THESIS

MINIMIZING IMPRACTICALITY OF COMPLEX MODELING VIA WEB SERVICE EXECUTION WITHIN A

CLOUD-BASED PLATFORM

Submitted by

Jeffrey Kwon Ditty

Department of Civil and Environmental Engineering

In partial fulfillment of the requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Spring 2015

Master's Committee:

      Advisor: Mazdak Arabi
      Co-Advisor: Peter Allen

      Ellen Wohl
      Brian Bledsoe

ABSTRACT

MINIMIZING IMPRACTICALITY OF COMPLEX MODELING VIA WEB SERVICE EXECUTION WITHIN A

CLOUD-BASED PLATFORM

Assessing the stability of stream banks is an important issue for those living within a floodplain, participating in activities such as hunting, rafting, and fishing, and for transporting water from one place to another safely. A stream's stability is influenced by frequency and magnitude of streamflow, particle size, sediment transport, bed slope, and depth of the water (Allen, P.M., Arnold, J.G., Stinchcomb, 2012). Streams of all sizes adapt to streamflow regimes by continually altering channel properties such as channel's shape and channel depth. As a result, assessing stream stability requires quantifying hydrologic factors that influence the stream's shape and depth, especially under conditions of urbanization.

Changes in climate and land use further complicate assessing a stream's stability. This complication is due partly to humanity's knowledge gaps or negligence in maintaining solutions of existing infrastructure (Vijay P. Singh, 2002). The most common cause of stream instability is the increase in surface runoff entering a stream due to urbanization. Engineers use models as a tool to help manage, address, or mitigate problems such as increasing surface runoff entering the streams.

The primary goal of this research is to integrate an existing model (SWAT-DEG) as a web-based modeling tool to assess the hydrologic response to changes in climate and land use.

Specific objectives of the study are:

1. To compliment an existing model by integrating the model into service oriented architecture.
2. To compliment an existing model by deploying a model utilizing web-services.
3. To explore the benefits of deploying an existing model as a web-based tool within a service-oriented architecture.

Web services provide access to SWAT-DEG at any internet-accessible location, limits version control, independent of operating system, and eliminate desktop installations. Currently, SWAT-DEG is hosted by eRAMS.com' (eRAMS) platform. eRAMS is an internet site that hosts and executes various models. The model is executed within a cloud environment to ensure the model can scale to increasing, simultaneous users.

The cloud environment obtains scalable models by reducing the total time per request for a given model. The total time per request is defined as the time that elapses from when the user executes a function till the function is completed. For example, the total time per request spent for executing a model is the time spent between clicking run and the return of output from the model. To accomplish a reduction in total time per request, the execution of SWAT-DEG is parallelized within Cloud Services Innovation Platform (CSIP) (Lloyd, W, David, O, Lyon, J, Rojas, K.W., Ascough II, J.C., Green, T.R., Carlson, J.R., 2012) CSIP is a cloud infrastructure trying to implement modeling-as-a-service. Modeling-as-a-service is an attempt to provide the ability to scale various web-based models for large amounts of simultaneous users and/or expensive computational models.

SWAT-DEG is broken up into two separate main functions within the source code, stochastic and deterministic. The stochastic part of SWAT-DEG executes the Monte Carlo (Ronald Christensen, Wesley Johnson, Adam Branscum, 2011; Yang, 2011) approach as the means to do uncertainty analysis. The deterministic approach is the same as the desktop version of SWAT-DEG. Scalability testing was applied to SWAT-DEG's stochastic and deterministic services.

The deterministic service showed clear benefits when executing multiple simultaneous users within the CSIP environment. However, a single user executing a deterministic service on the cloud is slower than a local computer. The stochastic service also saw a slight benefit from executing within the CSIP environment. The stochastic service did not receive a greater benefit due to a bottleneck in the output map reduction. Output map reduction is the process of taking multiple files and reducing them to a manageable size. SWAT-DEG processes this phase within one virtual machine (VM). A virtual machine is an online computer processor unit. Thus, by removing the limitation of executing map reduction within one VM, executing stochastic analysis within the cloud environment will result in far smaller total time per request.

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ix

CHAPTER 1: INTRODUCTION

1.1 River Restoration

Stability of river networks is influenced by climatic and hydrologic factors. Climate variability and land use change, specifically urbanizing areas, could have profound effects on stream's sedimentation processes. Urbanization not only causes an increase in surface runoff primarily due to lowering the potential infiltration of the soil, but urbanization also causes indirect changes to the river network's ecosystem. River networks adapt to climatic and hydrologic factors by altering channel properties and the ecosystem. Channel properties include but are not limited to the channel's shape and the depth of water, whereas, the ecosystem is the culmination of all the living organisms interacting with the channel. Thus, changing a stream in any manner causes indirect complications in many manners. These changes to streams and river networks are a necessity to accommodate the demands of society. Demands range from extracurricular activities to drinking water. For this reason, river restoration/rehabilitation projects are very common but also very difficult to handle.

Many river restoration/rehabilitation projects revolve or at the very least involve the stability of the river network. Cost of individual projects ranges from $200 to $150,000,000 with a median cost of $50,257 (Kondolf, G. M., Anderson, S., Lave, R., Pagano, L., Merenlender, A., Bernhardt, E. S., 2007). The United States spends in excess of 1 billion dollars each year compensating for these river restoration/rehabilitation project costs, (Palmer, Allan, Meyer, & Bernhardt, 2007) As time continues, an ever-increasing number of experts are involved with river restoration/rehabilitation projects (Bernhardt, Emily S., Sudduth, Elizabeth B., Palmer, Margaret A., Allan, J. David, Meyer, Judy L., Alexander, Gretchen, Follastad-Shah, Jennifer, Hassett, Brooke, Jenkinson, Robin, Lave, Rebecca, Rumps, Jeanne, Pagano, Laura, 2007).

Despite the continual increase in experts required for river restoration/rehabilitation projects, many rivers are likely to have different sediment and flow regimes then currently (Lettenmaier, Wood, & Palmer, 1999; Poff, Brinson, & Jr, 2002). However, it is not clear whether the rivers are capable of adapting to the new flow and sediment regimes without negatively interfering with human activity (Bernhardt, Emily S., Sudduth, Elizabeth B., Palmer, Margaret A., Allan, J. David, Meyer, Judy L., Alexander, Gretchen, Follastad-Shah, Jennifer, Hassett, Brooke, Jenkinson,

Robin, Lave, Rebecca, Rumps, Jeanne, Pagano, Laura, 2007). To minimize the negative impacts, engineers require modeling tools that analyze the hydrologic deviations and their effect on society.

1.2 Modeling

Modeling is the common solution for hydrologic assessments. The first hydrologic model was developed in the 1960s, when the Standford Watershed Model (SWM), developed by Crawford and Linsley, attempted to model the entire hydrologic cycle (Vijay P. Singh, 2002).

Simulating natural processes has inherent difficulties due to the limitations of modeling, such as inability to simulate all processes, unknown processes, and justifying the need for coding a certain process into the model. These limitations are under constant review from every discipline to ensure the best models. However, models are still a fragmented representation of reality.

