



FOSS4G

D E N V E R 2011

12-16 SEPTEMBER 2011 SHERATON DENVER DOWNTOWN
2011.FOSS4G.ORG

Conference Proceedings

**Free and Open Source Software for
Geospatial Conference**

OSGeo Journal
Vol. 10 - Feb 2012



Volume 10 Contents

FOSS4G 2011 Conference Proceedings	2	Using GRASS and R for Landscape Regionalization through PAM Cluster Analysis	26
Editorial - FOSS4G 2011 Academic Chair	3	Functional Coverages	32
Open source based online map sharing to support real-time collaboration	5	Opticks Open Source Remote Sensing and Image Processing Software	45
An innovative Web Processing Services based GIS architecture	15	Implementation, challenges and future directions of integrating services from the GIS and decision science domains	50
		A Vivid Relic Under Rapid Transformation	56

Welcome from the Conference Chair



Welcome to this special edition of the OSGeo Journal, featuring selected papers from the academic track that were presented at the FOSS4G (Free and Open Source Software for Geospatial) 2011 conference in Denver.¹ The conference was the largest FOSS4G yet, with 914 attendees from 42 countries. Feedback from attendees was very positive, with the post-conference survey giving it an overall rating of 4.32 out 5. The attendance reflects the strong growth in interest in open source software that we are currently seeing in the geospatial industry.

We made a conscious effort in 2011 to enhance the academic track at the conference by providing improved publishing opportunities. We did this through publishing papers both in “Transactions in GIS” and in this edition of the OSGeo Journal. I would like to thank Rafael Moreno for leading this effort, as well as the rest of the organizers of the academic track who Rafael recognizes below.

*Peter Batty, Ubisense
FOSS4G 2011 Conference Chair*

¹FOSS4G: <http://foss4g.org>

FOSS4G 2011 Conference Proceedings

An innovative Web Processing Services based GIS architecture

For global biogeographic analyses of species distributions

Jeffery A. Caver, Aimee M. Stewart, Charles J. Grady, James H. Beach

University of Kansas

jcavner@ku.edu, astewart@ku.edu, cjgrady@ku.edu,
beach@ku.edu

Abstract

Spatial patterns and properties of species richness in natural communities are of keen interest to biogeographers and conservation biologists as they describe key features of the location and distribution of the earth's biological diversity, but species richness tools are scattered across specialty software and are underrepresented in distributed approaches for GIS for work with large datasets. We describe an ongoing development effort for producing macroecology and biogeography tools dealing with large species presence data structures using the Web Processing Service (WPS) specification and Quantum GIS (QGIS) as a WPS client. The creation of species presence/absence matrices is one approach for linking range size and richness patterns and these spatial patterns are well suited for GIS analysis. This paper presents our efforts to date on the development of the Lifemapper Range and Diversity (LmRAD) analysis suite for Lifemapper and on its potential contribution to global biodiversity research and conservation. LmRAD is being engineered as a job based infrastructure that is portable across compute environments for exposing macroecology algorithms for biodiversity calculations as WPS services, a client library for GIS and Scientific Workflow environment software that is tailored for communication using WPS and other OGC standards, and a client plug-in for QGIS. The practical importance of bringing a standardized spatial data processing standard into a distributed GIS environment for macroecology is that larger institutional computer resources can be brought to bear on large problems at vast scales, i.e. continental to global extents at high resolutions for thousands of species. Additionally the decoupled nature of the Web Services approach can allow scientists to mix and match tools in user defined workflows where meta-data can be produced that allow experiment repeatability.

Introduction

The two fundamental units of biogeography are species diversity and the distributional range of species. These two fundamental concepts can be summarized by a basic analytical tool first introduced by Simpson in assessing the diversity of North American mammals (Simpson, 1964). The data for Simpson's study were records of species of recent mammals in quadrants of equal area covering North America,

consisting of a rectangular grid not oriented with respect to physiographic features or other known zoogeographic features. The presence or absence of each species was noted for each quadrant (Simpson, 1964). In this Presence-Absence Matrix (PAM), one axis represents species and the orthogonal axis represents geographic localities or samples. Each geographic site is coded for the presence (1) or absence (0) of each of hundreds or thousands of species resulting in a binary matrix. The PAM combines species richness of sites, the number of species summed across an individual site with range sizes, each species' range expressed as a sum of their presence values across all the sites that they occupy. The PAM has become a basic method used to test ecological and evolutionary hypotheses about the spatial patterns of biological diversity on continental and global scales (Arita et al. 2008), but very few software applications have been developed to aide in PAM construction, analysis and visualization for very large datasets.

We are addressing these challenges by engineering the Lifemapper Range and Diversity (LmRAD) tool. LmRAD is an analysis suite within the current Lifemapper (www.lifemapper.org) platform that will use the computational power of a high-throughput computer cluster to execute macroecology algorithms exposed as Open Geospatial Consortium (OGC)¹⁴ Web Processing Services (WPS) (Open Geospatial Consortium, Inc., 2007b). LmRAD uses a GIS client for for web service communication and data visualization. In this way LmRAD describes the composition of natural communities using PAM's and species range maps as primary data inputs. The primary software components of LmRAD are a set of macroecology WPS services, a multi-platform Python client library for interacting with the WPS services and a Python¹⁵ plug-in for QGIS.¹⁶

The analysis of diversity patterns at bio-geographical scales, i.e. continental to global extents, significantly increases the size of the PAM when describing whole taxa over such ranges and presents computational challenges during the construction or intersection of the PAM with species inputs, randomization of the PAM, and the generation of PAM statistics of diversity and range with linear algebra operations on large matrices. PAM's are often hand constructed requiring intersecting thousands of species range maps with a data grid. Construction of PAMs can be an extremely time consuming data management task for researchers when using a heterogeneous collection of GIS tools and statistical packages each with a required learning curve.

The role of species association at larger biogeographic scales remains an important question (Arita, et al. 2008). General assembly rules of species interaction have been called into question since Connor and Simberloff (Connor and Simberloff, 1979) showed that Diamond's (Diamond, 1975) assembly rules for species association and the determination of the composition of natural communities were not different than

¹⁴<http://www.opengeospatial.org>

¹⁵<http://www.python.org>

¹⁶<http://www.qgis.org>

could be expected by chance. Since Connor and Simberloff, null model generation using randomized PAM's has become a standard but computationally intensive methodology for testing species associations and their contribution to diversity patterns. LmRAD offers two randomization algorithms. In one type of null model analysis, the matrix is randomized to produce patterns that would be expected minus prohibitive species interaction (Gotelli and Graves, 1996), while keeping species richness and range size marginal totals intact. To see why randomization of potentially sparse binary matrices can become computationally intensive you could consider the proportional fill of the matrix, p , given as the total of presence values in a matrix $\delta_{i,j}$ divided by the product of total columns S and total rows N , where, $p = \frac{\sum_i \sum_j \delta_{i,j}}{NS}$. A swap algorithm that keeps range sizes and richness totals intact must search for sub matrices of the form $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ or $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ with any consistent distance between the vertical and horizontal components, therefore the probability of finding a swap becomes $P(A) = 2p^2(1-p)^2$. So the probability for finding a swap in a random fill of $p = 0.5$ is $P(A) = 0.125$, and represents the upper limit of the probability of finding a swap (J. Soberon, personal communication, Feb. 4, 2011). More typically fills are much lower. In two examples, global terrestrial mammals and Kansas flora, the proportional fills are $p = 0.014$ and $p = 0.12$ respectively. In the first case this leads to the probability of a swap occurring once out of approximately every 2624 searches of a matrix with 6.3×10^8 elements requiring several thousand swaps for randomization. Current popular desktop software for randomizing PAMs have proven insufficient when randomizing such large PAMs (J. Soberon, personal communication, Jan. 21, 2011). The computational constraints of operating against large matrices prompted the design of LmRAD as a client-server architecture using WPS, as opposed to a strict desktop software.

