Spatial biodiversity data can be important in informing decision-making processes, so it is necessary for data users to be aware of the possible causes of misalignments and where possible follow the guidance provided below.

Practical guidance for data users:

1. *Spatial misalignment, especially in biodiversity datasets, is common and does not necessarily indicate the data are not fit for purpose.*
2. *It is worth checking in the first instance whether just one feature is misaligned or if all features in the area of assessment are misaligned.*
3. *Although it may be common practise, clipping all datasets to one layer in order to correct misalignments is not recommended because it is not always obvious which layer is correctly aligned.*
4. *Obvious errors in datasets provided by UNEP-WCMC should be reported because assistance with problems in particular regions can be provided and the error can often be corrected.*
5. *Data is likely to be most reliable if it has been verified by ground-truthing.*



Introduction

Spatial biodiversity datasets are a useful resource for a wide range of organisations and vital in defining future conservation priorities. Datasets may include the location of protected areas, areas of biodiversity importance, as well as threatened marine, terrestrial and coastal habitats, ecosystems and species.

Due to the increasing use of spatial biodiversity data across sectors, and importance in informing decision-making processes, it is necessary to understand their potential limitations. Analyses using these datasets rely on accurate and reliable information, but errors and inaccuracies are inevitable.

The potentially negative effects of spatial errors, in particular data misalignments, are commonly overlooked. Possible consequences may include over or underestimation of the reported extent of protection for species and ecosystems. This may lead to an inaccurate assessment of risk and could have serious implications for companies.

This briefing note outlines common causes of data misalignments in key global biodiversity datasets such as the World Database on Protected Areas (WDPA), the International Union for the Conservation of Nature (IUCN) Red List of Threatened Species, and other global scale datasets in the marine and terrestrial realm.

Why are data misalignments an issue?

Misalignments in biodiversity datasets can lead to uncertainty in results and may have implications for risk assessments or environmental/biodiversity management plans.

The following present a number of scenarios leading to misalignments in spatial biodiversity datasets:

- Data from multiple sources
- Use of different reference layers
- Datasets based on different coordinate systems
- Scale (resolution) of dataset
- Digitisation methodologies

The main reason these scenarios can lead to a spatial misalignment is because the person using the dataset, and the person who created the dataset, often do not have the same references. These spatial issues have been discussed in the scientific literature (e.g. Chape et al. 2005; Visconti et al. 2013). Although spatial misalignment of datasets is common it is important to note that this does not indicate the data are not fit for purpose. Examples of spatial misalignment in datasets available through the Proteus Partnership are described in Table 1.

Data from multiple sources

Combining data from multiple sources is a leading cause of spatial data misalignment. As an example, the misalignment between the Global Distribution of Coral Reefs (2010) and Seagrasses (2005) in Belize is notable, but only on certain coastlines (**Figure 1**). In some locations these datasets are aligned, but in others they are not. This is most likely due to the year in which these datasets were created, or the reference layer/coordinate system upon which each is based. When using data from multiple sources it is valuable to firstly check if one or all features in the dataset are misaligned.

Table 1. Examples of possible spatial misalignment in datasets available through the Proteus Partnership.

Example dataset	Data type	Spatial misalignment
Global Distribution of Sea Turtle Nesting sites (1999)	Lines & Polygons	<ul style="list-style-type: none"> • Data is from 1999, so the reference layer upon which this is based may be out of date.
Marine Ecoregions of the World (2007)	Polygons	<ul style="list-style-type: none"> • Spatial misalignment may arise in some locations due to changing seasonal and climatological conditions.
Global Distribution of Coral Reefs (2010)	Polygons	<ul style="list-style-type: none"> • Data is from multiple sources, so in some locations it could be misaligned due to utilisation of different scales and conflicting reference layers.
Global Distribution of Northern Fur Seals (2013)	Raster	<ul style="list-style-type: none"> • Scale and coarseness of raster datasets can often lead to spatial misalignment with other data types.
World Database of Protected Areas (2015)	Points & Polygons	<ul style="list-style-type: none"> • Data is from multiple sources, so in some locations it could be misaligned due to utilisation of different scales and conflicting reference layers. Point datasets can also be subject to large georeferencing errors leading to misalignment with other layers.



Figure 1. Spatial misalignment when using data from multiple sources. An example from Belize using data on the Global Distribution of Coral Reefs (2010) and Global Distribution of Seagrasses (2005).

As a further example, the Global Distribution of Saltmarsh (2013) is known to be compiled from multiple data sources. To ensure maximum data usage, all known sources are compiled to create the global distribution. However, this results in the use of data with different resolutions and therefore potential misalignments.

Although combining data from multiple sources is likely to effect the precision of habitat extent or protection at a local scale, particularly those along coastlines, it is unlikely to alter statistical results at a regional or global scale (Chape et al. 2005). Clipping to one dataset is not advised because it is not always obvious which dataset is correctly aligned and it may result in loss of information for one of the datasets.