Traditionally, engineers have approached river restoration through two approaches: a reach-focused approach and a species/habitat driven approach (Clarke, Bruce-Burgess, & Wharton, 2003). Both approaches benefits from scenario analysis. Scenario analysis is used to assess the impacts of climatic variation and land use changes. Climatic variation and land use changes directly correlate to the amount of surface water entering a river network. The amount of water in the stream dictates the outcome of the sedimentation processes, mainly in the form of degradation or aggradation. Specifically, higher flows are more likely to cause erosion and downcutting, which in turn causes stream instability. Urbanization is one common factor in increasing the flow of rivers to a point of causing stream instability.

The Soil and Water Assessment Tool (SWAT) is a common tool or model used for continuous simulation of hydrologic processes at various temporal and spatial scales (J. Arnold, Allen, Muttiah, & Bernhardt, 1995; J. G. Arnold & Fohrer, 2005; J. Arnold & Srinivasan, 1999; Gassman, Reyes, Green, & Arnold, 2007). SWAT is an extremely complex model in regards to the amount of processes simulated within SWAT. SWAT contains modules for hydrology, erosion/sedimentation, nutrients, pesticides, land management, stream routing, plant growth, pond/reservoir routing, weather simulator, soil temperature, and agricultural management (J. Arnold & Srinivasan,

1999; Gassman et al., 2007; Jain, 2010; Li, Yu, Xin, & Sun, 2012; Van Griensven., 2012). As time progresses, SWAT continue to increase the number of processes it can simulate and continually tests whether SWAT represents reality adequately.

Numerous previous studies demonstrate, analyze, enhance, modify, or research SWAT and its current discretization scheme (C. Brown & Hollis, 1996; Li et al., 2012; Neumann, Western, & Argent RM, 2010). Previous projects of SWAT include: Conan, de Marsily, Bouraoui, & Bidoglio, (2003), who verified that SWAT adequately simulated the change from wetlands to dry land and (Srinivasan R., Ramanarayanan T, Arnold J, 1998) who concluded that SWAT sediment accumulation predictions were adequate. SWAT's ability to simulate multiple processes naturally increases the complexity of SWAT.

The complexity of a model is reflected in the amount of input required, the amount of time to setup a model, and the difficulty of changing the model to obtain meaningful results. Increasing the complexity of a model can also reduce the accessibility of the model. Accessibility, for a model, is defined as the ease to which a modeler can obtain, use, and output meaningful results from a given model. As seen, it is common for complex models to be less accessible. In the case of SWAT, accessibility is limited due to the large amount of input data required and that the model requires some time to execute successfully. There are many methods of increasing accessibility for any model, one of which is deploying a model utilizing web-services and service-oriented architecture.

1.3 Model Facilitation/Infrastructure

Traditional modeling requires the modeler to select the software, locate and parameterize the input data from multiple sites, and then run the software. This process, for desktop applications, reduces the accessibility of a model because of the large amount of time required from obtaining input data to obtaining meaningful output from the model.

Model preparation time and lack or hard to find model information are common reasons for why many published models do not make it to industry (Mineter, Jarvis, & Dowers, 2003). The issue of locating documentation can be addressed via web-services. Typically finding documents requires navigating through endless archives of documents

3

on various desktop computers. Web-services helped eliminate this by storing documents on websites. Developing models into web-services will help localize documentation and the model into a singular location. Thus, web-services help minimize the effort required to find documentation while also providing a natural organization system.

Organization is extremely important for the current modeling community as models become more complex. The modelling community are focused upon model integration and decreasing the size of data (R. Argent, Grayson, & Ewing, 1999; Beran & Piasecki, 2009; Goodall, Horsburgh, Whiteaker, Maidment, & Zaslavsky, 2008; Granell, Díaz, & Gould, 2010; McIntosh, Jeffrey, Lemon, & Winder, 2005; Mineter et al., 2003). Essentially, the modelling community would like to have models that can analyze a system holistically, while maintaining short run-times and accessibility.

Web-services not only increases accessibility by centralizing data to a singular location, but also enables anyone access to common hydrologic models. Global access to these models will increase interactions between industry, academia, and stakeholders (R. Argent et al., 1999). To keep web-services practical, each model's source code must also incorporate the concepts modularity and reusability (Goodall et al., 2008; Granell et al., 2010; Mineter et al., 2003). Modularity is the process of making functions within the code to serve a single purpose and reusability is the concept of making a function capable of having to enable multiple applications access to the given function. Atkins, (2003) stated that creating a "comprehensive cyberinfrastructure" to support scientific research would revolutionize the engineering community, as well as other related communities. A "comprehensive cyberinfrastructure" refers to a platform, such as the web, that enables everyone to access multiple models.

Increasing the accessibility of a model via web-services gives global access to models, global input about the model, and determination of the model's future on a global scale. As a result, stakeholders, industry, and academia can agree on the worth and utility of a model as it is being developed. This would also allow modeling efforts to receive securer funding, increase popularity, and increase trust from both stakeholders and modelers. Thus, accessibility must be a priority in model evolution; however, accessibility is constrained if not complemented by scalability.

To address expensive computational models or large amount of users simultaneously accessing a model on the web, scalability must be applied to the model. Scalability is defined as the flexibility of a service to handle models in the form of computationally expensive/more model runs or larger/extensive consumer use (La, Oh, & Kim, 2010). The

need for scalability is demonstrated in the large computation time required to run the complex model SWAT, which is still increasing as SWAT is further modified to better simulate processes or simulate more processes.

Scalability is addressed within web-services by deploying hydrologic modeling tools within service-oriented architecture. Service-oriented computing is defined as decomposing a software system into independent distributed components. Components refer to individualized functionalities within the software. For example, if a model provides multiple methodologies for calculating surface runoff, each of these methodologies would be an individual component. These components are then deployed on a multi-core processing platform to provide a scalable utilization of each individual component.

Technological advances like multi-core processors and online servers are revolutionizing current modeling practices. A multi-core processor is essentially having multiple computers within a single computer. Thus, multi-core processors enable modelers to run large models or modules on a single machine or computer.

This study provides a demonstration of both scalability, via CSIP (Castronova, Goodall, & Elag, 2013; Huhns & Singh, 2005; M. Papazoglou, 2006) and accessibility, via eRAMS, by reconfiguring an existing model, SWAT-DEG, to utilize both CSIP and web-services. SWAT-DEG is a simplified version of SWAT focusing on simulating downcutting and widening in small alluvial and threshold streams (Allen, P.M., Arnold, J.G., 1999; Neitsch, Arnold, Kiniry, Williams, & King, 2002). eRAMS (erams.com) is a website that utilizes service-oriented architecture to enable any user access to accessible and scalable engineering models. CSIP is a third party outsourcing its equipment to provide a web-based, cloud environment for executing hydrologic tools. SWAT-DEG is a model fully integrated into eRAMS and demonstrates accessibility and scalability via web-services with service-oriented architecture. Benefits reaped from SWAT-DEG's deployment within CSIP are discussed via analyzing the results of test cases executed within CSIP. In addition, the development of models, as done with SWAT-DEG, will increase the public and engineer's understanding of urbanization's effect on stream's stability ,thus, yielding more sustainable and realistic solutions to river restoration/rehabilitation problems (Vijay Singh, 2006).

