University of Nevada, Reno

Data Processing Toolset for the Virtual Watershed Project

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Computer Science & Engineering

by

Lisa Palathingal

Dr. Sergiu Dascalu/Thesis Advisor

May, 2016
We recommend that the thesis prepared under our supervision by

LISA PALATHINGAL

Entitled

Data Processing Toolset For The Virtual Watershed Project

be accepted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Sergiu Dascalu, Ph.D., Advisor

Frederick C. Harris, Jr, Ph.D., Committee Member

Laurel Saito, Ph.D., Graduate School Representative

David W. Zeh, Ph.D., Dean, Graduate School

May, 2016
Abstract

The proposed Data Processing Toolset for the Virtual Watershed Project (DPT-VWP) is a user-centric web application that includes a set of new data processing tools for the NSF EPSCoR-supported Western Consortium for Water Analysis, Visualization and Exploration (WC-WAVE) project. The application includes an information section presenting an overview of the DPT-VWP, a tool selection section providing data processing tools, a tool request section enabling the user to request for a new tool to the developers, and a feedback section, where the users can submit their feedbacks and/or suggestions regarding the tools available in DPT-VWP. The tool selection section provides tools that convert input and output files of the Precipitation Runoff Modeling System (PRMS) model from text format into NetCDF format. The files are converted into NetCDF format for visualization purposes. There also exists tools that convert NetCDF format files back to text format. The text format files are required to run the PRMS model. The other tools available in the tool selection section can download the datasets of watersheds available from the Geographic Storage, Transformation and Retrieval Engine (GSToRE) and visualize the files in CSV format on a bar chart.

The thesis presents background on the WC-WAVE project, Ground-water and Surface-water FLOW (GSFLOW) model and data processing, outlines the specification and the requirements of the DPT-VWP, presents the application’s design and implementation, and provides details of DPT-VWP’s evaluation performed by peer software developers. It also contains a comparison with related work and provides several suggestions for future work.
Acknowledgements

First and foremost I would like to thank my advisor Dr. Sergiu Dascalu for inspiring and encouraging me to complete this work. Thank you for all your intellectual and emotional support throughout this thesis work especially during hard times. I am sure your advices will help me in my personal and professional life in the long run. I also thank my committee members Dr. Frederick Harris and Dr. Laurel Saito for their valuable intellectual inputs in improving the thesis work.

I would also like to thank the members who are working in this project, especially Matthew Turner, Chao Chen, Chase Carthen, John Volk, Moinul Hossain Rifat, and John Erickson.

I would like to thank my family members who, from across the miles, provided tremendous encouragement and support throughout my studies at UNR. I would also like to acknowledge my lab colleagues for their support, in particular Rui Wu, Eric Fritzinger, Vinh Le, and Royal Stewart.

Finally, I am thankful to the National Science Foundation for the support provided during my Master’s studies. The material presented in this thesis is based upon work supported by the National Science Foundation under grant number IIA–1329469. 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.
# Table of Contents

1  INTRODUCTION .................................................................1

2  BACKGROUND .......................................................................5
   2.1  WC-WAVE .....................................................................5
   2.2  GSFLOW ......................................................................6
   2.3  Data Processing..............................................................16
   2.4  NetCDF .......................................................................18

3  SPECIFICATION AND REQUIREMENTS OF DPT-VWP .............20
   3.1  Functional Requirements .................................................20
   3.2  Non-functional Requirements ...........................................22
   3.3  Use Case Modeling..........................................................23
      3.3.1  Use Case Diagram .....................................................23
      3.3.1  Detailed Use Case Templates ......................................27
   3.5  Requirement Traceability Matrix .......................................30

4  DESIGN AND IMPLEMENTATION OF DPT-VWP .................31
   4.1  System Level Diagram ....................................................31
   4.2  Data Design ..................................................................35
   4.3  Detailed Design .............................................................37
      4.3.1  Pseudo-code .............................................................37
      4.3.2  Flow Chart ..............................................................41
   4.4  Implementation of DPT-VWP ............................................43
      4.4.1  Home Page ..............................................................45
      4.4.2  Tool Selection Page ..................................................45
      4.4.3  Tool Request Page ....................................................50
4.4.4 Feedback Page ........................................................................................................51
4.4.5 Developer Page ....................................................................................................52
4.4.6 List of New Tool Requests Page ........................................................................54

5 DPT-VWP EVALUATION AND DEVELOPES’ FEEDBACK ...............................56
6 COMPARISON WITH RELATED WORK ...............................................................60
7 CONCLUSIONS AND FUTURE WORK ...............................................................66
REFERENCES .............................................................................................................68
**List of Tables**

Table 3.1  Functional Requirements of DPT-VWP ......................................................... 21
Table 3.2  Non-Functional Requirements of DPT-VWP ..................................................... 22
Table 3.3  Detailed Use Case Template for ConvertFile .................................................. 27
Table 3.4  Detailed Use Case Template for VisualizeFile .................................................. 28
Table 3.5  Detailed Use Case Template for SubmitRequest .............................................. 29
Table 3.6  Requirement Traceability Matrix ................................................................. 30
Table 5.1  ASQ Results ............................................................................................... 58
Table 6.1  Comparison between Data Processing Tools ................................................... 65
List of Figures

Figure 2.1 Hydrological Processes Simulated by the Precipitation Runoff Modeling System [13] ........................................................................................................................................7
Figure 2.2 Spatial Distribution of HRUs [Received from Chao Chen, PhD Student, Water Resource Engineering, University of Las Vegas, Nevada] .......... 8
Figure 2.3 Data File ........................................................................................................................................10
Figure 2.4 Parameter File ................................................................................................................................12
Figure 2.5 Statistic Variables File ......................................................................................................................14
Figure 2.6 Animation File ................................................................................................................................14
Figure 2.7 A Discretized Hypothetical Aquifer System [14] ............................................................................. 15
Figure 2.8 Stages in Data Processing ..................................................................................................................16
Figure 3.3 Use Case Diagram of DPT-VWP Showing Actors and Common System Usage Scenarios .................................................................25
Figure 4.1 DPT-VWP System Level Diagram ....................................................................................................31
Figure 4.2 DPT-VWP Modules ..........................................................................................................................32
Figure 4.3 Flow Chart of DPT-VWP Conversion Tool ......................................................................................41
Figure 4.4 Parameter NetCDF File ..................................................................................................................42
Figure 4.5 Visualization of Space and Time Related Parameter ......................................................................43
Figure 4.6 Flow Chart that Visualizes CSV File on a Bar Chart ........................................................................44
Figure 4.7 DPT-VWP Home Page ....................................................................................................................46
Figure 4.8 Tool Selection Page ..........................................................................................................................47
Figure 4.9 Conversion Tool in DPT-VWP ..........................................................................................................47
Figure 4.10 List of Watersheds Available from GSToRE ................................................................................49
Figure 4.11 List of Datasets Available from GSToRE ......................................................................................49
Figure 4.12 Visualization of a CSV File on a Bar Chart.................................50
Figure 4.13 Tool Request Page in DPT-VWP .................................................51
Figure 4.14 Feedback Page in DPT-VWP.......................................................52
Figure 4.15 Tool Request(s) Tab in Admin Home.........................................53
Figure 4.16 Feedback(s) Tab in Admin Home...............................................54
Figure 4.17 List of New Tool Requests Page in DPT-VWP............................55
Figure 5.1 Details of Participants ....................................................................56
Figure 5.2 After Scenario Questionnaire.........................................................57
Figure 6.1 Mr. Data Converter [40] ..................................................................60
Figure 6.2 Bar Chart Visualization in MATLAB [41] ......................................61
Figure 6.3 VISTED Home Page [43] .................................................................63
Figure 6.4 Bar Chart Visualization in Microsoft Excel....................................64
# Acronyms and Glossary