Use of different reference layers

Spatial misalignments are notable when comparing datasets based on different reference layers. Reference layers outline administrative boundaries, coastlines and/or landscape features. There are many reference layers used globally, but biodiversity datasets available to Proteus Partners are usually based on one of three: the Digital Chart of the World (DCW), UN boundaries or the Global Self-consistent

Hierarchical High-resolution Geography database (GSHHG).

Mialignment between these is largely due to the time of production: DCW was produced in 1992, UN boundaries in 2012 and GSHHG in 2013. In **Figure 2**, which uses Northern Madagascar as an example, one of the islands is substantially misaligned between DCW and the UN boundaries. Slight differences are observed between GSHHG and UN boundaries as well.

The majority of marine datasets hosted on Ocean Data Viewer are based on the GSHHG reference layer and have been expertly assessed. Exceptions are the Global Distribution of Sea Turtle Nesting and Feeding Sites layers which were produced in 1999 and are based on the DCW reference layer from 1992. These datasets are no longer maintained by UNEP-WCMC and should therefore be used with caution. An example of misalignment in these datasets is shown in **Figure 3**.

Due to the wide range of data providers for the WDPA, protected area boundaries are likely based on a number of different reference layers. This may cause particular problems during analysis of protected areas in close proximity to islands, for example, because islands are usually isolated,

complex polygons, so the chance of error is substantially greater.

Species range maps provided by the IUCN Red List may also be based on different reference layers, which will normally depend on the IUCN unit carrying out the assessment, for example mammals, plants etc. This may therefore lead to misalignments with datasets based on different reference layers.

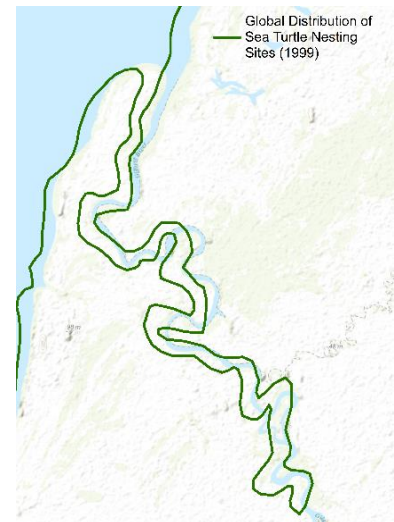


Figure 3. Misalignment of the Global Distribution of Sea Turtle Nesting Sites (1999) with the Australian shoreline.

Datasets based on different coordinate systems

Spatial data misalignments may also arise due to the use of different coordinate systems. Coordinate systems define the locations of geographic features. Geographic coordinate systems, such as latitude-longitude, are based on spherical surfaces, whereas projected coordinate systems, such as Universal Transverse Mercator, are based on two-dimensional surfaces.

The WDPA is supplied in a geographic co-ordinate system (World Geodetic System 1984). If misalignments occur between these features and another dataset, or reference layer, it may be due to differing geographic coordinate systems (Visconti et al. 2013; UNEP-WCMC, 2016). It is fairly common for

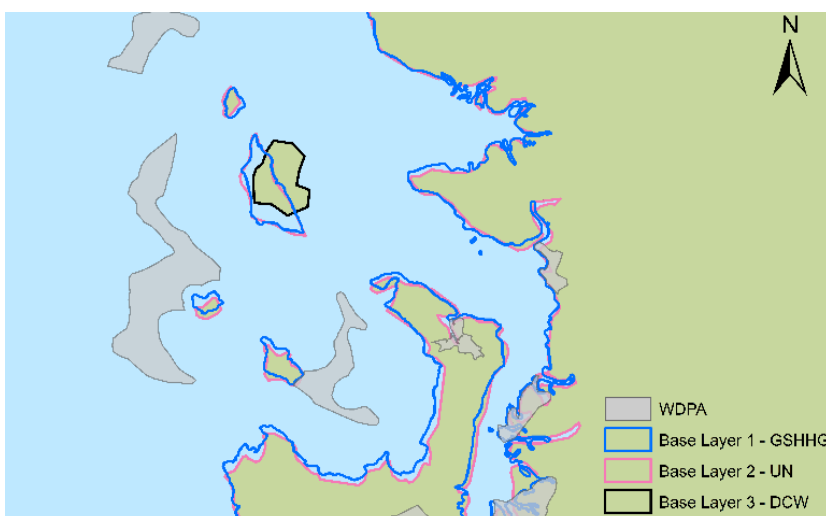


Figure 2. Spatial misalignment when using different reference layers. Data from the World Database on Protected Areas (WDPA) are shown in grey. GSHHG (blue) = Global Self-consistent Hierarchical, High-resolution Geography database; UN (pink) = United Nations boundaries; DCW (black) = Digital Chart of the World.

different countries to use different coordinate systems, therefore when combining national datasets in global analyses this should be considered; especially when using data from the USA which is likely to be based on the State Plane Coordinate System.

