

AVISTED—Analysis and Visualization Toolset for Environmental Data

Likhitha Ravi*, Eric Fritzinger*, Sergiu M. Dascalu*, Frederick C. Harris, Jr*.
University of Nevada, Reno, NV, USA

Abstract

Data analysis and visualization are two effective ways to unveil possible rules hidden in the environmental data. Environmental data is usually very large and can be stored in a format not commonly used, such as NetCDF. Most existing data visualization and interaction tools cannot handle environmental data efficiently because of these characteristics. Furthermore, most existing environmental tools require a user to download the data and execute software locally. This is not very convenient if the user's computer does not have high computing power and the data size is very big. To address these concerns, in this paper we present AVISTED, a generic web-based toolset used by the environmental scientists to analyze, convert, extract and visualize large datasets. This approach involves executing most of the operations on the server side, provides different visualization techniques and supports multiple environmental data formats such as NetCDF and HDF5. Additionally, in this paper we also present a new approach for software design, GUI-Enhanced Activity Diagrams (GEAD) that links UML activity diagrams to user interface snapshots. As such, it enhances the collaboration between teams and reduces the development time. This paper provides detailed descriptions of the design aspects of AVISTED by leveraging the concepts and strategies of GEAD.

Key Words: GUI-enhanced UML activity diagrams; large datasets; software design; data visualization; web application.

1 Introduction

Nowadays, data is produced at unprecedented rates. In particular, in environmental sciences, climate data is being collected, stored and transformed using various modeling techniques to predict the future climate. For example, The Nevada Research Data Center (NRDC) [18] stores sensor-based climate data of Nevada and the modeled datasets. The National Oceanic and Atmospheric Administration (NOAA) [17] provides the data about the weather and climate in the US. Similarly, Cal-Adapt [9] provides the climate information of California. Although the climate information is easily available, it is hard to get an insight into the data without specialized analysis and visualization tools.

The most common steps performed by a climate scientist are extracting data, applying models on the extracted data, converting the resulting model to the data format which is supported by the visualization tool, and visualizing it using data analysis tools, such as Matlab and R [28]. However, most of these require a user to install the software on the computer locally and some tools even require programming skills. These can be a problem for scientists who would like to analyze a large amount of data with a standard personal computer and do not know how to program. In order to address these concerns, in this paper we present a new approach entitled Analysis and Visualization Toolset for Environmental Data (AVISTED) aimed at addressing scientific data analysis needs.

AVISTED is a web-based visualization toolset. It helps climate researchers to find the trends, changes of variables with time, and dependencies between variables. AVISTED incorporates features such as User Authentication, User Authorization, Model Output Management, Model Operation, Tools Upload, and Archives. It supports datasets in NetCDF, HDF5, ASCII and CSV formats. Users can perform extraction, visualization, view, download and save operations on a dataset. Visualizations that are developed using the toolset can also be saved in the user accounts for accessing them in the future. Also, to support extensibility users are allowed to upload their own tools for data processing with the administrator's approval. AVISTED is implemented using the client-server architecture. It executes all the operations except visualization on the server side, which reduces the computing burden on the client side. Therefore, even if a normal computer is used, complex operations can be performed efficiently.

When we designed a supporting toolset for the AVISTED approach, a new design method titled GUI – Enhanced Activity Diagrams (GEAD) has been used. More details about GEAD is presented in our previous work [23]. The purpose of GEAD is to provide a comprehensive view of the application to the developers and testers. GEAD provides a user interface design for each significant activity in a UML activity diagram. This approach helps the developers in identifying activities that are unique and common between the user interfaces. Identifying these connections at design time are very helpful in the applications that are implemented using the MVC architecture. Furthermore, this technique helps users and stakeholders in getting a meaningful link between the execution and the user interface. GEAD is applicable in many fields such as web development and video game development. In this paper, we

* Department of Computer Science and Engineering. College of Engineering. Email: {ravi, ericf, dascalu, fred.harris}@cse.unr.edu.

showcase the GEAD approach by using its concepts in designing the activity diagrams of AVISTED.

The rest of the paper is structured into six sections. Section 2 presents the background from the scientific perspective. Section 3 gives an overview of AVISTED's software requirements. Section 4 presents the new GEAD technique and its usage in designing AVISTED. Section 5 presents an application scenario of AVISTED. Section 6 provides a comparison of AVISTED with related tools. Lastly, Section 7 concludes the paper by outlining several directions of future work for the AVISTED methodology.

2 Background on Scientific Need

AVISTED has been created to address the scientific needs of the environmental researchers. It is now integrated with NRDC 0 and is in the testing phase. The application uses a web interface to perform its functionalities by interacting with RESTful APIs for data conversion and data extraction. After moving the AVISTED toolset to the NRDC server, we have conducted functionality testing, use-case based testing, and interface testing to check if the application is functioning as per the specifications, navigating across different pages correctly, and connecting through the application layers, respectively.

Furthermore, in the future we intend to conduct more testing of the application with the participation of the end users in order to obtain valuable user feedback. Also, it is essential to evaluate the performance of the application with an increased number of concurrent users. However, as the application uses the RESTful API's for time intense tasks such as data extraction and conversion, these components can be placed in different Docker containers as needed. And, to improve the performance when the number of users increases, the computation tasks can be distributed to worker containers and processed in parallel using a job queue as discussed in our previous work [29] which mainly presents a new elastic scale hybrid server method for dynamically adjusting the web hosting resources based on queueing theory and user feedback.

This section gives the background of NRDC and explains why we developed the AVISTED toolset.

2.1 The Nexus Project & NRDC

NRDC serves in a critical role of science cyberinfrastructure for sensor-based and modelled data management. It facilitates the acquisition, transport, storage, query, and dissemination of observational data created by automated digital sensor systems. The NRDC participates in cutting-edge software and system development to enhance the next-generation science that leverages the Internet of Things (IoT). The goal of NRDC is to transform the scale, quality, impact, and bottom-line cost of research projects in Nevada that seek to deploy automated sensor systems as part of their scientific workflow. NRDC is dedicated to providing services and curation for project-level datasets and needs 0. The center was born out of a data portal that was developed during a Track 1 NSF EPSCoR project on climate change that included a cyberinfrastructure component

called Nevada Climate Change Portal (NCCP). NCCP has been developed with support from a large, 5-year (2008-2013) NSF EPSCoR Research Infrastructure Improvement (RII) grant to provide data resources and computing support for scientists researching the long-term effects of climate change in Nevada. To achieve its goals, NCCP contains not only environmental datasets but also software tools and services created by Nevada computer engineers to help data acquisition and processing. One such tool is the new web application AVISTED that we developed, which enables data selection, extraction, download, conversion, and visualization of the modelled environmental datasets.