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Front-end</td>
<td>Front-end is visible to the user in the form of an interface. The main purpose of the front-end is to interact with the user in real time.</td>
</tr>
<tr>
<td>Back-end</td>
<td>Back-end is never visible to the user. The main purpose of the back-end is to interact with the server.</td>
</tr>
<tr>
<td>Data Processing</td>
<td>Data processing is the collection of data from various sources and producing meaningful information. It includes various activities such as validation, sorting, summarization, aggregation, analysis, reporting, and classification of data.</td>
</tr>
<tr>
<td>Watershed</td>
<td>A watershed or a drainage basin is the area of land where water from precipitation or snowmelt flows down to a point on a water body.</td>
</tr>
<tr>
<td>WC-WAVE</td>
<td>WC-WAVE is the short form for Western Consortium for Watershed Analysis, Visualization and Exploration. The goal of WC-WAVE is to create a Virtual Watershed (VW) framework for developing and integrating watershed models, thereby allowing the researchers to procure and integrate the data, visualize the data to obtain the results, and understand the environmental consequences of hydrologic changes.</td>
</tr>
<tr>
<td>GSFLOW</td>
<td>GSFLOW is the short form for Ground-water and Surface-water FLOW. GSFLOW is a coupled ground water and surface water flow model based on the integration of the USGS PRMS and the USGS MODFLOW models.</td>
</tr>
<tr>
<td>PRMS</td>
<td>PRMS is the short form for Precipitation Runoff Modeling System. PRMS is a deterministic, distributed-parameter, physical-process-based modeling system developed for evaluating the response of different combinations of climate and land use on streamflow and general watershed hydrology.</td>
</tr>
<tr>
<td>Term</td>
<td>Description</td>
</tr>
<tr>
<td>------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>MODFLOW</td>
<td>MODFLOW is the short form for Modular Groundwater Flow. MODFLOW is a three-dimensional finite-difference groundwater model that simulates steady and non-steady flow in an irregularly shaped flow system in which aquifer layers can be confined, unconfined, or a combination of confined and unconfined.</td>
</tr>
<tr>
<td>GSToRE</td>
<td>Geographic Storage, Transformation and Retrieval Engine (GSToRE) is a data framework developed by University of New Mexico aimed at data discovery, delivery and documentation.</td>
</tr>
<tr>
<td>NetCDF</td>
<td>NetCDF is the short form for Network Common Data Form. This format supports the creation, access, and sharing of array-oriented scientific data.</td>
</tr>
<tr>
<td>CSV</td>
<td>CSV is the short form for Comma Separated Values. It is a simple format for representing a rectangular array (matrix) of numeric and textual values.</td>
</tr>
<tr>
<td>D3</td>
<td>D3 is the short form for Data-Driven Documents. It is a JavaScript library for visualization of the data.</td>
</tr>
<tr>
<td>Flask</td>
<td>Flask is a cross-platform micro web application framework written in Python.</td>
</tr>
<tr>
<td>Python</td>
<td>Python is general-purpose, high-level programming language which emphasizes code readability.</td>
</tr>
<tr>
<td>SQLite</td>
<td>SQLite is a software library that implements SQL database engine. It is a cross-platform relational database management system written in C.</td>
</tr>
</tbody>
</table>
1 INTRODUCTION

A watershed or a drainage basin is the area of land where water from precipitation or snowmelt converges to a single point on a water body that is often a river, lake, or reservoir [1, 2]. Watersheds provide water and ecosystem services. To understand the hydrologic changes in the watersheds due to climate change [3,4], the Experimental Program to Stimulate Competitive Research (EPSCoR) [5] jurisdictions of Idaho, Nevada, and New Mexico have established the Western Consortium for Watershed Analysis, Visualization and Exploration (WC-WAVE) [6]. Various computer related modeling tools are used by the watershed scientists in order to model the physical processes in the watersheds [7]. The goal of WC-WAVE is to create a Virtual Watershed (VW) framework for developing and integrating watershed models that allows researchers to acquire and integrate data, visualize results obtained after running the hydrologic models, and understand the environmental consequences of hydrologic changes and effects of environmental changes on hydrology. In other words, the aim of this research is to provide a software environment for the watershed scientists to run their models in the cloud through web services and share their data, models, and results [8, 9].

To implement and evaluate the VW platform, project members selected four watersheds in the tri-state region to focus on: Dry Creek and Reynolds Creek Experimental Watershed in Idaho; Jemez watershed in New Mexico; and Lehman Creek in Nevada. One of the models to be run in the VW framework is the Ground-water and Surface-water FLOW (GSFLOW) model [10, 11, 12]. GSFLOW is based on the integration of the United
States Geological Survey (USGS) Precipitation Runoff Modeling System (PRMS) model and the USGS Modular Groundwater Flow (MODFLOW) model [13, 14].

The data required for running the models have to be stored in a suitable format in the VW. The users can then extract the input data, run the model, generate the output data and visualize them. NetCDF (Network Common Data Form) is considered as a self-describing data format which contains its own metadata. It is a group of software libraries and self-describing, machine-independent data formats which support the creation, access, and sharing of array-oriented scientific data [15, 16]. Because of its unique features, we planned to store all the data required for running the models in NetCDF format.

Data Processing Toolset for the Virtual Watershed (DPT-VWP) is a web application incorporating tools that aids the researchers, students and others in the processing of the data. The tools provided by DPT-VWP are useful for: i) converting PRMS input and output files into NetCDF format, ii) downloading the datasets of the watersheds available from the Geographic Storage, Transformation and Retrieval Engine (GSToRE) [17] framework in a Georeferenced Tagged Image File Format (GeoTIFF) [18], and iii) visualizing the files in Comma Separated Values (CSV) format [19] on a 2D bar chart. All these tools are available in the tool selection section in DPT-VWP. The tool request section in DPT-VWP permits the users to submit requests for the development of a new tool to the developers. The developers may either accept or reject their requests based on whether he or she would be able to develop the requested tool or not. DPT-VWP also includes a feedback section where users can submit feedback and/or suggestions regarding the tools available in DPT-VWP. Both the users and the developers can view all the feedbacks and/or suggestions submitted by the users.
Before starting our work, we interviewed three students (Chen, Volk, and Carthen) working on the virtual watershed project to understand the users’ goals and demands from the system because customer satisfaction is critical to the success of every project. We asked their opinion as a general user about developing the website, DPT-VWP and their experience of using data processing tools that are presently available as desktop/web-based applications. The interviewees have experience in using various types of data processing tools for the WC-WAVE project as well as their course projects. Some of the tools are climate data downloader, CSV to JavaScript Object Notation (JSON) data converter, Geospatial Data Abstraction Library (GDAL), Blender, ArcGIS and Microsoft office. Although the interviewees are satisfied with the results obtained from the data processing tools, Chen, one of the graduate students, indicated that searching for the right tools in the web browser is time consuming. “If the useful data processing tools are available from a single web application, that would be great”, she said. When we asked them about the input file formats that they would like to download from the web application for processing, NetCDF was the answer from all of them. They also mentioned the need of data converters that can generate NetCDF output files. Other formats mentioned by them are CSV, JSON, GeoTIFF and XML. Volk’s response to our question about any difficult experience of using tools because of the lack of tutorial was “I’ve had trouble trying to make 3D maps of a watershed. The documentation of Matplotlib, which is used for making 3D figures with multiple layers of data in Python, does not give all of the necessary information. Sometimes, I have to do a lot of experimenting or searching on stack overflow/exchange. Mayavi may be more suited for these types of maps or 3D figures but I have not had the time to learn it yet”.
When we told the students our attempt to create a web application that includes new as well as existing data processing tools, they were really excited. Carthen’s response was “It would be good if you can include data converters that can generate NetCDF files from PRMS input files of text format. Also, adding existing tools to the website is a good idea because finding the right tools for data processing takes time”.

The DPT-VWP application is developed using Flask [20, 21], a micro web application framework written in Python [22] and HTML5. D3 JavaScript library [23] is used for the visualization of CSV files. The tools for the conversion of PRMS files into NetCDF format are written using netcdf4-python library [24]. SQLite3 [25] is used as the back end [26]. Because the VW platform for the WC-WAVE project is being developed using Flask, the same framework is used for implementing DPT-VWP for maintaining the team compatibility.

All the above are further detailed in this thesis. In its remaining parts, the thesis is structured as follows: Chapter 2 provides a background of the WC-WAVE project, GSFLOW model, data processing, and NetCDF; Chapter 3 presents the specification and requirements of DPT-VWP; Chapter 4 depicts the design and implementation of DPT-VWP, including through a variety of screenshots of the application’s user interface; Chapter 5 details the software developers’ evaluation and professional feedback; Chapter 6 contains a comparison with the related work; and, finally, Chapter 7 completes the thesis with details on planned future work and the main highlights of our work.
2 BACKGROUND

Because this thesis is focused on developing data processing tools for the virtual watershed, a brief background on WC-WAVE, GSFLOW, along with PRMS and MODFLOW models, data processing, and NetCDF are presented.

2.1 WC-WAVE

Watersheds provide water and ecosystem services to communities. Climate change may affect the hydrological services provided by watersheds. The WC-WAVE is a tri-state consortium composed of researchers and students in Idaho, Nevada, and New Mexico. Its overall aim is to understand the localized impact of climate change on high-mountain catchments by establishing a lasting and flexible watershed modeling framework to easily acquire and integrate data, use an integrated suite of models to discover processes linking components of the hydrologic cycle, to identify environmental consequences of hydrologic changes, and to visualize and interpret data and model results. The virtual watershed (VW) will be tested from four watersheds: Jemez, NM; Lehman Creek, NV; Reynolds Creek, ID; and Dry Creek, ID [27].

The three integrated components of WC-WAVE are Watershed Science, Visualization and Data Cyberinfrastructure (CI), and Workforce Development. Participants in these three components are collaborating to better understand interactions between precipitation, snow-pack, ground water flow, and other watershed properties within mountain catchments. The goal of Watershed Science component is to understand hydrologic interactions and their effect on ecosystem services using a VW framework. CI
Visualization team, along with the CI Data team, aims at accelerating collaborative, interdisciplinary watershed research and discovery through visualization environments and data management, discovery, and access. The CI Data team provides a user friendly VW platform that supports data analysis, conversion, modeling, and visualization activities. The goal of Workforce Development component is to engage university faculty and graduate students in interdisciplinary team-based watershed research [28].

