Data Management Challenges in Species Distribution Modeling

Colin Talbert[1] Marian Talbert[1] Jeff Morisette[1] David Koop[2]
[1] U.S. Geological Survey, Fort Collins, CO
[2] New York University, NY

Abstract

An important component in the fields of ecology and conservation biology is understanding the environmental conditions and geographic areas that are suitable for a given species to inhabit. A common tool in determining such areas is species distribution modeling which uses computer algorithms to determine the spatial distribution of organisms. Most commonly the correlative relationships between the organism and environmental variables are the primary consideration. The data requirements for this type of modeling consist of known presence and possibly absence locations of the species as well as the values of environmental or climatic covariates thought to define the species habitat suitability at these locations. These covariate data are generally extracted from remotely sensed imagery, interpolated/gridded historical climate data, or downscaled climate model output. Traditionally, ecologists and biologists have constructed species distribution models using workflows and data that reside primarily on their local workstations or networks. This workflow is becoming challenging as scientists increasingly try to use these modeling techniques to inform management decisions under different climate change scenarios. This challenge stems from the fact that remote sensing products, gridded historical climate, and downscaled climate models are not only increasing in spatial and temporal resolution but proliferating as well. Any rigorous assessment of uncertainty requires a computationally intensive sensitivity analysis accounting for various sources of uncertainty. The scientists fitting these models generally do not have the background in computer science required to take advantage of recent advances in web-service based data acquisition, remote high-powered data processing, or scientific workflow systems. Ecologists in the field of modeling are in need of a tractable platform that abstracts the inherent computational complexity required to incorporate the burgeoning field of coupled climate and ecological response modeling. In this paper we describe the computational challenges in species distribution modeling and solutions using scientific workflow systems. We focus on the Software for Assisted Species Modeling (SAHM) a package within VisTrails, an open-source scientific workflow system.

1 Introduction

The objective of this paper is to demonstrate the utility of scientific workflow tools in addressing the increasingly complex data management issues associated with species distribution modeling (SDM). We start with a brief overview of SDM and the currently available software options. We then explain two major factors that are contributing to the increasing modeling complexity. First there is the growing array of SDM algorithms

Bulletin of the IEEE Computer Society Technical Committee on Data Engineering

available, each having its own parameters, input file formats, and native software. Here the data complexity resides in the metadata associated with the model runs, such as software version, parameter files, and input and output that are not necessarily tracked in the SDM software. Second there is the complexity of data access; which has become a particular challenge with the growing interest of projecting SDM results into the future by using climate projections. We go on to explain how scientific workflow software, in particular VisTrails:SAHM has helped ecological modelers deal with this complexity of the first factor and how we hope to use it to address the second. The organizational capacity and provenance inherent to VisTrails:SAHM serve both to tie previously disparate tools together and maintain detailed records of the input used, output generated, and parameter specifications and allows the modeling to take advantage of machine service architectures to construct models using predictor layers stored on remote servers. We have found that organizing SDMs within the context of scientific workflow simplify data and model complexities, thus allowing analysts more time to concentrate on the ecological implications and useful application of their species distribution modeling activities.

2 Species Distribution Modeling: Overview and Challenges

Species distribution models (SDMs) are used to link species locations to spatially defined environmental variables. There are several algorithms available for fitting species distribution models which vary widely in their origins and complexity from statistical algorithms such as classic generalized linear models (GLMs) [17] and generalized additive models (GAMs) [13] to machine learning algorithms such as random forest [4] and boosted regression trees [11] to network models such as artificial neural networks (ANNs) [19]. The primary focus of the SDM literature and of the SAHM software is on observation-based determination of the correlative relationships between an organism and its environment. Once quantified, these correlative relationships can be used to gain insight into a variety of important ecological and evolutionary questions and model output can be projected onto spatial layers to produce maps of the species' ecological niche. Unfortunately, a large number of sources contribute to uncertainty in the prediction of species distribution especially when projecting to new spatial or temporal extents and when the input layers are often themselves output from other models. Approaches capable of partially quantifying this uncertainty such as ensembles and maps or partitions of quantifiable uncertainty have been recommended in the literature [8]. Proper model assessment often involve a sensitivity analysis on several groups of factors including the models and the environmental data.