To further our understanding of the software needs for macroecology we worked with ecologists from University of Kansas to answer questions pertaining to what methods were currently used for PAM construction. We sought to answer what the outstanding issues were concerning modern methods for randomizing PAMs, why computational constraints were present, what dissatisfaction existed with current software, and how best to operationalize methods for analysis of PAMs. We also sought to answer questions pertaining to specific research questions of range and diversity and how data from PAMs were best visualized. To answer these questions we began by modeling a range and diversity experiment as Python objects that could be built from information stored in a PostgreSQL¹⁷ database and exposing methods against those objects as simple WPS services. We began by attempting to adapt PyWPS¹⁸ (Cepicky and Becchi, 2007) to our current architecture but found it easier to start with the asynchronous architectural components already developed for our current Lfemapper modeling services and adapt our own WPS implementation to those specifications.

The web services approach also known as Service Oriented Architecture (SOA) provides a more 'democratic' mechanism for accessing geostatistical algorithms for ecology that would otherwise be bound within specialist GIS and spatial

statistics software. SOA promotes interoperability and loose coupling across remote systems allowing discovery and consumption of processes and chaining of those processes. By exposing spatial and statistical algorithms as Web Processing Services for generating PAMs of species data, as well as to create species and range indices and null hypothesis data inputs for these indices we are bringing together a set of tools into a single framework that will allow users to supply their own species data in addition to gaining access to species data from Lfemapper predicted range maps and use these data to populate PAMs and build range diversity plots by species and by geographic site following (Arita et al.'s 2008), methodology for the mathematics of range-diversity plots and finally to visualize these outputs across geographic and phenetic data spaces.

While there are reports in the literature about SOA and GIS with respect to WPS (Friis-Christensen et al. 2007) and others that are strict distributed GIS approaches (S.S. Wang and D.Y. Lili, 2004) fewer efforts have been detailed in the literature that address grid computing and distributed computing with GIS analysis using WPS. (Meng, Xie, and Bian, 2010, Lanig and Zipf, 2009, Muller et al. 2010) Other studies focus on WPS and its applicability to ecology, (Graul and Zipf, 2008), though the use of WPS with most ecology tools is sparse. Macroecology and biogeographic applications that have adopted WPS seem to be in their infancy with some notable exceptions. Forecasting biomes of protected areas using WPS is currently addressed by eHabitat which is conceptualized as part of the Model Web. (Dubois, et al. 2011, Skoien et al. 2011). The Model Web seeks to improve the predictive capacity of ecological models by increasing the connectivity and interoperability of models on the Web so that they can work together using web services (Geller and Turner, 2007). As part of the Model Web, eHabitat enhances its forecasting capabilities by using a discovery broker service to search for modeling services to provide inputs to other models such as climate models, to support an extended set of considerations for modeling impacts to critical habitats (Dubois, et al. 2011). Some aspects of data integration for environmental science collaboration using WPS have also been addressed by the USGS Geo Data Portal (Blodget, et. al., 2011). The Geo Data Portal allows modelers and other data users across disciplines to extract accurate spatially weighted coverage statistics from gridded data sources.

Performance issues encountered by (Friis-Christensen et al. 2007) in their description of SOA's in Spatial Data Infrastructures (SDI) stem from a fundamental limitation of such systems where the transport of data in the service chain becomes necessary if the data and the processing services are provided on different nodes in the service chain. An interesting approach for service based SDI suggests that it may be more efficient not to move the data to the processing instance, but rather move the algorithmic code closer to where the data resides when large amounts of data need to be processed, (Friis-Christensen et al. 2007, Muller et al. 2010). This involves moving away from the data driven paradigm to processing-centric infrastructures which share geoprocessing functionality through the "moving code" approach (Muller et al. 2010). Muller et al. 2010 also review Grid computing ap-

¹⁷<http://www.postgresql.org>

¹⁸<http://pywps.wald.intevation.org>

proaches for SDI's as a type of 'moving code' approach where processing tasks are passed to a middle-ware that shares the code with different nodes of the grid. This approach still has the constraint of requiring the transfer large datasets to each processing node at runtime. They conclude by pointing to the need for more research on Grid infrastructures and Geoprocessing services. In such systems a processing service interface and a deployment service provide access to the Grid environment. It would be possible to create a service based Grid, where each node would host a processing service creating a service array for parallel processing (Muller et al. 2010). Muller et al. 2010 describe an architecture following similar approaches from Friis-Christensen et al. 2007, for 'translucent' chaining of services, allowing a user to define work flows as a chain of operations and then pass this chain as a single execute request to a service instance, allowing for ad-hoc algorithm deployment.

LmRAD is being incorporated into the existing Lifemapper architecture at the Biodiversity Institute at the University of Kansas. Lifemapper is a robust archival and species distribution modeling platform consisting of a computational pipeline, specimen data archive, predicted species distribution model archive, and an Open Geospatial Consortium (OGC) and Representation State Transfer (REST) based suite of Web Services for on-demand modeling using openModeller.¹⁹ Species distribution models are created by using climate scenario data and aggregated specimen occurrences from the Global Biodiversity Information Facility (GBIF).²⁰ Existing Lifemapper computational resources provide data through web services for species occurrence, environmental data, and predicted habitat maps for various research and education laboratories.

Components that have been completed for LmRAD include web services to 1.) define and construct a gridded study area; 2.) intersect species input layers with that grid; 3.) construct a PAM from the intersected layers; 4.) compress and analyze the PAM; 5.) randomize the PAM; 6.) assess the significance of the PAM against the randomized null model; 7.) create Range Diversity plots; and 8.) a client library to interact with those services. The QGIS plug-in client library connectors, have also been completed, and services are producing preliminary plots, and the outputs can be visualized in the QGIS client. (See Figure 1.) We have implemented synchronous mode execution for smaller jobs on a web server and the processes are returning data products for continental sized PAMs at reasonable response times, while asynchronous WPS execution will be achieved through extensions to our current computational pipeline.

For our first client we chose the open source desktop GIS software, Quantum GIS (QGIS) as a WPS client (See Fig. 1). Quantum GIS was chosen because of its full featured GIS capabilities and its open and extensible architecture, but LmRAD has been built with an open, decoupled approach that should allow a variety of standards based clients, such as the scientific workflow environment software VisTrails. The QGIS client interface consists of a Python plug-in and a client library that abstracts the client functionality generalizing it for use across clients. Each Lifemapper web service is capable of self documentation by producing standard EML for re-

sulting datasets containing a process metadata section that is readable by the Lifemapper EML reader. This extended EML fully documents experiment results, allowing an experiment to be repeated across clients.

The remainder of the article will start by introducing one type of PAM analysis dealing with range-diversity plots, requiring a particular methodology for analyzing the data in the PAM, and requiring a specific method for future directions in interactive visualization of the data in the PAM that require a linked spatial representation in a GIS. We compare this to current species diversity visualization software and note that this specific type of visualization is not addressed by most species diversity software packages. This is followed by a description of the steps and elements of a range and diversity experiment and how the range and diversity data objects for an experiment that follow from those steps have been implemented in LmRAD. Methodologies for testing null hypotheses against the PAM data are described and how two of these algorithms were implemented using OpenSource libraries and were exposed as WPS services. Next, a description of our current implementation of WPS is discussed which has been designed to allow us to move to a more distributed computing approach with WPS on our compute cluster. A discussion of client architecture and some of the GIS functionality for PAM construction with QGIS is discussed. Finally the early results of testing, limitations, considerations and directions for further work are discussed.

