



Project no. 018340

**Project acronym: EDIT**

**Project title: Toward the European Distributed Institute of Taxonomy**

Instrument: Network of Excellence

Thematic Priority: Sub-Priority 1.1.6.3: "Global Change and Ecosystems"

## **C5.040 Application to select vectorial units by environmental and spatial criteria**

Due date of Component: Month 45

Actual submission date: Month 46

Start date of project: 01/03/2006

Duration: 5 years

Organization name of lead contractor for this component: CSIC

<b>Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)</b>		
<b>Dissemination Level</b>		
<b>PU</b>	Public	X
<b>PP</b>	Restricted to other programme participants (including the Commission Services)	
<b>RE</b>	Restricted to a group specified by the consortium (including the Commission Services)	
<b>CO</b>	Confidential, only for members of the consortium (including the Commission Services)	

This report provides a summary of ideas and procedures to check if occurrence data (records) follow a distribution pattern caused by an environmental variable. It has been compiled by Marcin Gašior (MIZPAN) with a generous input from Pere Roca Ristol (CSIC), Patricia Mergen (RMCA) and the colleagues from the EDIT developer mailing list.

### ***Introduction***

Due to the lack of quality and incompleteness of significant part of biological information research on biodiversity can be biased. Additional surveys directed to fill in all gaps in knowledge and recover a reliable picture of the distribution for highly diverse taxa are impossible to conduct because of limited time and resources. That is why alternatives to assess the reliability and/or enhance the information currently stored in biodiversity databases are preferred (*deliverables 5.35* & 5.38)

One of these alternatives can be the EDIT mapViewer tool which should allow to “evaluate whether the areas in which reliable (complete) inventories have been detected cover the range of environmental conditions of all the analyzed area (polygonal layer selected by user)”.

This would be a method of allocation with selected environmental data as friction parameter.

Allocation models try to minimize costs and maximize efficiency (in that case collecting data different enough to be considered a reliable inventory) assigning to some variables a specific importance (user selected environmental data).

### ***Description of mapViewer functionalities***

The analyzing tools module can be enabled after user data have been uploaded. This data must be in simple .csv (comma separated values) format and have at least latitude and longitude in WGS84 datum and a field (genus, species).

Occurrence data, uploaded and displayed on the map as points, are then overlaid with environmental raster layer (Fig. 1) and the raster value is extracted for each point. Environmental data which are available on EDIT mapViewer cover:

- climate (temperature, precipitation, seasonality, etc.)*
- topography (relief for ocean and land areas, elevation)*
- soil and vegetation*
- land cover (in categories)*
- human population*

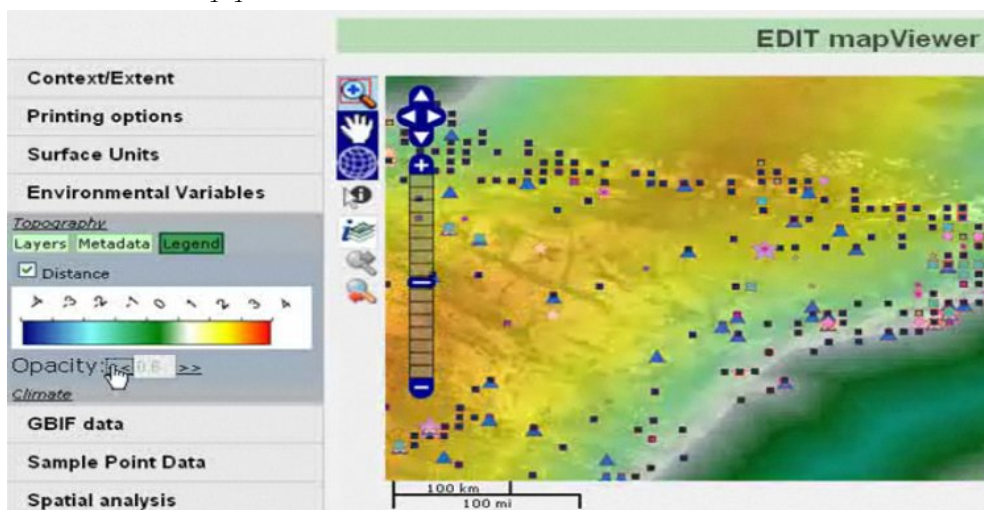


Figure 1. Occurrence points overlaid with environmental variable.

A more detailed description of the GIS layers of surface units and environmental variables can be found at [http://edit.csic.es/web/docs/EDIT\\_GIS\\_layers.rtf](http://edit.csic.es/web/docs/EDIT_GIS_layers.rtf) and <http://edit.csic.es/GISdownloads.html>

When user's points have assigned raster values, than it is possible to check if the data follows an environmental pattern. For example if the major part of occurrence points representing the same genus/species fall in raster cells with high values of precipitation and temperature than it can suggest existence of distribution pattern.

The data assigned to each point is used to calculate media for the specific polygonal feature (UTM squares, provinces, grid cells... selected by user when performing previous inventory completeness analysis) it belongs to an environmental value. This value will be used as friction parameter on allocation analysis.

In the case we find a "good collected area" (with reliable inventories) that clearly follows a distribution pattern caused by our selected environmental pattern, the tool should give to this environmental value a high importance on the method.

It will be useful to clearly evaluate the next polygonal areas to center our efforts (taking into account the same environmental layer).

The evaluation will be different if we find a clear environmental pattern in areas with non-reliable inventories. Then, the importance given to areas with similar environmental values will be lower. The same applies when there is not a clear environmental pattern on the selected area.

Because of lack of stable tools to store raster data on PostGIS database, automatic calculation can not be performed for the moment. That's why selection of data to be analyzed can only be made manually from the map displayed by the graphical interface and can only be done in "extreme" cases: highly reliable inventories following its area an environmental pattern.

### **Conclusions and suggestions**

Selecting vector units by environmental and spatial data can reveal a distribution pattern caused by environmental variable. User can check if occurrence points follow this pattern and then assess the reliability of this data. Finding the areas with reliable inventories facilitate a selection of the next regions to research taking into consideration the same environmental value.

Due to lack of tools to handle raster data in PostGIS database, it was proposed to transform raster data into vectors and then derive pixel values from vectorial objects instead of raster cells. Such approach would require to create a huge vector datasets as one cell is represented as one polygon. That is why a generalization of the environmental variable datasets would be necessary to provide a sufficient performance.