

**MAKING SPECIES DISTRIBUTION MODELS AVAILABLE ON THE WEB FOR REUSE
IN BIODIVERSITY EXPERIMENTS: *EUTERPE EDULIS* SPECIES CASE STUDY**

**Modelos de distribuição de espécies disponíveis na Web para reutilização em experimentos de
biodiversidade: Estudo de caso com a espécie *Euterpe edulis***

Karla Donato Fook

Instituto Nacional de Pesquisas Espaciais - INPE / CEFET-MA
São Luis/MA – Brasil
karla@dpi.inpe.br

Silvana Amaral

Instituto Nacional de Pesquisas Espaciais – INPE
São José dos Campos/SP – Brasil
silvana@dpi.inpe.br

Antônio Miguel Vieira Monteiro

Instituto Nacional de Pesquisas Espaciais – INPE
São José dos Campos/SP – Brasil
miguel@dpi.inpe.br

Gilberto Câmara

Instituto Nacional de Pesquisas Espaciais – INPE
São José dos Campos/SP – Brasil
gilberto.camara@inpe.br

Arimatéa de Carvalho Ximenes

Instituto Nacional de Pesquisas Espaciais – INPE
São José dos Campos/SP – Brasil
arimatea@dpi.inpe.br

Luciana Satiko Arasato

Instituto Nacional de Pesquisas Espaciais – INPE
São José dos Campos/SP – Brasil
luarasato@dpi.inpe.br

Artigo recebido para publicação em 30/10/2008 e aceito para publicação em 16/02/2009

ABSTRACT: *Currently, biodiversity conservation is one of the most urgent and important themes. Biodiversity researchers use species distribution models to make inferences about species occurrences and locations. These models are fundamental for fauna and flora preservation, as well as for decision making processes for urban and regional planning and development. Species distribution modelling tools use large biodiversity datasets which are globally distributed, can be in different computational platforms, and*

are hard to access and manipulate. The scientific community needs infrastructures in which biodiversity researchers can collaborate and share knowledge. In this context, we present a computational environment that supports the collaboration in species distribution modelling network on the Web. This environment is based on a modelling experiment catalogue and on a set of geoweb services, the Web Biodiversity Collaborative Modelling Services – WBCMS.

Keywords: Collaboration, Species modelling, Species distribution modelling, E-science.

RESUMO: *A conservação da biodiversidade é uma das questões mais urgentes e importantes da atualidade. Pesquisadores da biodiversidade usam modelos de distribuição de espécies para fazer inferências sobre ocorrência e localização de espécies. Estes modelos são fundamentais para a preservação da fauna e flora mundiais, bem como nos processos de tomada de decisão durante o planejamento e desenvolvimento urbano e regional. Ferramentas para modelagem de distribuição de espécies usam grandes conjuntos de dados de biodiversidade que estão globalmente distribuídos, podendo estar em diferentes plataformas computacionais, o que dificulta seu acesso e manipulação. A comunidade científica precisa de infra-estruturas onde pesquisadores de biodiversidade possam colaborar e dividir conhecimento. Neste contexto apresentamos um ambiente computacional que suporta a colaboração em uma rede de modelagem de distribuição de espécies na Web. Este ambiente baseia-se em um catálogo de experimentos de modelagem e em um conjunto de serviços web geoespaciais, o Web Biodiversity Collaborative Modelling Services – WBCMS.*

Palavras-chave: Modelagem de distribuição de espécies, Biodiversidade, Serviços web, Colaboração.

1. Introduction

Biodiversity information is essential for decision making processes. Scientists working with biodiversity information use a variety of data sources, statistical analysis, modelling tools, and presentation and visualization software. Among these tools, we highlight species distribution models that allow researchers to make inferences about the diversity, abundance and spatial distribution of species over different geographical areas. The study about species distributed on Earth in space and time has a long history which has inspired many biogeographers to seek explanations (GUISAN and THUILLER, 2005).

The developed models to predict the distribution of plants and animals relate species occurrence and environmental factors that limit their distribution quantitatively. These factors are called environmental variables. This relationship is based

on the concept of ecological niche and it can be visualized as a multidimensional space (HUTCHINSON, 1957). Each dimension represents the interval of a certain environmental condition that indicates the species distribution in the geographical space. Biodiversity researchers should identify environmental factors to determine the threatened species distribution in order to plan mitigation policies of the population decline or to locate areas where the new individuals can be reintroduced (RUSHTON *et al.*, 2004). Species distribution models are also used to predict effects on climate change, to handle invasive species problems, and to predict the best places to set up new protected areas.