NRDC is currently funded exclusively through NSF-EPSCoR programs as part of large interdisciplinary research efforts 0. NRDC operates remote high-speed digital data links as part of the Nevada Seismological Laboratory wide-area research data network. Part of the cyberinfrastructure team of the NEXUS project [6], we maintain a cluster of servers that house its primary data storage and software services. NRDC maintains near-real-time archival of both flat-file and relational data structures of small-to-moderate scale. Its individual project databases range from a few hundred to billions of individual data points. NRDC exposes project data to researchers and collaborators using high-performance query interfaces as well as aggregated file archives. The data is shared directly with the Western Regional Climate Center (WRCC), the Southern Nevada Research Center at UNLV, the PRISM Climate Group at the Oregon State University, and the DataONE distributed data curation clearinghouse [8]. The NRDC develops solutions for research cyberinfrastructure (CI). The CI team members are active participants in the Earth Science Information Partners (ESIP) Federation, and work as national community members to improve standards, practices, and technologies in scientific data management. Their research includes a significant student education and training component, where computer science and engineering student's interface with domain scientist "clients" and seek to aid their sensor-based research using CI tools. We believe that this cross-disciplinary integration of engineering with domain science can ultimately lead to higher-quality science and research in Nevada.

2.2 Need of Scientists

NRDC provides scientists with quality data but the scientists also need tools to further explore the data and find valuable rules and patterns within the data. Extraction, visualization, and conversion are three common data analysis operations in the scientific field. Data extraction allows a user to extract a certain parameter from the dataset based on a particular time slot or location. This operation avoids loading irrelevant data and simplifies the data analysis. Data visualization can highlight the data trend of a parameter and display possible relations between two parameters. It is very important to find "gold nuggets" from the raw data 0. Data conversion is another significant operation. It means converting a data from one format into another format. A scientist usually uses multiple tools to analyze a dataset and these tools may each require input in a different file format.

Therefore, the scientist needs to write glue codes or use a data conversion tool. Since most scientists that use NRDC are environmental scientists and do not necessarily know how to program, data conversion is implemented in AVISTED for their convenience. Usually, the environmental scientists need to process a large amount of data and this requires a powerful machine or cluster. To reduce the burden, a client-server architecture is leveraged. The architecture does not require a powerful machine on the client side because most of the operations are executed on the server side. Our previous work proves that this architecture works well in this scenario [14], [20], and [27].

One use case of AVISTED is that a scientist can find possible connections between two environmental variables by extracting and drawing scatter plots. For example, the scientist can extract humidity and temperature values of the last ten years of Snake Range West Montane from a huge NetCDF file and visualize these two variables in a scatter plot by using AVISTED. Because the extraction step is done on the powerful server side, the operation is executed very fast even if the file size is huge. If the humidity and temperature variables are highly correlated with each other, the scatter plot should be approximately in a linear relationship, which means if one of these two parameters are known, a scientist can estimate the other parameter by using the correlation relationship.

The users of AVISTED are researchers, educators, students, and the general public. Using AVISTED, researchers will be able to quickly find the trends and patterns in data. Educators can use the toolset in the classroom while teaching the students about climate change. It also helps the general public in getting the information about the impacts of climate change.

3 Software Requirements

Initially, the main goal of the project was to visualize the climate modeling data provided by the Nevada Climate Change Portal. In this process, we first developed the Visualization Toolset for Environmental Data (VISTED). VISTED [24] is a non-generic version of AVISTED with limited functionalities. This tool operates on the output of a hind cast NCAR/WRF-based model developed by Dr. John Mejia at the Desert Research Institute, Nevada [16]. This output contains daily values for six environmental variables and extends over 30 years (1980–2009), so it is fairly large, and data operations like the ones mentioned above are needed by scientists who investigate the results of running this model. Extension plans of VISTED have led to the development of AVISTED that include expanding it to operate on a broader range of model outputs, adapting it to individual user preferences, and incorporating more diverse visualization capabilities.

Figure 1 shows the traditional activity diagram of VISTED with the actions involved in displaying the data extracted by the VISTED based on the user's selection. User selects the required variables/parameters, time period, location and submits the form. The user request is then sent to the application server. If the user input is valid, data is extracted and presented to the user or else an error message will be sent to the user. User can either

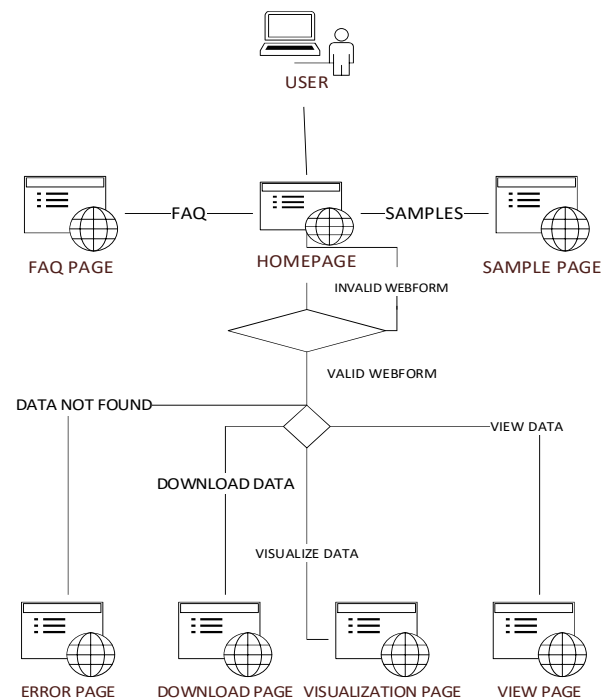


Figure 1: Traditional activity diagram of VISTED

view, download or visualize the extracted data.

Although VISTED helped environmental scientists in analyzing the NCAR dataset, based on the interviews conducted with researchers, an application that supports different file formats, modeled datasets and visualization techniques is much more beneficial for the broader community. Currently, a single generic tool that provides all of these features is not available for environmental researchers. We aim for the AVISTED approach to address these current challenges. AVISTED is an extended version of VISTED. The extension facilitates the application to be applicable to other model datasets in environmental sciences.

The new approach, AVISTED, for data analysis and visualization of environmental data incorporates: (i) a combination of key features that include extraction of datasets based on user requests, conversion of the extracted datasets to various data formats, a flexible user-friendly interface for visualizing the extracted data with several visualization options including zoom in and filter, capability to upload and manage model datasets represented in diverse data formats, and facilities to archive and download the work of authenticated users; (ii) environmental datasets; (iii) support for variety of environmental data formats such as NetCDF, HDF5, CSV, and ASCII; (iv) extensibility of the supporting environment, both in terms of the flexibility of the steps users can take for analyzing the datasets and in terms of available data processing features; and (v) applicability to other domains where geospatial information is represented using time series charts.