1.4 Objectives and Methods

The main objective of this study is to deploy the SWAT-DEG model as a web tool using a service-oriented architecture. Specific objectives of the study are:

- To compliment an existing model by integrating the model into service oriented architecture.

- To compliment an existing model by deploying a model utilizing web-services.

- To explore the benefits of deploying an existing model as a web-based tool within a service oriented architecture.

2.1 Model Description

SWAT is a physically based, lumped parameter, continuous-time hydrological model (Jain, 2010; van Griensven et al., 2012) created by integrating three U.S. Department of Agriculture (USDA) models: (1) the Chemicals, Runoff, and Erosion from Agricultural Management Systems (CREAMS) model (Knisel, 1980), (2) the Groundwater Loading Effects on Agricultural Management Systems (GLEAMS) model (R. A. Leonard, W.G. Knisel, 1987), and (3) the Environmental Impact Policy Climate (EPIC) model (Gassman, Williams, & Benson, 2005; Izaurralde, Williams, McGill, Rosenberg, & Jakas, 2006), as shown in figure 1.

SWAT Developmental History
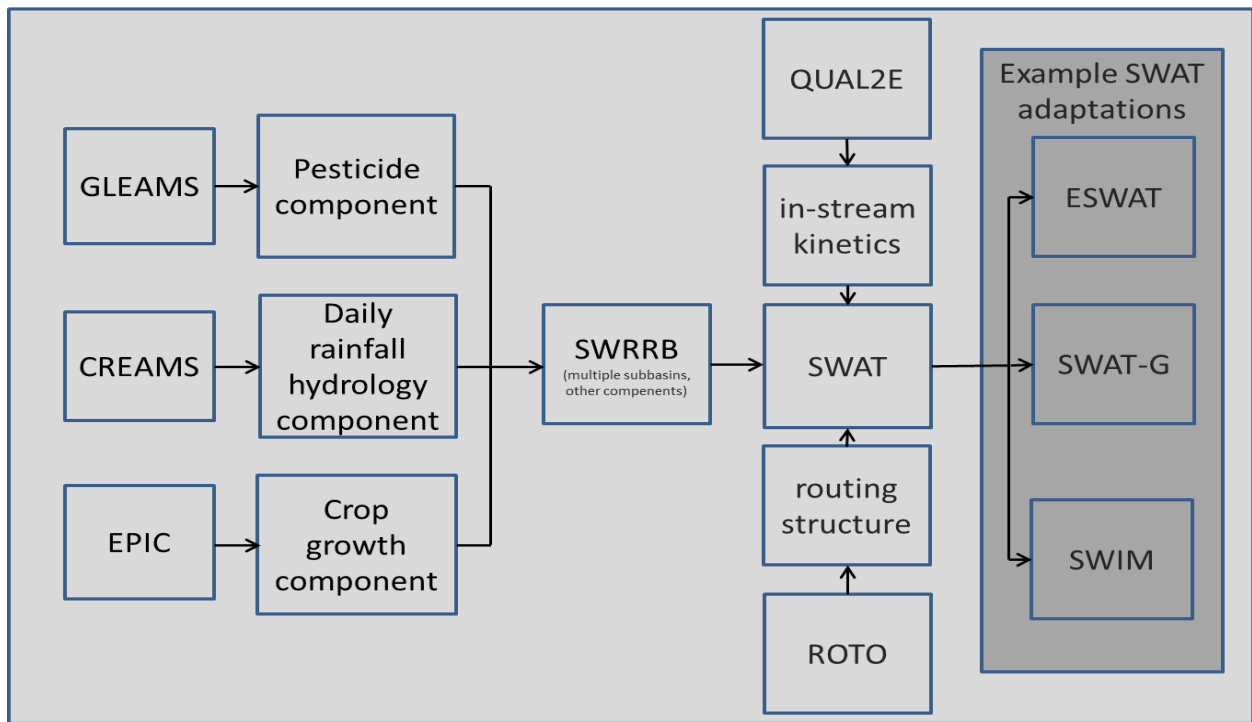
Figure 1: Schematic of SWAT developmental history, including selected SWAT adaptations (Gassman et al., 2007)

The aggregation of these three models not only created a holistic model but also a complicated model. Further detail of the modifications to SWAT can be found by J. G. Arnold & Fohrer, (2005), S.L. Neitsch, J.G. Arnold, J.R.

Kiniry, R. Srinivasan, (2005), and (L. Brown & Barnwell, 1987). One modification in particular stripped SWAT of functionality to focus the model upon analyzing hydrologic impacts on streams due to urbanization.

SWAT-DEG is a modified version of the Soil Water Assessment Tool (SWAT) to simulate downcutting and widening of small alluvial and threshold streams. Channel downcutting is calculated through an erodibility coefficient from submerged jet testing and widening is calculated by local width-depth ratios, demonstrated by figure 2.
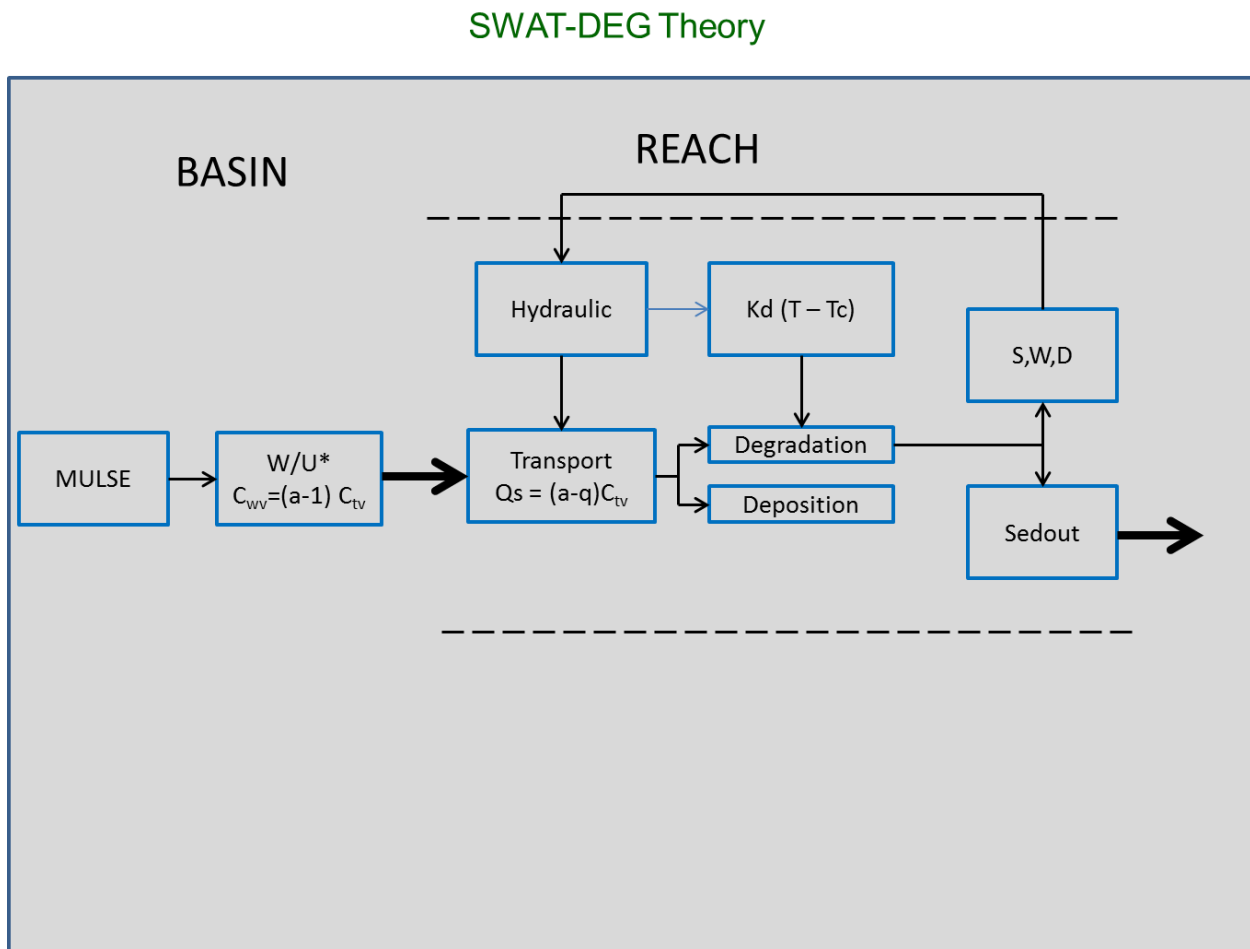
Figure 2: SWAT-DEG approach to calculating width, slope, and depth of channel during urbanization of watershed (Peter Allen, Jeff Arnold, Stephanie Capello, 2010).

