Enabling online geospatial isotopic model development and analysis

Hyojeong Lee[1], Lan Zhao[1], Gabriel J. Bowen[1], Chris C. Miller[1], Ajay Kalangi[1], Tonglin Zhang[1], Jason B. West[2]

1: Purdue University, West Lafayette, Indiana, USA

2: Texas A&M University, College Station, Texas, USA

{lanzhao}@purdue.edu

ABSTRACT

In recent years, there has been a rapid growth in the amount of environmental data collected over large spatial and temporal scales. It presents unprecedented opportunities for new scientific discovery, while in the same time poses significant challenges to the research community on how to effectively identify and integrate these datasets into their research models and tools. In this paper, we describe the design and implementation of IsoMAP - a gateway for Isoscapes (isotopic landscapes) modeling, analysis and prediction. IsoMAP provides an online workspace that helps researchers access and integrate a number of disparate and diverse datasets, develop Isoscapes models over selected spatio-temporal domains using geo-statistical algorithms, and predict maps for the stable isotope ratios of water, plants, and soils. The IsoMAP system leverages the computation resources available on the TeraGrid to perform geospatial data operations and geostatistical model calculations. It builds on a variety of open source technologies for GIS, geospatial data management and processing, grid computing, and gateway development. The system was successfully used to teach a tutorial in the 2011 conference on the Roles of Stable Isotopes in Water Cycle Research. A post-tutorial survey was conducted. We review the users' feedback and present a future development plan based on that.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Online Information Services – *Data sharing, Web-based services.* H.5.2 [**Information Interfaces and Presentation (e.g., HCI)**]: User Interfaces – *User-centered design.*

General Terms

Design, Reliability, Performance.

Keywords

Isoscapes, TeraGrid, Geospatial data processing, Dynamic map rendering, Geostatistical model, Stable isotope distribution.

1. INTRODUCTION

In the past decade, there has been a rapid growth in the amount of environmental data collected over large spatial and temporal scales. Multiple continental-scale data-gathering networks are in development, build-out, or operational stages, including the National Ecological Observation Network (NEON), EarthScope,

CLEANER, and the WATERS network [14, 4, 3, 20], and will further add to this data deluge. The wide availability of station-based and continuous gridded environmental data presents great opportunities for scientific discovery in a number of biological and environmental disciplines. On the other hand, researchers face a critical challenge on how to effectively identify, access, and utilize the vast amount of data available. Furthermore, these datasets often come from distributed sources with different geospatial and temporal attributes and non-uniform metadata. It is not easy for a researcher to integrate these diverse, multidisciplinary datasets into their existing modeling tools in order to solve new ecological and environmental problems.

Such data challenges call for the development of cyberinfrastructure (CI) systems that support seamless integration of spatial data and modeling tools by automating the tasks of data access and transformation, which are time consuming and require familiarity in that area. There are many existing CI systems that enable integrated data and application access through a web interface, including GEON (GEOscience Network), LEAD (Linked Environments for Atmospheric Discovery) portal, QuakeSim, and NEES (Network for Earthquake Engineering Simulation) [15, 13, 16, 17]. Each of these gateways differs in the specific scientific domain to which they cater. The LEAD portal enables users to run automated workflows including data assimilation, atmospheric model simulation, data mining and visualization using the large computational resources provided by the TeraGrid. The OpenEarth Framework developed by GEON 2.0 project provides software services and tools to facilitate the integration and sharing of a variety of multi-dimensional data for geosciences researchers.

An area of emerging CI need exists in the field of environmental isotope chemistry. Isotopes (both stable and radiogenic) are powerful, inherent tracers of environmental processes, and their use has been rapidly growing, from their initial development in the geosciences to new disciplinary and interdisciplinary applications in biology, anthropology, environmental science, and forensic science. Isotope analysis is technically challenging, and much of the growth in isotope science has been driven by the continuous evolution and improvement of analytical instrumentation. In the past 5 years the field has experienced a step change as a new class of more stable, less expensive laser spectroscopy-based instruments have emerged that will further accelerate the proliferation of isotope data. Together with the adoption of stable isotope measurements core data products of new monitoring networks [14, 20] there is an existing and growing need for CI tools that allow the access, integration, and analysis of large isotope datasets in both disciplinary and interdisciplinary contexts.

The need for and power of web-based isotope modeling resources has been demonstrated by the success of a simple geostatistical modeling package provided through the WaterIsotopes.org website. This resource was an outgrowth of our early work with statistical modeling of spatial water isotope patterns in the environment [2], and featured a simple form-based query to an interpolation algorithm. The algorithm operated on a database of monitoring station observations to return isotope ratio predictions for environmental waters at any location on Earth. Over time somewhat more sophisticated interfaces were developed, including script allowing queries to be run through Google Earth, and the site now receives more than 1,200 visits and 1,000 prediction requests per month. The site has made unique contributions to science and society, being referenced in more than 80 publications and engaging new communities that have utilized the intuitive, map-based framework represented by the WaterIsotopes.org concept to access isotope data. Despite this success, the content of WaterIsotopes.org is limited and static, and the site is not conducive to providing a gateway for access to and analysis of the coming generation of high-volume environmental isotope monitoring data.