It has been argued that climate is often the most basic determinant of a species fundamental niche in that it limits the species' range at the broadest spatial scale [2]. There has been a rapidly growing interest in projecting species distribution models onto future climate scenarios to inform management decisions under a changing climate. In this arena, the quantification or exploration of uncertainty based on the various sources is critical and the most favorable approach involves producing species distribution maps for various combinations of SDM models, emissions scenarios, and coupled Atmosphere-Ocean Generalized Circulate Models (AOGCMs), downscaling methods, as well as any other known sources of uncertainty. This situation creates an explosion in computation time and the data storage requirements for fitting SDMs. One recent paper, for example, generated 8400 projections to describe the uncertainty in fish assemblage projections [5]. The modules we are developing within SAHM are designed to remove the programming and data storage challenges that limit the use of recommended practice for many ecologists.

The scientific data management challenges encountered in species distribution modeling stem from two main factors. First is the number of disparate tools required for a typical SDM workflow. The second stems from the difficulties encountered preparing and using the gridded environmental inputs. While neither of these factors is unique in computational science, certain aspects of SDM make it a good candidate for implementing tools and practices for robust scientific data management. First is the widespread acceptance and use of SDM methods for many ecological research and management questions. The second is lack of experience and training among many practitioners of SDM in the tools and techniques used for scientific data management. Additionally, with

SDM modeling the results can be used to make controversial decisions with respect to endangered or invasive species and proper documentation of the modeling workflow is helpful to review and defend results.

2.1 Integrating SDM Tools

Several tools are currently available for modeling species distribution. We were able to determine the most commonly utilized tools based on a recent survey of SDM software use by ecologists [15] and use this information to review the available tools. Due to the complex and multidisciplinary workflows required for SDMs it is unlikely that the entire workflow could be carried out in any single software package with the exception of possibly R, but this would require significant programming skill. Also, due to the various components, it seems practical that species distribution modeling requires expertise in climate models, remote sensing, GIS, the species being considered, and statistics techniques. Further, if the model output is to be applied to natural resource management questions, expertise in that management domain is also required. All of the software packages with GUI interfaces demonstrated limited flexibility in terms of either modeling for only one type of response, with one modeling algorithm, not providing opportunities to explore the data, or not producing detailed graphics for model evaluation. Most packages with GUI interfaces focused on the model fitting and evaluation components of the workflow but require a separate set of tools and expertise to perform the preprocessing or data layer preparation steps. These issues are only compounded for the ecologist hoping to understand the nuances of how models differ or hoping to map patterns in uncertainty or determine when the spatial projections are extrapolating beyond the environmental calibration space.

It is important to emphasize that the full SDM process with climate inputs will generally need to integrate a variety of tools at different points in the process. For example, a scientist might use a web interface to select and download the required climate inputs such as the USGS Geo Data Portal (GDP) [3]. She might then use a custom script to process them into the format required for their model such as a toolbox in ArcMap. The model is then run using a specific GUI tool such as Maxent. The steps to select the final model and analyze model performance might be done with custom code, perhaps written in R. The modeling is also likely to involve decisions and deliberations by the analyst(s); which may or may not be rigorously documented. The final model output might then be reformatted and visualized in a final GIS program such as ArcMap. This type of workflow often presents challenges with tractability, transportability, documentation of methodology, and repeatability.

2.2 Handling Input Data

A primary data need of species distribution modelers are the geospatially referenced gridded environmental data often referred to as layers which are used as inputs to SDMs. These data can include traditionally mapped GIS layers such as land cover, remotely sensed products such as those from Moderate Resolution Imaging Spectroradiometer (MODIS)[1], and the subset that we will focus for the remainder of this paper on historic and modeled climate products. These climate data can represent interpolations of historical readings such as PRISM (Parameter-elevation Relationships on Independent Slopes Model [12]) or modeled hindcasts or projections of future climate such as those produced under CMIP5 (Coupled Model Intercomparison Project Phase 5 [18]). The native formats and schemas of these data generally fall into a limited number of standards including NetCDF (network common data form) with CF (Climate and Forecast) metadata conventions and HDF5 (Hierarchical Data Format, Version 5). Due to the large data volumes inherent to global and high resolution downscaled regional climate models, the data often come tiled into discrete files either spatially or temporally. The size of individual climate datasets is generally reasonable but cumulatively they can become quite large (100s of GBs to 10s of TBs). Although these data generally adhere to common data schemas, scientists and analysts still often run into novel formats or nuances that require ad hoc tools and workflows. The ability to utilize these non-standard data formats is often beyond the capabilities of researchers without strong technical and programing skills.