AVISTED has four types of users. A new user, registered user, admin user, and guest user. A new user will be able to register and log in to the application. A registered user can

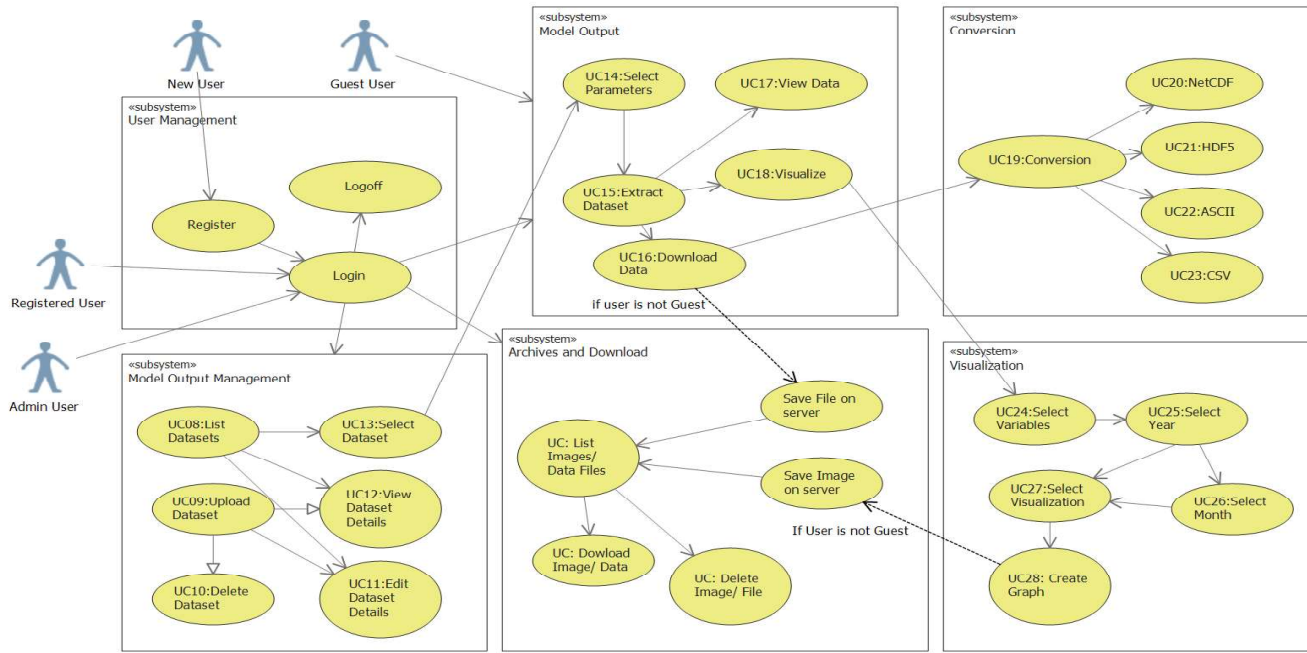


Figure 2: Use case diagram of AVISTED

access most of the features except some functionalities such as deleting a dataset uploaded by another user. On the other hand, a registered user can delete a dataset if it is uploaded by him/her. An admin has access to the all the features of the application. Finally, guest users have very limited accessibility to the application. They can only access the Model Output page. Figure 2 shows the use case diagram which depicts the interaction of different users with AVISTED. Each use case defines an action that AVISTED performs in collaboration with users.

The key requirements of AVISTED are noted based on the interviews conducted with individuals from different disciplines. Each requirement is given a unique ID and description. Non-functional requirements specify the standards that are used in evaluating the functionalities of the system. The main non-functional requirements of AVISTED are listed in Table 1.

Functional requirements define the functions of the software

Table 1: Non-functional requirements of AVISTED

ID	Description
T01	AVISTED shall be platform independent.
T02	AVISTED shall support different browsers.
T03	AVISTED shall support devices like desktops, laptops, tablets, and mobile phones.
T04	AVISTED shall be scalable.
T05	AVISTED shall be extensible and reusable.

system. The main functional requirements of AVISTED are presented in Table 2.

4 Architecture and Design

4.1 Design Solution

AVISTED is a web application that uses a 3-tier client-server architecture. The first tier holds the presentation logic, the second tier the business logic, and the third tier provides the data services for the application. This architecture makes the application easily manageable and scalable.

The architectural diagram of AVISTED is shown in Figure 3, which provides the description of how the application is organized. The top layer is the Web User Interface, which serves the front-end logic. The front-end logic of the application presents the user with the webpages that allow the user to interact with the web application via a web browser by sending requests from client to server using Hypertext Transfer Protocol (HTTP). It also handles the operations for data visualization using JavaScript libraries such as C3 0 and D3 0. This layer is implemented using HTML5, ASP.NET Razor Pages, and JavaScript. These cutting-edge technologies are supported by most of the modern web browsers.

The next layer is the Application Programming Interface, which provides the business logic for the application. The business logic takes care of the communication between the client requests and the database server. This communication is handled through libraries, functions, objects and variables. The layer is implemented using C# and the ASP.NET Core MVC

Table 2: Functional requirements of AVISTED

ID	Description
R01	AVISTED shall support datasets of different file formats such as ASCII, CSV, HDF5 and NetCDF.
R02	AVISTED shall provide user authentication and authorization.
R03	AVISTED shall support 4 modes: User Mode, Admin Mode, New User Mode, and Guest Mode.
R04	AVISTED shall provide Model Output Management.
R05	AVISTED shall allow the user to upload a dataset of up to 3GB from a single file or multiple files.
R06	The administrator of AVISTED can add a large dataset of size larger than 3GB directly to the repository in the back end.
R07	AVISTED shall allow the user to delete, update and select a dataset.
R08	AVISTED shall provide Model Output featuring the parameters, the date range, and the location details of the selected dataset.
R09	AVISTED shall allow users to select a graph over a specific month and/or year of a variable from the extracted data.
R10	AVISTED shall allow the user to select from a line chart, area chart, bar chart, or scatter chart.
R11	AVISTED shall allow users to compare two or more variables during a specific time period using the charts.
R12	AVISTED shall allow the user to save visualizations in his or her accounts.
R13	AVISTED shall allow users to download the extracted data in ASCII, CSV, HDF5 or NetCDF.
R14	AVISTED shall allow the user to save the extracted data in the user's account.
R15	AVISTED shall allow the user to download or delete the saved archives from his or her account.