As an effort to address the data challenges for the growing isotopic research community, we have developed IsoMAP – a gateway for Isoscapes modeling, analysis and prediction. The long term goal of the IsoMAP project is to provide powerful models and tools which transform the way that scientists can interact with the data and conduct their research in the isotopic science community. IsoMAP features four focus application areas: precipitation isotope ratios, surface water isotope ratios, plant water isotope ratios, and plant tissue hydrogen, oxygen, and carbon isotope ratios. The system serves as an online workspace that enables scientists to easily access and integrate a number of disparate and diverse datasets and use them to develop and implement statistical and process-based models for stable isotope distributions in water, plants, and soils. IsoMAP is a GIS-enabled platform that provides services for end-to-end isotopic data and modeling workflows ranging from geospatial data identification, transformation, and integration, to geostatistical model development, assessment, prediction and visualization. It builds on existing CI software and uses TeraGrid computation resources for model execution and raster data processing.

The first version of IsoMAP system was released in March, 2011. The main contributions of the IsoMAP system include (1) automation of geospatial data processing which requires technical expertise and is error-prone; (2) provision of a set of spatial and geostatistical tools for isotopic model assessment, prediction, and data analysis; (3) dynamic visualization and dissemination of spatial data products; and (4) comprehensive metadata tracking and cataloguing for models and datasets.

The rest of this paper is organized as follows: Section 2 provides an overview of the IsoMAP system as well as design issues involved. Section 3 describes implementation details including geospatial data processing, modeling, rendering, and metadata documentation. A use case study is described in section 4. Section 5 discusses the future plan and concludes the paper.

2. IsoMAP System Overview

In this section we first describe an end-to-end workflow that allows users to use IsoMAP as a flexible geospatial data synthesis and model development environment. Then, we present the overall system architecture. At the end, we describe the data and metadata used by the IsoMAP system and several design challenges and requirements involved.

2.1 IsoMAP workflow

A typical isotopic model development, evaluation, and prediction workflow includes four steps as shown in Figure 1: develop a model, explore models, generate a map, and display maps. We will describe each step in details next.
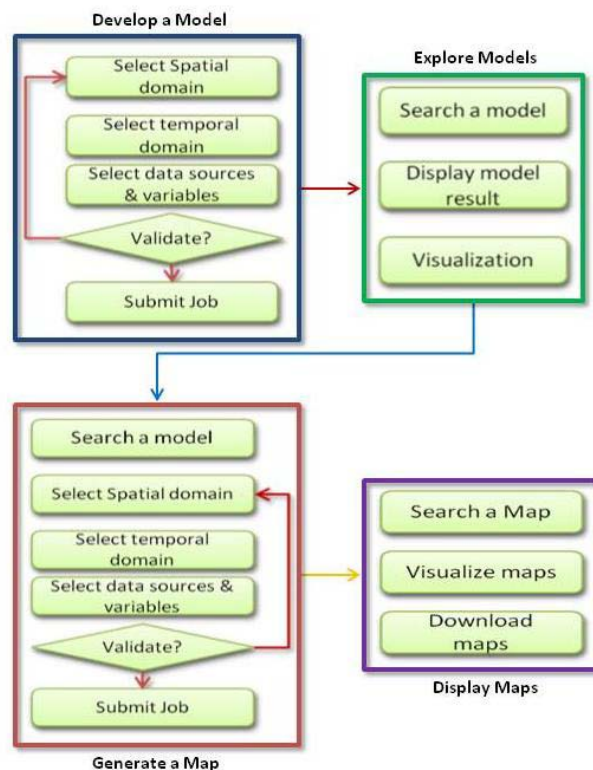

Figure 1. IsoMap workflow

2.1.1 Step1: Develop a Model

The goal of this step is to dynamically develop an isotopic model based on a user's project-specific requirements. A user can interactively set up a number of model parameters including the spatial and temporal ranges, independent variables, transformation formula for each variable, as well as the corresponding data sources. The gateway continuously monitors the input selections entered by the user, and only returns valid selection options for the next step. For example, after a user selects the spatial and temporal ranges, the gateway dynamically checks its backend database and only presents a set of variables and data sources that are valid for the selected spatial and temporal domains. Once a user completes the selection of model parameters, the gateway checks whether there are enough valid isotopic station data points that meet the user's requirement and may be used to develop the model. This process is known as *pre-validation*. If a predefined validation criterion (as described in section 3.1.2) is met, the user can move on and submit a model development job to the TeraGrid. The job constructs two models – with (geostatistical) and without (regression) consideration of spatial effects - automatically fitting all model parameters to the observational data selected by the user.