Motivation and Methodology

Range-Diversity Plots

Most traditional analysis of PAMs focus on the distribution of species or their ranges represented by the number of sites where species occur, or on the diversity of sites, the number of species within sites. These approaches ignore the relationships between the axes of the data in the PAM. Arita et al. 2008, have shown that a link between these two variables can take into account information from both axes of the PAM, describing two different correlations between species diversity, and range size. The link is the correlation between the species diversity of sites and mean range size of species occurring in the site. A second correlation is between the range sizes of species and the mean species diversity within those ranges where the species occurs. Arita et al. 2008, show that both correlations are mirror images of the same patterns reflecting fundamental mathematical and biological relationships represented by the PAM. The per-site mean range size is a measure of a 'dispersion field' which is the set of geographic ranges of species occurring in a given site. The dispersion field volume of a site, based on the average covariance of a site with all the sites within the extent of the PAM, can be interpreted as an index of similarity between sites. Comparing the range size of a species occurring in a site is equivalent to studying the covariance of that site with all sites in the PAM. The analogue to that measure but by species is the 'diversity field' which quantifies the species diversity of all sites in which a particular species occurs. Analyzing the diversity field within the range of a species is equivalent to studying

¹⁹<http://openmodeller.sourceforge.net>

²⁰<http://www.gbif.org>

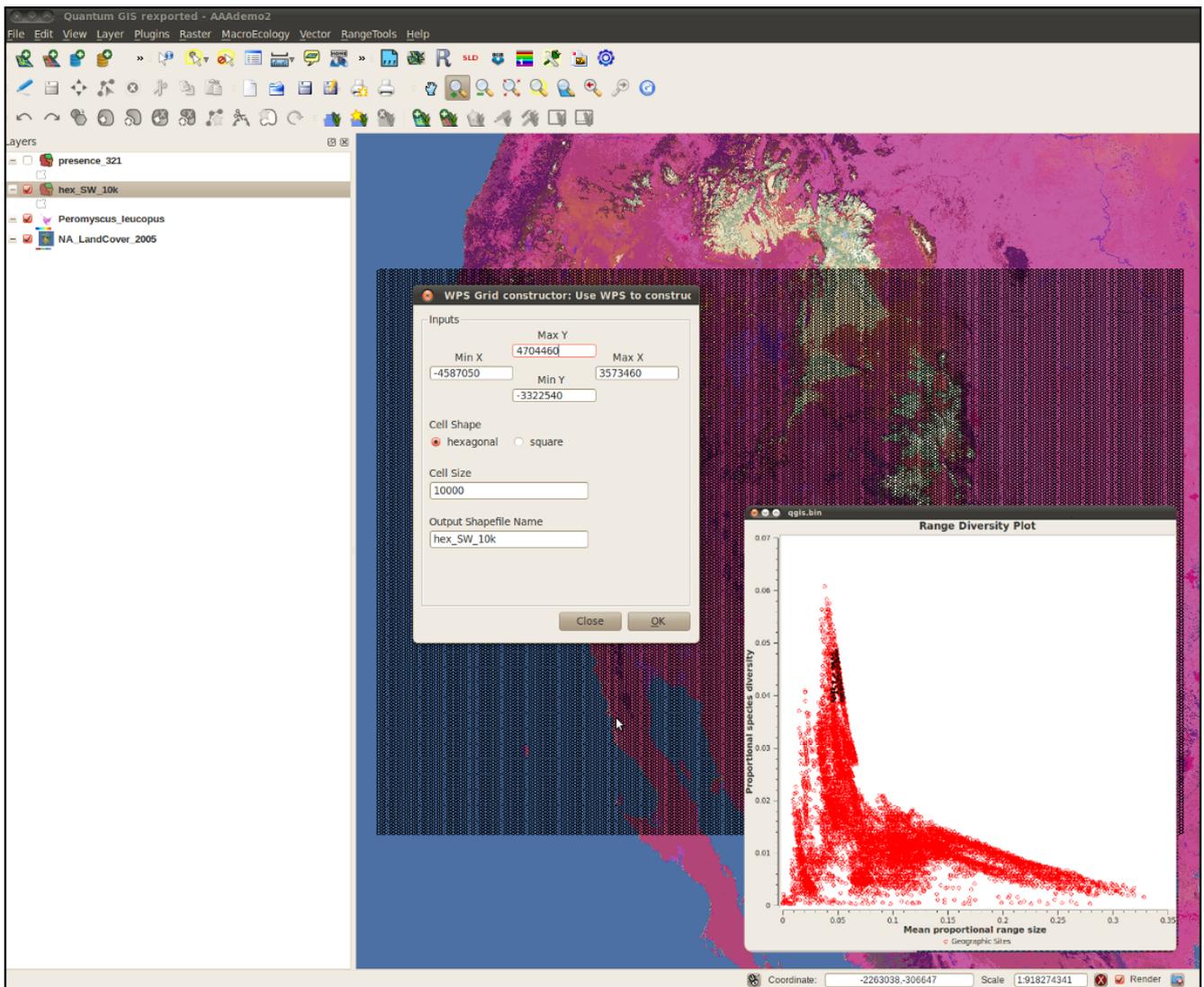


Figure 1: QGIS WPS client plug-in

its covariance with all the species. These relationships inform a "by-species" and "by-sites" view of the fundamental relationships present in the PAM. Arita et al. 2008, introduce the range-diversity plot as a way to depict species diversity of sites and geographic ranges of species simultaneously in two different types of plots. The "by-species" plot describes the relationship between the mean proportional species diversity of sites in which a particular species occurs and the proportional range size of a species. The "by-sites" plot describes the relationship between the mean proportional range size of the species within a site and the proportional species diversity of sites (Arita et al. 2008).

The visualization and analytical goal of the LmRAD platform will be to depict species diversity of sites calculated from the PAM along with species views of the PAM by using panes for different data views within QGIS. The dispersion of sites in an interactive "by-sites" range-diversity plot will be linked geographically to a map where the quadrants or

cells of the PAM are represented visually as a grid. Three data spaces will be linked and data points will be able to be selected and shown across maps, dendrograms, and plots. For species associations the data will depict association of species and link average proportional species diversity of sites containing a particular species correlated with range size in an interactive range-diversity plot with their association in a dendrogram, representing their phylogenetic relationship or morphological attributes. The range-diversity plots will be linked in the client software allowing for 'brushing' of datasets by species or location across tree data space and geographic data space, respectively. This will allow for cross-view interactivity, where selection of data points in one view of the data will highlight the data points from a different perspective in a related visualization space. (see Figure 2.)