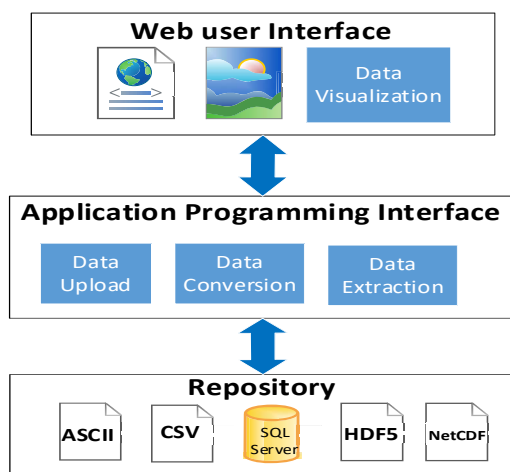


Figure 3: Architectural diagram of AVISTED

Framework. The logic for operations such as data conversion and data extraction is provided in this layer. These functionalities and the statistical computations are implemented using the ASP.NET 4.5 framework. CSV, ASCII, HDF5 data are handled through the RESTful API's and C# libraries such as HDF5DotNet Library 0 and Scientific Dataset Library (SDS) 0 are used for working with the HDF5 and NetCDF files, respectively.

The last layer is the Repository layer, which holds all the data needed for the application. All the files of the large datasets are stored in the repository. These files follow a naming convention that helps the application to retrieve the user-requested data more efficiently. Furthermore, for saving the information related to the users, SQL Server 2016 models and archives have been used.

All the libraries and frameworks used for building AVISTED are open source. The source code of AVISTED application is available in the GitHub repository [2], which makes the application extensible and reusable. Because we used the ASP.Net Core framework and HTML5, the application is platform-independent and supports several devices. Furthermore, we have hosted AVISTED on the IIS server to make it scalable.

4.2 A New Software Engineering Design Notation (GEAD)

The process of building AVISTED involved several phases. After collecting the requirements, we have designed several web user interfaces and drawn UML diagrams [10]. In traditional UML diagrams as shown in Figure 1, there is no link between the GUI designs and the UML activity diagrams. Without this connection it is time consuming for the team to search for the related user interface designs associated with the activity diagrams.

GEAD [1] is a new technique that allows connections between actions in the activity diagrams and their corresponding GUI designs of the user interface. Linking the GUI designs to the actions of the activity diagram allows the software engineering team to check related interface designs and activities quickly. It also helps the developers to identify actions that are common between different GUIs.

If an action has more than one interface design linked to it, it is considered a common task for these designs. Thus, the implementation of such actions can be placed in a global scope for reuse. In an MVC architecture, the common actions can be placed in the same controller and the actions that are specific to a GUI can be placed in the view components. This, in practice, saves a lot of development time.

To illustrate the GEAD approach we present an activity diagram of AVISTED in Figure 4. The small rectangular box at the bottom of each action has a link to the corresponding GUI design of the application. All the actions in the activity diagram have this link thus enabling the team to find the associated GUIs easily. For example, the action Display Model Output Management with User Settings is linked to the interface design #202 shown in Figure 5.

Similarly, the actions Model Output Management, Model

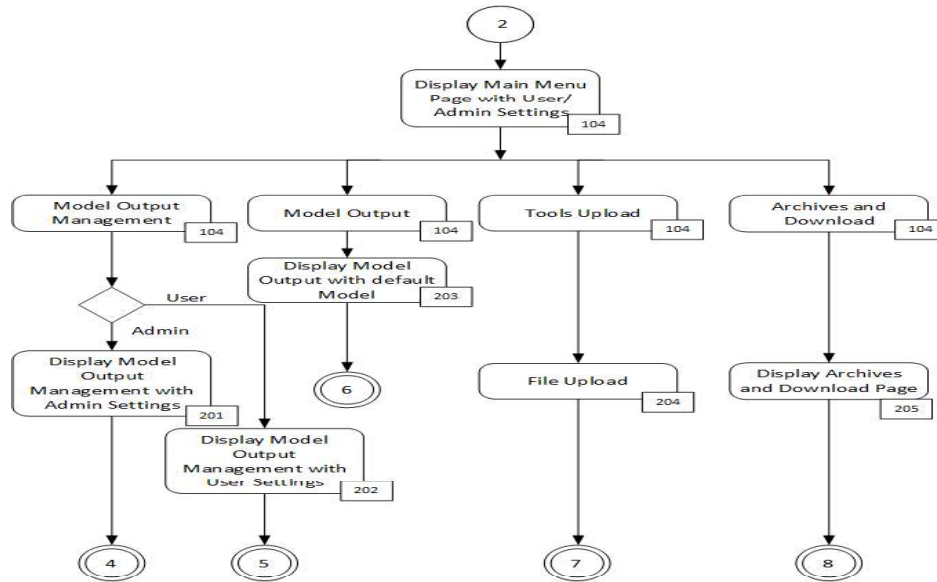


Figure 4: Activity diagram with four modes of AVISTED



Figure 5: Design of model output management page in user mode

Output, Tools Upload, and Archives/Download in the activity diagram are all associated to interface design #104. Since AVISTED is designed using the Model-View-Controller framework 0 the tasks that are related to these activities are placed in the same controller and, respectively, the same view. The actions that are different, for example Display Model Output with Default Model has a different design, #203, so it will be placed in a different controller and view.

This approach provides the developers with a more comprehensive view of the system, so the team can find if there are no designs linked to a significant activity. For example, in a gaming application where the design of the game changes with the state of the game the team can easily identify if more interface designs are needed for a particular state of the game. As GEAD requires interface designs for all major actions the

probability of missing interface designs in a software model is much smaller.

In large organizations, GEAD allows the development team to plan the software production more effectively and helps improve the collaboration within the team. For instance, a back-end developer can easily find the front-end developer who is working on the GUI designs that are related to the actions that he or she is working in the back-end. Further, a manager can easily divide the tasks among the team members without any duplication in actions.

5 Application Scenario

In this section, we present an application scenario of a user working with a Nevada regional climate model dataset provided by the NCCP.

The actions required to perform this scenario include selecting a dataset, extracting data, visualizing data and saving it on the local machine. In order to perform the above tasks, the user visits the home page of AVISTED and clicks on the login button. Then the user is required to enter his or her email ID and password on the login page. After a successful login, the user is navigated to the home page, with all the features of AVISTED available, including Model Output Management, Model Output and Archives/Download. For example, the user can select the Model Output Management page, as shown in Figure 6, which provides the user with the list of available models. This feature also allows the users to upload a new model of up to 3GB, view details of existing models, and select a model for further exploration. A model that is uploaded by a user is in the model-under-validation state. This model is checked by the administrator and its state is changed to ready after its configuration and validation are completed.