2.1.2 Step2: Explore Models

At this step, a user can browse and assess the result of the models he has developed. The user may track the status of his jobs from the gateway interface. After a job is completed, he can explore the model fitting result including parameter values, mean squared error, coefficient of determination, Moran's I as well as visualizing residual and Q-Q plots for the regression and geostatistical models. The provenance information of the model construction process is recorded in a job-specific metadata XML file and indexed for easy access.

2.1.3 Step3: Generate a Map

After developing a model with satisfactory results, the user may apply that model to map stable isotope ratios. He may use the same spatial and temporal domains and datasets as those selected when developing the model. Otherwise, he has the option to change these inputs. A pre-validation check is performed again after the user completes his selection. If successful, the gateway submits a map prediction job to the TeraGrid. When the prediction job completes, several stable isotope distribution maps will be generated.

2.1.4 Step4: Visualize Maps

At the final step, the user may search for a map prediction job and then dynamically visualize the map products online. He can also download map products in standard raster grid data format together with complete job metadata that describes how the map is generated. These data products enable researchers to solve problems in biogeochemical, ecological and anthropological systems in new and innovative ways.

In addition to providing a platform and tools for isotopic model and map development, another goal of IsoMAP is to serve as a repository of these resources. As part of the workflow, a user is able to publish a model or map he developed and share with other users. As a result, it is important to be able to track the metadata and provenance of the models and maps developed by the system so that it is easy for others to discover and use these resources in their scientific studies or publications.

This end-to-end workflow is the primary usage scenario that drives the design and implementation of the IsoMAP system.

2.2 System Architecture

The overall system architecture for IsoMAP is shown in Figure 2. It includes four layers. At the bottom are the computation and data resources. Model input data are stored either in a community software area on the TeraGrid or in a PostGIS spatial database. Geospatial data transformation and model computation are performed on TeraGrid cluster nodes provided by Purdue resource provider. In the middleware layer, we use a number of existing software applications for metadata cataloguing, indexing, geospatial data operations, visualization, and grid job submission/management. On top of the middleware is a group of software modules developed to directly support the services provided by the gateway. The gateway software components are implemented in a modular way so that some common functionality such as pre-validation and spatial data processing may be reused and shared by multiple workflows. The gateway interface was developed using GridSphere portal development framework [9]. AJAX, jQuery library [12] and OpenLayers map clients are used extensively to produce a highly interactive GIS-enabled user interface. Currently users can perform the end-to-end experiments described in section 2.1 including development of an isotopic model for a specific spatial-temporal domain and a flexible array of independent variables, evaluation of the model,

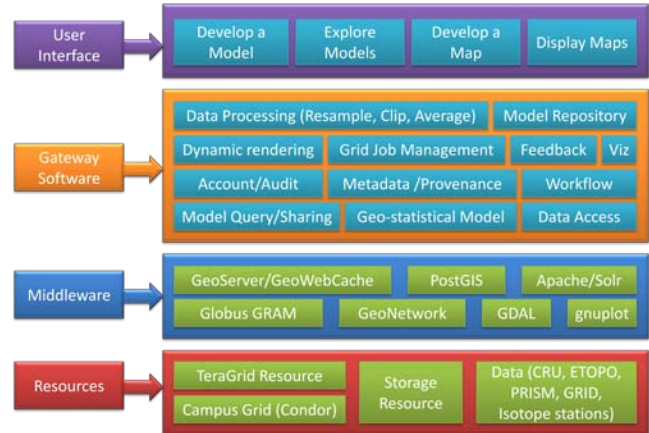prediction of isotope distribution, and dynamic map visualization, all done through the gateway.

Figure 2. IsoMAP system architecture

IsoMAP is also a model/data repository and collaboration environment where scientists can share and exchange the models they developed with a broad user community. Complete metadata and provenance about a model and its input and output data are automatically tracked and recorded by the system. All metadata are indexed using the Apache Solr search engine [1] and may be easily accessed to evaluate the quality of the model/data, reproduce experiment results, support collaborative model development, as well as for data aggregation and inter-operation by third party applications.

2.3 Data and Metadata

IsoMAP provides a transparent interface that spares scientists the tedious tasks of identifying, accessing, and preprocessing disparate datasets that are needed to drive the geospatial models for isotopic ratio mapping. As a beginning step, we have integrated point and raster data from four commonly used and high quality environmental datasets into IsoMAP:

- CRU 2.1: It includes continuous time-series of monthly values interpolated on $0.5°$ continental grid excluding Antarctica. The dataset is developed by the Climate Research Unit in University of East Anglia and are available at http://www.cru.uea.ac.uk/cru/data/hrg/cru_ts_2.10/newly_gridded/data_all/. This dataset is stored in a custom format. The variables used in IsoMAP models include average daily temperature, average daily minimum temperature, average daily maximum temperature, precipitation, and elevation.

