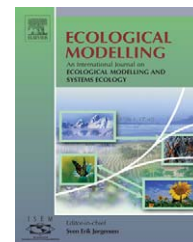


available at www.sciencedirect.comjournal homepage: www.elsevier.com/locate/ecolmodel

The use of the GARP genetic algorithm and Internet grid computing in the Lifemapper world atlas of species biodiversity

David R.B. Stockwell^{a,*}, James H. Beach^b, Aimee Stewart^b, Gregory Vorontsov^b, David Vieglais^b, Ricardo Scachetti Pereira^c

^a San Diego Supercomputer Center, 9500 Gilman Drive, La Jolla, CA 92037, United States

^b Biodiversity Research Center, 1345 Jayhawk Boulevard, University of Kansas, Lawrence, Kansas 66045, United States

^c Centro de Referencia em Informacao Ambiental, Av. Dr. Romeu Tortima, 388, Campinas, Sao Paulo, 13084520, Brazil

ARTICLE INFO

Article history:

Published on line 20 December 2005

Keywords:

GARP

Lifemapper

Species distribution

Genetic algorithm

Biodiversity

ABSTRACT

Lifemapper (<http://www.lifemapper.org>) is a predictive electronic atlas of the Earth's biological biodiversity. Using a screensaver version of the GARP genetic algorithm for modelling species distributions, Lifemapper harnesses vast computing resources through volunteers PCs similar to SETI@home, to develop models of the distribution of the world's fauna and flora. The Lifemapper project's primary goal is to provide an up-to-date and comprehensive database of species maps and predictive models (i.e. a fauna and flora of the world) using available data on species' locations. The models are developed using specimen data from distributed museum collections and an archive of geospatial environmental correlates. A central server maintains a dynamic archive of species maps and models for research, outreach to the general community and feedback to museum data providers. This paper is a case study in the role, use and justification of a genetic algorithm in development of large-scale environmental informatics infrastructure.

© 2005 Elsevier B.V. All rights reserved.

1. Introduction

Humans have explored the life of the planet for the past 250 years. That knowledge is documented by millions of original specimens of plants and animals in the world's natural history museums and herbaria. However, we have yet to achieve the goal of identifying and mapping the distribution of the species present on earth. The vision of the Lifemapper system was to use the Internet to retrieve records of species locations from museum collections, to compute the ecological profile of each species and predict where each species could potentially live. This would then be the basis of a database of species distribution maps. The approach of developing models of species'

habitat provides not only maps of where Earth's species of plants and animals live, but also predictions of where Earth's species of plants and animals could potentially live under different scenarios such as climate change, and where and how introduced species could spread across different regions of the world.

The primary data for species distribution analysis are species location records. These localities for millions of plants and animal specimens are recorded in the world's natural history museums. There is a growing recognition of the usefulness of these records for spatial modelling of biodiversity (Peterson and Stockwell, 2002). These data have also become easily accessible via the integration of the world's museum

* Corresponding author. Tel.: +1 858 8220942; fax: +1 858 8223631.

E-mail address: davids@sdsc.edu (D.R.B. Stockwell).

0304-3880/\$ – see front matter © 2005 Elsevier B.V. All rights reserved.

doi:10.1016/j.ecolmodel.2005.11.016

collection into a distributed database called the Species Analyst (Vieglais et al., 1998).

Developing maps of species distributions using multivariate models of species occurrence points with environmental variables is now a widespread practice in biodiversity science (e.g. Scott et al., 2002). The primary challenge of biodiversity mapping is to develop accurate maps based on statistical relationships rather than actual observations, particularly using ad hoc museum collections data rather than controlled survey data. The Lifemapper team has chosen to use an evolutionary algorithm called the genetic algorithm for rule-set production (<http://www.biodi.sdsc.edu>), or GARP (Stockwell and Noble, 1992; Stockwell and Peters, 1999; Stockwell, 1999). In comparison with other algorithms, the GARP algorithm has shown to be the best available method for reliable species predictions using small sets of ad hoc data typically returned from museum databases (Peterson and Stockwell, 2002; Stockwell and Peterson, 2002a,b). The robustness of the algorithm has contributed to its use in projects requiring the development of large numbers of species distribution models, such as change in ecological communities due to climate change (Peterson et al., 2002). One of the purposes of this paper is to describe how the qualities of the GARP genetic algorithm contribute to the success of the project.