Species distribution modelling tools access large sets of geospatial data such as environmental layers or variables that may be archived by different institutions. It creates computational challenges of data collection integration, management and storage

systems, knowledge extraction, and access to distributed geospatial data. In addition, “*species distribution model results should be easily accessible to decision makers*” (BEST *et al.*, 2007). These features involve computational resources to enable scientists to share experiments in a distributed environment. This scenario calls for infrastructures to support local and global research and to disseminate information. Collaborative environments on the Web present resources to supply these aspects. These environments have become an important dimension of the scientific method. They complement theory, experimentation, and simulation in various applications.

This paper reports a collaborative environment to support modelling experiment sharing, and its reuse on the Web. This environment is based on a species distribution modelling experiments catalogue, and on a set of geospatial web services, the Web Biodiversity Collaborative Modelling Services – WBCMS. For an early discussion of WBCMS, see (“autor” *et al.*, “ano”). The WBCMS architecture is part of an international project for building biodiversity models, the OpenModeller Project (<http://openmodeller.cria.org.br/>) (MUÑOZ, 2004; GIOVANNI, 2005; OPENMODELLER, 2005).

The present work is organized as follows. Section 2 presents the theoretical foundation for the collaborative environment. Section 3 describes an example of the model instances catalogue usage. Section 4 presents final comments.

2. Background

2.1. Species distribution models

This section briefly describes a species distribution model, highlighting those points that facilitate the understanding of the rest of the paper. Generally, researchers make field work to get ecological information and localization related to species under study. Other sources are museums and herbariums. However, sometimes it is highly costly

to accomplish systematic studies to know the true species distribution. In addition, institutions lack biodiversity data. Researchers build a predictive statistical model to approximate to potential species distribution. This model, named as species distribution model, results from relationship analysis between georeferenced species occurrence data and environmental variables related to species distribution such as vegetation covering, temperature, and topography. The process continues by projecting the model onto a map of the study region (GRILO, 2006).

There are several algorithms used to produce species distribution models, such as Genetic Algorithm for Rule-set Production – GARP (STOCKWELL and PETERS, 1999), Bioclimatic Envelope – BIOCLIM (BUSBY, 1991), and Maximum Entropy Method (PHILLIPS *et al.*, 2006), among others. Each algorithm has its own features and parameters, which are outside the scope of this paper. However, modelling processes have input data, algorithms and output data in common. For more details about species distribution models, see (GUISAN and ZIMMERMANN, 2000).

2.2. OpenModeller desktop

The OpenModeller Desktop is a modelling application that offers a user-friendly front end to the OpenModeller library. It provides an environment where aspects of data preparation and local model running can be carried out. Algorithms for predictive species distribution modelling such as Genetic Algorithm for Rule-set Production – GARP (STOCKWELL and PETERS, 1999), and Maximum Entropy Method (PHILLIPS *et al.*, 2006) are available in OpenModeller Desktop. For more details, see (SUTTON *et al.*, 2007).

The OpenModeller Desktop is part of OpenModeller Project, a thematic project for collaborative building of biodiversity models. There are several development areas within this Project besides OpenModeller Desktop: the OpenModeller library and the OpenModeller Web Service (OMWS). The library provides a uniform method for modelling

distribution patterns using various modelling algorithms. The OMWS is a web version that is available for remote execution of openModeller jobs (GIOVANNI, 2005; SUTTON *et al.*, 2007).

2.3. Related work

Trends point to collaborative environments on Web to support biodiversity research. Serique *et al.* (2007) have proposed Mo Porã tool (www.lba.inpa.gov.br/mopora), a web collaborative environment to share files and data in research groups in LBA Program (Large-Scale Biosphere-Atmosphere Experiment in Amazonia – www.lba.inpa.gov.br/lba). The WeBIOS Project (Web Service Multimodal Tools for Biodiversity Research, Assessment and Monitoring) provides scientists with a system that supports exploratory multimodal queries over heterogeneous biodiversity data sources (WEBIOS, 2005).

BioGeomancer Project (www.biogeomancer.org) is a collaborative project that aims to improve the quality and quantity of biodiversity data. This project develops products such as workbench, web services, and desktop applications that provide georeferencing for collectors, curators and users of natural history specimens (BIOGEOMANCER, 2005). Beija-flor Project (www.lba.cptec.inpe.br/beija-flor) involves an internet-based approach for sharing scientific data. It provides a data search by harvesting and organizing metadata (HORTA and GENTRY, 2008). The Aondê Web service supports ontology sharing and management on the Web in biodiversity domain (DALTIO and MEDEIROS, 2008). O'Connor *et al.* (2005) propose Spatial Information Exploration and Visualization Environment – SIEVE, an online collaborative environment for visualizing environmental model outputs in 2D and 3D.