- PRISM: It includes continuous time-series of monthly values interpolated on 2.5 arc minute continental grid for the contiguous USA provided by the PRISM group at Oregon State University. The files are in Arc ASCII format and there is one file for each variable per month. The variables used in IsoMAP models include precipitation, average daily minimum temperature, and average daily maximum temperature. This dataset is available for download at http://www.prism.oregonstate.edu/

- ETOPO: It is 5-minute gridded elevations/bathymetry data for the world provided by National Geophysical Data Center (NGDC). The data is available at

The spatial extent is the entire globe except South Pole.

- Precipitation isotope database: It is a compilation of precipitation isotope monitoring data from a wide range of literature sources, anchored by the dataset of the Global Network for Isotopes in Precipitation [10]. It includes point data consisting of monthly average precipitation isotope, temperature, and precipitation amount measurements at >800 sample sites.

In order to apply a geospatial model across a spatial temporal domain, all the parameters used by the model need to be available for every grid cell within the domain – this requirement is supported by the three gridded data resources (CRU, PRISM, ETOPO) plus a set of regular grids provide cell center latitude and longitude values for use in calculations (GRID). The input data are of different formats, with different map projections, spatial boundaries and resolutions. It is necessary to preprocess the input data to have the same common spatial attributes. Such tasks are time-consuming and error prone. They also require certain expertise in geospatial data processing which is often a barrier for researchers. IsoMAP helps automate these data processing steps.

3. System Implementation

In this section, we describe in details how IsoMAP gateway is implemented, including geospatial data processing, modeling, rendering and metadata documentation.

3.1 Spatial data processing

There are several places where geospatial data processing is performed behind the scenes during a workflow. First, the datasets are downloaded from different data sources and pre-processed to some standard formats. Second, each time a user selects a spatial temporal range and the data sources for a model (for either model fitting or map prediction), a pre-validation step is run that extracts relevant data based on user's selection and calculates whether there will be enough data produced to support the model. Third, a series of raster data processing steps is needed to prepare the input datasets for the map prediction code. Finally, additional minor processing occurs during map visualization. We will describe how the data are processed in details in the following sections.

3.1.1 Input pre-processing

The goal of this step is to convert the heterogeneous datasets downloaded from disparate data sources into standard formats that are supported by Geospatial Data Abstraction Library (GDAL) which is a library for raster geospatial data processing [6]. For example, raw CRU data is stored in its own format. We first converted CRU data into NetCDF format using an existing python tool called *cru2nc* (http://www.cru.uea.ac.uk/~timm/grid/software.html). The data was then converted again into GeoTiff format which is commonly-used for raster data. The final result is a group of yearly GeoTiff files ranging from 1901 to 2002.

All the raw raster datasets are stored in a file system and the point isotope station data is entered into a PostGIS database for easy management. To speed up the data processing processes that happen each time a user makes a selection from the UI, several PostGIS tables are pre-created with data extracted from CRU, ETOPO, PRISM, and GRID which matches the locations of isotope stations (Figure 3). Those tables will be used repeatedly later by the pre-validation process. The system keeps a record of about the raw datasets and the intermediate database tables. Additionally at this stage, proper spatial metadata (described

below) are written for all source data and stored in an extant metadata system operated by Purdue Libraries.
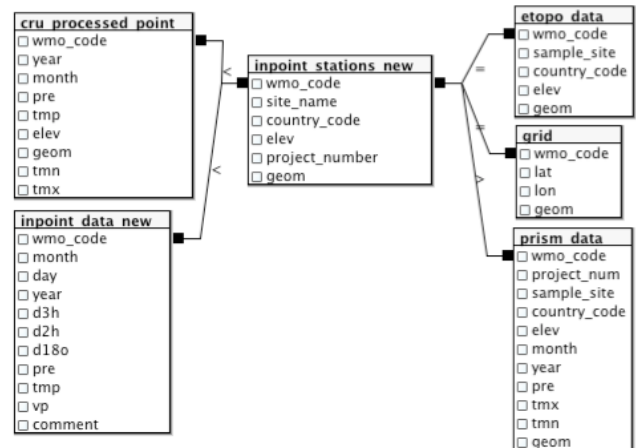

Figure 3. Structure for the pre-created database tables

3.1.2 Pre-validation

The pre-validation process is executed each time a model fitting or map prediction job is constructed. It includes the following steps:

1. The web GUI captures user's selection in an xml document and sends it to the pre-validation module. Figure 4 shows a snapshot of the dataset selection interface which allows users to choose a dependent isotope variable, several independent variables, data sources, and transformation formula.

2. The pre-validation module performs a series of database queries and operations. It generates a set of temporary database records which include the averaged values for the dependent and independent variables selected by the user. Only the data at the locations of the isotope stations in the selected spatial extent are extracted. A unique job ID is used to identify the temporary records created for each user session.

3. Transformations may be applied to the independent variables if selected by the user. Currently two transformations are supported – *Square()* and *Absolute()*.