2.2 GSFLOW

As described in the GSFLOW manual [10], “GSFLOW is an integrated hydrologic model that was developed to simulate coupled groundwater and surface-water resources”. The models that couple two or more components of the hydrologic cycle are required to assess the effects of variability in climate, biota, geology, and human activities on water availability and flow. GSFLOW is based on the integration of the USGS PRMS and the USGS MODFLOW. GSFLOW can be used to verify the factors such as land use change, climate variability and groundwater withdrawals on surface and subsurface flow [29].

As described in PRMS manual [13], “PRMS is a deterministic, distributed-parameter, physical-process-based modeling system developed for evaluating the response of different combinations of climate and land use on streamflow and general watershed hydrology”. PRMS model is developed to simulate and evaluate the watershed response of different combinations of climate and land use. In other words, it is a watershed scale, physically based, distributed-parameter model designed to simulate precipitation and snowmelt runoff. Response to rainfall and snowmelt can be simulated to estimate changes in water-
balance relations, soli-moisture relations, and groundwater recharge. Inputs to this hydrologic model are daily time-series values of precipitation, minimum and maximum air temperature, and short wave solar radiation [13]. Figure 2.1 shows the hydrological processes simulated by the PRMS.

**Figure 2.1: Hydrological Processes Simulated by the Precipitation Runoff Modeling System [13]**

The watershed area is divided into Hydrologic Response Units (HRUs). This division depends upon the hydrologic and physical characteristics such as drainage boundaries, land surface altitude, slope, and aspect; plant type and cover; land use; distribution of precipitation, temperature, and solar radiation; soil morphology and geology; and flow direction. Each HRU is identified by a numerical index. The HRU index starts with 1 [13]. Figure 2.2 shows the spatial distribution of HRUs in PRMS output and input files. As shown in the figure, there are 4704 grid cells (means that the model is on 49 x 96 cell grid)
Figure 2.2: Spatial Distribution of HRUs [Received from Chao Chen, PhD Student, Water Resource Engineering, University of Las Vegas, Nevada]

where each cell represents one location of the watershed, and there is one value for each cell.

The input data for a PRMS simulation are specified in various files that should be prepared before simulation. The input files required for the PRMS simulation are the control file, the data file, and the parameter file. All the files are in text format. The data file and the parameter file are required for running the model. In order to visualize these files, the files are converted into NetCDF format.

The control file controls the overall simulation of the PRMS model. This file sets the basic administrative data values related to the PRMS modules. It specifies the input and output file names, the file path, the simulation starting and ending dates, and the active modules. The data file includes the measured time-series climate data used in a PRMS simulation. The values of the one-dimensional data vary with time. The time-series data may include the daily precipitation, maximum and minimum air temperatures, solar
radiation, pan evaporation, measured streamflow, humidity, wind speed, and snowpack-water equivalent [30]. Figure 2.3 shows an example of a data file. As shown in the figure, the data file has a header, input variable declaration items, and time series data items.

The header describes the name of the file. The input variable declaration items specify two values: i) a character string that is the name of the input variable, and (ii) an integer value that is the number of values (or columns) specified in each time series data item for the input variable. The last section represents the time series data items. The first six columns represent the time step of each time-series data item as integer values in the order: year, month, day, hour, minute, and second. The values for hour, minute, and second must be specified as zero (that is, columns 4 through 6 must be 0 0 0). The remaining columns for each time-series data item specify the data values. The number of remaining columns is equal to the sum of integer values specified in the input variable declaration items. For example, in Figure 2.1, the sum of integer values in the input variable declaration items is 4. So, the number of remaining columns is equal to 4 [30].

Along with this, the file also includes metadata such as unit, station ID, latitude, longitude, and elevation values for each input variable [30]. The variables created in the NetCDF file are time and the input variables specified in the data file. Because the values of the variables change with time, the NetCDF file includes time as the dimension.
The parameter file contains the parameters that control each of the hydrologic processes. It includes the values of parameters specified for each module that do not change during a simulation. In the parameter file, the values of the one-dimensional data vary with space. Some parameters are space related (For instance, carea_min). There will be 4704 values for the space related parameter in total. These parameters do not have time features.

Some parameters are time related (For instance, wrain_intcp). There are 12 months in a year, so there is a value for each month. There will be 12 values for the time related parameter in total. There parameters do not have space features. Some parameters are both space and time related (For instance, rain_adj and snow_adj). So, for each month, there is a value for each of the cell. There will be 12 x 4704 values for the space and time related parameters in total. These parameters have both space and time features.
Figure 2.4 shows an example of a parameter file. The parameter file consists of three sections: header, dimensions, and parameters. The header section indicates that the file is a parameter file. The dimensions section define the size of dimensions that are used for allocating memory for parameters required by the PRMS modules. The dimensions section begins with an identifier (** Dimensions **). Each dimension is declared in three lines. The first line has a string of four pound signs (####) which indicates the beginning of a dimension declaration. The second line is a character string that is the name of the dimension. The third line is the dimension size, specified as an integer value [30].

Similar to the dimensions section, the parameters section begins with an identifier (** Parameters section **). In the parameter declaration, the first line has a string of four pound signs (####) which indicates the beginning of a parameter declaration. The second line is a character string that is the name of the parameter. The third line is the number of dimensions for the parameter, specified as an integer value. Line 4 specifies a character string that is the name of the dimension. The fifth line indicates the total number of values for the parameter. Line 6 indicates the data type of the parameter values: 1 (integer), 2 (real), 3 (double), and 4 (character string). The remaining lines contain the parameter value(s) [30]. For instance, in Figure 2.4, after indicating the beginning of a parameter declaration with four pound signs, the name of the parameter is tmax_allrain. The number of dimensions is 1. The name of the dimension is nmonths. The total number of values for the parameter is 12, which indicates that tmax_allrain is a time related parameter. The type of the parameter values is 2 which implies that the values for the parameter are real. The remaining 12 lines represents the parameter values.
Most of the parameters are single dimensional arrays (that is, the number of dimensions for the parameter is always 1). It also includes two dimensional arrays (that is, the number of dimensions for the parameter is 2). Some examples of parameters consisting of two-dimensional arrays are rain_adj and snow_adj. The variables created in the NetCDF file are latitude, longitude and the parameters specified in the parameter file. Because the
values of the variables change with time, space or both time and space, the NetCDF file includes time, latitude, and longitude as the dimensions.

The output data from a PRMS simulation can be written into several files. The output files that are generated after running the PRMS model are the statistic variables file, the animation variables file, and the water budget file. The statistic variables file (aka statvar file) includes one dimensional data that vary with time. Figure 2.5 shows an example of statvar file.

Line 1 specifies the number of variable values that are written in the file. The next few lines indicate the names and array indices of each output variable. In Figure 2.5, the first line has the value 4. Hence, the next section indicates the names and array indices of 4 output variables. The remaining section provides the model calculated values. They are year, month, day, hour, minute, second, and the value of each variable in the order specified by the list of variable names [30]. The variables created in the NetCDF file are time and the variables specified in the statvar file. Because the values of the variables change with time, the NetCDF file includes time as the dimension.

The animation file generates the two dimensional temporal and spatial data (means that the data changes with time and space). Figure 2.6 shows an example of animation file. The first few lines indicate the metadata such as the file format and the contents. The lines with metadata starts with pound characters (#). The next line indicates the list of variable names that are separated by tab spaces. In Figure 2.6, the list of variable names are shown in line 9. The remaining section provides the model calculated values. The timestamp is specified as a 19 character string in the format YEAR-MO-DY:HR:MN:SE. The value of each variable is written in the order specified by the list of variable names [30]. The variables
created in the NetCDF file are time, latitude, longitude, and the variables specified in the animation file. Because the values of the variables change with time, space or both time and space, the NetCDF file includes time, latitude, and longitude as the dimensions.
The water budget file generates a summary table of the water budget for a PRMS simulation. There are three types of summary tables: i) a listing of the measured and simulated flow; ii) a table of water balance computations that includes the area weighted average of net precipitation, evapotranspiration from all sources, storage in all reservoirs, and the simulated and measured stream flows; iii) a detailed summary including several important states and fluxes [30].

As described in the manual [14], “MODFLOW is a three-dimensional (3D) finite-difference groundwater model that simulates steady and non-steady flow in an irregularly shaped flow system in which aquifer layers can be confined, unconfined, or a combination of confined and unconfined”. An aquifer system, as shown in Figure 2.7, is discretized with a finite difference grid that represents the computational cells used to calculate groundwater heads and flows.

![Figure 2.7: A Discretized Hypothetical Aquifer System [14]](image)
Finite-difference cells are numbered according to the indices $i$, $j$, and $k$, where $i$ is the row index, $j$ is the column index, and $k$ is the layer index. Horizontal discretization is defined as a rectangular grid of rows and columns. Rows are aligned parallel to the $x$ axis. Columns are aligned parallel to the $y$ axis. Vertical discretization is defined by layers. Layers of cells are aligned parallel to the horizontal plane. Layers can have uniform thickness or can have variable thickness among cells in a layer [14].