The second main challenge of developing a fauna and flora for the world is the computational scale of the project. This problem is computationally intensive in many aspects:

- The number of species in the world that could be mapped, e.g. for birds alone already 9600 species (Sibley and Monroe, 1990).
- The fine scale of maps that could be produced, e.g. a map of the world at a resolution of 1 km at the equator has approximately 1,000,000,000 cells.
- The number of replicates that need to be produced for each species for estimating statistical variance.
- The need for computing distributions under alternative scenarios such as climate change and geographic invasion.
- The need to recompute the maps when new data becomes available.

The need to maintain an up-to-date resource by recomputing maps when new data becomes available ensures a perpetual demand for computing resources. The combination of millions of idle computers around the world connected to the Internet forms the infrastructure for Internet grid computing which seeks to exploit otherwise idle workstations, PCs and bandwidth to create powerful distributed computing systems. Internet computing was popularized by the SETI@home project that enlisted personal computers to analyze data for indications of extraterrestrial intelligence. As SETI@home is now running on half a million PCs and delivering 1000 CPU years per day, it is currently the fastest (admittedly special purpose) computer in the world.

The goal of the Lifemapper project is to become a major component in the biodiversity informatics infrastructure. The archive of maps and models for the world's species developed through Lifemapper will be an invaluable resource for researchers in terrestrial, marine and freshwater environments. This paper describes the Lifemapper resource with

particular reference to the genetic modelling algorithm that serves as the core computational component of the screen-saver modelling program.

2. Methods

The general approach to producing maps of species' distributions is to develop a multivariate statistical model of known species' occurrence records and environmental variables. The species occurrence data are gathered from a number of biological collections housed at several museums and herbaria worldwide. Those institutions have their specimen databases linked and integrated through the Species Analyst (Vieglais et al., 1998) project. The environmental information is composed of a set of global geographic coverages, called environmental layers. Each layer represents one particular environmental parameter, such as temperature, rainfall, land use and elevation, among others. The layers are continuous grids, where each cell contains the value of an environmental parameter at a location.

A model for estimating the probability of occurrence of a species is developed from the species occurrences and the environmental variables. Using the values of the environmental variables at each grid cell to predict the probability of the occurrence of the species over the entire grid then produces a map of the distribution of the species. An example is given in Fig. 1.

The Lifemapper system is composed of components for harvesting species occurrence data from museum databases, developing models via the screen-saver program, storing the results in an archive, and enabling access to information and membership management functions. The schematic diagram for Lifemapper is shown in Fig. 2. This is described in three stages, pre-processing, processing and post-processing.

2.1. Pre-processing

The Lifemapper program retrieves georeferenced data from a network of biodiversity databases that are accessible over the Internet. At this time, almost all sources utilize the Species Analyst architecture, which uses Z39.50 protocol, an international (ISO 23950) standard defining a protocol for computer-to-computer bibliographic information retrieval from participating databases over the Internet. The first stage consists of periodically scanning through the data sources, building and updating the list of unique scientific names that are currently available for retrieval. The second stage builds a list of georeferenced records, using the list of names. The environmental layers consist of a set of geographical coverages that describe the main environmental parameters that may affect species' geographical distributions: temperature, precipitation, rainfall, solar radiation, terrain elevation and slope, and tree coverage, among others. The original climate data used by Lifemapper was obtained from the Intergovernmental Panel on Climate Change (<http://www.ipcc.ch/>) at a scale of 0.5° and processed to 1 km resolution to match the other data sets. All variables are year averages from 1961 to 1990, of the averages for the months of January and July for the same period: cloud cover, diurnal temperature range, ground-

Download English Version:

<https://daneshyari.com/en/article/4379098>

Download Persian Version:

<https://daneshyari.com/article/4379098>

[Daneshyari.com](https://daneshyari.com)