These demands have led to the development of technologies aimed at facilitating data use by scientists through distributed storage and remote web service based access. Examples of these technologies include OPeNDAP, (Open-source Project for a Network Data Access Protocol) [22] , THREDDS (Thematic Realtime Environmental Distributed Data Services) [24], and Open Geospatial Consortium (OGC) Web Coverage Service (WCS)[14]. In general these technologies provide web service protocols for remote data access and discovery and provide capabilities for spatial and temporal grid subsetting. These data delivery protocols can be further extended by providing data processing services which leverage remote computational resources, for example, The OGC Web Processing Service (WPS) Interface Standard [14] or the USGS Geo Data Portal [3].

There are currently a number of excellent tools available for researchers who need to consume scientific data from these types of services. These include web-based interfaces, GUI tools, application specific toolkits, and APIs for a variety of programing languages and libraries (Matlab, R, Python, GDAL, etc). While these technologies have gained wide acceptance and use within the earth science community, their use is not currently widespread among ecologists and biologists. This might be partly due to the lack of support for these technologies in many of the current widely-used ecological modeling tools, including those for SDM. Integrating service-based climate data acquisition or processing currently involves either custom code to obtain the data or a multi-tool workflow that can be cumbersome to run and document. By better integrating these technologies with existing tools and workflows, the process could be streamlined and standardized while at the same time realizing better computational efficiencies for species distribution modelers as well as increasing the integration of climate science into ecology and natural resource management.

It is important to note that the temporal format, i.e., discrete daily or monthly time steps, of most climate data must generally be summarized prior to its use in SDMs. This temporal summarization can involve one of several common algorithms, for example, the 19 Bioclimatic (BioClim) variables which capture annual temperature and precipitation patterns needed to model species distributions. Alternately modelers might need custom climate temporal summary algorithms based on the life history of the species being modeled. The flexibility required in generating these custom summaries precludes pre-calculating and caching the intermediate results. But at the same time these summarizations result in data volume reduction by several orders of magnitude.

3 Scientific Workflows and Provenance for Species Distribution Modeling

Scientific workflow systems provide a context in which to specify computational processes which integrate existing applications according to a set of rules [6]. Within this context, scientific workflows allow researchers to model complex analysis processes at various levels of detail and systematically capture the required information necessary for reproducibility and sharing; collectively referred to as workflow ”provenance” [7, 21]. When applied to species distribution modeling, scientific workflow software provides an opportunity to integrate existing SDM modeling routines seamlessly with software to ingest data, preprocess those data, and visualize model results.

3.1 The VisTrails:SAHM System

One current implementation of SDM integrated with scientific workflow systems is the SAHM package which runs within VisTrails [9, 10, 25]. VisTrails provide an approachable and extendable graphical user interface, and also provides sophisticated tools for workflow provenance capture and visualization. VisTrails is written in Python, a programing language which has many powerful and mature packages for scientific data processing including SciPy, NumPy, and matplotlib. Additionally there are efforts underway to enable complex climate data acquisition, processing, and visualization in VisTrails, namely the VisTrails Package UV-CDAT (Ultrascale Visualization - Climate Data Analysis Tools) [20, 25].