### 2.3 Data Processing

Data processing is the conversion of raw data into a meaningful information [31]. The various stages of data processing cycle are: i) collection, ii) preparation, iii) input, iv) processing, v) output, and vi) storage [32]. The data processing stages are shown in Figure 2.8.

![Figure 2.8: Stages in Data Processing](image)
The collection of data is important because the quality of data will impact heavily on the output. Preparation of data is the manipulation of data into a form appropriate for further analysis and processing. Input is the stage where verified data is converted into a machine readable form so that it can be processed through a computer. In processing stage, the data is subjected to various means and methods of manipulation. Output is the stage where processed information is transmitted to the user in various formats. The last stage, storage is where the data, instruction and information are held for future use. The various activities that are involved in the data processing include [33]:

- Selection – selecting the raw input data from various sources.
- Organization – grouping the data into various categories.
- Sorting – arranging the data in a logical order.
- Validation - ensuring that the data is clean.
- Summarization - Reducing the detailed data into key points.
- Addition, selection, modification, extraction, and deletion of the data from the relational databases.
- Conversion of the data from one format to another as required by the intended recipient.
- Downloading the data in various formats.
- Data visualization - the visual representation of the data in order to communicate the information clearly and efficiently to the users.
- Data compression - transmitting or storing the same amount of data in fewer bits.

The various methods of data processing include batch processing, online processing, real time processing, multiprocessing, multiprogramming, and time sharing [33]. Batch
processing is also known as sequential, serial, offline or stacked/queued processing. In batch processing, large loads of data are handled at specific times. It runs the program only once for many transactions. Online processing is also known as direct access or random access processing. In online processing, a job is processed at the same time when it is received. Real time processing is quick and efficient. In this processing, receiving and processing of transactions are performed simultaneously. Multiprocessing is the simultaneous execution of several instructions. It distributes processing load among various CPUS. Multiprogramming is the simultaneous execution of programs on one computer system. In time sharing, CPU time is divided among all the users on a scheduled basis [33].

DPT-VWP website deals with conversion, download and visualization of data. It converts the files required for running the watershed models into NetCDF format for visualization and vice-versa. It downloads datasets of watersheds available from GSToRE. It also visualizes CSV data on a bar chart.

2.4 NetCDF

A NetCDF file is created in python using Dataset constructor. A NetCDF Dataset is a collection of dimensions, variables, and attributes. This collection is used to describe the meaning of data stored in a NetCDF file. The size of all the variables are defined in terms of dimensions. Hence, dimensions used by the variables should be created before the variables are created. A python string is used to set the name of the dimension, and an integer value is used to set the size of the dimension. The createDimension method of
Dataset is used to create a NetCDF dimension. A variable can have more than one dimension [24]. The files that contain the temporal (time related) data include time as a dimension in the NetCDF file. The files that contain the spatial (space related) data include latitude and longitude as the dimensions in the NetCDF file.

The createVariable method of Dataset is used to create a NetCDF variable. The arguments of the createVariable method are the variable name, variable datatype and dimension names of the variable. A python string is used to set the name of the variable. The datatype can be ‘f4’ (32-bit floating point), ‘f8’ (64-bit floating point), ‘i4’ (32-bit signed integer), ‘i2’ (16-bit signed integer), ‘i8’ (64-bit signed integer), ‘i1’ (8-bit signed integer), ‘u1’ (8-bit unsigned integer), ‘u2’ (16-bit unsigned integer), ‘u4’ (32-bit unsigned integer), ‘u8’ (64-bit unsigned integer), or ‘S1’ (single-character string). The dimensions of the variable are given by a tuple containing the dimension names. The dimensions themselves are usually also defined as variables, called coordinate variables [24].

The two types of attributes in a NetCDF file are global and variable. The global attributes provide detailed information about the entire dataset. The variable attributes provide detailed information about one of the variables in a dataset. Global attributes are set by assigning values to Dataset instance variables. Variable attributes are set by assigning values to Variable instances variables [24].
3 SPECIFICATION AND REQUIREMENTS OF DPT-VWP

3.1 Functional Requirements

In this section, functional requirements are provided. Functional requirements help to define what the system should accomplish in operation [34, 35, 36]. These functional requirements as well as a brief description of what each requirement entails are provided to help the reader understand the web application. The functional requirements are shown in Table 3.1. Those listed with a label of ‘1’ were implemented in the DPT-VWP web application. Requirements with a label of ‘2’ may be added in the web application. A label of ‘3’ indicates features that have not yet been implemented, but would be useful in future work.

Here are some examples about the functional requirements: If the user wants to convert or visualize a file and upload a file of incorrect format, the system will pop up a warning message as described in R12 of Table 3.1. If user wants to visualize a file, he/she should upload file of CSV format. After uploading the CSV file, the user can visualize the data on a bar chart. Similarly, the user should upload a file of text format to convert it into NetCDF format as described in R02 of Table 3.1.

DPT-VWP permits the users to submit requests for the development of a new tool to the developers (R08). The developers may either accept or reject their requests based on whether he or she would be able to develop the requested tool or not (R07). The user can view the status (Accepted / Rejected / To be reviewed) submitted by the developers (R13).
### Table 3.1: Functional Requirements of DPT-VWP

<table>
<thead>
<tr>
<th>Requirement</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>R01 [1]</td>
<td>DPT-VWP shall allow the user to upload an input file for conversion and visualization.</td>
</tr>
<tr>
<td>R02 [1]</td>
<td>DPT-VWP shall convert PRMS and MODFLOW input and/or output files in text (.txt) format to NetCDF (.nc) format.</td>
</tr>
<tr>
<td>R03 [1]</td>
<td>DPT-VWP shall download the output file after conversion.</td>
</tr>
<tr>
<td>R04 [1]</td>
<td>DPT-VWP shall visualize files in CSV format on a bar chart.</td>
</tr>
<tr>
<td>R05 [1]</td>
<td>DPT-VWP shall allow the user to submit request for a new tool to the developers.</td>
</tr>
<tr>
<td>R06 [1]</td>
<td>DPT-VWP shall display all the tool requests to the users as well as the developers.</td>
</tr>
<tr>
<td>R07 [1]</td>
<td>DPT-VWP shall allow the developers to either accept or reject the requests submitted by the users.</td>
</tr>
<tr>
<td>R08 [1]</td>
<td>DPT-VWP shall allow the user to submit feedback and/or suggestion regarding tools available in the website to the developers.</td>
</tr>
<tr>
<td>R09 [1]</td>
<td>DPT-VWP shall display all the feedbacks and/or suggestions to the developers.</td>
</tr>
<tr>
<td>R10 [1]</td>
<td>DPT-VWP shall display a video showing an overview of the website to the users in the home page.</td>
</tr>
<tr>
<td>R11 [1]</td>
<td>DPT-VWP shall display the names of data processing tools provided by the website to the users.</td>
</tr>
<tr>
<td>R12 [1]</td>
<td>DPT-VWP shall display an error message if the user uploads an input file of incorrect format for either conversion or visualization.</td>
</tr>
<tr>
<td>R13 [1]</td>
<td>DPT-VWP shall display a status (Accepted / Rejected) indicated by the developers for each question submitted by the user.</td>
</tr>
<tr>
<td>R14 [1]</td>
<td>DPT-VWP shall allow the developers to login into the website.</td>
</tr>
</tbody>
</table>
R15 [1] DPT-VWP shall provide existing data processing tools in the tool selection section.

R16 [2] DPT-VWP should be able to visualize files in CSV format on a line chart.

R17 [3] DPT-VWP may visualize files in NetCDF format on a line chart and bar chart.

R18 [3] DPT-VWP may include 3D visualization tools in the tool selection section.

### 3.2 Non-Functional Requirements

Non-functional requirements presents restrictions and constraints on the system’s implementation [34, 35, 37]. These requirements, which can be found in Table 3.2 below, define performance, security, and other technical qualities of the system. We used the Flask, a python framework, to build DPT-VWP. Therefore, all the backend codes are written in python. JavaScript libraries for data visualization, such as D3 is used for implementing DPT-VWP. We developed the website using JavaScript and HTML for using these libraries.

**Table 3.2: Non-Functional Requirements of DPT-VWP**

<table>
<thead>
<tr>
<th>NR01</th>
<th>DPT-VWP shall be implemented with Flask, HTML5, JavaScript, and D3 library.</th>
</tr>
</thead>
<tbody>
<tr>
<td>NR02</td>
<td>DPT-VWP shall maintain a simple web interface.</td>
</tr>
<tr>
<td>NR03</td>
<td>DPT-VWP shall run on the Linux operating system.</td>
</tr>
<tr>
<td>NR04</td>
<td>DPT-VWP shall use SQLite3 as the backend.</td>
</tr>
</tbody>
</table>
3.3 Use Case Modeling

Use cases describe the desired behavior of a system through interaction with a user. Use cases can be demonstrated both graphically and textually. This section includes a use case diagram, detailed use case templates, and a requirements traceability matrix.