Model Management Output

[Create New](#)

Author	StartDate	UploadDate	Description	Format	Name	Parameters	Size	Status	
Dr. John Mejia	1/1/0001 12:00:00 AM	1/1/0001 12:00:00 AM	<input type="button" value="Description"/>	NetCDF	NCCP Climate Modeling Data	<input type="button" value="Parameters"/>	6GB	READY	Edit Details Delete Select
Goddard Earth Sciences Data and Information Services Center	1/1/0001 12:00:00 AM	1/1/0001 12:00:00 AM	<input type="button" value="Description"/>	HDF5	Precipitation Processing System	<input type="button" value="Parameters"/>	669 MB	READY	Edit Details Delete Select
National Renewable Energy Laboratory	1/1/0001 12:00:00 AM	1/1/0001 12:00:00 AM	<input type="button" value="Description"/>	CSV	National Solar Radiation Database	<input type="button" value="Parameters"/>	36.9 MB	MODEL- UNDER- VALIDATION	Edit Details Delete
Dr J. Scott Hosking	1/1/0001 12:00:00 AM	1/1/0001 12:00:00 AM	<input type="button" value="Description"/>	ASCII	Amundsen Sea Low (ASL) index	<input type="button" value="Parameters"/>	97.1 KB	MODEL- UNDER- VALIDATION	Edit Details Delete

Figure 6: Model management page of AVISTED



Please follow the below steps to visualize the data.

1 * Select the parameters

precip tmin tmax tmean qmean u10mean v10mean sdownmean

2 Select the time period

* Select start date
* Select end date

3 Select the latitude and longitude

[Note: Select a point from the highlighted region]



4 * Select the required statistics

Maximum Minimum Mean Variance Standard Deviation

5 Select an output format

CSV NetCDF ASCII HDF5

Save file on the server

Figure 7: Model output page

If the user selects the NCCP climate modeling data, he or she will be redirected to the Model Output page shown in Figure 7. In this figure an NCAR climate model data that has been generated by applying regional climate models upon historical data, which is based on the NCAR re-analysis and CCSM3 is displayed. The size of the NCAR/WRF summary output dataset is about 6 GB. First, climate variables such as precipitation (*precip*), temperature, (*tmin*, *tmean*, *tmax*), solar radiation (*qmean*, *u10mean*, *vmean*), and snow pack (*sdownmean*) are selected

by the user. Then as shown in Figure 7, the user selects a time interval, in this case 10/01/2006 to 12/31/2009. Next, the user selects a location from the polygon area of the map, which indicates that data is available in those regions. Also, the statistics required to be performed on the extracted data must be selected, in this scenario *standard deviation*.

Finally, after selecting the parameters, time range, location, and statistics, the user can either visualize, download, or view the data. In this scenario, the user chooses to visualize the data and

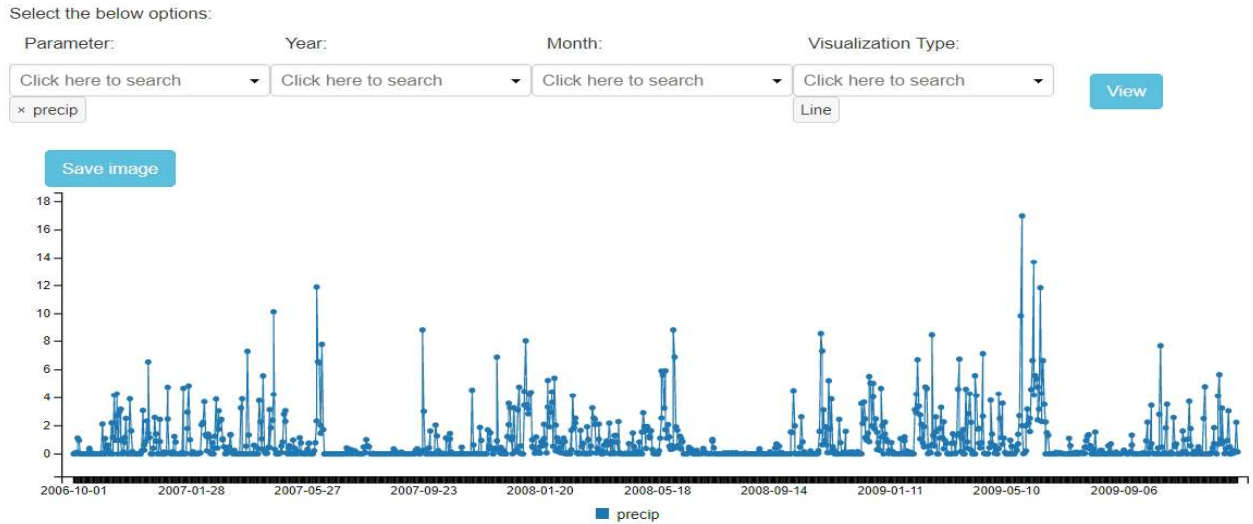


Figure 8: Line chart showing the precipitation parameter

submit the form. Consequently, AVISTED checks the input request and if it is valid the requested data is extracted by making a call to the AVISTED data extractor RESTful API. The API extracts the data as instructed and sends the result back to the web interface. An example of the visualization page is shown in Figure 8 where a line chart of the Precipitation variable is displayed.

AVISTED’s visualization tool is flexible and unique. The users can select a combination of different parameters, years or months for time interval, and various types of visualization. AVISTED filters and aggregates the extracted data based on the user’s selection of data, time interval, parameters, and visualization type. The resulting data is transformed to JSON format and sent back to the calling JavaScript module. For example, data is aggregated for each year of a parameter when a bar chart is selected.

AVISTED allows users to compare variables, which is a tool very useful for researchers. In the case presented in Figure 9, the user compares the minimum and maximum daily

temperatures (*tmin* and *tmax*) in the month of October 2007 using line charts for visualization. The user can also compare several parameters to identify potential trends. To do this, the user selects all parameters of interest and bar chart as visualization type, as shown in Figure 10. Each parameter is aggregated for the selected years.

Finally, the user can find outliers using the scatter plot. For example, the user selects u-radiation (*u10mean*) and v-radiation (*v10mean*) from the parameter list and then chooses a scatter plot for visualization, as shown in Figure 11. Using this plot, an outlier is found in the month of May, 2009 with a very high value for *v10mean*. Similarly, the user can further investigate the dataset by selecting different combinations of parameters and time periods to find the results of the applied model and check his or her hypotheses. Also, the user can save the chart in his or her local machine or on the server under the user’s account. To do this, he or she can use the Save Image option provided on the top of each visualization. Once the image is by using the Archives and Download feature of AVISTED.