$$QS \sim W_s D_{50} \hspace{4cm} \text{(Eqn. 1)}$$

Where Q is the channel forming discharge, S is the channel gradient, W_s is the bed material discharge, and D_{50} is the median grain size (Allen P.M., Arnold J.G, Skipwith W., 2008,; Peter Allen, Jeff Arnold, Stephanie Capello, 2010).

Equation 1 implies that if changes in the system cause a decrease in sediment loads and an increase in annual or peak discharges, concentration of flow, or increase in channel gradient then degradation will occur (Simon & Rinaldi, 2006). SWAT-DEG attempts to help managers address these occurrences by outputting data on downcutting, widening, slope, depth, and the flow duration curve. The outputs should then be used to aid in timeseries prediction of channel degradation due to urbanization in small urban watersheds. As a result, the impact of river restoration/rehabilitation solutions due to timing can be better assessed. In addition, SWAT-DEG will help provide insight to which current river restoration/rehabilitation projects require prioritization due to urbanization (Allen, Arnold, & Skipwith, 2002; Allen P.M., Arnold J.G, Skipwith W., 2008).

SWAT-DEG also provides a flow duration curve to provide a more holistic analysis. The flow duration curve (FDC) uses historical data to determine the percent of time a stream discharge value has been met or exceeded. The flow duration curve can then be used to derive a load duration curve from which Total Maximum Daily Loads (TMDL) can be derived ("An Approach for Using Load Duration Curves in the Development of TMDLs," n.d.). In addition, the flow duration curve provides a method to predict changes in bankfull discharges.

2.2 GIS Capability

SWAT-DEG's users interface is integrated with free ARCGIS-like software provided by its host eRAMS.com (eRAMS).The ARCGIS-like functionality is contained within two main categories, normal ARCGIS functionality and SWAT-DEG functionality, a short list of functionalities and their categories is shown below in table 1.

Table 1: eRAMS GIS Functionality

Function Name	Brief Description	Category
Interact with United States Consensus Layers	Query layers demonstrating population demographics via spatial partitions	Normal ARCGIS functionality
Interact United States geographical data layers	Access to interactive layers, such as National Hydrography Dataset (NHD)	Normal ARCGIS functionality

Interact with United States Environmental Protection Agency (EPA) Water Layers	Access to interactive layers of varying data, such as STORage and RETrieval (STORET) Water Monitoring stations	Normal ARCGIS functionality
Interact with land use and land cover layers	Access to National Agricultural Statistics Service (NASS) and National Land Cover Database (NLCD) datasets	Normal ARCGIS functionality
Measure tools	Calculate distance and area	Normal ARCGIS functionality
Data Downloads	Access to multiple organization/agency's stations data	Normal ARCGIS functionality
Digitize/Edit	Manually manipulate or create layers.	Normal ARCGIS functionality
Geoprocessing	Zonal Statistics, Proximity calculations, dissolving	Normal ARCGIS functionality
HRU Generation	Merge and dissolve land cover, soil, and slope shapefiles to generate unique polygonal shapefile	SWAT-DEG interactive functionality
Land cover, soil, and slope extraction	Download corresponding data from various organizations to further input generation for SWAT-DEG	SWAT-DEG interactive functionality
Automated CN estimates	If HRU generation is done, eRAMS will calculate an initial CN from the HRU generation outputs	SWAT-DEG interactive functionality

The last three are specific to SWAT-DEG and provide automations to minimize user error. The hydrologic response unit (HRU) generation (S.L. Neitsch, J.G. Arnold, J.R. Kiniry, R. Srinivasan, 2005) is a method developed and utilized by the model SWAT. The method requires three GIS layers: soil, land use, and slope. These three layers are

intersected with each other to create polygons with a unique soil, land use, and slope classification. SWAT-DEG automatically calculates curve number (CN) and automatically classifies each HRU via a table within SWAT-DEG's source code.

The other functionalities listed above are within eRAMS but are not specifically integrated with SWAT-DEG. These functions still provide indirect benefits to SWAT-DEG and are accessible via any device, mobile or local. Specifically, the other functions include supplementary models, automated data queries, and model input manipulation, all of which the user can use in conjunction with SWAT-DEG for a holistic analysis.

2.3 Scenario Analysis

Scenario analysis is vital to any type of river restoration/rehabilitation analysis or channel stability analysis. Scenario analysis' main purpose is to provide the user the capability of analyzing a given area of interest in regards to futuristic land use conditions due to rapid urbanizations and climate change. The scenario analysis should still be applied to small threshold rivers to keep in scope of SWAT-DEG's analysis.

The scenario land use analysis is dependent upon the creation of HRUs, and therefore upon the availability of land use, soil, and slope data. Table 2 provides logistics of eRAMS capabilities in obtaining data for the layers required to generate an HRU layer.

Table 2: Brief Description of Necessary Inputs for HRU Generation within SWAT-DEG User Interface.