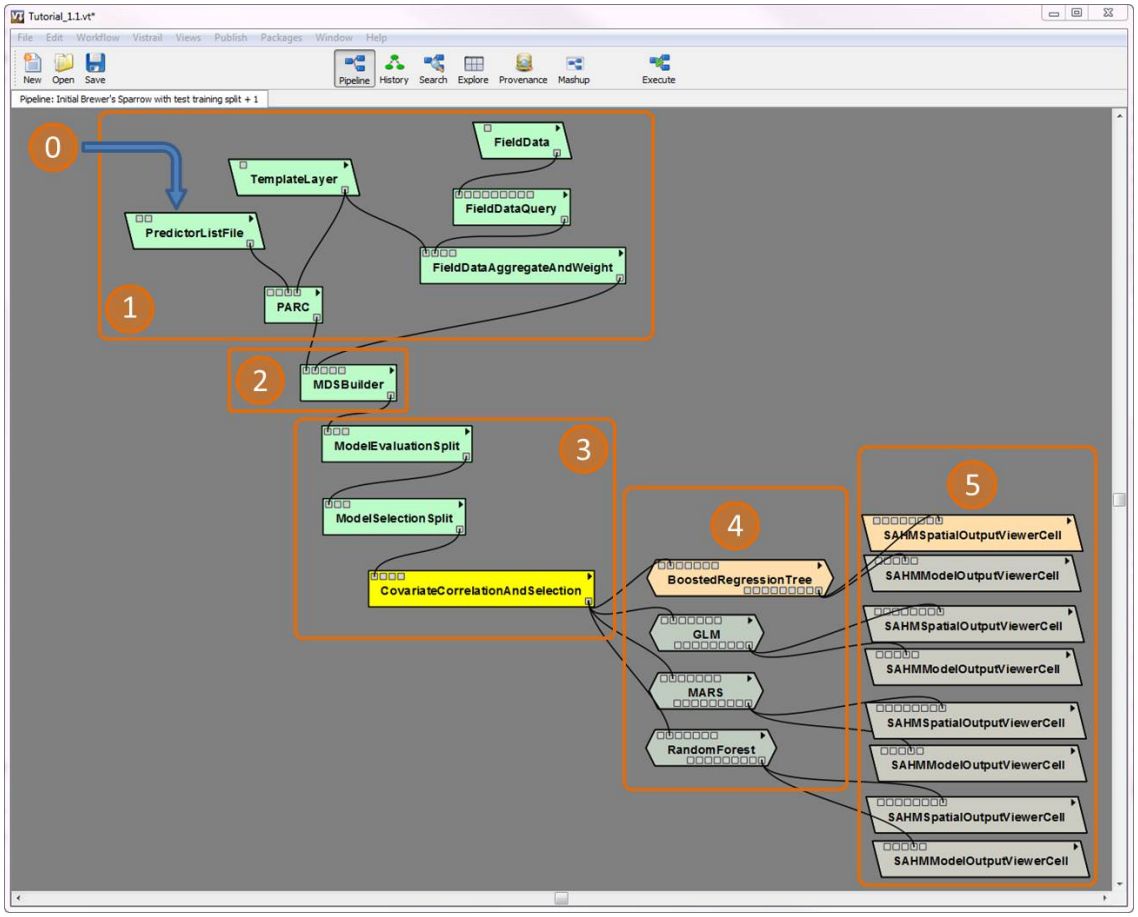

Figure 1: The VisTrails:SAHM User Interface. 1. preprocessing and data clean up, 2. merging input observational data with available covariates, 3. preliminary data exploration, 4. fitting correlative models, 5. model evaluation, comparison, and selection. Item 0 in the figure represents the remote data access and climate data summarization which is not currently implemented in VisTrails:SAHM

SAHM was developed as a VisTrails package to expedited the process and formalize the disparate tools required to build complex SDM workflows. Developing our habitat modeling software within VisTrails offered several advantages over previous tools. We are now able to develop template workflows that ecologist can use as starting points in their own analysis. Modifications to workflows are accomplished by dragging and dropping new modules onto the canvas and connecting these to the existing workflow components rather than by modifying scripts. These modifications can be annotated and tracked through the history view and we can easily compare differences between workflows. Within this framework much of the complexity of data management and multiple tool workflow integration and are hidden from the user. By hiding this complexity we were able to facilitate the use and coordination of fairly disparate technologies currently used in SDM. Fortunately, many of the most commonly used tools for SDM are script-based (Matlab, R, Python), have a scripting API (ArcMap, IDL), or can be run from the command line (Maxent and GDAL), which facilitates their incorporation into scientific workflow software. Figure 1 shows an example SDM workflow in VisTrails:SAHM and will be used here to highlight the complexities of 1) combining multiple analysis tools and 2) ingesting and handling a large array of potential predictor layers. (Additional details on this example workflow and the VisTrails: SAHM package are given in Talbert and Talbert [23]). The tools incorporated in SAHM can be broken loosely into five components (see figure 1): 1. preprocessing and data clean up, 2. merging input observational data with available

