OPENMODELLER: A NEW COMPUTING ENVIRONMENT FOR MODELING SPECIES DISTRIBUTION

Vanderlei Perez CANHOS

Reference Center on Environmental Information (CRIA)

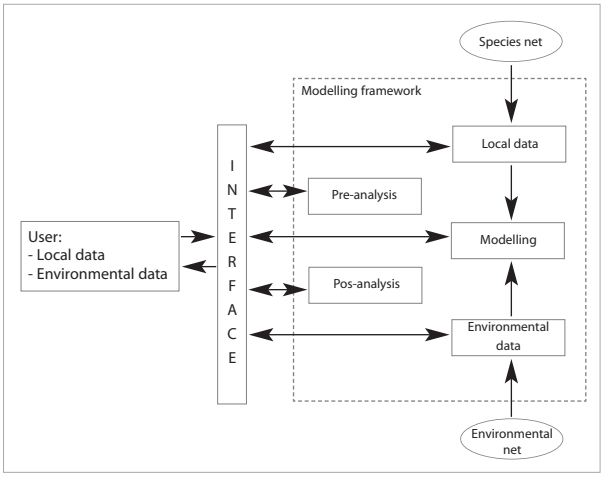

Figure 1. Original architecture of the openModeller *framework*

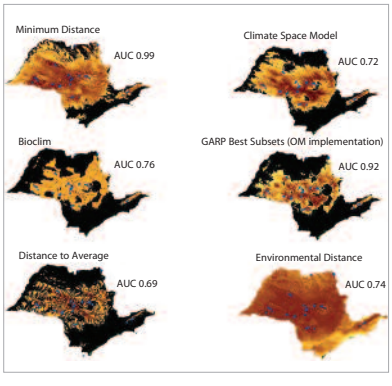

Figure 2. Results of six modeling experiments using the openModeller framework with the same points for Ouratea spectabilis, *the same raster layers with the respective AUCs, and comparing six different algorithms*

Conciliating social and economical development with environmental conservation is one of the biggest challenges today. There is a growing demand for quick answers to solve problems related to the occurrence and distribution of biological species, including threatened species, invasive species, and disease vectors. The formulation of strategies for sustainable development depends on these answers, which can only be obtained by accessing various types of data from different sources and by using advanced tools to analyze, visualize results, and build different scenarios.

This project aims to develop a new computing environment to model the potential distribution of species. It basically consists of a multiplatform framework specialized in performing modeling tasks, compatible with a service-oriented architecture. High-performance computing techniques are used, and the whole functionality is available to researchers through different interfaces. Study cases are being carried out to test and improve different modules of the framework and to compare the different algorithms that are being added.

Since the modeling task depends on different types of data, the scope of this project also includes activities related to species occurrence data (presence and absence) and environmental data stored in different formats and located in different places (local or remote).

The institutions involved in the project are: Reference Center of Environmental Information (CRIA), Polytechnic School of the University of São Paulo (USP) and National Institute for Space Research (INPE).

SUMMARY OF RESULTS TO DATE AND PERSPECTIVES

Species occurrence data. The speciesLink network is being maintained, improved and expanded as part of this project. Several tools have been created to help collections to improve the quality of their data. The modeling framework allows, not only direct access to speciesLink online database, but also to the Global Biodiversity Information Facility (GBIF) data.

Environmental data. Approximately 4,000 environmental layers have been cataloged. All layers are available in modeling servers and will soon be accessible through a new web service based on TerraLib, implementing a special protocol adapted to handle large volumes of raster data (Web Coverage Service – WCS).

Computing infrastructure. Modeling servers are accessible through a new protocol (OM-WS) that was developed through this project. One of the modeling servers available is a computing Cluster with 11 nodes (Core 2 Quad), which allows users to efficiently carry out complex experiments involving several species, several algorithms and a large number of layers.

Modeling framework. During the last 3 years, 12 new versions of the modeling framework have been released as part of this project. The framework has 8 modeling algorithms. It also offers pre-analysis tools to help the selection of the most relevant environmental layers, the post analysis tools, for models validation, and a command-line interface that can be used to perform the main functionalities of the framework. *Figure 1* shows the original architecture proposed.

Interfaces. The access to modeling functionalities is also available through advanced graphical user interfaces. A desktop interface – "OpenModeller Desktop" – is available for all 3 platforms (GNU/Linux, Mac OS X, and Windows) and allows the use of remote modeling servers. Recent versions of TerraView, a geographic information system developed by National Institute for Space Research (INPE), offers a new modeling plug in that was also developed through this project. The GBIF data portal has also developed a demonstration interface making use of one of the CRIA modeling servers. Lastly, a new web interface is currently under development and will be released before the end of the project in March, 2009.

Study cases. Study cases were carried out for São Paulo State, the Brazilian Savannah (Cerrado), and the Amazon region. New study cases are being carried out for other regions of interest to test technical aspects of the framework, such as the comparison of different algorithms and to understand the modeling process. Figure 2 shows some modeling results for the potential distribution of a tree species that occurs in São Paulo State.

Vanderlei Perez Canhos

Centro de Referência em Informação Ambiental (CRIA)
Av. Dr. Romeu Tórtima, 388 – Barão Geraldo
CEP 13084-791 – Campinas, SP – Brasil

+55-19-3288-0466
vcanhos@cria.org.br