ESRI Layer	Organization/Agency	Purpose
Area of interest	N/A	Define area of interest for clipping of all subsequent layers and calculating area for SWAT-DEG analysis
Land Use	National Agricultural Statistics Service (NASS)	Provide land use information
Soil	Soil Survey Geographic Database (SSURGO)	Provide soil information

11

	State Soil Geographic Database (STATSGO)	
Slope	National Hydrologic Database (NHD)	Provide soil information

After providing these necessary inputs to the HRU generation user interface, eRAMS executes an intersection and dissolve of the three layers. During the intersection and dissolve, a minimum area percentage may occur upon user specification. The minimum area percentage is to eliminate HRU polygons that are too small to have an effect upon the final analysis.

Upon completion of a new polygon ESRI layer, the newly created polygons contain the following attributes: land use, soil, slope, area, CN, and identification number. A simple programmed dictionary/map is used to classify a given land use via CN tables provided by Chow , 1959. Some CN classifications were slightly modified to fit the table due to a lack of automated information in terms of the type of managements applied to the given area of interest or HRU polygon. These initial CN estimates were then aggregated via a weighted average, defining the weight as the percent area a given HRU occupies within the area of interest the user uploaded or digitized within eRAMS' GIS functionalities.

The initial CN analysis may not be sufficient for the user due to rapid urbanization, other best management practices (BMP), or a lack of current data. Thus, SWAT-DEG enables the user the ability to further analyze a given area of interest via land use modifiers. These land use modifiers are in the form of sliders and correspond to the percentage of area within the area of interest a given classification occupies. The general classifications provided by SWAT-DEG are forest, crop land, urban and residential area, grassland or pasture, and other. Thus, after generating the initial HRUs, each slider will contain a percentage and that percentage corresponds to the total area of a given classification over the total area of analysis.

Applying further generalization also requires further generalizing the CN to provide the user a new estimate of the CN. The generalization step takes a non-weighted average of all the CNs within one of the general classifications provided by SWAT-DEG. After which, any modification to the percent area sliders for a given general classification will result in a new weighted average. This new weighted average utilizes the new percent areas provided by the

user as the weight and multiplies these weights by the non-weighted averages held in storage by eRAMS from the original HRU generation. When addressing land use variations and CN modifications, climate is always a highly influential factor.

SWAT-DEG attempts to address climate change scenarios by enabling the user easy manipulation of the climate data files. SWAT-DEG requires a climate data file containing the minimum temperatures, maximum temperatures, average temperatures, and precipitation for a given time period. Both these are easily modified by the climate change sliders underneath the "Climate" tab of SWAT-DEG's user interface. The precipitation slider modifies the original precipitation data via a percentage multiplier. Whichever percentage the user places on the modifier, SWAT-DEG's code will multiply and add back that percentage from the original amount. An incremental method is applied to the temperature. Thus, the modifier number displayed on the temperature slider is the increment that will be added or subtracted to the minimum temperature and maximum temperature.

2.4 Climate Station

Climate data is an essential aspect of SWAT-DEG's processes. In the original version of SWAT-DEG, uploading climate data was difficult and cumbersome. Uploading climate data required struggling through installing and figuring out how to upload a file. In addition, this struggle climaxes at the aspect of formatting the data into the exact format a FORTRAN compiler can read for SWAT-DEG purposes. The web services eliminate installation and helps minimize the frustrations in obtaining and formatting the data for SWAT-DEG.

SWAT-DEG not only provides instructional guidance and links to locations where data can be obtained for analysis purposes, but eRAMS supplements this with automated querying functionalities to obtain various data, including climate data. SWAT-DEG provides instructions to National Climatic Data Center (NCDC), as figure 3 demonstrates, and eRAMS provides data to Global Historical Climatology Network - Daily (GHCND) and Snowpack Telemetry (SNOTEL), the GHCND user interface is demonstrated in figure 4.

Instructions to Download NCDC Data

Step 1:
1. Download data from this website: Download Data Link
2. Copy and paste the "GHCND" tag of the desired station (located on the far right) into step "1" of the website.
3. Choose "Daily GHCND" from the select box under "2" of the website.
4. Now click the button "Search."

Step 2:
1. Click the button "Add to cart" for the desired station
   ○ NOTE: The first station listed should be the desired station.
2. Now click the button "Cart (Free Data)" in the upper right hand corner.
3. Choose the output date range in the upper right hand corner.
4. Choose "custom GHCN-Daily CSV" as the output format from the NCDC site
5. Choose "Daily GHCND" from the select box under "2" of the website.
6. Now click the button "Continue."

Step 3:
1. Now, click the "+" by the option "precipitation" and check "PRCP" only.
   ○ NOTE: If the option is not there, then that data does not exist for current station
2. Now, click the "+" by the option "Air Temperature" and check "TMAX" and "TMIN" only.
   ○ NOTE: If the option is not there, then that data does not exist for current station
3. Once the data has been downloaded, click **here** to see how to format the data.

Figure 3: NCDC Download Instructions Provided by SWAT-DEG User Interface

14
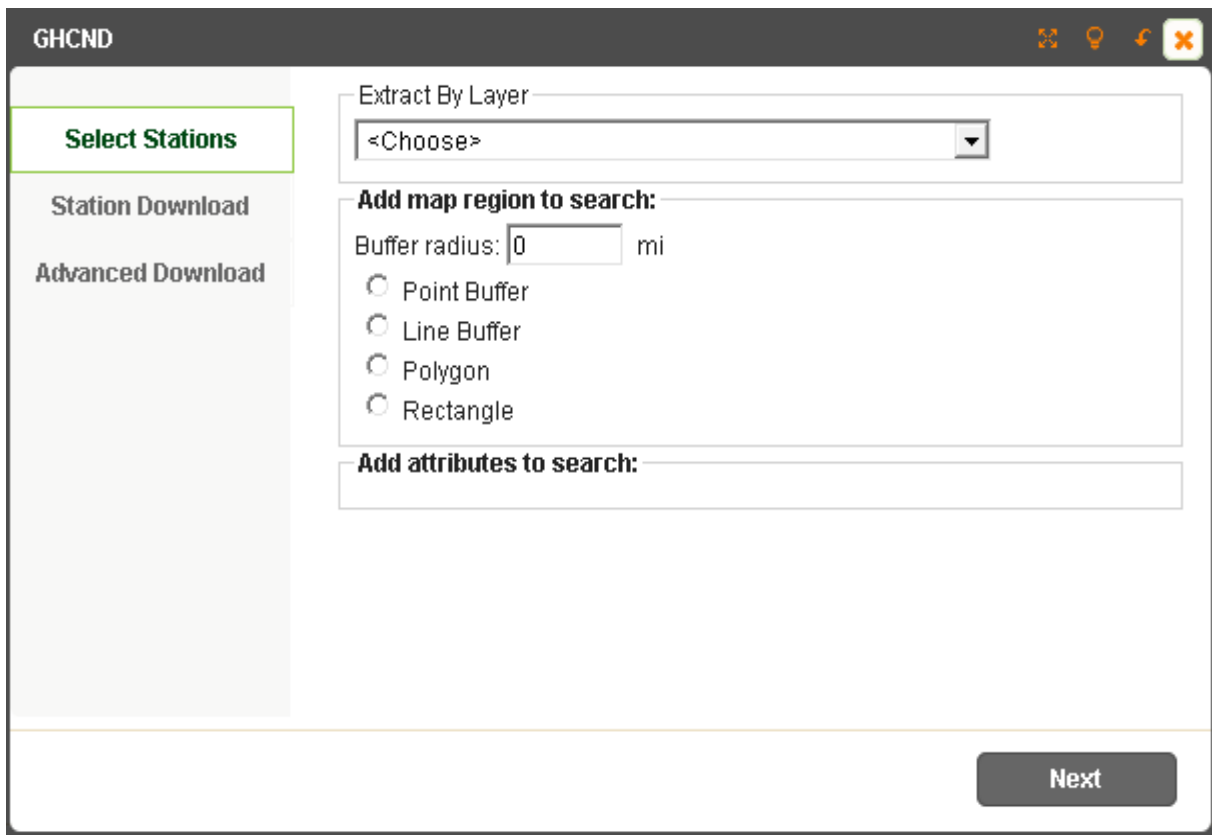
Figure 4: GHCND Data Downloader User Interface Hosted by Erams