3.3.1 Use Case Diagram

Figure 3.3 displays DPT-VWP use case diagram. It is very useful to introduce DPT-VWP from the users’ view. The actor on the left side of Figure 3.3 represents our system users. The big rectangle on the right side of Figure 3.3 represents DPT-VWP. All the ellipses in the rectangle are the functions offered to users by DPT-VWP.

UC01: ViewVideo
The home page of DPT-VWP gives an overview of the website along with a video to the users. The video gives the users an idea regarding how to use the various data processing tools provided by the website.

UC02: UploadFile
Before using the data processing tools, the user wants to upload an input file of specified format. When the user clicks the ‘upload’ button, DPT-VWP pops up a window that enables the user to choose a file to upload. If the file is of the supported format, the file is uploaded to the server.

UC03: ConvertFile
When the user uploads a PRMS file of text format, DPT-VWP converts the file into NetCDF format. The PRMS model has three input files (control, data and parameter) and
three output files (statistic variables, animation and water budget). All these files are in text format. The converted NetCDF file includes dimensions, variables, and attributes. If the uploaded file is of wrong format, error message is displayed.

**UC04: VisualizeFile**

When the user uploads an input file of CSV format, DPT-VWP visualizes the data in the file on a bar chart. If the file is of unsupported format, DPT-VWP displays an error message.

**UC05: DownloadFile**

After performing the conversion of PRMS files from text into NetCDF format, DPT-VWP allows the user to download the output file of NetCDF format. Also, DPT-VWP enables the user to download the datasets of the watersheds available from the GSToRE framework in a GeoTIFF format. Using the existing tools available in DPT-VWP, users can download file in JSON (converted from CSV format) and Tab Separated Values (TSV) (converted from CSV format) formats.

**UC06: UseExistingTools**

DPT-VWP provides existing data processing tools to the end user. All the existing data processing tools are available in the tool selection section. Some examples of existing tools available in the tool selection section are CSV to JSON and CSV to TSV converters.

**UC07: SubmitRequest**

If the user wants additional tools on the DPT-VWP website, he/she can submit a request for the new tools to the developers. When the user selects the ‘tool request’ option, he/she will be redirected to a web page which includes three questions – name, email and question.
The user wants to fill out the answers for these questions and select the ‘submit’ button. All the answers will be stored in the database.

**UC08: SubmitFeedback**

If the user wants to let the developers know his/her opinions and/or suggestions about the DPT-VWP website, he/she can submit the feedback to the developers. When the user selects the ‘feedback’ option, he/she will be redirected to a web page which includes two
questions – name and feedback. The user will also be asked to rate the website on a scale of 1 to 5. The user wants to fill out the answers for these questions and select the ‘submit’ button. All the answers will be stored in the database.

**UC09: AcceptOrReject**

The developers, after logging into the DPT-VWP website, can view all the tool requests submitted by the users. The developers can either accept or reject the tool requests. The status (Accepted / Rejected) of the tool requests will be added to the database. The user can then view the status of the tool requests by selecting the ‘list of new tool requests’ option.

**UC10: ViewStatus**

When the user selects the ‘list of new tool requests’ option, he/she can view the list of tool requests submitted by all the users. It includes the total number of requests, number of requests accepted and rejected by the developers, and the number of requests not viewed by the developers yet. Also, it will display the list of tool requests.

**UC11: LogIn**

The developers have the option to log in into the DPT-VWP website and view the tool requests and feedbacks submitted by the users. The developers has to submit the correct username and password to successfully log in and view the tool requests and/or feedbacks.

**UC12: ReceiveMessage**

DPT-VWP website displays messages to the user in various situations. Successful messages will be displayed after (i) converting the input file of text format into a NetCDF format (ii) downloading the datasets of the watersheds available from GSToRE. Error messages will be displayed if (i) the developers enter a wrong username and/or password, (ii) the user uploads a file of format not supported by DPT-VWP.
3.3.2 Detailed Use Case Templates

The use case templates express the flow of events as the user works with the system in different activities. Tables 3.3, 3.4 and 3.5 contain a primary scenario that details the steps taken when the DPT-VWP UC03, UC04, and UC07 are used as intended, with no complications, respectively.

Table 3.3: Detailed Use Case Template for ConvertFile

<table>
<thead>
<tr>
<th>Use Case ID</th>
<th>UC03</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actor</td>
<td>User, Developer</td>
</tr>
<tr>
<td>Precondition(s)</td>
<td>None</td>
</tr>
</tbody>
</table>
| Flow of Events | 1. The user selects the ‘tool selection’ page.  
2. The user selects the ‘data converters’ option.  
3. The user will be shown two categories – PRMS and Other Tools.  
4. The user selects PRMS.  
5. The user will be shown ‘text to NetCDF’ and ‘NetCDF to text options.  
6. The user selects ‘text to NetCDF’ option.  
7. The user will be shown ‘control/data/statistic variables/water budget’ and ‘parameter/animation’ options.  
8. The user selects ‘parameter/animation’ option.  
9. The user uploads a parameter file of text format, a location file which has latitude and longitude values of each HRU, and enter the number of rows and columns (49 and 96 |
respectively which means that the model is on a 49 by 96 cell grid).

10. The user selects the ‘convert’ option.

11. DPT-VWP converts the file from text format into NetCDF format.

12. After performing the conversion, the file in NetCDF format is stored in the local server.

| Postcondition(s) | The output file is stored in the local server. |

**Table 3.4: Detailed Use Case Template for VisualizeFile**

<table>
<thead>
<tr>
<th>Use Case ID</th>
<th>UC04</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actor</td>
<td>User, Developer</td>
</tr>
<tr>
<td>Precondition(s)</td>
<td>None</td>
</tr>
</tbody>
</table>
| Flow of Events | 1. The user selects the ‘tool selection’ page.  
2. The user selects the ‘data visualizers’ option.  
3. The user uploads an input file of CSV format.  
4. The user will be shown the labels of each column specified in the file in checkboxes.  
5. The user selects two labels from the checkboxes.  
6. The user selects the ‘visualize’ option.  
7. DPT-VWP visualizes the csv format file on a bar chart. |
| Postcondition(s) | None |
### Table 3.5: Detailed Use Case Template for SubmitRequest

<table>
<thead>
<tr>
<th>Use Case ID</th>
<th>UC07</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actor</td>
<td>User</td>
</tr>
<tr>
<td>Precondition(s)</td>
<td>None</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Flow of Events</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. The user selects the ‘tool request’ page.</td>
</tr>
<tr>
<td>2. The user will be asked to enter his/her name, email address, and a question (request for a new tool).</td>
</tr>
<tr>
<td>3. The user selects the ‘submit’ button.</td>
</tr>
<tr>
<td>4. The developers log in into the website and views the request(s) submitted by the user(s).</td>
</tr>
<tr>
<td>5. The developers either accept or reject the request(s).</td>
</tr>
<tr>
<td>6. The user can view the status of the request – Accepted/Rejected/ To be reviewed.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Postcondition(s)</th>
<th>None</th>
</tr>
</thead>
</table>
### 3.4 Requirement Traceability Matrix

The requirement traceability matrix, shown below in Table 3.14, maps the previously listed requirements to corresponding use cases that meet those requirements.

**Table 3.14: Requirement Traceability Matrix (UC = Use Case, R = Requirement)**

<table>
<thead>
<tr>
<th></th>
<th>UC 01</th>
<th>UC 02</th>
<th>UC 03</th>
<th>UC 04</th>
<th>UC 05</th>
<th>UC 06</th>
<th>UC 07</th>
<th>UC 08</th>
<th>UC 09</th>
<th>UC 10</th>
<th>UC 11</th>
<th>UC 12</th>
</tr>
</thead>
<tbody>
<tr>
<td>R01</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R02</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R04</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R05</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R06</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R07</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R08</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R09</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>R10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>R11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>
4 DESIGN AND IMPLEMENTATION OF DPT-VWP

4.1 System Level Diagram

DPT-VWP website consists of multiple subsystems that include various functionalities of the system. The system-level diagram of DPT-VWP is shown in Figure 4.1.

Information/overview subsystem provides an overview of DPT-VWP with a video describing the tools provided by the system. The tool selection subsystem includes three categories: data conversion, data download and data visualization. Data conversion section includes tools that perform the conversion of PRMS and MODFLOW files from text format.
to NetCDF format and vice versa. Some of the existing data converters are also available in DPT-VWP. Data download section includes tools that aids in downloading the datasets available from GSToRE. Data visualization section includes tools that performs 2D visualization of user input data in CSV format on bar charts. The tool request subsystem in DPT-VWP permits the users to submit requests for the development of a new tool to the developer. The developers may either accept or reject their requests based on whether they would be able to develop the requested tool or not. DPT-VWP also includes a feedback section where users can submit feedback and/or suggestions regarding the tools available in DPT-VWP. The developer, after logging in to the application, can view all the feedbacks and/or suggestions submitted by the users. Figure 4.2 shows the various modules of DPT-VWP followed by a detailed description of each module in Table 4.1.