The presented approaches aim to integrate and share biodiversity and geographical data and tools. However, they do not aim to share species modelling results. Our proposal holds a geoweb services based architecture that aims to support sharing descriptive

information about spatial data, and relevant information objects. In addition, it also enables researchers to reuse catalogued data. Our goal is also to extract implicit knowledge inserted in the modelling process and to make it available in an online catalogue.

3. Collaborative environment for sharing and reusing of species distribution modelling results on the Web

This section presents a collaborative environment to support biodiversity research. This environment is based on a modelling experiment catalogue. One species modelling experiment is represented by a model instance. To a better understanding of this section, we briefly describe the model instance idea. It aims to describe a species distribution modelling experiment as a whole and to capture information inserted into an experiment. The model instance includes information related to

- (a) *Model*: name, description, author, and modelled species (data and metadata);
- (b) *Model generation*: algorithms and their parameters, and input data, such as occurrence points (latitude and longitude) and environmental layers;
- (c) *Results*: reports, evaluation indexes, and georeferenced maps.

Besides the information above, the biodiversity researcher complements the model instance with extra data such as personal comments, and confidence degree. These data allow other researchers to assess the species distribution modelling experiments.

Now, let's consider that researchers from different institutions wish to share modelling experiments, to access experiments performed elsewhere, and to compare them. They can use the collaborative environment to publish their modelling experiments, to access experiments, and to run new models reusing published ones. This environment allows researchers to compare models and to make new discoveries. There is a model instance catalogue

available on the Web. Researchers can access this catalogue through a set of geospatial web services,

the Web Biodiversity Collaborative Modelling Services – WBCMS (see Figure 1).

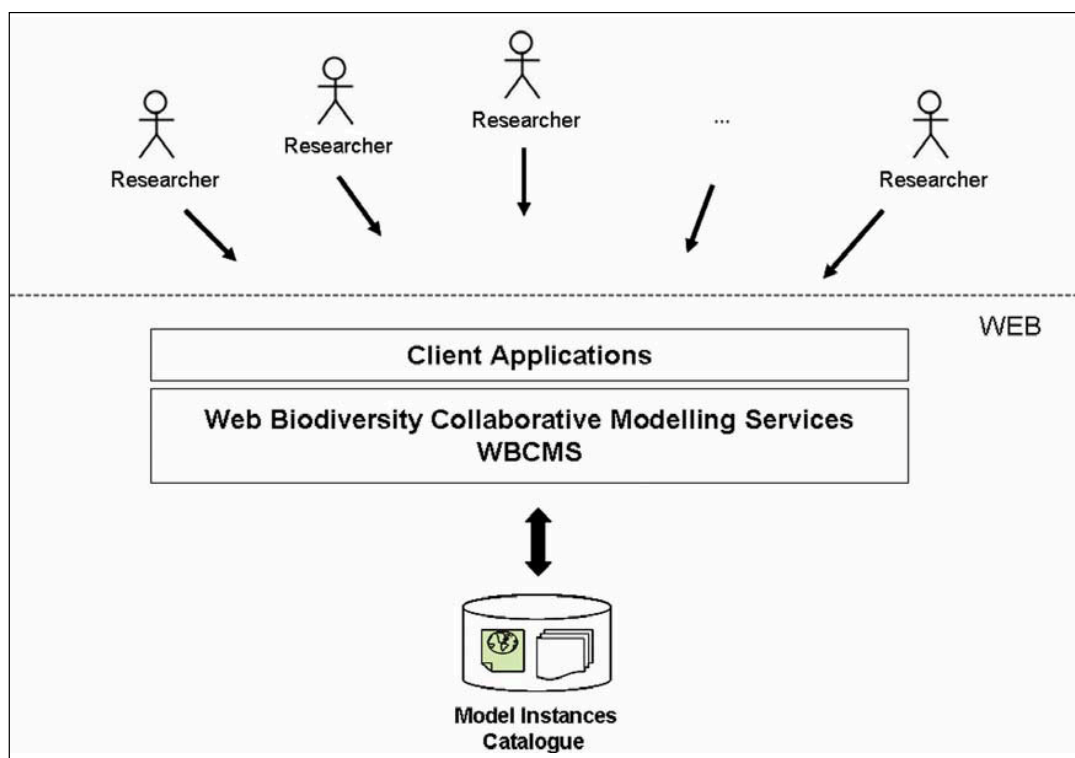


Figure 1. Model instance catalogue

The Figure 1 diagram shows that WBCMS enable researchers to share model instance, and to visualize it from catalogue. There are two client applications in WBCMS architecture that allow the researcher to access the catalogue. They are *Model Instance Catalogue* client application and *Model Instance Access* client application.