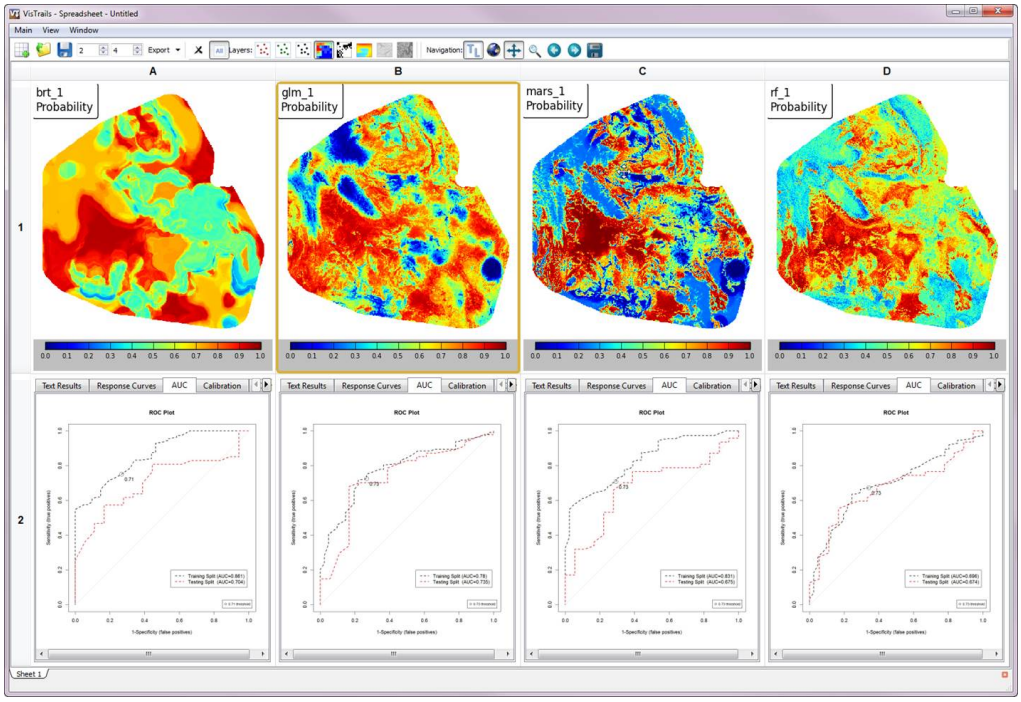

Figure 2: VisTrails:SAHM Output displays and compares continuous and binary maps as well as pertinent evaluation metrics from several models

covariates, 3. preliminary data exploration, 4. fitting correlative models, 5. model evaluation, comparison, and selection. Each module within the SAHM package allows the user to customize the step to meet their need and many modules allow the user to visualize and explore the data allowing subsequent decisions to be informed by previous modules in the workflow.

Another advantage to integrating current SDM tools in the VisTrails platform is the ability to incorporate coordinated visualization of multiple model outputs from multiple model runs. Figure 2 show the VisTrails:SAHM output from four model runs displayed in the built-in VisTrails spreadsheet. Output content is synced across multiple cells to facilitate model comparison. While many other SDM tools provide visualization options the ability to easily organize and traverse numerous outputs is unique to this platform.

4 Future Work

VisTrails:SAHM is a first step towards helping ecologists use increasingly complex SDM workflows and tool sets while also maintaining the crucial provenance and repeatability needed for defensible science. VisTrails also has the potential to provide tractable solutions for the current and emerging input data access and computation demands that SDM scientist encounter. The scope and context of each SDM research question can require specific data management strategies to maximize data acquisition and processing efficiency. These strategies are dependent on several factors including the spatial and temporal extent of the study, the number of different species being modeled, how many iterations will be run, the downloading and processing capacity available locally, and the technical expertise of the researcher. In this section we present four use cases which demonstrate different strategies for handling data acquisition and processing. The first strategy presented, maintaining a local copy of the input data, is the one most often used by species distribution modelers currently. It is presented here as basis for comparison with the other three.