Figure 4.2: DPT-VWP Modules
MainMenu

MainMenu module (DPT-VWP home page) includes the links of all the functions provided by the DPT-VWP website. These functions are information, tool selection, feedback, and tool request system.

Overview

Overview module provides the description of the website. It shows a video to the end user explaining the tools provided by DPT-VWP. When the user selects information link from the main menu, he/she is redirected to this module. When the user selects the play option in the video, the video starts playing.

DisplayVideo

When the user selects information section from the main menu, he/she is redirected to DisplayVideo module. When the user selects the play option in the video, the video starts playing.

SelectATool

When the user selects tool selection from the main menu, he/she is redirected to SelectATool module. The module displays three options: data downloader, data converter, and data visualizer. When the user selects one of these options, he/she will be redirected to the corresponding webpage.

ConvertData

ConvertData module displays the tools that aids in converting the PRMS and MODFLOW files in text format to NetCDF format and vice versa. It also includes tools that are already available online. When the user selects a tool, he/she will be redirected to the corresponding webpage of the selected tool.
**DownloadData**

DownloadData module displays the tools that aids in downloading the datasets from GSToRE. When the user selects a name of watershed, the web page displays the dataset names. When the user selects a dataset name, the dataset is downloaded in GeoTIFF format.

**VisualizeData**

VisualizeData module performs 2D visualization of files uploaded by the user. When the user uploads a file of CSV format, the data is visualized on a bar chart.

**TextToNetCDF**

In TextToNetCDF module, the user will be asked to upload an input file of text format. If the file is of the format supported by the module, the tool converts it to NetCDF format file.

**NetCDFToText**

In NetCDFToText module, the user will be asked to upload an input file of NetCDF format. If the file is of the format supported by the module, the tool converts it to a text format.

**DownloadFromGSToRE**

DownloadFromGSToRE module downloads datasets of the watersheds available from GSToRE in GeoTIFF format.

**VisualizeCSV**

VisualizeCSV module visualizes files in CSV format on a bar chart. The user will be shown the labels of each column specified in the file in checkboxes. When the user selects two labels from the checkboxes and clicks the ‘visualize’ button, DPT-VWP visualizes the csv format file on a bar chart.
SubmitAFeedback

SubmitAFeedback module displays a webpage asking the user to fill out his/her name, feedback, and rating of the website on a scale of 1 to 5. The developers after logging into the website can view all the feedbacks from the users.

RequestANewTool

RequestANewTool module displays a webpage asking the user to fill out his/her name, email address and request for new tool.

AcceptOrRejectRequest

The developer, after logging into the website, can view all the requests from the users. The developers can either accept or reject the requests. After the developers assign a status to the requests, the users can view the developers’ status.

4.2 Data Design

The back end used by DPT-VWP is SQLite3. The database includes four tables: watersheds, datasets, requests and feedbacks.

The watersheds table stores the details of the watersheds available from GSToRE. The primary key of the watersheds table is the id. The fields in the table are:

id | name | uuid

where name is the name of the watershed, uuid (Universally Unique Identifier) is the unique id of the watershed. When the user selects the ‘Data Downloaders’ option from the ‘Tool Selection’ section, the names of all the watersheds are displayed. When the user selects a name from the list, the user is redirected to the next web page.
The datasets table stores the details of the datasets of watersheds available from GSToRE. The primary key of the datasets table is the id. The fields in the table are:

| id | wuuid | name | duuid |

where wuuid is the unique id of the watershed, name is the name of the watershed, duuid is the unique id of the dataset. When the user selects a name of watershed from the list, the names of the datasets of the selected watershed are extracted from the ‘datasets’ table and are displayed on the web page.

The requests table stores the requests submitted by the users to the developer. The primary key of the requests table is the id. The fields in the table are:

| id | name | email | request | status |

When the user submits the request, the status will be stored as ‘N/A’. After the developers log into the website, they can either accept or reject the request. Then, the status will be updated to either ‘Accepted’ or ‘Rejected’.

The feedbacks table store the feedbacks submitted by the users to the developers about the tools available in DPT-VWP. The primary key of the feedbacks table is the id. The user will be asked to enter his/her name, feedback, and rating of the website on a scale of 1 to 5, 5 being the highest. The fields in the table are:

| id | name | feedback | scale |
4.3 Detailed Design

4.3.1 Pseudo-code

Main Menu:

Case 1: “Information” function selected

display a webpage showing the overview and video

Case 2: “Tool Selection” function selected

display a webpage showing data converter, downloader and visualizer functions

if ‘data converter” selected

display a webpage showing data converter tools

if ‘PRMS text to NetCDF” converter selected:

wait for user input file

verify the file format

if wrong format:

prompt user to upload correct format

else:

if ‘Convert’ option selected:

convert the file to NetCDF format

display ‘conversion completed’ message

download NetCDF output file

if ‘PRMS NetCDF to text’ converter selected:

wait for user input file
verify the file format
if wrong format:
   prompt user to upload correct format
else:
   if ‘Convert’ option selected:
      convert the file to text format
      display ‘conversion completed’ message
   download text output file
if ‘Other Converters’ selected:
   wait for user input file in CSV format
   verify the file format
   if wrong format:
      prompt user to upload correct format
   else:
      if ‘Convert’ option selected:
         convert the file to JSON format
         display ‘conversion completed’ message
      download JSON output file
if ‘data downloader’ selected:
   display the names of watersheds
   if ‘select’ button selected
      verify a name has been selected
      if no name selected:
prompt user to select a name
else:
  display the names of datasets
  if 'select' button selected
    verify a name has been selected
    if no name selected:
      prompt user to select a name
  else:
    download the dataset in GeoTIFF format
if 'data visualizer' selected:
  display a webpage showing the visualization tools
if ‘Visualize CSV” selected
  wait for user input file in CSV format
  verify the file format
  if wrong format:
    prompt user to upload correct format
  else:
    if ‘Visualize’ option selected:
      Visualize the file in a line chart
Case 3: “Submit A Feedback” function selected:
  display a webpage that asks the user to submit name, feedback, and rating
  if ‘Submit’ option selected:
check if any fields are empty
if empty field found:
    prompt user to fill in the empty field
else:
    save the feedback

Case 4: “Tool Request” function selected:

display a webpage that asks the user to submit name, email, and question
if ‘Submit’ option selected
    verify the form has been completed
    check if any fields are empty
    if empty field found:
        prompt user to fill in the empty field
    else:
        save the request with status = ‘N/A’

Case 5: “Developer” function selected:

display the login page
verify the username and password
if wrong format:
    prompt user to enter the correct username and password
else:
    accept and/or reject the tool request(s)
    view the feedback(s)
4.3.2 Flowchart

The flowchart illustrated in Figure 4.3 describes the conversion of PRMS parameter input file from text format into NetCDF format. The user wants to upload the file in text format and the location file which includes the latitude and longitude values of each HRU. The user should also input the number of rows, columns and HRUs of PRMS. The tool verifies if the user has uploaded a file of correct format and then converts the file into NetCDF format. After conversion, the NetCDF format is file is downloaded and a message is displayed to the user informing that the file has been successfully converted and downloaded.

![Flow Chart of DPT-VWP Conversion Tool](image)

*Figure 4.3: Flow Chart of DPT-VWP Conversion Tool*
Figure 4.4 shows an example of a parameter NetCDF file with dimensions, variables, attributes and data for each variable.

```c
netcdf parameter {
    dimensions:
        lat = 49;
        lon = 96;
        hru = 4704;
        months = 12;
    variables:
        double lat(lat);
          lat:long_name = "latitude";
          lat:units = "degree_north";
        double lon(lon);
          lon:long_name = "longitude";
          lon:units = "degree_east";
        int soil_type(lat, lon);
          soil_type:layer_name = "soil_type";
          soil_type:dimension = "hru";
          soil_type:layer_desc = "HRU soil type";
          soil_type:layer_units = "none";
          soil_type:grid_mapping = "crs";
        double hru_lat(lat, lon);
          hru_lat:layer_name = "hru_lat";
          hru_lat:dimension = "hru";
          hru_lat:layer_desc = "HRU latitude";
          hru_lat:layer_units = "degrees";
          hru_lat:grid_mapping = "crs";
    // global attributes:
        :title = "HRU parameter file for Lehman Creek";
        :version = 1.7;
        :number_of_hrus = 4704;
        :number_of_rows = 49;
        :number_of_columns = 96;
    data:
        lat = 39.028015, 39.027084877551, 39.02615075102, 39.0252166326531,
            39.024205102041, 39.0233463877551, 39.0224142653061,
            39.0214801428571, 39.0205460204081, 39.0196118979592, 39.0186777755102,
            39.0177436530612,
            39.0168095306122, 39.0158754081632, 39.0149412857143,
            39.0140071632653, 39.013073548163, 39.0121369183673, 39.0112047959183,
```

**Figure 4.4: Parameter NetCDF File**

Fig. 4.5 is an example of visualization of a space and time related parameter, tmaxf (Maximum air temperature distributed to each HRU) in the NetCDF file using Quantum Geographic Information System (QGIS) software [38]. The figure is represented using two different colors. The red color indicates that the corresponding HRU has a low value and the green color indicates that the corresponding HRU has a high value.
The flowchart illustrated in Figure 4.6 describes the visualization of a file in CSV format on a bar chart. When the user uploads a CSV format file, DPT-VWP displays all the label names in the file on a check box list. The user will be asked to select 2 label names from the list. When the user selects ‘Visualize’ button, the selected column values are displayed on a bar chart. Error message is displayed if more or less than 2 label names are selected from the check box list.