²¹www.purl.org/biodiverse

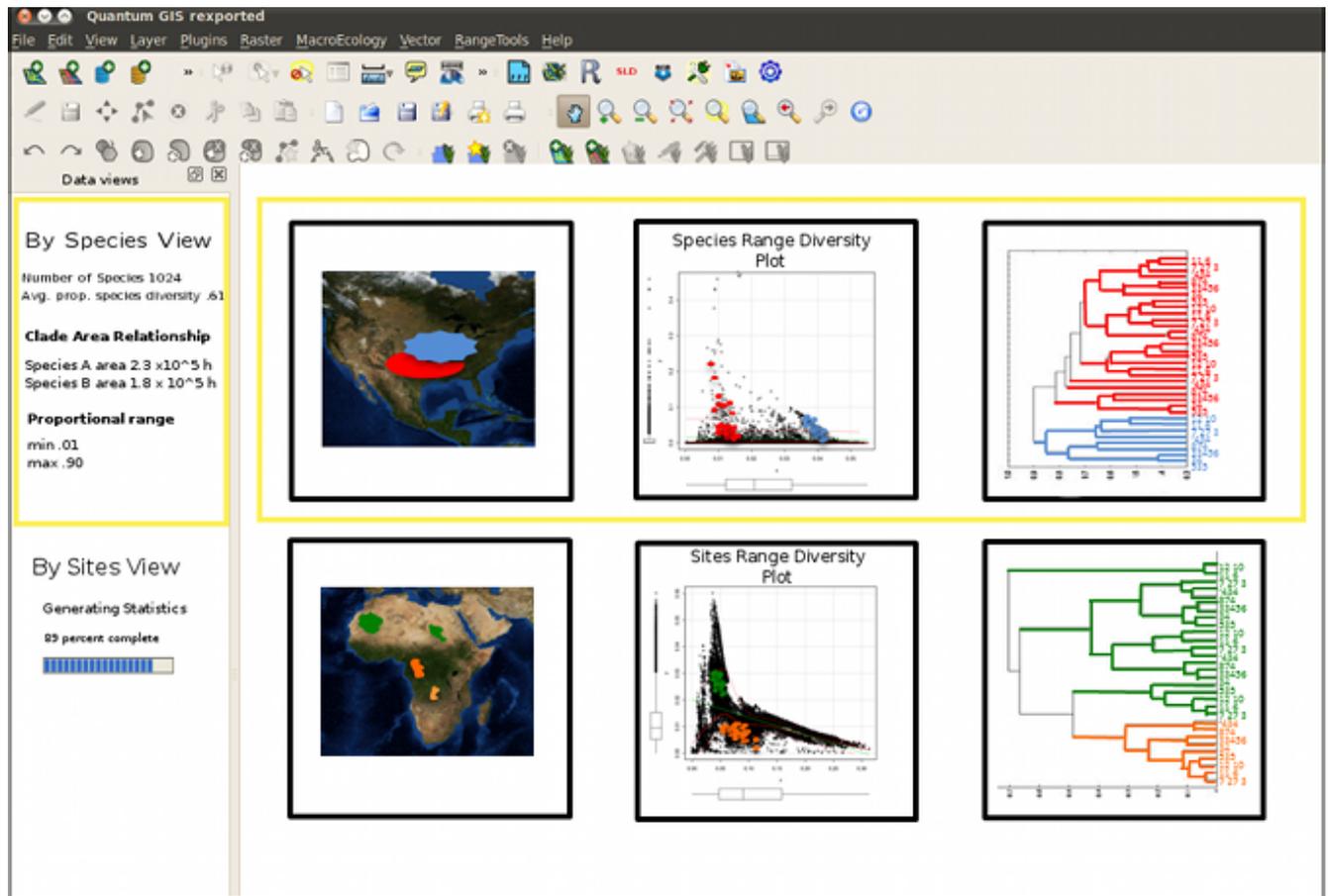


Figure 2: Proposed data panes in QGIS client showing three possible views of the data from a PAM.

Comparison of Approaches

One very interesting approach to investigating and visualizing phylogeographic relationships in a desktop software is BioDiverse²¹, (Laffan, Lubarsky, Rosauer, 2010). It provides linked visualization of data distributions in geographic, taxonomic, phylogenetic, and matrix spaces, providing the calculation of diversity indices with a focus on moving window/cluster and neighbor analysis, along with randomization for hypotheses testing. Similar to a future direction for LmRAD, Biodiverse provides tools through its windowing analysis for demonstrating the effect of scale on different diversity statistics. The set of indices that can be calculated can be customized through a scripting language. With Biodiverse, as opposed to our approach, data and results are stored in a native format and can then be exported for use in a GIS or statistical package. LmRAD starts out by extending a GIS desktop environment with statistical services and the data can then be returned and analyzed further with the existing GIS tools in QGIS.

The core difference between other software that explore phylogeographic relationships, is that LmRAD will incorporate range-diversity plots into the visualization and data space. The range-diversity plot depicts species diversity of sites and geographic ranges of species simultaneously. Phylogenetic software that have a geographic visualization space

focus on clade-area relationships, but our motivation for building a suite of tools for dealing with PAMs seeks to explore more complex mathematical and biological relationships present in the presence absence matrix, dealing with the correlations between the species diversity of sites and mean range size of species occurring in those sites and the converse relationship between range sizes of species and the mean species diversity within those ranges where the species occurs. This focus lends itself to both a "by-sites" and a "by-species" data view that can be expressed through different types of range-diversity plots, one linked to the geographic representation and the other to a phylogenetic or phenetic tree data visualization. This is more than a clade-area view of the data and represents more complex relationships of the dispersion of data points in the range-diversity-plots in terms of the degree of association among species and similarity among geographic sites.

Spatial Analysis in Macroecology, Rangel et al. 2010²² offers another software alternative for biodiversity experiments. SAM offers a comprehensive set of tools for spatial statistics, some simple mapping tools and advanced spatial auto-regression models. It uses extremely optimized linear algebra libraries for large matrix operations. The data table in SAM accommodates PAM data, and can be formatted as ESRI shapefiles. This data structure can also be used in ad-

²²SAM: <http://www.ecoevol.ufg.br/sam>

dition to PAM data to store and analyze species trait data to compute richness patterns according to different trait criteria. LmRAD uses associated data matrices to achieve the same purpose. Just as in LmRAD, data grids can be prepared in SAM at any extent and resolution as shapefiles and PAMs can be generated directly from the shapefile. Additionally, as in LmRAD, environmental raster data can be summarized over the grid cells in the data grid and added as additional variables for analysis. Also very similar to future developments in LmRAD is a tool offered by SAM called Pattern Finder that allows a user to geographically link scatter plots and maps, where grid cells can be selected in a map and then correspondingly highlighted in a scatter plot or vice versa, allowing a user to detect outliers. SAM also has sophisticated tools for evaluating the changes in spatial correlation as they are affected by changes in scale. Currently LmRAD's data model is set up to enable the user to change the scale of the PAM by taking submatrices as individual samples from an existing PAM and recalculating diversity indices, a future direction is to be able to visualize these changes. Species Distribution Modeling (SDM) is touched on in the SAM package with a routine for logistic regression, however, it does not provide many of the different algorithms used for SDM, whereas LmRAD is built on top of Lifemapper, the core functionality of which is to provide SDM algorithms as Web Services and soon to be offered as WPS services. One difference that may be pointed to here is that SAM is a Windows dependent desktop software, where LmRAD provides a cross-platform library for interfacing with remote WPS services.

EcoSim²³ is another Windows based macroecology software built specifically for dealing with null model hypotheses testing and PAM data (Gotelli, and Entsminger. 2011). It provides a variety of randomization routines for each of its modules. The co-occurrence module randomization routine allows row and column constraints, including fixed-sum similar to LmRAD. EcoSim also allows equiprobable, proportional, and weighted constraints (Ellison et al. 2000). LmRAD currently has two randomization algorithms. EcoSim has four different ways of dealing with sparse or degenerate matrices (with empty rows and columns) (Ellison et al. 2000). LmRAD currently has a compression algorithm for compressing and re-expanding such matrices, but could gain from investigating alternative approaches for dealing with degenerate matrices. The limiting factor for EcoSim seems to be the size of the PAM; 240,000 cells, or approximately 800 by 300 rows and columns is an absolute limit. One of the core requirements first addressed by LmRAD was being able to work with much larger matrices. Initial tests in LmRAD for randomization algorithms were done for matrices in the range of 6.3×10^8 cells. The computational constraints inherent in platform dependent systems and desktop software are the chief gap that LmRAD strives to solve through WPS.

User defined workflows are addressed by LmRAD through the recording of Web Service calls in a process metadata extension to the Ecological Metadata Language.²⁴ These workflows can then be re-executed in client packages incorporating the Lifemapper EML reader (Grady et al. 2011). McFerren et al. 2010 investigate a similar approach for scien-

tific workflow environments and the possibility of utilizing FOSS4G libraries for the scientific workflow environment Kepler (McFerren et al. 2010). Issues for geospatial visualization in Kepler were encountered and McFerren et al. 2010 suggest that these issues could be resolved by exporting data back out of Kepler for use in QGIS (McFerren et al. 2010). These issues may be remedied in LmRAD by either embedding QGIS libraries into a more accommodating workflow environment, currently we have embedded OpenLayers²⁵ into VisTrails²⁶, or building workflows into QGIS using the Lifemapper EML reader. McFerren et al. 2010 found that workflow environments can be freed from having to deal with library dependencies through the use of standardized web service interfaces. One implication is the need for libraries that expose complete client implementations for all of the major OGC standards. (McFerren et al. 2010). The LmRAD client library deals with much of the abstraction for bringing OGC services into both QGIS as a client, and the VisTrails environment (Grady et al. 2011).

Range and Diversity Experiments

A Lifemapper range and diversity experiment starts with the construction of a PAM. Information about species input layers for PAM construction are stored with threshold parameters for presence and absence set by the user through a REST based interface to a PostgreSQL database along with information about the study area extent, resolution, and cell shape. The PAM has two related matrices that hold intersected environmental information, species trait information, and summary statistics for each site present in the grid and for each species for use in a "by-sites" analysis. Randomized PAM's for null hypotheses tests are stored on the server-side file system in a user data space and information used for their retrieval and data about the methodology used for their randomization stored in the database. A user may collect both species data and environmental data related to the geography of the PAM, but can also use OGC services as input parameters in a WPS execute requests to provide inputs from other providers. A user's experiments can be retrieved at different stages of analysis and recreated. Provenance and repeatability of experiments will be enabled by recording service-based 'events' in process metadata within an EML document by the WPS services (Grady et al., 2011). The QGIS client and the VisTrails client will incorporate an EML reader that will then allow other scientists to be able to recreate experimental results on the same data. The Lifemapper EML reader for VisTrails is currently in early testing and the same library will be used for reconstructing workflows in QGIS.