4.1 Input Data Complexity

The traditional model for data management in SDM involves downloading or otherwise obtaining the required gridded spatial inputs and then preprocessing and storing them on a local computer or network (see Figure 3). Once the store of local data has been created there is maximal flexibility in how the data are processed. Given adequate computational resources locally, the processing of large areas, numerous SDMs, and multiple iterations becomes feasible. One drawback of this strategy is that the effort to store and process the data locally can be substantial and the disk space required to maintain local copies of the data can be limiting. Each time a new set of climate models is released, this data management effort will need to be duplicated in full. Another drawback is that the resulting workflow is not transferable to researchers at other locations unless they also obtain an identical copy of the input datasets and directory structure. This can limit repeatability of individual experiments. Using scientific workflow software to manage the initial input data acquisition and preprocessing can solve issues related to transferability of the workflow but does not overcome limitations in data acquisition bandwidth, storage, or processing limitations.

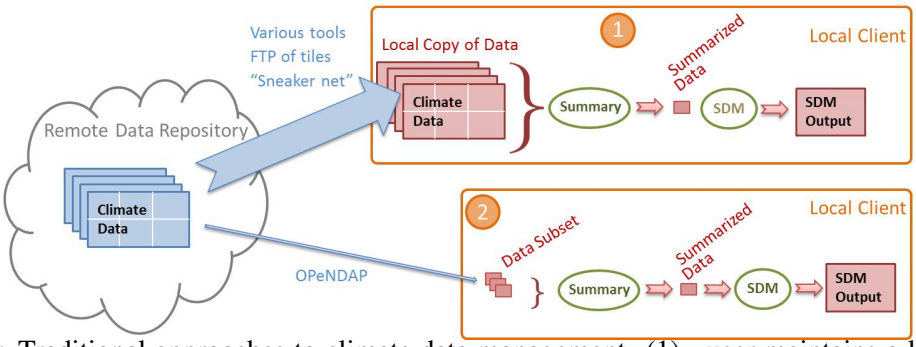

Figure 3: Traditional approaches to climate data management. (1) - user maintains a local copy of input data, (2) - small subsets of data are accessed remotely as needed using OPeNDAP.

A second approach to data acquisition in SDM is to obtain just the spatial and temporal subset of the gridded data needed for a given analysis. This is illustrated in Figure **??**. Many climate and remote sensing data are available in a tiled format which allows a user to download just the individual tiles needed. Identifying and obtaining the appropriate tiles can be cumbersome depending on the distribution method and available tools. One approach to facilitate getting a subset of the data is to use a service-based interface to specify the data needed, for example, WCS or OPeNDAP. While these technologies might be unfamiliar to many ecologists their implementation details can be hidden within user-friendly SDM applications. Implemented correctly, this tool could automatically download and process the appropriate data based on the spatial and temporal extent of the other SDM inputs. This type of solution works especially well for research questions on a local to regional scale. By implementing this strategy in a scientific workflow-based tool, the inherent complexity of the current tools can be hidden from researchers using SDM.

One concern with implementing a tool that automates service-based distribution of subsets of climate data is that it can lead to inefficiencies and duplicated data downloads if caching is not handled appropriately. For example, a single routine to produce a gridded climate summary might start with a call to a service to obtain the required input data. Once the appropriate summary has been generated this routine might delete the downloaded climate inputs to free up disk space. If there is only a single summary to produce, this strategy works fine, but if there are numerous summaries that need to be produced they should not each be downloading the required inputs redundantly. Within a single workflow, this level of optimization is relatively easy to set up nut across workflows and tools this data management can be much harder to manage. Optimally a tool would implement a persistent cache of previously downloaded inputs with a database that tracks multiple downloads of overlapping

but non-coincident spatial extents, total cache size, and differences between the cache contents and the data available from the data server [16].

Since SDMs generally require only a temporal summary grid of individual climate models, a third approach to data acquisition is to use a web service to deliver only this summary grid. Given that climate models can have hundreds to thousands of daily or monthly time steps, this strategy can represent several orders of magnitude reduction in data volume that must be downloaded. Some climate summaries such as decadal Bioclimatic Variables (Bioclim) are sufficiently general that they can be generated ahead of time and delivered using standard web data services such as WCS and OPeNDAP. Unfortunately, many SDM research questions require novel climate summaries which cannot be generated ahead of time. Using a Web Processing Service allows for these summaries to be calculated as needed on a remote server. If sufficient computational resources are available on the server hosting the remote processing, there is also the potential to realize a significant reduction in computation time versus calculating these summaries locally. One example of this type of WPS is currently being developed at the USGS Geo Data Portal (GDP) [3]. This WPS accepts custom parameters for the BioClim algorithms. These parameters are used to generate a server side processing job which produces and returns the desired custom summary grids. The ability of this strategy to scale to non-trivial problems given current infrastructure has yet to be tested. But potentially it could extend the subset of SDM research problems handled by the second option above to those covering a much larger spatial extent. If the processing was happening locally to the data many of the issues of data caching would not be a problem.

