# eHabitat: Large scale modelling of habitat types and similarities for conservation and management of protected areas.

Javier Martínez-López, Jon Skøien, Gregoire Dubois

European Commission - Joint Research Centre, Institute for Environment and Sustainability, Via Fermi 2749, Ispra, 21027 (VA), Italy.

Email: {javier.martinez-lopez; jon.skoien; gregoire.dubois}@jrc.ec.europa.eu

## 1. Introduction

Protected areas need to be assessed systematically according to biodiversity values and threats so as to support decision making and fund allocation processes. Characterizing protected areas according to their species, ecosystems and threats is therefore required. While species based conservation approaches are the most commonly used, assessing natural habitats is also important. Among other ecosystem services, natural habitats offer refuge for species and can be mapped at a global scale by means of remote sensing in a harmonized way, not being biased by sampling efforts related to study location or taxa.

## 2. Methods

eHabitat, which is one of the services supporting the DOPA, the Digital Observatory for Protected Areas[1] (Dubois et al. 2013a,b), proposes a habitat replaceability index (HRI) which can be used for characterizing each protected area worldwide. More precisely, eHabitat computes for each protected area a map of probabilities to find areas within the corresponding ecoregion (Olson et al. 2001) presenting ecological characteristics that are similar to those found in the selected protected area. The HRI is then computed as the ratio between similar areas outside park and the park area itself.

Several environmental variables are used for identifying similar habitats through multivariate analysis using the Mahalanobis distance (Mahalanobis 1936): percentage of tree cover, percentage of grassland cover, elevation, slope, aridity, biotemperature, precipitation, Normalized Difference Vegetation index (NDVI) and Normalized Difference Water index (NDWI), some of the representing long term annual averages and all of them mapped at 1 km$^2$ globally (for more details see Dubois et al. 2013b).

There are two recent implementations of eHabitat, one programmed in R and a second one in Python. The first one uses some of the main spatial libraries in R, such as the 'sp', 'rgdal' and 'raster' libraries, has parallel computing capabilities and is integrated as part of some Web

---

[1] http://dopa.jrc.ec.europa.eu/

Processing Services (WPS)[2] (Skøien *et al*, 2013, Dubois *et al*., 2013b). The underlying codes are available online as an R library from GitHub[3]. The Python version, not yet integrated within the WPS, was developed for improving stability and increasing computational speed. It uses several numerical and scientific python libraries, such as NumPy, SciPy, Multiprocessing, Scikit-learn and the source code is also available online[4].

## 3. Improvements to eHabitat

One of the main limitations of the current eHabitat version is that protected areas with heterogeneous landscapes would lead to an overestimation of the probabilities to find similar areas elsewhere because the statistical approach considers an "average habitat" over the whole surface of the analysed protected area. The variables characterizing the "average habitat" may be represented by a range of values that is too broad, leading consequently to a high variance in the final results. This problem can be illustrated by computing the HRI over the Odzala-Kokoua National Park, a large (13,600 km$^2$) protected area located in the North of the Congo Republic.
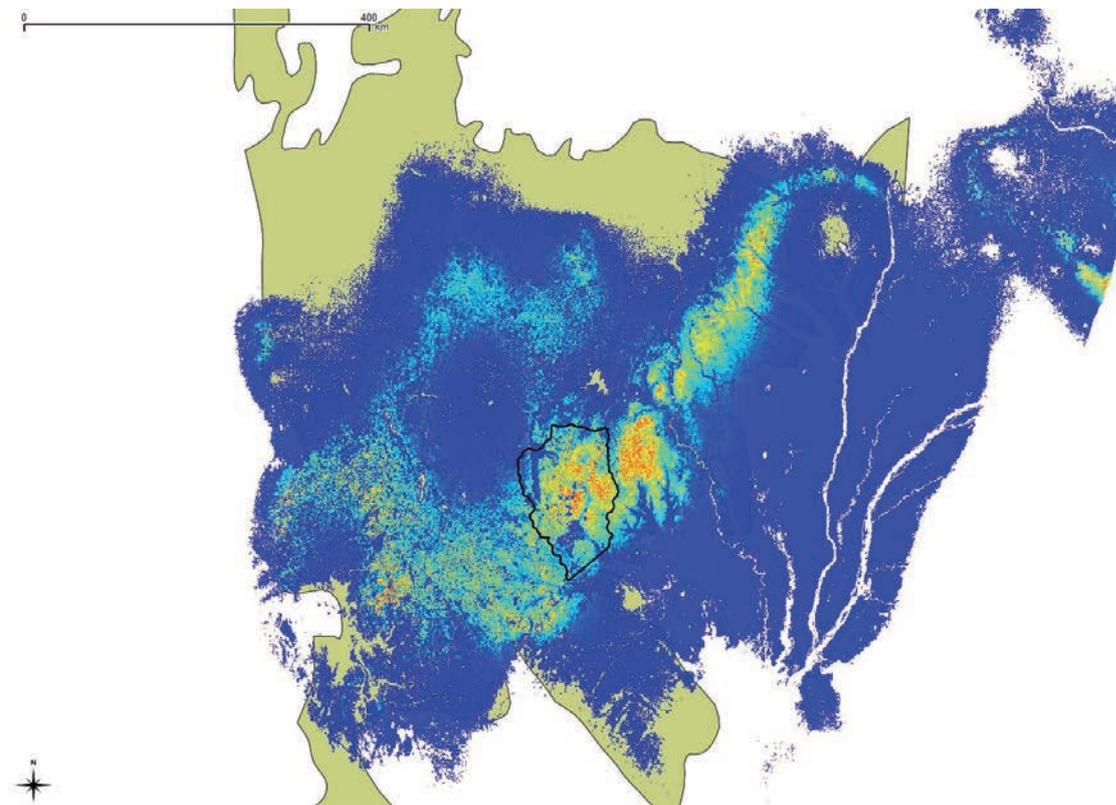


Figure 1: Map of habitat similarities to a protected area (delimited by the black central polygon boundary) within the Northwestern Congolian lowland forests ecoregion where it is present

[2] http://ehabitat-wps.jrc.ec.europa.eu/ehabitat/
[3] https://github.com/javimarlop/eHabitat
[4] https://github.com/javimarlop/eHabpy

(delimited by the green polygon). Similarity values range from 0 (blue; high dissimilarity) to 1 (red; total similarity).