4.4 Implementation of DPT-VWP

After completing design work, the process of implementation was begun. The DPT-VWP was constructed using Flask – Python micro framework. HTML, along with Bootstrap, CSS, JavaScript, and JQuery were used to create the web pages. D3 JavaScript
library is used for the visualization of CSV files on a bar chart. SQLite3 is used as the database for storing the details of watersheds, datasets, user requests and user feedbacks.

The structure of the website is simple. All the webpages - home, tool selection, tool request, feedback, list of new tool requests, and developer - are loaded at runtime and displayed based on the user selection.
In the remainder of this chapter we review the major components of the DPT-VWP website, including certain details of implementation and screen shots of the user interface design.

### 4.4.1 Home Page

DPT-VWP website starts with the home page. The home page of DPT-VWP gives an overview of the website along with a video to the users. The video gives the users an idea regarding how to use the various data processing tools provided by the website. The links to all other features in the website are provided at the top of the web page. The home page is shown in Figure 4.7.

### 4.4.2 Tool Selection Page

The tool selection page provides various tools to the users. The different categories of data processing tools, as shown in Figure 4.8 are data converters, data downloaders and data visualizers.

Data converters include tools that perform the conversion of PRMS input files (control, data, and parameter) and output files (animation, statistics variables, and water budget) files from .txt (text) format to .nc (NetCDF) format. It also converts the input files – parameter and data – from .nc (NetCDF) format to .txt (text) format. Files are converted into NetCDF format for visualization purposes. Files are then converted back to text format for running the PRMS model. Once the tool completes the conversion, the converted file automatically gets downloaded into the user’s computer.
Figure 4.7: DPT-VWP Home Page

Figure 4.9 shows the webpage of PRMS parameter and animation file conversion. The user has to upload the parameter input file in text format and the location file which includes the latitude and longitude values of each HRU. The user should also input the number of rows, columns and HRUs of PRMS. The tool verifies if the user uploaded a file of correct format and then converts the file into NetCDF format. Error message is displayed if the user selects ‘Convert’ option without uploading the input file or if the user uploads a file of format not supported by the tool. Data converters also include CSV to JSON, CSV to TSV and TSV to CSV conversion tools that are already available online.
Figure 4.8: Tool Selection Page

Figure 4.9: Conversion Tool in DPT-VWP
Data downloaders allow the users to download datasets of watersheds available from GSToRE. When the user selects the ‘Data Downloaders’ option, he/she will be redirected to a page which displays a list of all the watersheds available from GSToRE. Because the GSToRE platform does not include the details of the watersheds now, Figure 4.10 shows a list of collection names available from GSToRE. When the user selects a name from the list and clicks the ‘Select’ button, he/she will be redirected to a page which displays a list of all the datasets, as illustrated in Figure 4.11, available for the selected collection name. When the user selects a name of dataset from the list and clicks ‘Download’ button, the dataset gets downloaded into the user’s computer as a compressed file which includes a GeoTIFF format file and a XML file. The user will receive a ‘Successfully downloaded’ message once the dataset download has been successfully completed.

Data visualizers allow the users to visualize the files in CSV format on a bar chart. When the user uploads a .csv file and select ‘Upload’ button, DPT-VWP displays the column labels on different checkboxes. The user wants to select two checkboxes and select ‘Visualize’. The user has to select the first checkbox because the tool takes the column label represented in the first checkbox as the x label for visualization. The y label would be the second checkbox selected by the user. Figure 4.12 shows the visualization of a .csv file on a bar chart.
Figure 4.10: List of Watersheds Available from GSToRE

Figure 4.11: List of Datasets Available from GSToRE
4.4.3 Tool Request Page

The tool request web page in DPT-VWP permits the users to submit requests for the development of a new tool to the developer. The user should provide his/her name, email address, and the request. When the user clicks the ‘Submit’ button, the details provided by the user are stored in the database and a ‘Successfully submitted’ message is displayed to the user. The web page displays an error message if any of the fields (name, email address and/or request) are left blank by the user. The tool request web page is illustrated in Figure 4.13.
Figure 4.13: Tool Request Page in DPT-VWP

4.4.4 Feedback Page

The feedback web page, as illustrated in Figure 4.14, allows the users to submit their feedback and/or suggestions regarding the tools available in the website to the developer. The user should provide his/her name, feedback, and the rank/rating of the website on a scale of 1 to 5. When the user clicks the ‘Submit’ button, the details provided by the user are stored in the database and a ‘Successfully submitted’ message is displayed to the user. The web page displays an error message if any of the fields (name, feedback and/or rank/rating) are left blank by the user.
4.4.5 Developer Page

When the developers select the ‘developer’ link, he/she is redirected to the login page. The developers have to type the correct username and password to view the admin home page. The web page displays an error message if the developers submit a wrong username and/or password.

The developers’ home page includes two tabs – tool request(s) and feedback(s). In the tool request(s) tab, the developers can view all the requests submitted by the users. There will be ✓ (Accept) and ✗ (Reject) symbols along with each request. The developers may either accept or reject the requests based on whether they would be able to develop the
requested tool or not. In the feedback(s) tab, the developers can view all the feedbacks and/or suggestions submitted by the users. The tool request(s) and feedback(s) tabs are displayed in Figure 4.15 and Figure 4.16 respectively.

![Figure 4.15: Tool Request(s) Tab in Admin Home](image)
4.4.6 List of New Tool Requests Page

The ‘List of New Tool Requests’ web page displays all the requests that are submitted by the users. The requests that are accepted and rejected by the developers are displayed in separate tabs. The requests that are submitted by the users and are yet to be reviewed by the developers are displayed in a different tab in the web page. Figure 4.17 displays the accepted request(s) tab in DPT-VWP.

![Figure 4.16: Feedback(s) Tab in Admin Home](image)
### DPT-VWP

Data Processing Toolset for the Virtual Watershed Project

**Total Number of Questions:** 8

- **Accepted:** 2
- **Rejected:** 4
- **N/A:** 2

<table>
<thead>
<tr>
<th>Request Type</th>
<th>Accepted Requests</th>
<th>Rejected Requests</th>
<th>Requests to be Reviewed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Include CSV to TSV Converter</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Include NetCDF to XML Converter</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Include CSV to JSON Converter</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Include CSV to XML Converter</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Include 3D Visualization tools</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Include more 2D Visualization tools</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Include more existing tools</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Include NetCDF visualization tools</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 4.17:** List of New Tool Requests Page in DPT-VWP
5 DPT-VWP EVALUATION AND SOFTWARE PROFESSIONAL FEEDBACK

The next step after completing the implementation of the DPT-VWP was to evaluate the website. Evaluation is useful for a variety of reasons. It allows for exploration of new features in the website, assessment of current implementation, or verification of features in the website [39].

We conducted proper usability evaluation with the users working on the WC-WAVE project as well as graduate students with similar research interests. The details of the participants are shown in Figure 5.1.

![Figure 5.1: Details of Participants](image-url)
After using the website, the participants were asked to rate their experience by completing the After Scenario Questionnaire, or ASQ. The ASQ is used to judge the user’s perception of effectiveness, efficiency, and satisfaction in using the website. The ASQ can be found in Figure 5.2. During work with the program, participants were encouraged to “think aloud,” verbalizing any confusion or difficulty. Time was also devoted for any additional comments the participant wished to write down.

**After Scenario Questionnaire** - Please check the box that reflects your immediate response to each statement. Don't think too long about each statement. Make sure you respond to every statement.

<table>
<thead>
<tr>
<th>1. I am satisfied with the conversion tools available in DPT-VWP</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strongly Disagree</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Strongly Agree</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2. I am satisfied with the download tool available in DPT-VWP</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strongly Disagree</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Strongly Agree</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>3. I am satisfied with the visualization tool available in DPT-VWP</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strongly Disagree</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Strongly Agree</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>4. I am satisfied with the tool request and feedback sections available in DPT-VWP</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strongly Disagree</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Strongly Agree</td>
</tr>
</tbody>
</table>
5. The video helped me to understand how to use the tools

<table>
<thead>
<tr>
<th></th>
<th>Strongly Disagree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1 2 3 4 5

Additional Comments:

*Figure 5.2: After Scenario Questionnaire*

The ASQ asks participants to rate their experience on a five-point scale ranging from *Strongly Disagree*, which is worth one point, to *Strongly Agree*, which is worth five points.