After downloading the data, SWAT-DEG requires the user to reformat the data; however, reformatting the data is not as complicated as the original requirements. Instead, SWAT-DEG's user interface requires a different format then SWAT-DEG's source code. The reason is to make implementation of SWAT-DEG easier for the user. The format required by SWAT-DEG's user interface is shown in figure 5.

Year	Julien Day	Precip	Avg Temp	Max Temp	Min Temp
		mm	(^{0}C)	(^{0}C)	(^{0}C)
1974	1	0	-3.8	5	-8.3
1974	2	1	6.8	8.9	4.4
1974	3	0	2.5	10	-3.4
1974	4	0.5	1.1	5	-1.2
1974	5	0	16.2	24.4	10.4

ClimateData.csv Format

1. Place the data (using Excel) in their respective columns as demonstrated below
2. Save the Excel file as a ".csv"
3. NOTE: exclude the first two rows in your Excel File

Figure 5: Required format for SWAT-DEG Climate Data Uploading

As seen, the format is fairly simple and requires little effort on the user's part. The column titled "Julien Day" is referring to each day as a single integer that is continuous till the end of the year. For example, January 1st for any year is Julien day 1 and January 2nd is Julien day 2, and so forth. Uploading a climate data file, with the format specified above, results in SWAT-DEG regenerating the climate data on the fly when the user executes SWAT-DEG. The reason for generating the climate data on the fly is to easily apply any modifications to the climate data, as specified in section 2.3.

2.5 Visual Graphical User Interface

Previously, SWAT-DEG had a visual graphical user interface, or the ability to graph the results of a given run. However, the graphs did not lend itself to user interaction. User interaction includes but is not limited to zooming features, printing features, hovering features, and etc. Web services provides the platform for such interactions and are implemented within SWAT-DEG. SWAT-DEG utilizes JS HighCharts (AS, 2014b). JS HighCharts is a website

that provides a pure JavaScript (JS) library for visual graphing capabilities, a demonstration of the graph is shown in figure 6.
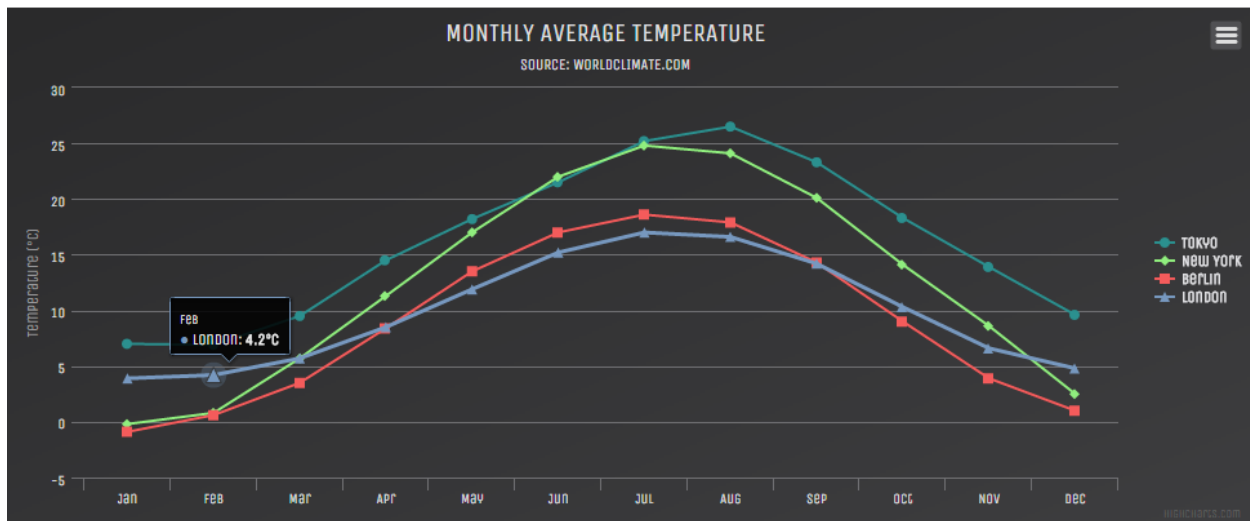
Figure 6: Demonstration of JS HighChart's Functionality (AS, 2014a)

JS HighChart provides a plethora of graph styles, graph types, and graph features that are all customizable and free. SWAT-DEG consumes JS HighCharts' library for visual aids in analyzing the results of a simulation. SWAT-DEG also provides the user with statistical options to represent the data in a meaningful manner for a given objective.

Different representation of data comes in two forms within the SWAT-DEG interface, timestep and a statistic. Thus, any output parameter from SWAT-DEG can be visualized in any variation of timestep and statistic. SWAT-DEG has a daily, monthly, and yearly timestep and provides a minimum, maximum, summation, mean, and various percentile options for visualizing a given output parameter. The output parameters include Evapotranspiration (mm), runoff (mm), flow (mm), groundwater (mm), velocity (m/s), width (m), slope (m/m), depth (m), sediment in (tons), degradation (tons), deposition (tons), sediment out (tons), wash load (tons), bed load (tons), and flow duration curves. SWAT-DEG's graphing features for each variation of timestep and statistic includes multiple scenario graphing, zooming features, hovering features, and print capabilities.

Graphing a single statistic and parameter may or may not have utility or meaning to the user. Only when this graph is seen in conjunction with another graph will any meaning or utility come about. Thus, SWAT-DEG provides the user the capability to graph multiple scenarios for a given parameter, timestep, and statistic simultaneously. The scenarios are various SWAT-DEG simulations. Variations include but are not limited to land use scenario changes,

climate scenario changes, and input parameter changes. As a result, a graph can contain the effect of progressive urbanization on degradation and width of a stream. The graph also enables the user to print or save the graph in multiple formats, zoom into certain time periods, and hover over data points for a quick view of a given data point.