Scale (resolution) of datasets

The scale (or resolution) of datasets is likely to cause spatial misalignments especially near country boundaries.

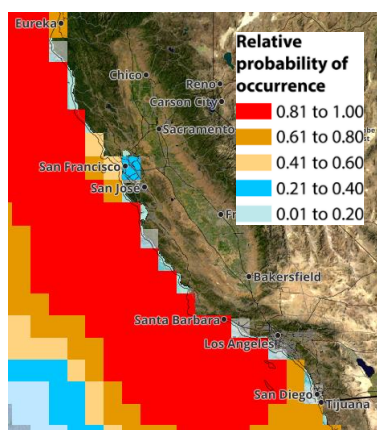


Figure 4. Misalignment of the Northern Fur Seal distribution with the North America shoreline.

Raster datasets (composed of grid cells of varying resolutions), such as those on

Ocean Data Viewer which represent global modelled distributions of marine mammals, are most likely to cause scale-related misalignments due to their gridded nature. As an example the Northern Fur Seal distribution is shown in **Figure 4**. With these distributions represented at 0.5 degree resolution (~50 x 50 km grid cells at the equator), alignment with the shoreline, or with other polygon datasets, is poor.

Digitisation methodologies

Different digitisation methodologies used by data providers can also lead to spatial misalignments. Manual digitisation using coordinates and spatial references can be time-consuming, but can accurately copy maps in poor condition. ‘Heads-up’ digitisation involves scanning of maps or images and although can be faster and more accurate, with less chance for human error, it is reliant on high-quality inputs. Digitising maps from paper without taking any spatial references can give highly inaccurate results.

There tend to be numerous data providers for biodiversity datasets. For example, the WDPA has over 500 sources from different countries and

territories around the world. Due to different capacities and resource availability to digitise boundaries there are likely to be misalignments with other data layers (UNEP-WCMC, 2016). There is also likely variability in the number of nodes used to create protected areas, the scale at which they have been digitised and the reference layer used. This information is partly captured in the WDPA source table which contains a ‘SCALE’ field, reporting the scale of the dataset when it was originally created i.e. digitised or surveyed.

Working with data misalignments

As with all biodiversity datasets, on-site verification, in the form of ground-truthing, is a vital part of the decision making process. This may not always be required for regional or global analyses, but at a minimum the misalignment errors discussed in this technical briefing note should be considered by users of biodiversity datasets, especially when complex decisions are to be made. It is recommended that data users report any obvious misalignments in UNEP-WCMC datasets so that assistance with problems in particular regions can be provided and in some cases the misalignment can be corrected.

References

Chape, S., Harrison, J., Spalding, M. & 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.

UNEP-WCMC (2016). World Database on Protected Areas User Manual 1.2. UNEP-WCMC: Cambridge, UK. Available at: http://wcmc.io/WDPA_Manual

Visconti, P., Di Marco, M., Álvarez-Romero, J.G., Januchowski-Hartley, S.R., Pressey, R.L., Weeks, R. & Rondinini, C. (2013). Effects of errors and gaps in spatial data sets on assessment of conservation progress. *Conservation Biology* (27): 1000-1010.

Citation: UNEP-WCMC (2016). Out of line, out of mind? Working with data misalignments in global biodiversity spatial datasets. UNEP-WCMC. Cambridge, UK.

Authors: Katie leach, Matt Jones.

Available online at: www.proteuspartners.org/

Contact: businessandbiodiversity@unep-wcmc.org

Copyright: 2016 United Nations Environment Programme



The United Nations Environment Programme World Conservation Monitoring Centre (UNEP-WCMC) is the specialist biodiversity assessment centre of the United Nations Environment Programme (UNEP), the world’s foremost intergovernmental environmental organization. The Centre has been in operation for over 30 years, combining scientific research with practical policy advice.

Copyright release: This publication may be reproduced for educational or non-profit purposes without special permission, provided acknowledgement to the source is made. Reuse of any figures is subject to permission from the original rights holders. No use of this publication may be made for resale or any other commercial purpose without permission in writing from UNEP. Applications for permission, with a statement of purpose and extent of reproduction, should be sent to the Director, UNEP-WCMC, 219 Huntingdon Road, Cambridge, CB3 0DL, UK.

Disclaimer: The contents of this report do not necessarily reflect the views or policies of UNEP, contributory organizations or editors. The designations employed and the presentations of material in this report do not imply the expression of any opinion whatsoever on the part of UNEP or contributory organizations, editors or publishers concerning the legal status of any country, territory, city area or its authorities, or concerning the delimitation of its frontiers or boundaries or the designation of its name, frontiers or boundaries. The mention of a commercial entity or product in this publication does not imply endorsement by UNEP.