The researcher uses the *Model Instance Catalogue* application to publish his experiments, and the *Model Instance Access* application to visualize model instances available on catalogue. The next subsection describes the Collaborative Environment usage from a simple case study.

3.1. *Euterpe edulis* Mart. Model Instance – a simple case study

Briefly, the researcher can use the WBCMS

architecture to: (a) publish his model instance; (b) access model instance catalogue, and (c) produce new species distribution models. In this example, the researcher creates the *Euterpe edulis* Mart. species distribution model using the OpenModeller Desktop. The researcher uses the *Model Instance Catalogue* application to publish his modelling experiment into model instance catalogue.

- **Publishing the model instance**

The *Model Instance Catalogue* application captures model generation process information from result files, allows the researcher to inform personal comments about the experiment, and sends model instance data to catalogue. Figure 2 shows the *Model Instance Catalogue* application form.

The screenshot shows a web browser window with the URL `http://localhost/vs-php/cd_catalog/cd_catalog.php`. The page title is "Model Instance". On the left, there is a vertical sidebar with four small images: a flower, a bird, a landscape, and another bird. The main content area is divided into three columns. The first column, "Experiment Information", contains a "Title" field with the value "Model Instance Euterpe e", a "Description" field with the text "The algorithm Environmental Distance was used to develop a model of potential distribution for Euterpe edulis (Arecaceae family) in the Brazilian territory.", a "Question" field with the text "What is the potential distribution to Euterpe edulis species in Brazil?", and an "Author" dropdown menu showing "luarasato - Luciana Arasato" with a "New" button. The second column, "Additional Information", contains a "Researcher Confidence Degree" section with radio buttons for "High", "Medium", and "Low", and a "Confidence degree justification" text area containing the text "Modelling performed without previous selection of environmental variables.". The third column, "Authorization", contains a text box with the text "The data liberation will allow the experiment reproduction by other researchers." and a checkbox labeled "Do you make available the data (occurrence points) for storage into repository?" with "Yes" selected and "No" unselected.

Figure 2. Model Instance Catalogue application form

The researcher uses this form to publish the modelling experiment. He indicates general data related to modelling such as confidence degree, comments, and occurrence data publish authorization. This authorization makes the model instance available to reuse. Model generation data are extracted from OpenModeller result files. Therefore, result file paths are also informed by scientist.

- **Visualizing the model instance**

Remember that the researcher can visualize catalogued model instances using the *Model Instance Access* application. All model instance elements are

available in this application. WBCMS has a number of predefined queries that enables the researcher to get answers for the following questions: “*What species are being modelled?*”, “*Where does the data come from?*”, “*What are the environmental variables?*”, and “*What are the algorithms?*”. Figure 3 displays predefined queries, and parameterized queries available to use.

After selecting the model instance, the researcher can access its general information, modelled species data, algorithm parameters and information, as shown in Figure 4.