4. Rows with missing data are marked as invalid.

5. A data completeness test is calculated for each isotope station. The formula being used is *sum(cru_p_valid)/sum(cru_p_all)*, where *sum(cru_p_valid)* represents the sum of CRU precipitation values of valid rows and *sum(cru_p_all)* represents the sum of CRU precipitation values over all rows. It the ratio is greater than 0.8, the station is considered as a valid station.

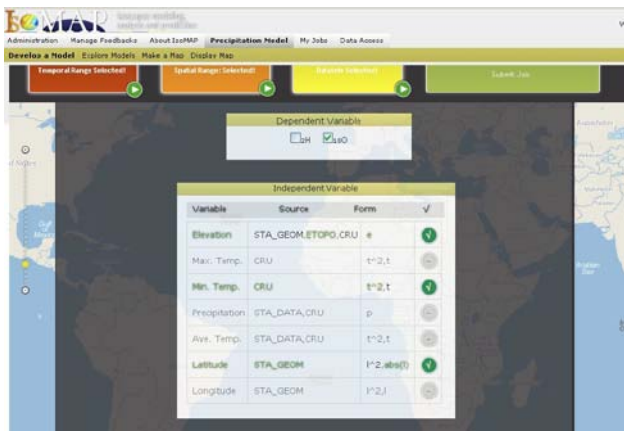6. The pre-validation module returns success if there are at least 15 valid stations.

Figure 4. A snapshot of the data selection interface.

One challenge of pre-validation is to complete all the data processing operations within a reasonable amount of time to ensure an interactive user experience. As described in 4.1.1, we pre-created several intermediate database tables and significantly reduced the amount of time it takes at runtime. We also encountered a scalability issue with pre-validation which involves many database intensive operations. There is a known issue with increased likelihood of deadlocks in PostgreSQL when the number of transactions that need to modify the same table increases. This is mainly caused by race conditions in row level locking. To solve this problem, we plan to use advisory locks to protect from deadlock conditions to happen while in the same time minimize the potential degradation on performance.

3.1.3 Raster data processing

The raster datasets (CRU, PRISM, ETOPO, and GRID) are processed to generate the input data for map prediction code. There are several issues to solve: First, each dataset is formatted differently. For time-independent data such as elevation in ETOPO, there is only one file that needs to be retrieved. For time-dependent data such as average daily temperature in CRU, each file includes a year's of data with 12 bands corresponding to 12 months. On the other hand, PRISM data are downloaded as monthly files ranging from 1895 to 2009 and grouped into folders where each folder contains ten years' of data. Such format information is stored in the metadata associated with each input data source. The system dynamically determines how to fetch data based on the metadata.

Second, the IsoMAP raster datasets have different resolutions and spatial attributes. They need to first be re-sampled to the same resolution, and then clipped to the spatial extent selected by the user. These operations are done using GDAL command line tools.

Third, after the actual values for each grid cells are extracted, we need to perform an average operation for time-dependent variables. There are no existing averaging tools and as a result we developed a tool using the JAVA API for GDAL library.

The processing time and memory requirement for the raster data processing procedures increases when the size of the raster data gets big. To improve the overall system performance and scalability, we wrap the raster data processing operations together with the map prediction code and submit it to the TeraGrid cluster nodes instead of executing it locally on the IsoMAP web server.

## 3.2 Geo-statistical modeling tools

A hierarchical statistical modeling code was developed to interpolate precipitation isotope ratios at continental to global scales using a flexible array of dependent and independent variables. The statistical model is composed of a function of independent variables and a spatiotemporally dependent error term. The code includes three modules: a statistical model estimation and testing procedure, a universal kriging prediction procedure, and a cross validation procedure [21]. The first two modules are run independently through the *Develop a Model* and *Generate a Map* components of the IsoMAP workflow, respectively. The third is not currently used given its relatively high computational demand, but is tested and available for future implementation. The model code, written in C++, can be run independently of the IsoMAP interface, but is seamlessly coupled with the data processing and post-job result handling and visualization capabilities of IsoMAP. Information exchange between the code and other IsoMAP CI occurs through the exchange of simple, standardized files, making the IsoMAP system blind to the details of the modeling (and vice versa) and increasing the modularity of the system. The automated geostatistical algorithms developed here will be useful for a wide range of applications within and beyond the environmental chemistry field.

The gateway leverages TeraGrid, a national high end computation infrastructure, to run its modeling code and raster data processing programs. The jobs are submitted to the PBS queue of Steele cluster at Purdue using Globus GRAM API. There is a *JobListener* class associated with each job. It has callback functions at various stages of job execution. The gateway uses a shared community account to access the computation resources on behalf of the gateway users. After a job is completed, its metadata is updated and submitted to the Solr server for indexing.