2.6 Stochastic Analysis

River restoration, and thus river stability analysis, still have two challenges : (1) incorporating the effect of natural variability and (2) estimating the uncertainty within stream parameters (Jha, Western, Rutherfurd, & Grayson, 2005). SWAT-DEG implements stochastic analysis via the Monte Carlo methodology (Ronald Christensen, Wesley Johnson, Adam Branscum, 2011; Yang, 2011) to minimize uncertainty of the SWAT-DEG's input parameters. Each parameter of SWAT-DEG can undergo stochastic analysis with a Constant, Beta, Binomial, Chi-square, Exponential, Extreme Value, F, Gamma, Generalized Pareto, Geometric, GEV, Hypergeometric, Log Normal, Negative Binomial, Noncentral chi-square, Noncentral t,Normal, Poisson, Rayleigh, Students t, Uniform, or Weibull distribution. SWAT-DEG's user interface will show the parameters required to define the chosen distribution and provide default values if the distribution is given no definition. In addition, SWAT-DEG's user interface provides brief descriptions as to what should be inputted where.

When running a simulation involving stochastic analysis SWAT-DEG's source code contains three triggers. The triggers are as follows: generate (1) JavaScript Object Notation (JSON) object, (2) send JSON object to Cloud Services Innovation Platform (CSIP) (Lloyd, W, David, O, Lyon, J, Rojas, K.W., Ascough II, J.C., Green, T.R., Carlson, J.R., 2012), (3) execute Monte Carlo Methodology via Java. The first two triggers are explained in detail in chapter 3 and the third trigger's relationship with cloud computing is also defined in chapter 3. The following description of the third trigger is a high level description.

The third trigger is composed of three processes. The first process is generating a user defined distribution object within the source code. With the newly created distribution object, SWAT-DEG's Java code generates random values within the bounds of the given distribution. The number of parameters created is dependent upon how many simulations the user inputs and how many parameters are undergoing stochastic analysis. Thus, if the user inputs 500 simulations for two separate parameters, then 1000 random values are generated, 500 for each parameter undergoing stochastic analysis.

The second process is to merge the newly created random values into files for SWAT-DEG to run. Thus, continuing with the example of 500 simulations, the Java code generates 500 sets of input files for SWAT-DEG to execute. Each set of input files contains one set of stochastic parameters. One set of stochastic parameters is one random value for each input parameter undergoing stochastic analysis. Thus, continuing with our previous example, a set would contain two random values.

The third process is the actual execution of the SWAT-DEG model and map reduction. Multiple instances of SWAT-DEG are substantiated to efficiently run all the input file sets. After set has ran fully, SWAT-DEG employs the concept of map reduction, which for the purposes of SWAT-DEG is the action of reducing many outputs to one output. Thus, SWAT-DEG will only return one set of output files to the user.

2.7 User Interface Features

The user interface previously used for SWAT-DEG simply contained the textboxes for data input. The new interface, via web services, inherited the old user interface but also enhanced the old interface. These enhancements came in the form of reorganization of data, user error checks, climate data upload and reformat functionality, eliminating desktop installation, online data storage, and supplementary features provided by eRAMS.
The layout or organization of the actual user interface is vital in societies' acceptance or rejection of anything. This is particularly true in the engineering world because there is a plethora of models but only a select few are commonly used. Thus, a tab format approach is applied to SWAT-DEG to help organize the data into the following categories: setting, climate, scenario, management, hydrology, bed material, channel, aquifer, topography and soil, and outputs. Separating the data as such makes finding inputs easier for the user, and thus, makes fixing user error easier alleviate.

User error checks is the source code determining if everything is inputted correctly to have a meaningful run. SWAT-DEG employs a few user checks but the most common is forgetting a data input or entering an invalid input, figure 7 demonstrates an example of SWAT-DEG's error message.
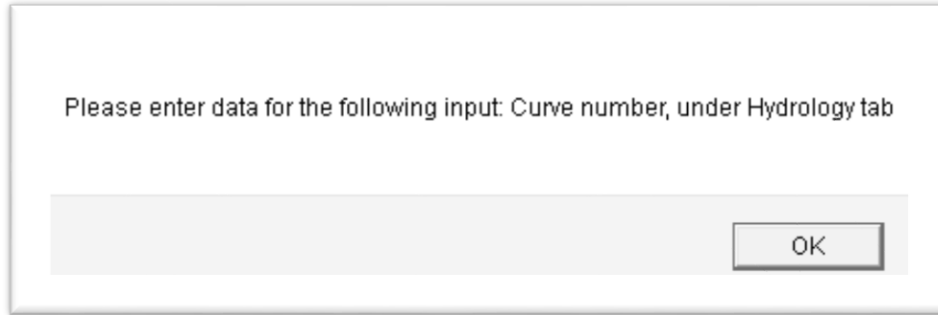
Figure 7: Example of an Error Catch within SWAT-DEG User Interface

Figure 7 states that the user forgot to enter a value for the parameter "Curve number" and this parameter is located under the "Hydrology tab." When the user successfully runs SWAT-DEG, the data is stored in an online database for the user's convenience, a glimpse of this feature is demonstrated in figure 8.
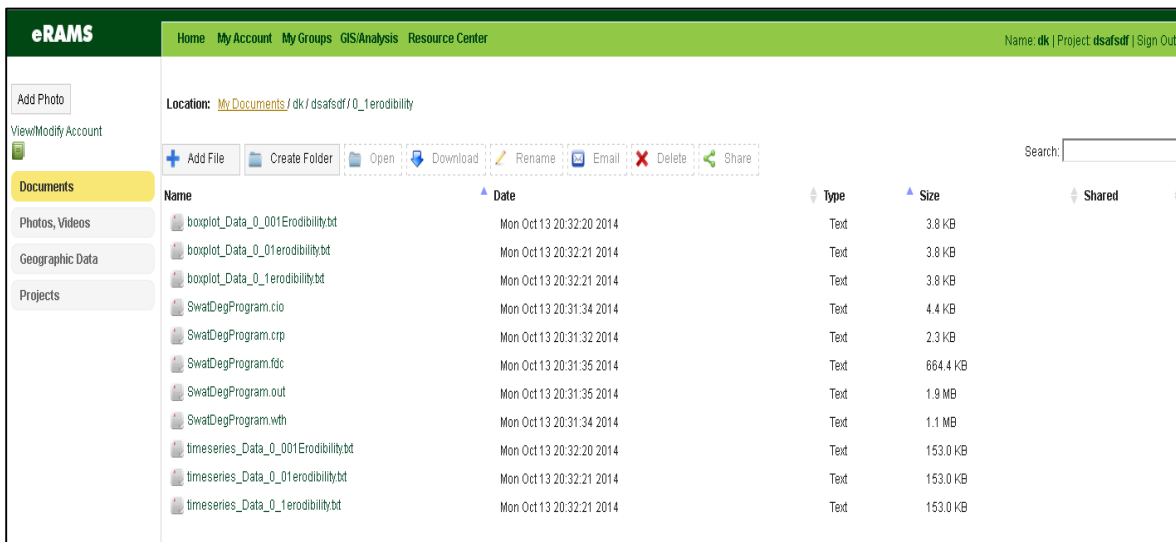
Figure 8: Glimpse into the Documentation Feature provided by eRAMS for the Model SWAT-DEG

As shown in figure 8, there are eleven documents within this given directory. These 11 documents correspond to a scenario ran within SWAT-DEG and the user can download, share, or store all these documents until they are needed.