In summary, the ASQs revealed that DPT-VWP was mostly effective, efficient, and satisfying. The core results of the ASQ can be found in Table 5.1.

*Table 5.1: ASQ Results*

<table>
<thead>
<tr>
<th></th>
<th>Q1</th>
<th>Q2</th>
<th>Q3</th>
<th>Q4</th>
<th>Q5</th>
</tr>
</thead>
<tbody>
<tr>
<td>User 1</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>User 2</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>User 3</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>User 4</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>User 5</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>User 6</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Average</td>
<td>4.8</td>
<td>3.8</td>
<td>3</td>
<td>5</td>
<td>3.8</td>
</tr>
</tbody>
</table>
As shown in Table 5.1, all the participants were satisfied with the conversion tools available on the website. Chase and Chao knew about the files that are used for conversions. As all other students were not involved in the project, we described to them about the files that are used for conversions. They all mentioned to include additional existing conversion tools on the website. The average score received for the conversion tools section is 4.8 out of 5.

The participants were also satisfied with the download tool available on the website. The tool downloaded an XML file along with the file in GeoTIFF format. The download tool section received an average score of 3.8 out of 5.

The visualization tool received the lowest score, 3 out of 5, and will require the most attention for improvement. From the comments the participants wrote on the ASQ, it was mentioned that it would be useful to visualize the data on a line chart along with the bar chart. They have also mentioned the use of including 3D visualization tools in the website.

The tool request and feedback sections received a full score. The comment written was ‘it’s always good to include a feedback section on the websites. Also, a feature that enables users to request for a new tool seems like an innovative idea’.

The average score received for including videos in the website is 3.8 out of 5. The tutorials are good for first time users. It explains how to use the tools provided by DPT-VWP.

Additional feedbacks were provided by the users about the developer page. The tool request and feedback sections in the admin home page were pretty impressive to the users. The use of Bootstrap along with JavaScript, JQuery and HTML helped us to develop those sections in the admin home page.
6 \hspace{1em} COMPARISON WITH RELATED WORK

In this section, we compare our work with other data processing tools; advantages and disadvantages are listed.

Mr. Data Converter [40] is a web application that converts a file in CSV format into various web-friendly formats such as HTML, JSON and XML. The user wants to copy and paste the input data on the space provided and select the output format required from the dropdown list. The application is available online and is free to use. However, the application does not provide an option to upload and/or download the file(s). If the user wants to convert a file, he/she has to copy and paste the data from the file on the space provided in the website. After the data conversion, the user again has to copy and paste the

![Figure 6.1: Mr. Data Converter [40]](image-url)
converted data into a file. Also, the tool takes a large amount of time to convert large file(s) (file of size 500 KB). The tool also does not show the progress of file conversion which would affect the user satisfaction. Figure 6.1 shows Mr. Data Converter.

MATLAB (MATrix LABoratory) is a desktop application used by millions of users worldwide for data visualization and processing. MATLAB can visualize data in 2D and 3D because it provides built-in 2D and 3D plotting functions [41]. Figure 6.2 shows a bar chart visualization using MATLAB.

![Figure 6.2: Bar Chart Visualization in MATLAB [41]](image)
The users have to write MATLAB scripts to visualize and process the data which would be difficult for the researchers and the students who have never learned programming. A well written documentation about MATLAB is available online which includes scripts and built-in functions required for visualization and processing. The MATLAB license is around $100. Because it is a desktop application, MATLAB cannot be accessed by the users online. The performance of the MATLAB software heavily depends upon the users’ computers. If the computers are not good enough, the amount of time required to process and visualize the data will be large [42].

VISTED (VISualization Toolset for Environmental Data) is a client/server web application, the goal of which is to aid the researchers in the visualization of the climate data [43, 44]. The features of VISTED are data selection, extraction, download, conversion, and visualization. The tool is easy to use and the visualization results are interactive. It is built using HTML5, JQuery, CSS5, D3 JavaScript library, and C#. The users will be able to download the selected data in ASCII, binary and NetCDF formats. The users can visualize and download datasets generated by the NCAR/WRF climate model [45], which is from 1980/01-2009/12 NCEP/NCAR reanalysis and 2040/01-2069/12 for CCSM3 based on the A2 Scenario. Figure 6.3 shows the VISTED home page.
Microsoft Excel [46] is a spreadsheet developed by Microsoft and is widely used in various fields. Users can visualize the data in various charts such as line charts, pie charts, bar charts, 3D surface, radars, and so on. Users do not have to write any scripts for data visualization. They have to input the data and choose the graph on which he/she wants to visualize the data. The icons indicating the various graphs are displayed. The graph gets automatically updated based on the changes in the input data. The users can process the data using Microsoft Excel. Microsoft license is around $80. Similar to MATLAB, the
performance of Microsoft Excel depends upon the users’ machines [47]. Figure 6.4 shows a bar chart visualization using Microsoft Excel.

Several other tools are available online that converts files from CSV into JSON, XML, and HTML formats [48]. Files in JSON format can also be converted into CSV, XML, and HTML using some web applications [49]. Table 6.1 summarizes the features of Mr. Data Converter, MATLAB, VISTED, Microsoft Excel and DPT-VWP.

![Figure 6.4: Bar Chart Visualization in Microsoft Excel](image-url)
Table 6.1: Comparison between Data Processing Tools

<table>
<thead>
<tr>
<th>SOFTWARE</th>
<th>Mr. Data Converter</th>
<th>MATLAB</th>
<th>VISTED</th>
<th>Microsoft Excel</th>
<th>DPT-VWP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Available for free</td>
<td>YES</td>
<td>NO</td>
<td>YES</td>
<td>NO</td>
<td>YES</td>
</tr>
<tr>
<td>Performance based on machine</td>
<td>NO</td>
<td>YES</td>
<td>NO</td>
<td>YES</td>
<td>NO</td>
</tr>
<tr>
<td>Programming skills required for users</td>
<td>NO</td>
<td>YES</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
</tr>
<tr>
<td>Supports Visualization</td>
<td>NO</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>Supports Conversion</td>
<td>YES</td>
<td>NO</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>Supports Download</td>
<td>NO</td>
<td>NO</td>
<td>YES</td>
<td>NO</td>
<td>YES</td>
</tr>
<tr>
<td>Supports NetCDF</td>
<td>NO</td>
<td>NO</td>
<td>YES</td>
<td>NO</td>
<td>YES</td>
</tr>
<tr>
<td>Supports CSV</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
</tbody>
</table>
7 CONCLUSIONS AND FUTURE WORK

In this thesis, we proposed a new integrated expandable online software solution to help watershed scientists with data processing. This software is significant in that it aids the researchers in the WC-WAVE project.

The major contribution of our work is the integration of data processing tools: data conversion, data download, and data visualization. DPT-VWP supports conversion of the PRMS model input and output files from text format into NetCDF format, which is a widely used format in scientific research projects. It also reconverts the PRMS model input files from NetCDF format back to text format. The reason for including these converters is that text format files are required to run the PRMS model. The software includes tools to download datasets of watersheds available from the GSToRE platform. The application also supports visualization of CSV format files on bar charts. It also includes several conversion tools that are already available online such as CSV to JSON, CSV to TSV, and TSV to CSV.

Along with these functionalities, the tool request feature in DPT-VWP enables the users to request the availability of new tools from the developers. They can also submit their feedback and/or any suggestions regarding the improvement of the website to the developers. The developers, after logging into the system, can view all the feedbacks and tool requests from the users. After viewing the requests, the developers can either accept or reject the requests based on the time available. After indicating a status to the user requests, the users can view the status on the website.
The work presented in this thesis could be furthered beyond its current state, with a goal of developing DPT-VWP into a generic web application with more features. Currently, the conversion tools are very specific to the PRMS model. In the future, the website could support the conversion of MODFLOW model files from text to NetCDF format.

Based on the DPT-VWP’s evaluation performed by peer software developers, another area of improvement is the inclusion of additional data visualization techniques. Presently, DPT-VWP performs 2D Visualization of CSV format files on a bar chart. In the future, DPT-VWP will visualize data on a line chart and a pie chart using D3 and Dygraph JavaScript libraries. The website will also incorporate 3D visualization techniques. The visualization will also be performed in files of formats other than CSV such as NetCDF and JSON.

Also, in future versions, the developers will be able to indicate the status of the tool requests (Accepted/Rejected) to the users by sending an email to them. If the request is rejected, the developers will specify the reason for rejection to the user.

All the PRMS model files that were used for the conversion were of size less than or equal to 1 GB. The tools that are available now can process only files of size up to 1 GB. To handle larger files, we are planning to enhance the DPT-VWP tools such that they can process big data [50]. There are various frameworks that can handle big data, including Hadoop, MapReduce [51, 52], Apache Storm, and Apache Spark. Our web application will be enhanced with new capabilities to handle big data using these frameworks.
REFERENCES