Running the precipitation model and data processing programs on the TeraGrid provides several benefits. First, the memory requirement and execution time of the model development code increases quickly when the number of valid isotopic stations in the selected spatial and temporal domains grows (Figure 5a). Although the map prediction code is not as memory intensive, its execution time also grows when the number of grid cells in the selected spatial domain increases (figure 5b). By running these jobs asynchronously on the TeraGrid cluster nodes, the performance of the gateway server and end user experience are not affected by the memory consumption of multiple user requests. It improves the overall scalability and reliability of the entire system.
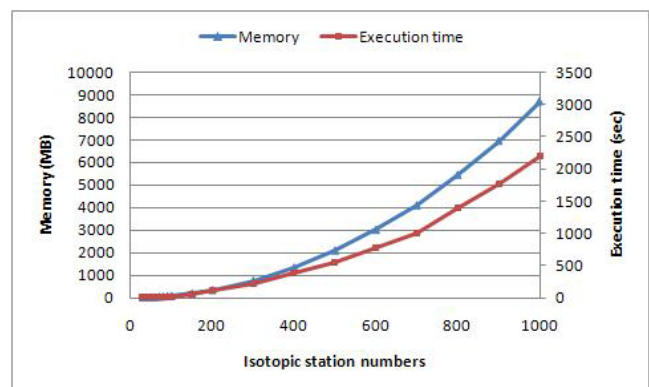

5(a)
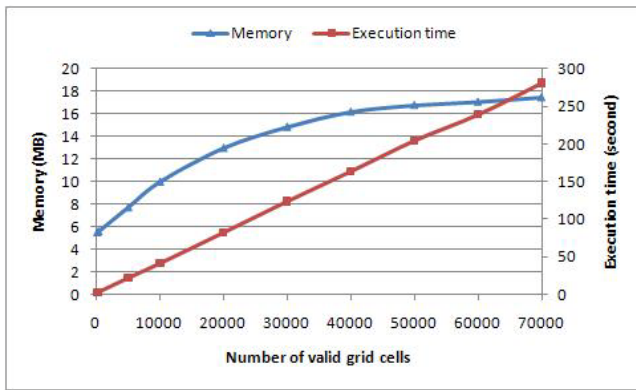
5(b)

Figure 5. Memory requirement and execution time for precipitation model fitting code (5a) and prediction code (5b).

Secondly, GDAL library is used extensively by the gateway web application for geospatial data processing. It is a known issue that GDAL library is not completely thread-safe due to the way it manages its global cache and error context. Instead of implementing locks which reduces the system throughput and adds delays when there are multiple users, running the raster data processing jobs asynchronously on distributed cluster nodes provides a simple solution to avoid race conditions.

3.3 Output exploration and dynamic map rendering

After a model fitting job completes, users may explore the output and assess the uncertainty of the model. A number of parameters for the regression model and geostatistical model are reported including ANOVA (Analysis of Variance), coefficient of determination, mean square error, Moran's I, and parameter values. A QQ plot and residual plot are generated dynamically for both models so that users can directly visualize them online (Figure 6). Users can also download both the raw data and the plots from the web site.
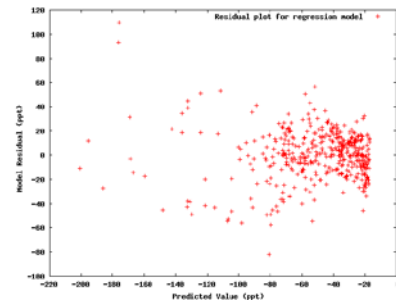
6(a)
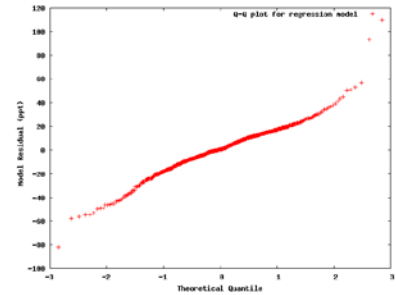
6(b)

Figure 6. Visualization of model fitting results. 6(a) A residual plot for regression model; 6(b) A Q-Q plot for regression model.