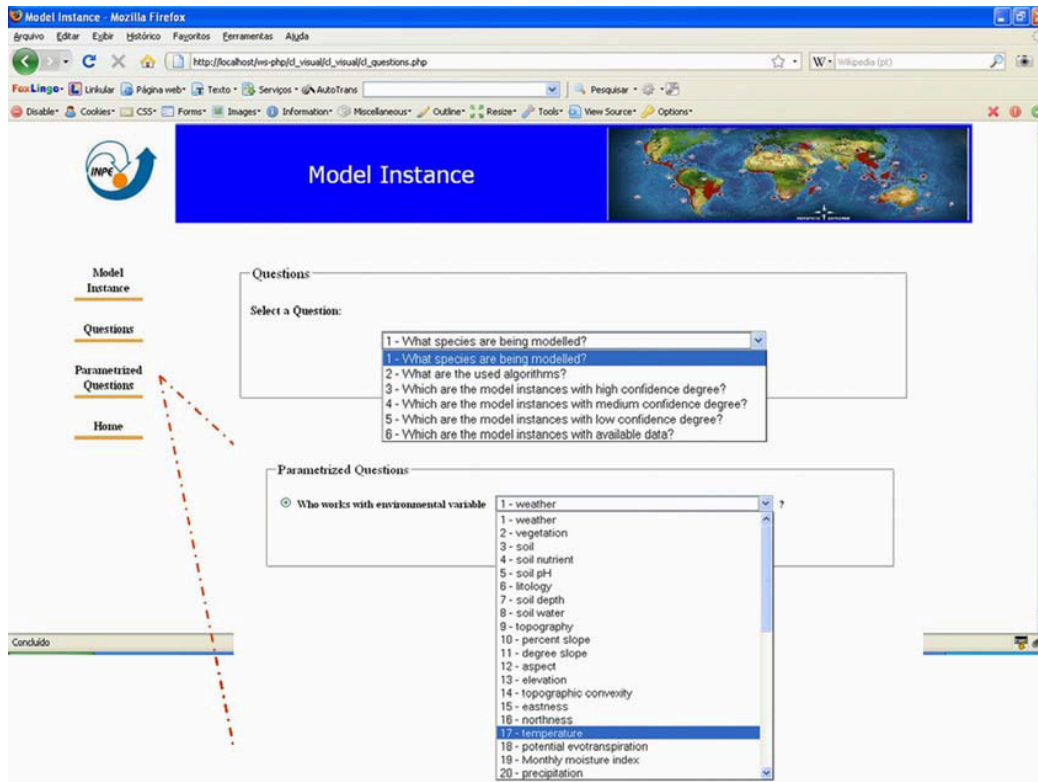


Figure 3. List of available queries

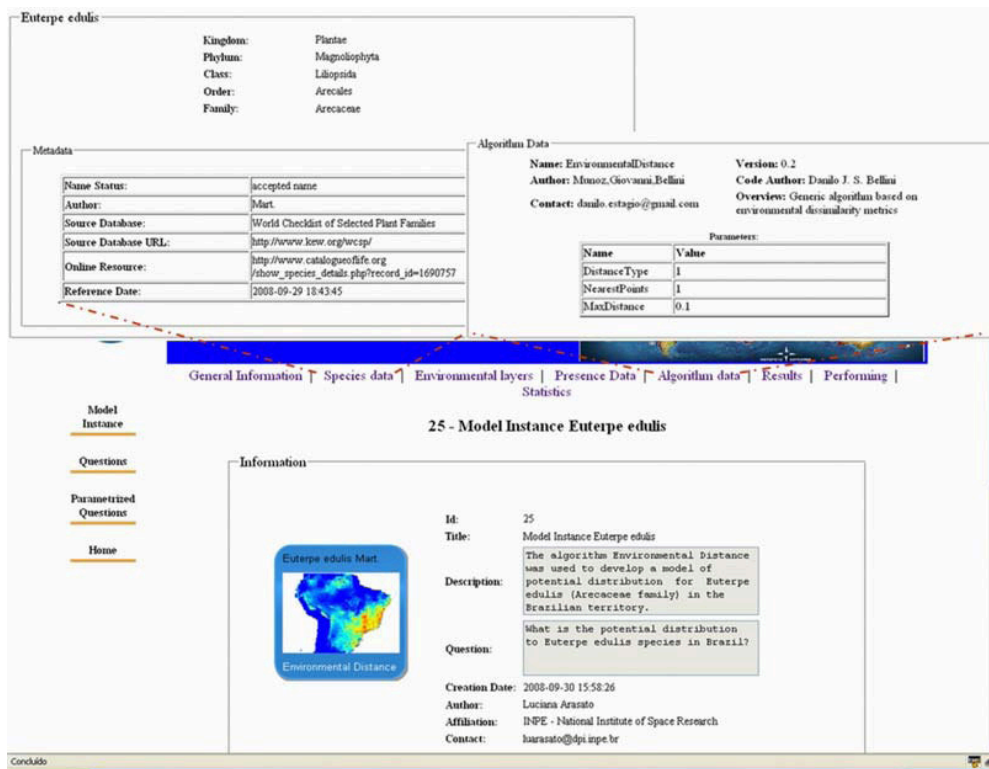


Figure 4. Model instance *Euterpe edulis* Mart. visualization

Besides model instance general information, the modelled species information is presented (Figure 4). Considering that species-occurrence records have different sources and methods, they present different reliability degree to biodiversity researchers. Therefore, making it available is not sufficient to

assure their use by the community. The minimum requirements for a species occurrence record are its geographical positioning, and its taxonomic identification together with metadata such as details of when and where the specimen was collected (GURALNICK *et al.*, 2007).