The park, with a surface of 13,600 km$^2$, shows, according to Figure 1, that the northwest and the south of the protected area are dissimilar to the centre of the area. Continuity of the estimated "average habitat" outside of the park is shown along an axis of 30° across the ecoregion. A detailed analysis using land cover maps and expert feedback will confirm that the area is densely forested in the northwest while the south of the region presents a forest-savannah mosaic. The largest part of the park is covered by open forest in the south and the east. By implementing a preliminary segmentation step to eHabitat 2.0, we allow these main ecological features of the protected area to be identified automatically prior to a further individual processing for generating the similarity maps (Figure 2).
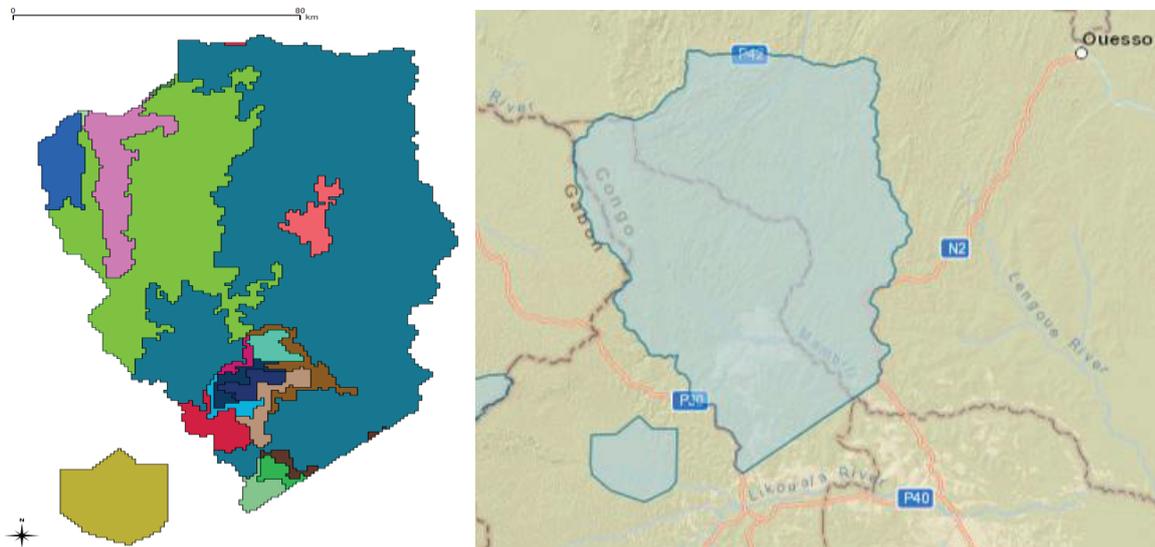


Figure 2: Automatic habitat segmentation (left) of the Odzala-Kokoua National Park (right) based on several environmental indicators using GRASS GIS 7 (using *i.segment* module).

The automatic segmentation of parks prior to HRI computation allows for a discrimination of different habitats types inside of protected areas. By reducing the variability within landscape patches, similarity values can be considered to be more accurate. This approach should also further improve the associated niche modelling tools as proposed by Skøien et al. (2013).

## 4. Further developments

The above method will still require case studies to be validated by means of expert knowledge (parks managers, research community) and comparisons with different regional, national and global land cover maps. Still, the underlying tools allow for the automation of large scale analyses using continent wide consistent datasets, allowing results to be easily compared.

In addition to the new automatic segmentation step implemented using GRASS GIS 7, various landscape metrics such as patch area, fragmentation and shape indices, are considered for further improving the characterization of each protected area.

Technically, our developments based on self-written codes used in combination with Free and Open Source Software tools, should benefit in the future from links with the rasterEngine R library, Scidb or Hadoop which could significantly reduce the processing time of large datasets and allow us to implement web processing services capable to provide such functions for larger datasets to a larger group of simultaneous end-users.

## References

Dubois G, Schulz M, Skøien J, Bastin L and Peedell S, 2013a, eHabitat, a Multi-Purpose Web Processing Service for Ecological Modeling. *Environmental Modelling & Software* 41, 123 – 133.

Dubois G, Schulz M, Skøien J, Cottam A, Temperley W, Clerici M, Drakou E, et al., 2013b, *An introduction to the Digital Observatory for Protected Areas (DOPA) and the DOPA Explorer (Beta). EUR 26207 EN, EC. Luxembourg: Publications Office of the European Union 72 p.*

Mahalanobis PC, 1936. On the generalized distance in statistics. *Proc. Nat. Inst. Sci. India*, 12, 49.

Olson DM, Dinerstein E, Wikramanayake ED, Burgess ND, Powell GVN, Underwood EC and Kassem KR, 2001. *Terrestrial Ecoregions of the World: A New Map of Life on Earth: A new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. BioScience*, 51(11), 933 – 938.

Skøien J, Schulz M, Dubois G, Fisher, I, Balman, M, May I and Tuama, É Ó, 2013. A Model Web approach to modelling climate change in biomes of Important Bird Areas. *Ecological Informatics* 14, 38 – 43.