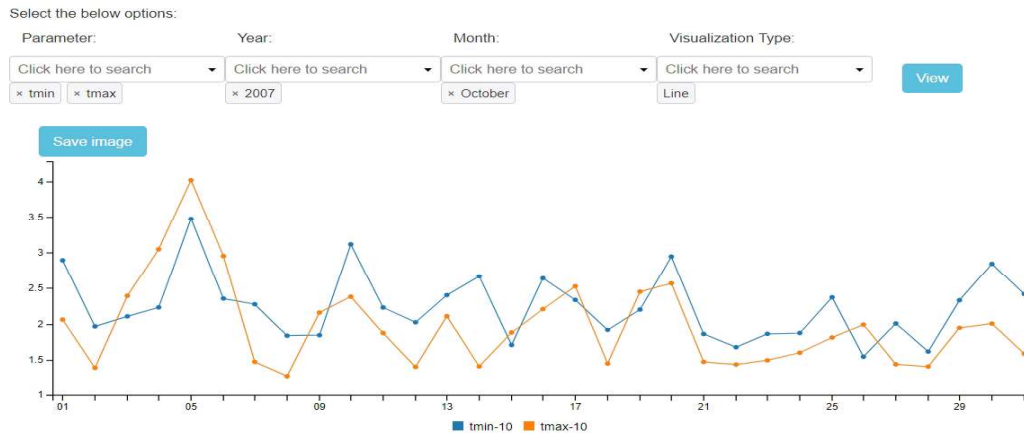


Figure 9: Line chart with min and max temperatures in the month of October 2007

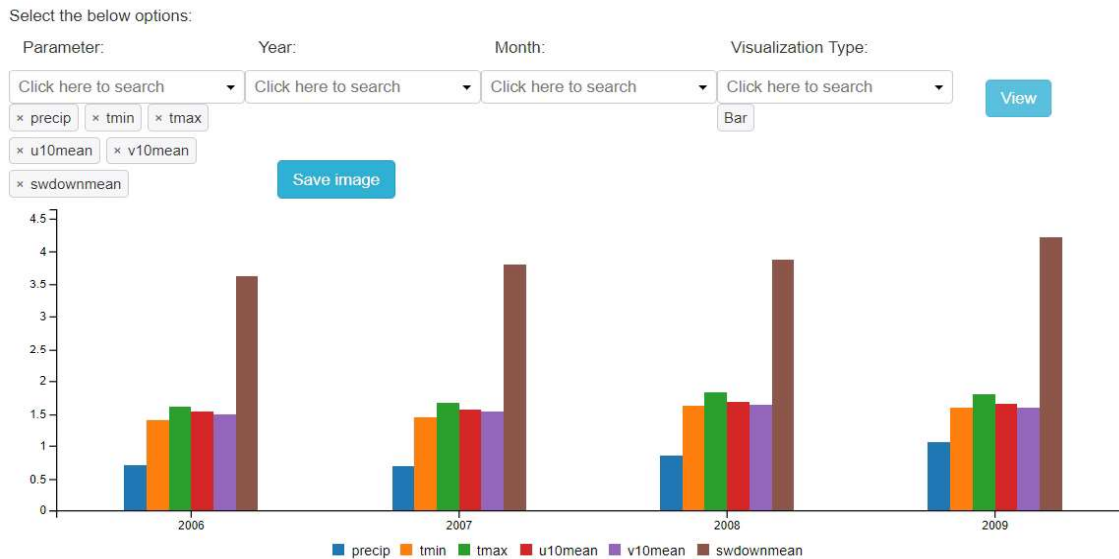


Figure 10: Bar chart with all the variables

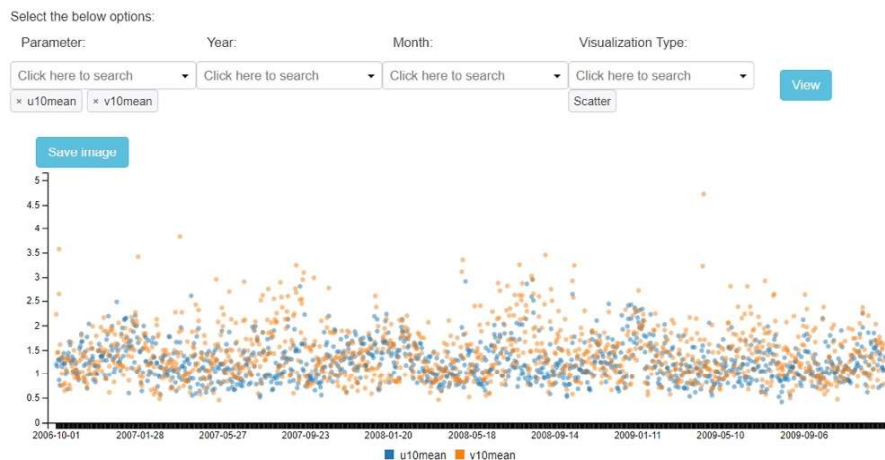


Figure 11: Scatter plot with the u10mean and v10mean variables from the NCAR dataset

6 Comparison with Related Work and Discussion

6.1 AVISTED Related Work

Certainly, there exist some works related to AVISTED, for example Cal-Adapt [9], Plotly [21], Climate Explorer [5] and DPT-VW 0 are used for data visualization and processing. These are described in some more detail next.

Cal-Adapt [9] is a web application that provides climate change information for California. The application presents research work done by California state agencies. Many datasets with climate projections for climate variables, such as snow pack, wildfires, flooding, and temperatures, from 1950 through 2099 are available on the Cal-Adapt website. The users can use the tools available to visualize the data for each model as well as compare the values of climate variables between the models. The users can also select a location on the map. The tool also allows users to calculate the statistics based on historical annual

mean or modeled projected annual mean. It provides many scenarios and uses different models for predicting the climate change data of California. These functionalities make the toolset more unique and sophisticated than many others, but Cal-Adapt's impressive features would have been even more beneficial if it was an open source project. As of now, it cannot be reused or extended by other communities who are doing similar research. Also, some functionalities such as data upload and editing are not supported by its toolset.

Plotly [21] is a generic web-based data visualization toolset with a collection of high-quality complex visualizations and dashboards. It provides free open-source libraries in JavaScript, Matlab, Python, and R languages for further customization of the plots. The users can upload data from a single or from multiple CSV files. If multiple files are uploaded, each file is placed in a separate grid. Once the data is ready, charts can be created by selecting the grid and chart type. A unique feature of Plotly is that it allows users to collaborate with others who

would like to view and edit the plots on their work through an email with link to the generated graph. Although the features of Plotly are very promising for users, it is not very suitable for large datasets. Another limitation of Plotly is that it allows data upload only in CSV and TSV formats.