A range and diversity experiment consists of an input layer set (both species range and environmental), a regularly spaced grid, the resulting PAM, randomized PAMs, and related environmental matrices and summary statistics matrices. A set of species input layers in the form of raster or vector species range maps are defined by the user with thresholds by for calculating presence and absence. Those data are cataloged as members of a layer set, then immediately available as OGC web services. Using these data the user is able to

²³<http://garyentsminger.com/ecosim>

²⁴EML: <http://knb.ecoinformatics.org/software/eml>

²⁵<http://openlayers.org>

²⁶<http://www.vistrails.org>

build the PAM by intersecting both species and selected environmental data layers against a geospatial grid. The PAM starts as an equal area grid defined by the user using a grid constructor WPS service. The service allows the user to define resolution, cell shape (square or hexagonal) and extent of the grid. WPS based intersection tools provide optimized intersection methods for intersecting multiple types of species and environmental inputs against the grid. Species data can be provided by the user, retrieved from repositories such as DataONE²⁷, or be chosen from a listing service that exposes species distribution model outputs from Lifemapper. Once species and environmental layers are assembled the PAM is populated with these inputs, where the presence and absence of a species are determined using the threshold parameters and are recorded as 0's and 1's in a binary matrix. A compression algorithm can be used to remove zero sum columns or rows insignificant to the matrix statistics and unnecessary for the sequential swap algorithms for randomizing the matrix. De-compression recreates the full PAM to re-randomize using different methods that require the entire geospatial extent or to add new species or environmental data to the matrix. Statistical algorithms for calculating range and diversity indices can then be performed on observed data or modelled data.

The Python NumPy²⁸ library built with the Basic Linear Algebra Subprograms (BLAS) library provides most of the matrix and statistical calculations against the PAM. Some of the diversity indices that can be derived from the presence/absence matrix include total richness or gamma diversity, local or alpha diversity, its distribution and average, and Whitaker's beta diversity (Arita et al. 2008, Borregaard and Rahbek, 2010). As a multi-species analysis covariance is central to the composition of the range-diversity plots derived from the PAM, including matrices of covariance of composition of sites, covariance of ranges of species, mean composition covariance, mean species covariance, and mean range covariance. The advantage of the PAM is that it combines information on range and diversity and co-occurrence of species measured by the degree of covariance in the matrix (Arita et al. 2008).

LmRAD will provide inputs to hypotheses dealing with assembly rules tested against species associations that keep marginal totals of occupancy and diversity intact while randomizing the presence absence data using Gotelli's recommendation for sequential swap algorithms. (Gotelli, 2001, Gotelli, 2000). Sequential swap algorithms are also the preferred method for generating null matrices in studies that analyze the degree of nestedness in a PAM. Nestedness is a structure found in the PAM where triangular submatrices of species presence 'nest' themselves into a corner of the PAM, this phenomenon, while debates range on how to measure its degree, represents the idea that less diverse niches represent subsets of more diverse niches (Brualdi and Sanderson, 1999).

A second algorithm for hypotheses-testing in LmRAD generates null models by using a dye dispersion algorithm, a 2-Dimensional geometric-constraints model that assumes range continuity. Range allocations are reassembled using a process of expansion where a species range is allowed to 'grow' to a prescribed size akin to dropping a volume of dye onto a 2-dimensional bounded surface (Jetz and Rahbek, 2001). Different methods of randomizing the PAM require

different compression states of the PAM. The swap algorithm leaves marginal totals for occupancy and diversity intact and requires searching the PAM for pairwise disjoints or checkerboard patterns with a random distance between its horizontal and vertical components. Columns or rows with no presences represent a zero sum marginal total and must be removed to increase the efficiency of the swap algorithm. This is achieved by a compression algorithm. By treating the PAM as a Boolean grid, bit-wise operations increase the efficiency of randomizing large PAMs. The Dye Dispersion algorithm for mixing of species ranges requires the PAM be uncompressed so a species range can grow within the geographic limits of the entire PAM. Statistical algorithms for calculating range and diversity indices are then performed on the randomized data.

LmRAD will provide several methods for dealing with user defined scales when constructing the PAM and comparison across PAM's. Spatial scale in PAM based studies includes both the extent of the area studied and the cell or quadrant resolution (Lira-Noreiga et al. 2007). Scale considerations are key in understanding biodiversity patterns, and are known to affect the patterns and processes that determine spatial patterns of species richness. Differences in scale directly affect estimates of average alpha diversity, gamma and beta. Whitaker's diversity measure and the fill of the matrix depend on the number of sites that are used describe the distribution of the species and therefore they are sensitive to the changes in the scale of analysis (Arita et al. 2008). LmRAD has been developed to accommodate changes in scale to enable a user to compare changes in diversity indices.

Informatics and Architecture

WPS on a compute cluster

The current Lifemapper architecture utilizes a set of RESTful web services written in Python for exposing environmental layers, modeling outputs, species occurrence points, and modeling services. When incorporating WPS into our existing architecture we wanted a Python centric solution and one that avoided Common Gateway Interface (CGI) implementations or implementations reliant on mod_Python and also one where processes can operate in a variety of different computing environments. Rather than adapt existing WPS implementations, LmRAD utilizes an autonomous job object paradigm for executing WPS requests. On the front end we use templating for document generation and a computational pipeline and PostgreSQL database to store asynchronous requests. The web architecture is generalized allowing us to switch in and out of web frameworks. The web server hosts a WPS module that registers all of the WPS services, exposes metadata classes for each one, and handles requests to individual processes, including GetCapabilities requests which return a description of all available processes. A metadata service module for each process registers inputs and outputs for the process, handles the DescribeProcess request and exposes an execute method. The execute method is responsible for entering the job into the database. Then the computational pipeline picks up the job, assembles a job object, serializes it and posts it to the head node of the cluster. On the cluster,

²⁷<http://www.dataone.org>

²⁸<http://numpy.scipy.org/>

the job is de-serialized and pickled so that the target node can request the job object from a web server on the head node. Worker threads simultaneously update experiment status and inputs, submit experiments to and retrieve results from the compute cluster. The head node of the cluster submits the job ID and job Type, and a pipeline ID to the Sun Grid Engine scheduler. The scheduler farms out the job to a node where a job factory initializes the appropriate job runner with the appropriate environment variables using classes of functions specific to the environment. One of these will be passed to a job runner when it is instantiated so that it knows where to store data, how to access data, and how to update the status. Job input requests for raster data for species layers are requested by the node using OGC Web Coverage Service (WCS) requests from mapping services that use MapServer (<http://mapserver.org/>) against the species distribution model archive in Lifemapper.

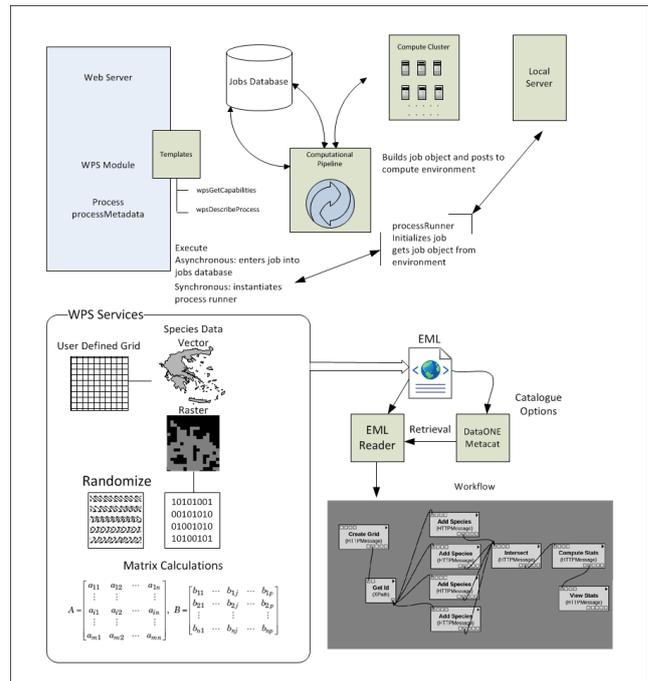


Figure 3: Overview of WPS and cluster architecture.

