

# Extracting oceanographic data via R:

## An application to habitat modelling of marine species

**Maite Louzao<sup>1,2,\*</sup>, Anne Goarant<sup>3</sup>, Clara Peron<sup>1</sup>, Jean-Baptiste Thiebot<sup>1</sup>, David Pinaud<sup>1</sup>, Philippe Koubbi<sup>3</sup>, Karine Delord<sup>1</sup>, Christophe Barbraud<sup>1</sup>, Henri Weimerskirch<sup>1</sup>, Charles-André Bost<sup>1</sup>, Guy Duhamel<sup>4</sup>, Patrice Pruvost<sup>4</sup>**

1. Centre d'Etudes Biologiques de Chizé, CNRS UPR 1934, 79369 Villiers en Bois, France
2. UFZ Leipzig-Halle, Permoserstrasse 15, 04318 Leipzig, Germany
3. LOV, CNRS, Station Zoologique BP28, 06230 Villefranche sur Mer, France
4. MNHN, DMPA -UMR 5178, 43 rue Cuvier, 75005 Paris, France

\* Contact author: [louzao@cebc.cnrs.fr](mailto:louzao@cebc.cnrs.fr)

**Keywords:** oceanographic data, Xtractomatic, habitat modelling, marine top predators

Thanks to the rapid development of remote sensing technologies, the availability of oceanographic data has dramatically increased during the last years. Oceanographic data available via website are very diverse in terms of sources (NOAA, NASA), file type (netcdf, hdf) and resolution (temporal and spatial). Large extraction of data for different time periods and areas can therefore be time-consuming and difficult to handle within the same format. The Environmental Research Division, Southwest Fisheries Science Center and US National Marine Fisheries Service has recently developed a software (Xtractomatic, <http://coastwatch.pfel.noaa.gov/xtracto/>) that simply make available environmental data (SST, chlorophyll, wind) within the R environment. This R-based tool allows the extraction of oceanographic data along (1) a series of input of time, longitude and latitude (e.g. a track of an animal or ship) specifying an extraction box and (2) a 3-Dimensional cube specified by limits of longitude, latitude and time.

Here, we present how Xtractomatic can be applied to the extraction of oceanographic data which are used for habitat modelling of marine species in the southern Indian Ocean, including top predators. We relied on two types of data: (1) tracking data for seabirds and (2) occurrence patterns of pelagic fishes. Both type of data were placed over a standard grid and the Xtractomatic function was used to extract oceanographic parameters, after adjusting time resolution and spatial scale to the species biology and locations accuracy. Once oceanographic variables were obtained, different regression techniques (Generalized Linear Mixed Models and Generalized Additive Models) were applied within the R environment in order to identify those variables which best explained the oceanographic habitat of the species and predict density or habitat use probability.

Given the high conservation concern of marine top predators and current major environmental changes, the standardization of the whole habitat modelling process (including the download of large amount of environmental data) makes much easier and faster the investigation of the oceanographic processes influencing marine species distribution patterns.