Climate Explorer [5] is a tool developed by Habitat Seven [11] and hosted by NOAA's National Centers for Environmental Information. The tool allows users to explore historical and projected climate change data for all counties in the US. Users can use the map to view the data for all the counties or select a county by entering the county name, city, state, or zip code. After selecting a county, users can explore that area or select other nearby weather stations using the map. Temperature and precipitation information for the county is provided. In spite of all these very useful features, the usage of the tool is somewhat limited because it does not allow data upload or data extraction, and the graphs provided by the tool are not interactive. Users are only provided with a line chart or a map. The application is not open source, and there is no API for customizing or extending the visualizations.

Data Processing Toolset for the Virtual Watershed (DPT-VW) is another tool designed and implemented by our group to process environmental datasets [0]. The toolset is fairly comprehensive and some ideas used in its design have also been used in AVISTED. However, the DPT-VW toolset has limitations. For example, it does not support some useful file format conversions, such as ASCII and HDF5. Also, the visualization component is not very powerful. For example, the users cannot select a particular month or year for visualization. Additionally, DPT-VW allows only a single data file upload, which is a limitation if the user is interested in data contained in multiple files. These aspects of DPT-VW were improved in AVISTED.

6.2 GEAD as a New Software Engineering Design Notation

At present, there are many tools available for designing software applications. They can be classified into categories based on their features like general purpose, special purpose, specific language, code generation, executable UML, desktop application, mobile application, online application, open source, commercial and freeware.

The following are some of the current prevalent UML drawing tools: Rational Rose, SmartDraw, Microsoft Visio, LucidDraw, and ArgoUML. Some of these have support for activity diagrams and GUI builders, some have for building GUIs, and some have support for both activity diagrams and creating GUIs, like Visio, but the latter are static. None of them provide links between activity diagrams and GUI, which is an original idea of this paper's authors. These links are possible by building an HTML-based UML tool that allows drawing UML diagrams, designing GUIs, and providing HTML links between them. A similar HTML web application that allows drawing UML diagrams online is Diagramo [15]. Specifically, it allows building UML designs online but it does not support links. Tools such as Dreamweaver, Sublime, Coda 2, and Brackets 0 are used in building such web applications. Tenzer 0 developed

an interactive game to improve the UML designs. The user of the game is allowed to explore several variations of the designs while playing the game. This allows the users to improve their designs interactively.

6.3 Discussion

After comparing AVISTED with similar tools, we conclude that AVISTED fills the gap between general purpose and climate-related data visualization tools. The strength of AVISTED lies in its support for datasets which are in climate data formats. Also, the flexibility it offers to users in selecting a dataset, data variables, statistics, conversion, and visualizations is unique in climatology. Moreover, creating user accounts and giving users a choice of saving their visualizations and data selections is also not observed in other tools in this field.

On the other hand, AVISTED does not allow users to edit a dataset after uploading; with the large amount of datasets it supports, this will be a very complex task. Furthermore, at present the toolset does not support features such as editing visualizations, sharing visualizations on social media, or embedding them into another web application. These features, however, can be relatively easy to add to the toolset with some add-ons to the client-side JavaScript functions in the future. More details about a comparison of AVISTED with similar tools can be found in the thesis completed in 2018 by the first author of this paper [22].

7 Future Work and Conclusion

In this paper, we have introduced the new software application AVISTED, which has four main components: data upload, data extraction, data visualization, and data conversion. The software requirements and design models of AVISTED were described together with its main capabilities, an application scenario has been presented, and a comparison with related work and a discussion have been provided. We have also introduced the new GEAD software design approach and illustrated it as we applied it to AVISTED's development.

In the future, we would like to add several other useful features and functions to AVISTED, such as data processing using executable environmental models and comparisons of their results. We also intend to optimize the visualization component of AVISTED. Some preliminary work has been done and presented in our recent papers [0, 0]. Also, we plan to support other data formats and aim to add more datasets to AVISTED from other repositories. Furthermore, we would like to improve the usability of the application by involving the climate researchers, getting their feedback, and incorporating it in a future version of the application. In addition, more testing shall be conducted to evaluate the performance of the application when the number of its concurrent users increases. Lastly, we plan to work on the development of a supporting tool

for GEAD and refine the GEAD technique by using the new tool in software-intensive development projects.

Acknowledgment

This material is based upon work supported by the National Science Foundation under grant numbers IIA-1329469 and IIA-1301726. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