GIS operations, visualization and client architecture

The work performed by our WPS processes is done by a job runner. These job runners are not aware of the environment where they physically exist. This is possible by abstracting the methods used to interact with their environment. When the job runner is instantiated, it is given a class that includes methods for interacting with the environment. These environment methods implement a standard interface for interaction with the job runner. The job runner expects a standardized object of a particular type from each method. If the environment is one of our core machines, the job object might be constructed by accessing the database and building the object from that response. On a cluster node, an HTTP request is sent to the front end of the cluster that sends back a serialized version of the object that is in-turn de-serialized and returned to the job runner. In the testing environment, an XML file is accessed on the file system and then restored to an object. Each environment can perform the tasks through different methods as long as the post condition is the expected object. This also applies to other methods, e.g. methods that return a file path for writing temporary data or updating the job's metadata. The job runner is decoupled from its environment, which allows the jobs to be agnostic of where they run and also allows them to be flexible enough to move. To move a job to a different environment only the registration needs to change. If a job type is found to be too resource intensive on the core machines, it can be registered to run on the cluster and the performance bottleneck can be alleviated quickly and simply. Figure 3 outlines the LmRAD architecture.

GIS operations are exposed as WPS services through a set of dialogues within a Python plug-in for QGIS. The plug-in can be used for building the data grid for the PAM and intersecting species and environment layers and returning PAM statistics. Geometries for the cells in the PAM created as a shapefile using the OSGeo OGR²⁹ library and returned by the WPS service to QGIS for visualization. The shapefile is also stored on the processing node of the service instance for use in further processing. A user can then choose from existing predicted species distribution maps in Lifemapper provided by a listing service, or upload raster or vector species maps. The presence and absence of a species is then calculated by intersecting the grid with the species layer and that data recorded in a Numpy matrix, using the regular grid as a place holder for the geographic data. The intersection algorithms use R-tree spatial indexing³⁰, matplotlib³¹, and NumPy matrix manipulations with the OSGeo GDAL (GDAL 2010) and OGR libraries to achieve efficiency and speed of operation. The user may expand or reduce the study area and add or subtract species columns. Visualizations of these changes will provide some challenges and further work must be done to bring these into QGIS.

The QGIS client architecture uses a threaded architecture that uses the PyQt widget toolkit³² with signals and slots to implement the Observer pattern (Gamma, et al. 1995) in order to handle asynchronous WPS requests. The Lifemapper client library abstracts the communication layer away from specific model implementations for a client. A model (in Mod-

²⁹<http://www.gdal.org/ogr>

³⁰<http://pypi.python.org/pypi/Rtree>

³¹<http://matplotlib.sourceforge.net>

³²<http://www.riverbankcomputing.co.uk/software/pyqt/intro>

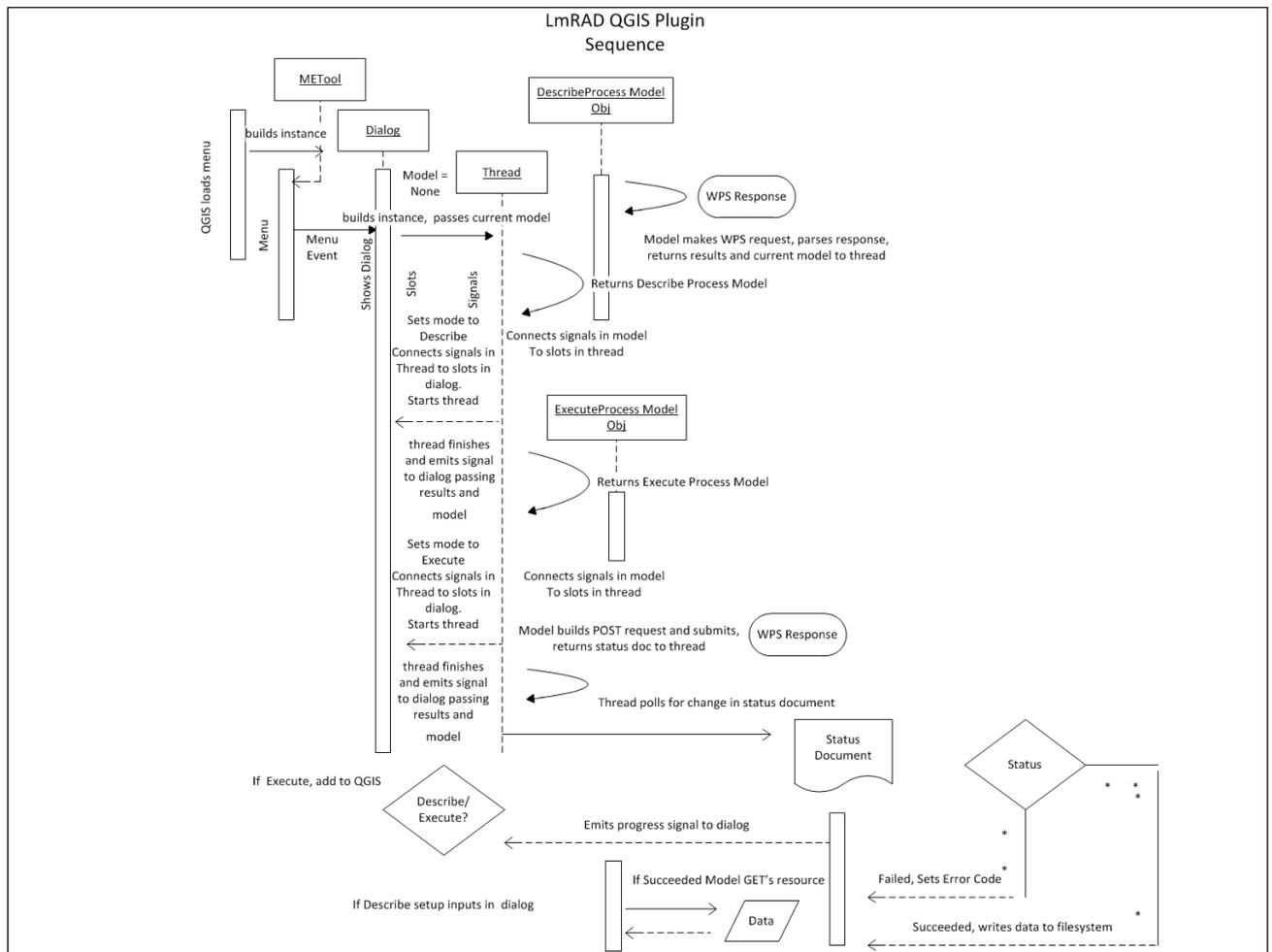


Figure 4: LmRAD QGIS Sequence Diagram for asynchronous requests

el/View/Controller) consists of a de-serialized Python object derived from the WPS XML, constructed using a recursive method in a wrapper to the Element Tree Python library. All dialogues for the suite of range and diversity tools inherit from a common set of dialogue classes, allowing different dialogues to be spun up readily. Figure 4 describes an asynchronous request from the QGIS client.

The Lifemapper client library, used for both LmSDM modeling and LmRAD experiments, is a Python module designed to interact with multiple clients to minimize code duplication. It communicates with the Lifemapper web services, both RESTful and OGC compliant. The library can be initialized as an anonymous user so that only public data from the Lifemapper web services may be accessed, or, the client may be initialized with user credentials allowing access to a private Lifemapper workspace. The client abstracts details of the exact interface to services and provides for a simpler interaction with the services for the developer. Hiding this underlying complexity does two things, first, application developers using the client library do not need to know the

details of different communication standards, and second, implementations can change without requiring modification to end-user software.