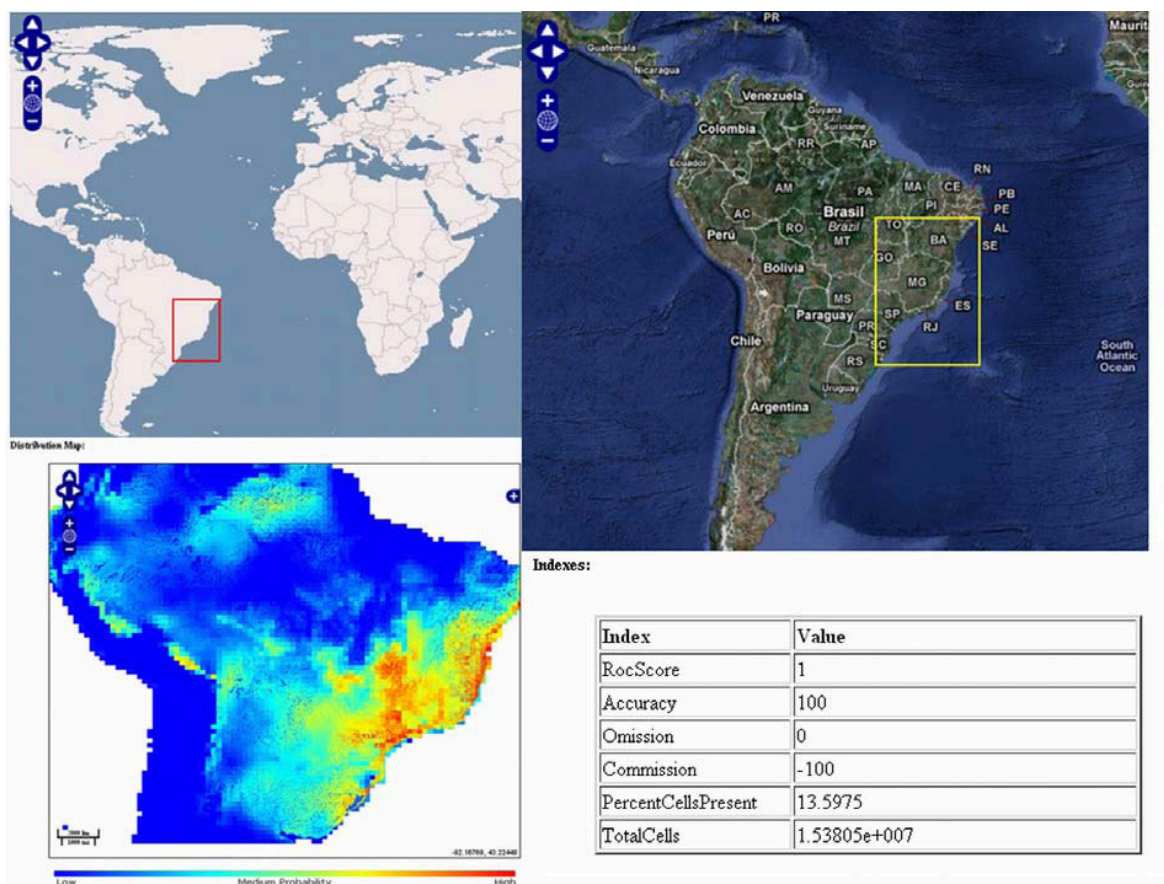


Figure 5. *Euterpe edulis* Mart. distribution map and evaluation indexes

Figure 5 displays model instance species distribution map and evaluation indexes. Maps and satellite images show the area where the species was found. The evaluation indexes and author comments about the experiment help the researcher to capture relevant aspects of the model. The *Model Instance Access* application also makes available data and metadata about modelling experiment authors.

- **Reusing model instance data to run new models**

The researcher can reuse catalogued model instance to run new models. Figure 6 displays the application form that enables the model instance reuse.

Figure 6 displays model instance algorithm information and parameters. The researcher can change algorithm parameters and select different environmental layers to run different models remotely. After this, new species distribution models are returned for comparisons (Figure 7).