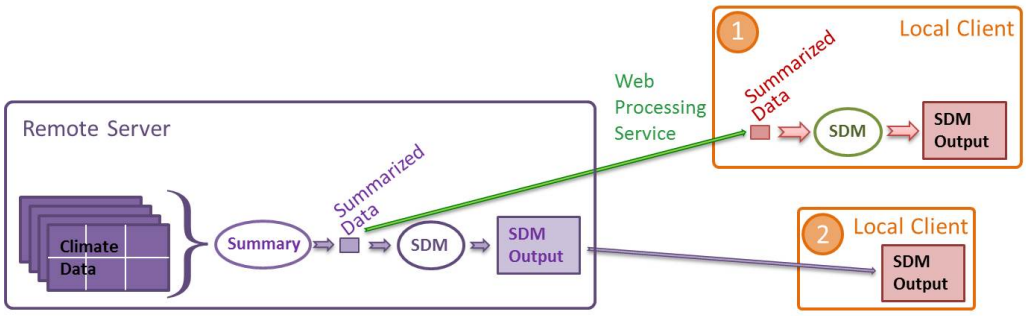

Figure 4: OGC Web Processing Service delivering custom climate summaries (1) or complete SDM model results (2)

A fourth approach to data management in SDM is to transfer even more of the processing workflow to a remote server. Instead of delivering custom climate summaries it is possible to engineer a WPS which runs an entire SDM workflow and returns the final model outputs. In this case, the species occurrence data becomes a parameter that is transfered to the remote server hosting the WPS along with detailed processing instructions. An advantage to having the full workflow details captured by a workflow management software such as VisTrails is the ease by which a detailed complex workflow can be packaged and transported. In this case, all relevant information about the workflow is stored in a compressed XML document. By scaling the data storage and computation resources available to the WPS it is feasible to use this strategy to solve many SDM data access issues. Implementing this solution has significant challenges though, including the trade off between computational flexibility and security on the remote server, managing large data hosted at multiple sites, and various other implementation details, scientific workflow software can be useful in in resolving some of these issues and hiding the unnecessary details from scientists.

5 Conclusions

Within the field of SDM the use of service based data acquisition and processing technologies is currently underutilized. This partly stems from the fact that the tools often require programing skills to be utilized effectively and partly because the need to work with large and unwieldy climate data is still emerging in the field of SDM. As more and more research questions require analysis of climate data the integration of these technologies will necessarily become more widespread. Given that ecologists and biologists often do not have the strong background in scripting required there is a developing need to extend current tools to take advantage of these technologies. In order to gain widespread usage in the community any tools or workflows developed should be user-friendly, transparent, and flexible. By building this next generation of SDM tools in a scientific workflow management system we can achieve this goal while also adding provenance capture, repeatability, and transferability.

Our development efforts have so far focused mainly on porting and optimizing exiting SDM tools for the VisTrails framework. This effort has been well received within the community. We are currently working to extend this framework to take advantage of data provided via a web service as well web processing services. The final phase of our work will extend this further to develop the four data acquisition approaches in which the vast bulk of the data management and processing is hosted remotely on high performance systems.

Acknowledgments: This research is funded by the U.S. Geological Survey and NASA's Applied Sciences Program. Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

References

[1] National Aeronautics and Space Administration. Moderate resolution imaging spectroradiometer data. https://lpdaac.usgs.gov/products/modis_products_table. Accessed: 2013-10-29.

[2] M. B. Araújo and A. T. Peterson. Uses and misuses of bioclimatic envelope modeling. *Ecology*, 93:1527–1539, 2012.