Statistical operations are exposed as WPS services through the plug-in in QGIS. Once an experiment has been setup to run using a constructed PAM, statistical services can be used to retrieve range and diversity indices derived from the PAM. Range-diversity plots are linked in QGIS using PyQt Widgets for Technical Applications (PyQt³³) allowing for 'brushing' of datasets by species or location across geographic space and in plot space and in the future, tree data space. Two different types of range-diversity plots map the relationships between the mean proportional species diversity against the proportional range size of a species; and the relationship between the mean proportional range size and the proportional species diversity. These two different 'views' of the data can then be linked for display in dendrograms in the case of the species specific plot and geographic space in the case of the locality specific plot.

³³<http://pyqt.sourceforge.net>

Findings, Considerations and Future Directions

This article has demonstrated the benefits of a distributed WPS approach within an SDI for macroecology tools. We have described our current efforts based on the needs described for working with PAM data which have become very large in current macroecology research efforts. We have addressed a need discussed in some of the literature (McFerren et al., 2010) for a library that can be used both from within QGIS and a scientific workflow environment with a common API for working with OGC standards. By building on an existing GIS platform we bring a more powerful set of tools to bear on macroecological problems and geospatially enable the scientific workflow environment software VisTrails, side stepping issues dealing with conflicting libraries by using WPS services.

Early results of this methodology have been very promising. Testing of the swap randomization algorithm on PAM matrices with cells and a very sparse fill provides one solution with 30,000 random searches for 15 suitable submatrices in less than one second. Increasing the search to 600,000 increases the number of swaps to about 155, with one solution found in 3.8 seconds. A more typical number of swaps is achieved with 12,000,000 searches on the dataset, with 3,617 submatrices found in 77.58 seconds. The vectors for the range-diversity plots are calculated against the same matrix in less than 3 seconds. Basic diversity indices are also being calculated but more computationally intensive matrix algebra algorithms still need to be developed and will benefit from being farmed out to the cluster environment. The intersection algorithms have also proved to be efficient. The current Python client library has been in use with the Lifemapper REST and OGC services for some time and has proved reliable. The implementation adjusted for WPS should prove to have the same reliability. Early testing on the library has been successful, and a future direction for it is to have it dynamically add functionality based on an online configuration document describing the services available in the Lifemapper system.

Deeper geostatistical functionality still needs to be developed. Future directions include interactive visualization for dendrograms or phentic trees in QGIS, along with visualization methods for quickly comparing changes in scale of the PAM. The methodology for handling asynchronous requests for WPS is not fully integrated into our computational pipeline. Range and Diversity experiments have been completely modeled in our database and the REST services retooled to account for WPS. The new WPS module has been used for range randomization in synchronous mode. Each compute environment will have its own computational pipeline, depending on the type of job, asynchronous jobs can be handled locally or on our compute cluster, but the WPS job object will be able to work in either environment, and we stress the agnostic nature of the WPS job for future implementation in a variety of environments. Work needs to be done to address the orchestration of jobs on the cluster which is currently handled in the database, this may require further research into WPS choreography or orchestration. A good deal of literature exists describing different approaches to WPS orchestration for chaining or nesting services and some within

Grid environments (Muller et al., 2010, Friis-Christensen et al. 2007, Meng et al. 2010, Diaz et al. 2010, Lanig and Zipf, 2009). Currently we leverage the high-throughput capacity of our compute cluster where one Lifemapper job runs on one node. Where possible jobs will be parallelized to run on multiple nodes using the Google MapReduce paradigm. Nodes on our cluster still must request data for potentially thousands of species range layers that can be pulled from repositories like DataONE, or our archives, but those data may also reside with a user, requiring them to upload all of their data inputs for the PAM. This represents a data-centric approach where data must be moved to the processing environment. The cluster environment provides future challenges and increased performance, that would allow us to address more of a 'moving code' approach described by Muller et al. 2010 to address the data access issue.

A different approach may prove beneficial if as summarized by Muller et al. 2010, algorithms frequently evolve, identical algorithms are shared among several service instances, and data can be shipped prior to execution, and algorithms have to be hosted at service instances close to the data to decrease bandwidth impact. (Muller et al. 2010). Data driven approaches are reasonable if one-time assembly and execution of workflows is the norm, and real-time response for complex service chains is not required (Muller et al. 2010). Michael and Ames, 2007 evaluate WPS for use within client-side GIS, and address similar concerns about moving data. They point out that WPS should not be used when operations on data can be completed more quickly locally than remotely, especially when factoring in the time to upload data, and download results. Conversely data should be sent to servers hosting service instances that have higher processing power when the time to process the data locally would be greater than the combined time to transmit the data. Similar to our approach they also note that building a WPS client into an existing GIS package takes advantage of the visualization and other GIS functionalities of the GIS. (Micheal and Ames, 2007). We were originally presented with performance issues with desktop software for operating against large PAMs. The processing time for individual PAMs fit the criteria for moving the data to a server with higher processing power. However, the data can also be local, and some outputs are best visualized and manipulated better in the local GIS client. Clearly a compromise has to be struck between the two approaches.

By employing WPS services with an established and extensible open source GIS platform we are beginning to provide intuitive, efficient and data rich biogeography tools for formulating biogeographical hypotheses and for simultaneously analyzing and visualizing species ranges and geospatial biological diversity patterns through range-diversity plots. A distributed approach for using WPS in different environments will allow current research practices of PAM assembly overcome large computational constraints and advance biodiversity research and education of the global impacts of climate change on species and biological communities.

Acknowledgements

We gratefully acknowledge the University of Connecticut (R. Colwell and T. Rangel), with co-funding from U.S. National Science Foundation Award BIO/DBI 0851290 and in

collaboration with Jorge Soberon, Narayani Barve and Andres Lira-Noreiga.