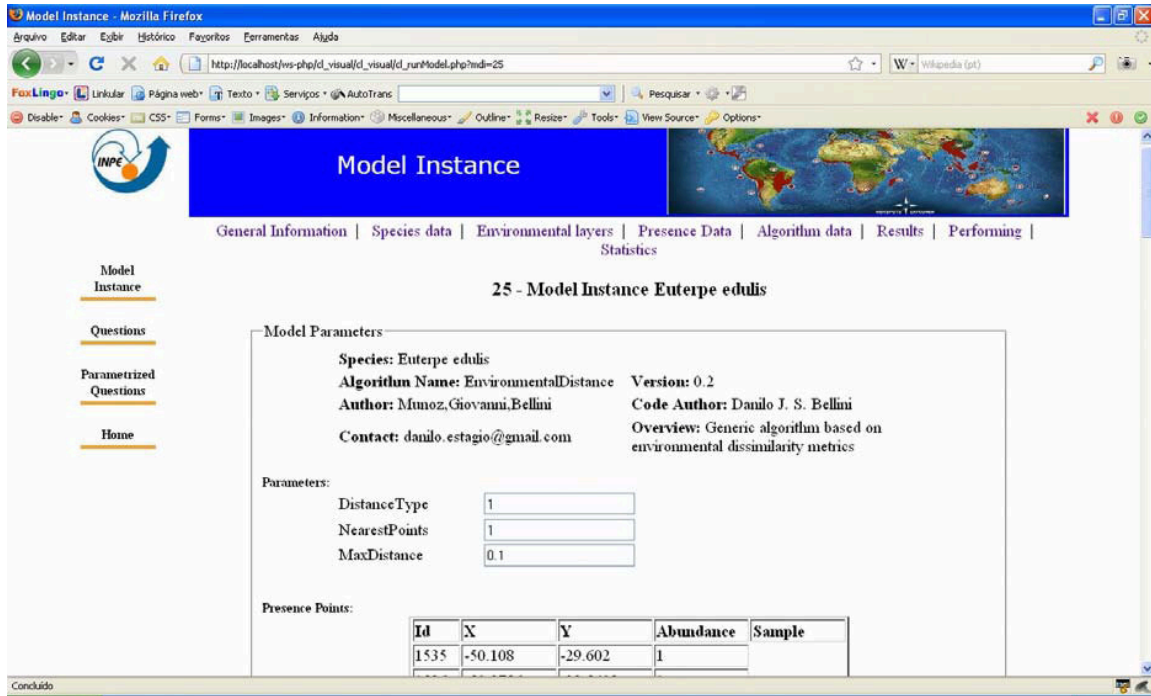


Figure 6. Reusing model instance data

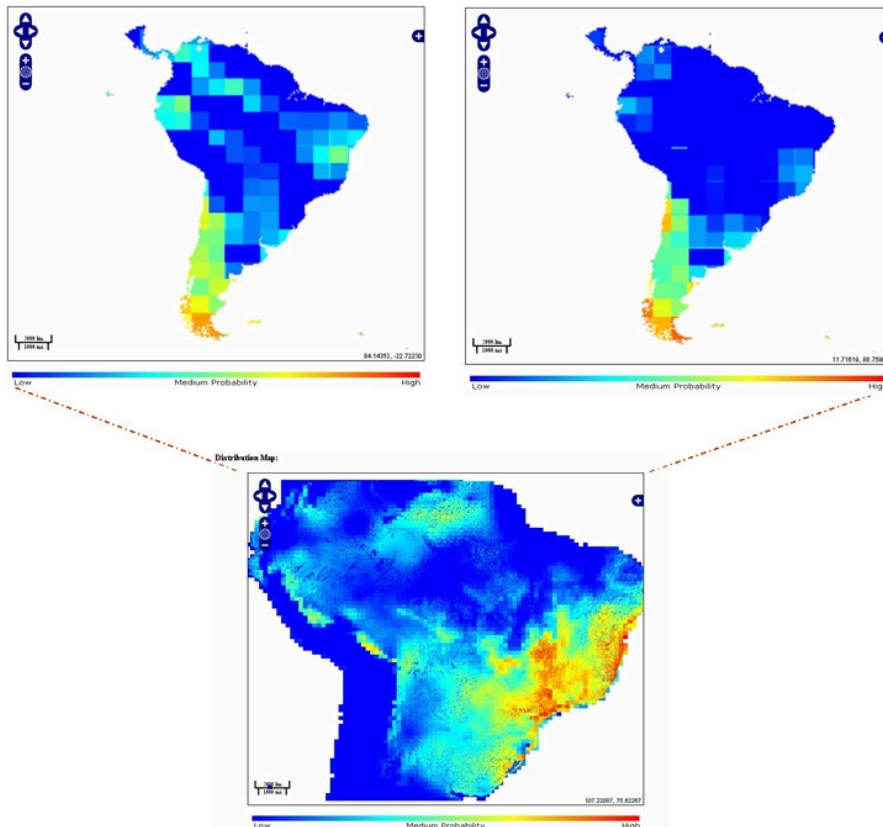


Figure 7. New distribution maps based on Model Instance *Euterpe edulis* Mart.

Figure 7 displays model instance distribution map, and two samples of species distribution maps based on catalogued model instance. Our main goal is to enable the scientist to compare different distribution models and to make new inferences about his studies.

4. Final Comments

This paper illustrated the use of a Collaborative Environment in a species distribution modelling network. The environment aims to support scientific research, planning, conservation, and management. The architecture is based on a model instance catalogue, and a set of geospatial web services, named as Web Biodiversity Collaborative Modelling Services – WBCMS. The approach makes modelling experiment results available on the Web, and enables biodiversity researchers to perform new models based on previous ones.