[3] D. Blodgett, N. Booth, T. Kunicki, J. Walker, and J. Lucido. Description of the U.S. Geological Survey Geo Data Portal data integration framework. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, 5(6):1687–1691, 2012.

[4] L. Breiman. Random forests. In *Machine Learning*, pages 5–32, 2001.

[5] L. Buisson, W. Thuiller, N. Casajus, S. Lek, and G. Grenouillet. Uncertainty in ensemble forecasting of species distribution. *Global Change Biology*, 16(4):1145–1157, 2010.

[6] A Dedeke. Workflow management: Models, methods, and systems., 2004.

[7] Ewa Deelman, Tevfik Kosar, Carl Kesselman, and Miron Livny. What makes workflows work in an opportunistic environment? *Concurrency and Computation: Practice and Experience*, 18(10):1187–1199, 2006.

[8] J. A. Diniz-Filho, R. D. Loyola L. M. Bini, T. F. Rangel, C. Hof, D. Nogués-Bravo, and M. B. Araújo. Partitioning and mapping uncertainties in ensembles of forecasts of species turnover under climate change. *Ecography*, 32(6):897–906, 2009.

[9] J. Freire, D. Koop, E. Santos, C. Scheidegger, C. Silva, and H. T. Vo. *The Architecture of Open Source Applications*, chapter VisTrails. Lulu.com, 2011.

[10] Juliana Freire, Claudio T Silva, Steven P Callahan, Emanuele Santos, Carlos E Scheidegger, and Huy T Vo. Managing rapidly-evolving scientific workflows. In *Provenance and Annotation of Data*, pages 10–18. Springer, 2006.

[11] J. H. Friedman, T. Hastie, and R. Tibshirani. Additive logistic regression: a statistical view of boosting. *Annals of Statistics*, 28:337–407, 2000.

[12] PRISM Climate Group. Prism climate data. http://www.prism.oregonstate.edu/. Accessed: 2013-10-29.

[13] T. J. Hastie and R. J. Tibshirani. *Generalized additive models*. London: Chapman & Hall, 1990.

[14] Open Geospatial Consortium Inc. Web processing service. http://www.opengeospatial.org/standards/wps. Accessed: 2013-10-29.

[15] L. N. Joppa, G. McInerny, R. Harper, L. Salido, K. Takeda, K. O'Hara, D. Gavaghan, and S. Emmott. Troubling trends in scientific software use. *Science*, 340(6134):814–815, 2013.

[16] David Koop, Emanuele Santos, Bela Bauer, Matthias Troyer, Juliana Freire, and Cláudio T. Silva. Bridging workflow and data provenance using strong links. In *Proceedings of the 22Nd International Conference on Scientific and Statistical Database Management*, SSDBM'10, pages 397–415, Berlin, Heidelberg, 2010. Springer-Verlag.

[17] P. McCullagh and J. A. Nelder. *Generalized linear models (Second edition)*. London: Chapman & Hall, 1989.

[18] WCRP World Climate Research Programme. Cmip - coupled model intercomparison project - overview. http://cmip-pcmdi.llnl.gov/index.html. Accessed: 2013-10-29.

[19] B. D. Ripley. *Pattern recognition and neural networks*. Cambridge Univ. Press., 1996.

[20] E. Santos, J. Poco, Yaxing Wei, Shishi Liu, B. Cook, D.N. Williams, and C.T. Silva. Uv-cdat: Analyzing climate datasets from a user's perspective. *Computing in Science and Engineering*, 15(1):94–103, 2013.

[21] Carlos Scheidegger, David Koop, Emanuele Santos, Huy Vo, Steven Callahan, Juliana Freire, and Claudio Silva. Tackling the provenance challenge one layer at a time. *Concurrency and Computation: Practice and Experience*, 20(5):473–483, 2008.

[22] Open source Project for a Network Data Access Protocol Inc. Opendap. http://www.opendap.org. Accessed: 2013-10-29.

[23] M. Talbert and C. Talbert. Tuturial for the software for assisted habitat modeling. https://www.sciencebase.gov/catalog/folder/503fbe63e4b09851b69ab463. Accessed: 2013-10-29.

[24] Unidata. Thredds data server. http://www.unidata.ucar.edu/software/thredds/current/tds/TDS.html. Accessed: 2013-10-29.

[25] VisTrails. http://www.vistrails.org.