- [1] "ASP.NET MVC Overview," available at: [https://msdn.microsoft.com/en-us/library/dd381412\(v=vs.108\).aspx](https://msdn.microsoft.com/en-us/library/dd381412(v=vs.108).aspx), [Accessed November 4, 2017].
- [2] "AVISTED," available at: <https://github.com/likhitharavi/AVISTED/tree/master/A.VISTED>, [Accessed August 20, 2018].
- [3] C3.js| "D3-Based Reusable Library", available at: <http://c3js.org/>, [Accessed November 4, 2017].
- [4] "Choosing the Right Text Editor | Brackets, Sublime Text, Coda, and More," available at: <http://blog.digitaltutors.com/brackets-coda-sublime-text-text-editor-choose/>, [Accessed November 4, 2017].
- [5] "Climate Explorer," available at: <https://toolkit.climate.gov/climate-explorer2/>, NOAA [Accessed 28 June 2017].
- [6] "Climate Nexus," available at: <https://climatenexus.org/>, [Accessed June 25, 2018].
- [7] D3.js – "Data-Driven Documents," available at: <https://d3js.org/>, [Accessed November 4, 2017].
- [8] "DataONE," available at: ww.dataone.org, [Accessed June 25, 2108].
- [9] M. S. Deas, "Cal-Adapt and the Usability of Climate Adaptation Tools," Massachusetts Institute of Technology, 2015.
- [10] M. Dumas, and A. H. Ter Hofstede, "UML Activity Diagrams as a Workflow Specification Language," In UML 2185:76-90, October, 2001.
- [11] "Habitat Seven," Available: <http://habitatseven.com/>, [Accessed 28 June 2017].
- [12] "HDF5DotNet," available at: <http://hdf5.net/>, [Accessed November 4, 2017].
- [13] M. Hossain, H. Munoz, R. Wu, E. Fritzing, S. M. Dascalu, and F. C. Harris, "Becoming DataONE Tier-4 Member Node:Steps Taken by the Nevada Research Data Center", *Proceedings of Optimization of Electrical & Electronic Equipment Aegean Conference on Electrical Machines & Power Electronics (OPTIM-ACEMP 2017)*, Brasov, Romania, pp. 1089-1094, May 25-27, 2017.
- [14] M. Hossain, R. Wu, J. T. Painumkal, M. Kettouch, C. Luca, S. M. Dascalu, and F. C. Harris Jr., "Web-Service Framework For Environmental Models," *Proceedings of the IEEE Conference on Internet Technologies & Applications 2017 (ITA 2017)*, Wrexham, North Wales, UK, 6 pp, September 12-15, 2017.
- [15] "HTML5 Diagram Editor," available at: <http://diagramo.com/editor/editor.php>, [Accessed November 4, 2017].
- [16]"Modeling Output," available at: <http://sensor.nevada.edu/NCCP/Downloads/Modeling%20Output.aspx>, [Accessed November 4, 2017].
- [17] National Oceanic and Atmospheric Administration, U.S. Department of Commerce, available at: <http://www.noaa.gov/>, [Accessed 8 December 2017].
- [18] Nevada Research Data Center, available at: <http://www.sensor.nevada.edu/NRDC/>, [Accessed November 4, 2017].
- [19] "Nevada System Sponsored Programs and EPSCoR," available at: <https://epscorspo.nevada.edu/>, [Accessed November 4, 2017].
- [20] L. Palathingal, R. Wu, R. Belkhatir, S. M. Dascalu, and F. C. Harris, "Data Processing Toolset for the Virtual Watershed," *Proceedings of the 2016 International Conference on Collaboration Technologies and Systems (CTS 2016)*, Orlando, FL, pp. 281-287, October 31-November 4, 2016.
- [21] "Plotly," available at: <https://plot.ly/dashboards-and-reports/>, [Accessed December 6, 2017].
- [22] L. Ravi, *AVISTED: Analysis and Visualization Toolset for Environmental Data*, PhD dissertation, University of Nevada, Reno, 2018.
- [23] L. Ravi, S. Dascalu, and F. C. Harris, "GUI-Enhanced Activity Diagrams with Application to the Design of AVISTED," *Proceedings of the 24th International Conference on Software Engineering and Data Engineering (SEDE-2015)*, pp. 436-442, October 2015.
- [24] L. Ravi, S. Dascalu, F. C. Harris, J. Mejia, and N. Belkhatir, "VISTED: A Visualization Toolset for Environmental Data," *Proceedings of the 2015 International Conference on Computers and Their Application (CATA-2015)*, pp. 335-342, March, 2015.
- [25] Scientific Data-Set Library, available at: <https://www.microsoft.com/en-us/download/details.aspx?id=52412>, [Accessed November 4, 2017].
- [26] J. Tenzer, "Improving UML Design Tools by Formal Games," *IEEE Explore Proceedings of the 26th International Conference on Software Engineering, (ICSE-2004)*, 2004.
- [27] R. Wu, C. Chen, S. Ahmad, J. Volk, C. Luca, F. Harris, and S. Dascalu, "A Real-time Web-based Wildfire Simulation System," *Proceedings of the 2016 IEEE Industrial Electronics Conference (IECON 2016)*, Florence, Italy, pp. 4964-4969, Oct 24-27, 2016.
- [28] R. Wu, S. Dascalu, and F. Harris, "Environment for Datasets Processing and Visualization Using SciDB," *Proceedings of the 24th International Conference on Software Engineering and Data Engineering (SEDE 2015)*, San Diego, CA, pp. 223-229, October 12-14, 2015.
- [29] R. Wu, J. Painumkal, S. M. Dascalu, and F. C. Harris Jr, "Self-managed Elastic Scale Hybrid Server Using Budget Input and User Feedback," *Proceedings of the 12th Workshop on Feedback Computing, as part of the*

Proceedings of the 14th International Conference on Autonomous Computing, Columbus, Ohio, 6 pp., July 17-21, 2017.

- [30] R. Wu, J. T. Painumkal, N. Randhawa, L. Palathingal, S. R. Hiibel, S. M. Dascalu, and F. C. Harris, "A New Workflow to Interact with and Visualize Big Data for Web Applications," *Proceedings of the 2016 International Conference on Collaboration Technologies and Systems (CTS 2016)*, Orlando, FL, pp. 302-309, October 31-November 4, 2016.
- [31] R. Wu, J. T. Painumkal, J. M. Volk, S. Liu, S. J. Louis, S. Tyler, S. M. Dascalu, and F. C. Harris, "Parameter Estimation of Nonlinear Nitrate Prediction Model Using Genetic Algorithm," *Proceedings of the IEEE Congress on Evolutionary Computation (CEC 2017)*, San Sebastian, Spain, pp. 1893-1899, June 5-8, 2017.



Likhitha Ravi received in 2018 a PhD degree in Computer Science and Engineering from the University of Nevada, Reno, USA. She also received in 2007 a Master's degree in Computer Engineering from Texas A&M University, Kingsville, USA and in 2005 a Bachelor's degree in Computer Science and Information Technology from Jawaharlal Nehru Technology University, Hyderabad, India. Her main research interests are in the field of data analysis and data visualization over the web.



Eric R. Fritzinger received in 2003 a BS and in 2006 an MS, both in Computer Science, from the University of Nevada, Reno. After spending several years in the field of medical robotics, he returned to UNR to participate in a state-wide project studying the effects of climate change in the Great Basin, Nevada. He has worked on model and data interoperability as well as management and organization of environmental data. He is currently the lead developer for the Nevada Research Data Center, based out of UNR's Computer Science and Engineering Department.



Department of Computer Science and Engineering at the University of Nevada, Reno, USA, which he joined in 2002. He received in 1982 a Master's degree in Automatic Control and Computers from the Polytechnic University of Bucharest, Romania, and in 2001 a PhD in Computer Science from Dalhousie University, Canada. His main research interests are in the areas of software engineering and human-computer interaction. He has published over 180 peer reviewed papers and has been involved in projects funded by industrial companies as well as federal agencies such as NSF, NASA, and ONR.



Frederick C. Harris, Jr. is currently a Professor in the Department of Computer Science and Engineering and the Director of the High-Performance Computation and Visualization Lab and the Brain Computation Lab at the University of Nevada, Reno, USA. He received his BS and MS in Mathematics and Educational Administration from Bob Jones University in 1986 and 1988 respectively, and his MS and PhD in Computer Science from Clemson University in 1991 and 1994, respectively. He is a member of ACM (Senior Member), IEEE, and ISCA (Senior Member). His research interests are in parallel computation, computational neuroscience, computer graphics, and virtual reality.