The output of a map prediction job includes several gridded data files. These data sets are first converted to Arc ASCII and GeoTIFF formats using GDAL library. They are then registered in a GeoServer as raster layers using the REST interface. A final requirement for the system – to dynamically scale the raster symbology based on a user's current zoom (and not, say, absolute min/max values) – was also accommodated using customized tools. A typical request from the OpenLayers-based browser client reaches GeoServer, which accesses the layer's coverage store, styles the data using an associated Open Geospatial Consortium Styled Layer Descripter (SLD) document [18], and returns a rendered image to the browser for display. For IsoMAP, we associate the call to GeoServer with a *different* SLD, which happens to be a custom-written php script. The script uses, again, the GDAL library to access the coverage store first and generate a VRT copy of the original. Consider this VRT an alias to the original, which can subset and mutate the original without actually storing copied data. A new SLD document is then written on the fly using the VRT subset of the original, which GeoServer is forced to use to render the image (instead of its own, default SLD) by OpenLayers' "mergeparams" function. As an aside to this dynamic styling, a similar process dynamically generates an accompanying legend graphic (using gnuplot). The metadata for map products are indexed by a Solr server and can be searched using geospatial attributes or keywords. All the map and model products are private by default. They become public if the owner chooses to publish and share them with the other gateway users. The maps may be viewed directly in the web browser as shown in Figure 7. All the associated metadata as well as the raw map files may be downloaded in standard formats and may be directly used in other models or applications.
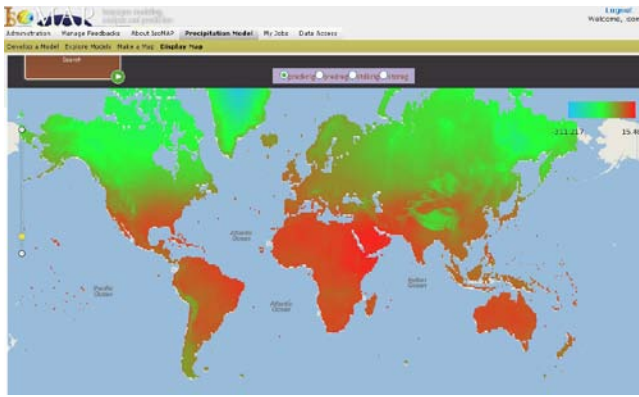
Figure 7. A dynamically rendered isotope prediction map for δ^{18}O.

3.4 Metadata documentation

There are two distinct elements in IsoMAP that can legitimately be referred to as "metadata." The first is "job metadata." This is the xml used to describe, in full, a given IsoMAP job -- its inputs, outputs, valid stations when relevant, etc. It is written in a custom xml schema, stored in MySQL, and indexed by Apache Solr to not only give the system and its users complete documentation of IsoMAP products, but to also allow for a rich and robust search of the entire system's public assets.

We chose a custom schema here because we needed the lightest but fullest document possible. The xml itself is constructed during a user's interaction with the IsoMAP GUI and is used, in effect, as a payload that accumulates vital data about jobs as it moves through the various stages of the IsoMAP workflow (including the final generation of output). As such the document serves several purposes. Primarily it documents a job so that it can be recreated if necessary or so the user can fully understand, document, and cite their IsoMAP output. Secondarily we have as-yet unimplemented plans to accommodate external job submission, whereby users can construct a job metadata payload outside of the GUI and submit the document to the back end for processing as usual. A thin layer of service software must be built to accommodate this submission, of course, which has been a lower priority during initial IsoMAP development. Tertiarily, the "job metadata" can be easily translated from its custom xml schema to a more formal schema (e.g. Dublin Core) for transfer to and from other systems. Although this is also an unimplemented, lower-priority function, we hope to make IsoMAP's repository of shared output as widely available as possible. The first form of this interoperability will be to allow other data access systems to harvest IsoMAP's jobs metadata so it can be searched for and found in other interfaces.

The other metadata behind IsoMAP is the material that describes the various data inputs to IsoMAP (ETOPO, PRISM, CRU, etc.) These metadata come from multiple sources and are written in official spatial metadata schema such as FGDC-CSDGM [5] or ISO19115 [11]. In some cases the source data were already accompanied by proper metadata documents, in which case they were edited and updated to reflect changes made to the data by the IsoMAP team (even to simply reflect the new location in Purdue-hosted storage). In other cases, no formal metadata existed and therefore had to be written by the IsoMAP team using any available, authoritative documentation from the source or elsewhere. The metadata were then submitted to an instance of the GeoNetwork metadata system already in operation at Purdue Libraries, where they are still available to the IsoMAP GUI during

data source-related queries (via an Apache Solr interface), but also sit as "collected" objects in a more generally available library of digital geospatial data built for the greater Purdue geoscience community.

4. Use Case

The initial version of the IsoMAP portal was released on March 30, 2011. The launch was scheduled to coincide with a NSF Research Coordination Network-sponsored meeting entitled "Roles of Stable Isotopes in Water Cycle Research", which brought together a subset of the potential user community for the IsoMAP precipitation isotope tools. As a component of the meeting agenda our team members organized and conducted two tutorial sessions that offered an introduction to the IsoMAP technology and functionality. In total 46 scientists participated in the sessions, or approximately 2/3 of the meeting attendees.

Tutorial participants were guided through the process of account creation, the precipitation isotope workflow, and job management capabilities of IsoMAP, and informal feedback was solicited. The hands-on sessions allowed new users to create and submit their own model and map jobs in real time, and thus provided a trial of the IsoMAP interactive experience. Users were able to successfully develop and execute jobs within the 45 minute tutorial, and most jobs completed and gave results for visualization within that time frame. The sessions also represented our first test of the IsoMAP system under relatively high-load (20 – 25 simultaneous users) usage conditions. Although most elements operated seamlessly, one important outcome of the testing was identification of the potential for database deadlocks during the pre-validation set, as described in section 3.1.2. Our ongoing work to resolve this issue will be important to support the future application of IsoMAP in high-load situations such as classroom exercises.