An example of species distribution modelling experiment cataloguing and reusing illustrated the collaborative environment usage from a researcher's point of view. Our experiments, have demonstrated the usefulness of the proposals and ideas presented in this work. We consider this line of work promising as a global tool to improve biodiversity research.

Acknowledgements

Special thanks go to OpenModeller Project (FAPESP process: 04/11012-0); and to FAPEMA (Fundação de Amparo à Pesquisa e ao Desenvolvimento Científico e Tecnológico do Maranhão) for partially supporting this research.

References

BEST, B. D.; HALPIN, P. N.; FUJIOKA, E.; READ, A. J.; QIAN, S. S.; HAZEN, L. J.; SCHICK, R. S. *Geospatial web services within a scientific workflow: Predicting marine mammal habitats in a dynamic environment*. *Ecological Informatics*, v. 2, p. 210-223, 2007.

BIOGEOMANCER, 2005, *Georeferencing reveals biological importance*. (BioGeomancer)(Brief Article) GeoWorld.

BUSBY, J. R. Bioclim : a bioclimate analysis and prediction system. *Plant Protection Quarterly* (Australia), v. 6, p. 8-9, 1991.

DALTIO, J.; MEDEIROS, C. B. *Aondê* : An ontology Web service for interoperability across biodiversity applications. *Information Systems*, v. 33, p. 724–753, 2008.

GIOVANNI, R. D. The OpenModeller project. In: *BiodiversityWorld GRID workshop*. e-Science Institute, Edinburgh, 2005. p.

GRILO, C. *Critérios para a selecção de zonas prioritárias para a conservação em áreas protegidas*. 2006. Article p. <http://www.naturlink.pt/canais/Artigo.asp?iArtigo=3245&iLingua=1>.

GUISAN, A.; THUILLER, W. Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, v. 8, n.9, p. 993-1009, 2005.

GUISAN, A.; ZIMMERMANN, N. E. Predictive habitat distribution models in ecology. *Ecological Modelling*, v. 135, p. 147–186, 2000.

GURALNICK, R. P.; HILL, A. W.; LANE, M. Towards a collaborative, global infrastructure for biodiversity assessment. *Ecology Letters*, v. 10, p. 663-672, 2007.

HORTA, L. M.; GENTRY, M., 2008, *Beija-flor User's Guide: An Internet-based Approach for Sharing Scientific Data in LBA, LBA / INPE-CPTEC*.

HUTCHINSON, G. E. Concluding Remarks. *Cold Spring Harbour Symposium on Quantitative Biology*, v. 22, p. 415-427, 1957.

MUÑOZ, M. openModeller: A framework for biological/environmental modelling. In: *Inter-American Workshop on Environmental Data Access*. Campinas, SP, Brazil, 2004. p.

O'CONNOR, A.; STOCK, C.; BISHOP, I. SIEVE: An Online Collaborative Environment for Visualising Environmental Model Outputs. In: *MODSIM 2005 International Congress on Modelling and Simulation*. Modelling and Simulation Society of Australia and New Zealand. 2005. p. 3078-3084.

OPENMODELLER. openModeller: Static Spatial Distribution Modelling Tool. CRIA/FAPESP, 2005. <http://openmodeller.cria.org.br/>.

PHILLIPS, S. J.; ANDERSON, R. P.; SCHAPIRE, R. E. Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, v. 190, p. 231–259, 2006.

RUSHTON, S. P.; ORMEROD, S. J.; KERBY, G. New paradigms for modelling species distributions? *Journal of Applied Ecology*, v. 41, n.2, p. 193-200, 2004.

SERIQUE, K. J. A.; SANTOS, J. L. C. D.; COSTA, F. S.; MAIA, J. M. F. Mo Porã – Um sistema gerenciador de repositórios distribuídos e colaborativos no ambiente científico da Amazônia. In: SBSR - Simpósio Brasileiro de Sistemas Colaborativos. *Anais do XXVII Congresso da SBC*. Rio de Janeiro, RJ - Brazil, 2007. p. 1801-1812.

STOCKWELL, D.; PETERS, D. The GARP modelling system: problems and solutions to automated spatial prediction. *International Journal Geographical Information Science*, v. 13, n.2, p. 143-158, 1999.

SUTTON, T.; GIOVANNI, R. D.; SIQUEIRA, M. F. D. Introducing openModeller - A fundamental niche modelling framework. *OSGeo Journal*, v. 1, 2007.

WEBIOS, 2005, *WeBios: Web Service Multimodal Tools for Strategic Biodiversity Research*, Assessment and Monitoring Project, <http://www.lis.ic.unicamp.br/projects/webios>.