References

1. Arita, H.T., Christen, J.A., Rodriguez, P., Soberon, J. 2008. 'Species diversity and distribution in presence-absence matrices: mathematical relationships and biological implications.' *Am. Nat.* 172: 519-532.
2. Borregaard, M. K. and C. Rahbek, 2010. Dispersion fields, diversity fields and null models: uniting range sizes and species richness. *Ecography* 000: 000-000.
3. Brualdi, Richard A. and Sanderson, James G. 1999. 'Nested species subsets, gaps and discrepancy.' *Oecologia*, 119:256-264.
4. Connor, E. F., and D. Simberloff. 1979. 'The assembly of species communities: chance or competition?' *Ecology* 60:1132-1140.
5. Cepicky, J. Becchi, L. 2007. 'Geospatial processing via Internet on remote servers – PyWPS, OSGeo Journal 1 (May 2007): 5p
6. Diamond, J. M. 1975. 'Assembly of species communities.' in M.L. Cody and J.M. Diamond, (eds.), *Ecology and evolution of communities*. Belknap, Cambridge, MA.
7. Diaz, L., Pepe, M., Granell, C., Carrara, P., Rampini, A. 2010. 'Developing and Chaining Web Processing Services for Hydrological Models' in Commission IV, WG IV/5: WebMGS 2010.
8. Dubois, G., Skoien, J.O., de Jesus, J., Peedel, A., Hartley, A., Nativi, S., Santoro, M., Geller, G. 2011. 'eHabitat: a Contribution to the Model Web for Habitat Assessments and Ecological Forecasting.' in :Submitted to the 34th International Symposium on Remote Sensing of Environment, April 10-15, Sydney, Australia.
9. Ellison, A. A. and Aaron M. Ellison. 2000 Reviewed work(s): 'EcoSim: Null Models Software for Ecology.' *Bulletin of the Ecological Society of America*. Vol. 81, No. 2 (Apr., 2000), pp. 125-127
10. Friis-Christensen, Ostlander, Lutz, Michael, Bernard, Lars. 2007. 'Designing Service Architectures for Distributed Geoprocessing: Challenges and Future Directions.' *Transactions in GIS*, Vol. 11, No. 6. pp. 799-81
11. Gamma, Erich; Richard Helm, Ralph Johnson, and John Vlissides (1995). *Design Patterns: Elements of Reusable Object-Oriented Software*. Addison-Wesley. ISBN 0-201-63361-2.
12. GDAL, 2010 – Geospatial Data Abstraction Library: Version 1.6.3., Open Source Geospatial Foundation, <http://gdal.osgeo.org>.
13. Geller, G.N., Turner, W. 2007. 'The Model Web: a concept for ecological forecasting'. *Geoscience and Remote Sensing Symposium, IGARSS IEEE International*, 2469-2472, 23-28 July 2007.
14. Gotelli, N. J. 2000. 'Null model analysis of species co-occurrence patterns.' *Ecology* 81: 2606-2621.
15. Gotelli, N. J. 2001. 'Swap and fill algorithms in null model analysis: rethinking the knight's tour.' *Oecologia* 129:281-291.
16. Gotelli, N. J., and G. R. Graves. 1996. 'Null models in ecology'. *Smithsonian Institution*, Washington , DC
17. Simpson, G. G. 1964. 'Species density of North American recent Mammals'. *Systemic Zoology* 12:57-73.
18. Gotelli, N.J. and G.L. Entsminger. 2011. *EcoSim: Null models software for ecology. Version 7*. Acquired Intelligence Inc. & Kesey-Bear. Jericho, VT 05465. <http://garyentsminger.com/ecosim.htm>.
19. Grady, C.J., Beach, J., Cavner, J., Stewart, A. 2011. 'Lifemapper, VisTrails and EML: Documented, Re-executable Species Distribution Models'. in: *Proceedings of the Environmental Information Management Conference 2011 (EIM 2011)*, eds. M.B. Jones, C. Gries.
20. Graul, C., Zipf, A. 2008. 'Putting Biogeographic Research on the Spatial Web: Towards Interoperable Analysis Tools for Global Change Based on the Web Processing Service (WPS)'. *Digital Earth Summit on Geoinformatics: Tools for Global Change Research*. Potsdam, Germany.
21. Jetz, W. and Rahbek, C. 2001. 'Geometric constraints explain much of the species richness pattern in African birds.' *Proc. Nat. Acad. Sci. USA* 98:5661-5666.
22. Laffan, Shawn W. Lubarsky, Eugene and Rosauer, Dan F. 2010. 'Biodiverse, a tool for the spatial analysis of biological and related diversity.' *Ecography* 33:643-647
23. Lanig, S., Zipf, A. 2009. 'Interoperable Processing of Digital Elevation Models in Grid Infrastructures' in *Earth Sci Inform* 2: 107-116
24. Lira-Noreiga, A., Soberon, J., Navarro-Siguenza, A. G., Nakazawa, Y. and Peterson, A. Townsend. 2007. 'Scale dependency of diversity components estimated from primary biodiversity data and distribution maps.' *Diversity and Distributions*. 13, 185-195.
25. McFerren, G., van Zyl, T., Vahed, A. 2010. 'FOSS Geospatial Libraries In Scientific Workflow Environments: Experiences and Directions.' *Applied Geomatics*. 21 July 2010
26. Meng, Xiaoliang , Xie, Yichun, and Bian, Fuling. 2010. 'Distributed Geospatial Analysis through Web Processing Service: A Case Study of Earthquake Disaster Assessment.' *Journal of Software*, Vol. 5, No. 6, June
27. Micahel, Christopher, and Ames, Daniel P. 2007. 'Evaluation of the OGC Web Processing Service for Use in a Client-Side GIS.' *OSGeo Journal* Vol 1. May 2007
28. OGC 2007b *OpenGIS Web Processing Service, Version 1.0.0*. Wayland, MA, OGC Document No 05-007r7
29. Rangel, Thiago F., Diniz-Filho, Alexandre F. and Bini, Luis Mauricio. 2010. 'SAM: a comprehensive application for Spatial Analysis in Macroecology.' *Ecography*. 33: 46-50.
30. Skoien, J.O., Dubois, G., De Jesus, J. 2011. 'Forecasting Biomes of Protected Areas'. *Procedia Environmental Sciences* 4 (2011) 44-49.
31. Wang, S.S. and Lilu, D.Y. 2004. "Spatial Query Pre-Processing in Distributed GIS" in *Grid and Cooperative Computing*, – GCC, H. Jin, Y. Pan, N.Xiao, and J.Sun (Eds.), (Berlin; Springer) p.p. 734-744.

Jeffery A. Cavner
Software Developer
KU Biodiversity Institute
1345 Jayhawk Blvd.
Lawrence, KS 66045 U.S.A.
jcavner@ku.edu

This PDF article file is a sub-set from the larger
OSGeo Journal. For a complete set of articles
please the Journal web-site at:

<http://osgeo.org/journal>

and ready for the future. TRLIB was only recently released under an open source license (and made available through <https://bitbucket.org/KMS/trlib>), but in the near future we hope to implement means for better interoperability with the more well established libraries in the open source geomatics field.

Acknowledgements: We thank Willy Lehmann Weng and Knud Poder for commenting on the draft of this paper.

Bibliography

- C. W. Clenshaw. A note on the summation of chebyshev series. *Math. Tables Aids Comput.*, 9(51):118–120, 1955. URL <http://www.jstor.org/stable/2002068>.
- Karsten E. Engsager and Knud Poder. A highly accurate world wide algorithm for the transverse mercator mapping (almost). In *Proc. XXIII Intl. Cartographic Conf. (ICC2007), Moscow*, page 2.1.2, 2007.
- Gerald I. Evenden. Cartographic projection procedures for the unix environment—a user’s manual, 1990. US Geological Survey Open-File Report 90–284.
- C. Gram, O. Hestvik, H. Isakson, P.T. Jacobsen, J. Jensen, P. Naur, B.S. Petersen, and B. Svejgaard. Gier - a danish computer of medium size. *IEEE Transactions on Electronic Computers*, EC-12(5):629–650, December 1963. URL http://www.datamuseum.dk/site_dk/rc/gierdoc/ieeeartikel.pdf.
- Charles Karney. Transverse mercator with an accuracy of a few nanometers. *Journal of Geodesy*, 2011(1):1–11, 2011. doi: <http://dx.doi.org/10.1007/s00190-011-0445-3>.
- R. König and K. H. Weise. *Mathematische Grundlagen der Höheren Geodäsie und Kartographie, Erster Band*. Springer, Berlin/Göttingen/Heidelberg, 1951.
- Knud Poder and Karsten Engsager. Some conformal mappings and transformations for geodesy and topographic cartography, 1998. National Survey and Cadastre, Denmark, Publications, 4. series vol. 6.
- Carl Christian Tscherning and Knud Poder. Some geodetic applications of clenshaw summation. *Bolletino di Geodesia e Scienze Affini*, XLI(4):349–375, 1982.

Thomas Knudsen
National Survey and Cadastre
kms.dk
[thokn AT kms DOT dk](mailto:thokn@kms.dtu.dk)

Imprint - Volume 10

Editor in Chief:

Tyler Mitchell, Locate Press - info@locatepress.com

Proceedings Editor:

Gary Sherman, GeoApt LLC - gsherman@geoapt.com

FOSS4G Academic Committee:

Chair - Rafael Moreno Univ. of Colorado, Denver, USA

Thierry Badard Laval University, Canada

Maria Brovelli Politecnico di Milano campus Como, Italy

Serena Coetzee University of Pretoria, South Africa

Tyler Erickson Michigan Tech Research Institute, USA

Songnian Li Ryerson University, Canada

Jeff McKenna Geteway Geomatics, Canada

Helena Mitasova North Carolina State University, USA

Venkatesh Raghavan Osaka City University, Japan

Acknowledgements

Gary Sherman, L^AT_EX magic & layout support

Various reviewers & writers

The *OSGeo Journal* is a publication of the *OSGeo Foundation*. The base of this journal, the L^AT_EX 2_ε style source has been kindly provided by the GRASS and R News editorial boards.



This work is licensed under the Creative Commons Attribution-No Derivative Works 3.0 License. To view a copy of this licence, visit: <http://creativecommons.org/licenses/by-nd/3.0/> or send a letter to Creative Commons, 171 Second Street, Suite 300, San Francisco, California 94105, USA.



All articles are copyrighted by the respective authors. Please use the OSGeo Journal url for submitting articles, more details concerning submission instructions can be found on the OSGeo homepage.

Journal online: <http://www.osgeo.org/journal>

OSGeo Homepage: <http://www.osgeo.org>

Mail contact through OSGeo, PO Box 4844, Williams Lake, British Columbia, Canada, V2G 2V8



ISSN 1994-1897