Formal feedback was gathered from tutorial participants through a short questionnaire administered using the Qualtrics software package. Users were asked to rate the perceived value of the current implementation of IsoMAP to their research, teaching, and outreach programs. They were also asked to provide comments on elements of the system they liked and changes they would like to see made, and estimate the potential value of the system if the requested content and functionality changes were incorporated. A total of 33% of the tutorial participants responded to the survey within 2 weeks. Of these, 53% said that they considered the current release of IsoMAP to be "Very Useful" or "Extremely Useful" to their research, teaching, and outreach programs. The perceived value of the site following further development was high, with 73% of participants giving a "Very Useful" or "Extremely Useful" ranking for research and 66% for teaching and outreach, with no respondents giving a ranking below "Useful". Survey participant comments also provided valuable guidance that will inform future IsoMAP development efforts – by for the most consistent suggestion (53% of participants) was to support additional data, including user-supplied datasets and data from other environmental isotope monitoring networks, in future versions of the gateway.

5. Conclusions

In this paper, we described our design and implementation of the IsoMAP gateway as an online workspace for browsing, modeling, and prediction of precipitation isotope ratios over user-specified spatiotemporal domains. The system has been successfully used in teaching a tutorial at a conference for water research community. Positive feedback has been received after the tutorial.

As a future step, we will extend the modeling framework to create models for other application areas such as prediction of surface water isotope ratio, leaf water and plant tissue hydrogen, oxygen, and carbon isotope ratios. We will work on the capability of enabling users to develop models using their own datasets which is a much desired capability and technically difficult to implement. Further, we plan to develop tools that enable users to model the geographic origin of samples based on their isotope composition combining the Isoscapes models developed within IsoMAP, their corresponding statistical uncertainties, and the sample values provided by the user. In the longer-term, our development efforts will focus on enhancing the data richness of IsoMAP by linking our services to the databases of emerging monitoring networks and providing functionality for users to upload and model their own data.

6. ACKNOWLEDGMENTS

7. REFERENCES

[1] Apache Solr: http://lucene.apache.org/solr/

[2] Bowen, G. J., and J. Revenaugh (2003), Interpolating the isotopic composition of modern meteoric precipitation, *Water Resources Research*, *39, 1299*, doi:10.1029/2003WR002086.

[3] CLEANER: Collaborative Large-scale Engineering Analysis Network for Environmental Research. http://cleaner.ce.berkeley.edu

[4] EarthScope: An Earth Science Program: http://www.earthscope.org/

[5] FGDC-CSDGM: http://www.fgdc.gov/metadata/csdgm/

[6] GDAL - Geospatial Data Abstraction Library: http://www.gdal.org.

[7] GeoNetwork: http://geonetwork-opensource.org

[8] GeoServer: http://geoserve.org

[9] Jason Novotny, Michael Russell, Oliver Wehrens: GridSphere: An Advanced Portal Framework. EUROMICRO 2004: 412-419

[10] IAEA/WMO (2004), Global network for isotopes in precipitation, the GNIP database, http://www-naweb.iaea.org/napc/ih/GNIP/IHS_GNIP.html.

[11] ISO 19115: Organización Internacional de Normalización. 2003. ISO 19115: geographic information : metadata. Geneva: ISO.

[12] jQuery: http://jquery.com

[13] Beth Plale, Dennis Gannon, Daniel A. Reed, Sara J. Graves, Kelvin Droegemeier, Robert Wilhelmson, Mohan Ramamurthy: Towards Dynamically Adaptive Weather Analysis and Forecasting in LEAD. International Conference on Computational Science (2) 2005: 624-631.

[14] National Ecological Observation Network (NEON): http://www.neoninc.org/

[15] Nambiar, U., Ludaescher, B., Lin, K., and Baru, C. 2006. The GEON portal: accelerating knowledge discovery in the geosciences. Workshop On Web Information And Data Management Archive, Proceedings of the eighth ACM international workshop on Web information and data management, 2006.

[16] NEES – Network for Earthquake Engineering Simulation: http://nees.org/

[17] Pierce, M. E., Fox, G. C., Aktas, M. S., Aydin, G., Gadgil, H., Qi, Z., and Sayar, A. The QuakeSim Project: Web Services for Managing Geophysical Data and Applications. PAGEOPH Special Issue for 5th ACES International Workshop, Island of Maui, Hawaii.

[18] SLD: http://www.opengeospatial.org/standards/sld

[19] Software for climate grids: *http://www.cru.uea.ac.u*k/~timm/grid/software.html

[20] WATERS Network: http://www.watersnet.org/.

[21] Zhang, T. et al. (2011), Abstract: An Automated Geostatistical Toolkit for Mapping Stable Isotope Ratios of Precipitation over Space and Time, *Roles of Isotopes in Water Cycle Research*, Keystone, CO, 3/29-30/2011.