Chapter 3: SWAT-DEG's Cloud Computing and Web Service

3.1 Web User Interface:

A web service has two overall components, the Uniform Resource Locator (URL) and Extensive Markup Language (XML). The URL is the identifying pointer and the XML is the public interface and bindings (Dustdar & Schreiner, 2005). These two concepts enable invoking internet protocols to interact with a web interface.

The overwhelmingly universality and accessibility of the internet and the flexibility to target audiences with tailored interfaces induces an attractive scheme for model-adaptation into web services Models' user interface can exist within the common platform of a browser with minimal hindrance from operating systems (OS) due to today's rich internet connectivity. In addition, the web assists in consolidating model information, model requirements, and different models/tools a single location. Thus, alleviating hindrances in tracking, obtaining, verifying validity, and sharing results between modelers and models (Aldrawiesh, Al-Ajlan, Al-Saawy, & Bajahzar, 2009; R. M. Argent, Perraud, Rahman, Grayson, & Podger, 2009; Gortmaker, Janssen, & Wagenaar, 2004). Such features assist modelers by attempting or diminishing hindrances occurred by desktop dependencies.

SWAT-DEG's current user interface is demonstrated in figure 9 and was designed to maximize flexibility by providing key features.
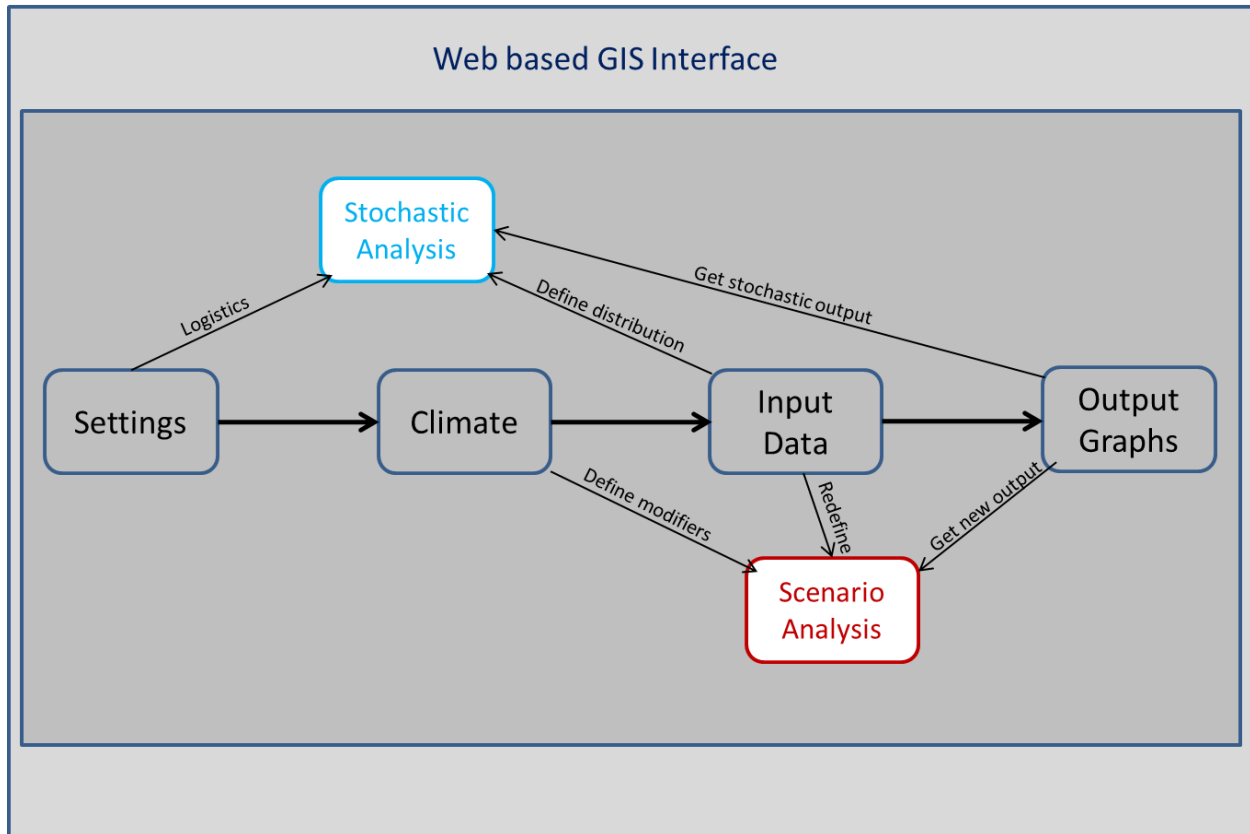
Figure 9: SWAT-DEG's User Interface Summarized.

There are four main features of SWAT-DEG's user interface that encourages audience interaction and modelling: (1) watershed generation, (2) climate station functionality, (3) data entry, and (4) visualization of results, demonstrated in figure 9. These four features compliment SWAT-DEG's purpose and enlarge the user base.

The "settings" icon represents the logistical operations of SWAT-DEG. These operations are simulation time, the model's start year, number of stochastic runs, and defining the watershed layer. SWAT-DEG defines a watershed layer by the user physically drawing the watershed boundary via predefined functions (polygon, rectangle, and circle) or defining the watershed via a Hydrologic Unit Code (HUC 8, 10, or 12). The HUC choices are automatically populated by the extent of the interactive map associated with the current project, represented by figure 9's "web based GIS Interface" icon. The next icon "climate" is tied to both the SWAT-DEG's settings and GIS interface. SWAT-DEG queries an eRAMS layer which contains the locations of National Climatic Data Center (NCDC) database and from the United States Geological Survey (USGS) database. This query is bounded by the

watershed defined previously in SWAT-DEG's settings and the buffer around the watershed boundary defined by the user's preference. After defining the watershed, the user is required to input the data for the model to run the model. The outputs can then be manipulated by predefined, interactive graphical visual options. Each of the basic premises of SWAT-DEG (i.e. settings, climate, input data, and output graphs) can be modified by either the stochastic or scenario analysis of SWAT-DEG.

Stochastic analysis requires extra input from the settings and input data aspect of SWAT-DEG. The user must input the number of simulation runs, from the settings tab and define the distribution for the parameter(s) undergoing stochastic analysis. After running SWAT-DEG's executable for the stochastic analysis, the pre-defined output visualizations change to display appropriate visualizations for stochastic analysis.

Scenario analysis enables analysis of climate change by modifying the uploaded climate data. If scenario analysis is used, the user is required to create a new scenario within the current SWAT-DEG project. Thus, the "default" scenario becomes the baseline and any new scenarios, created by the user, are then compared to the baseline. After simulating, SWAT-DEG's graphical interface enables overlaying the results of multiple scenarios within the same graph.

Models cannot only provide meaningful output but models also must incorporate an interface that supports the user in their objectives. Supporting the user is accomplished through limiting or expanding the user's flexibility in manipulating the model's parameters to communicate the results in an effective and appropriate manner. These results then must become transparent, where transparency is the continual action of presenting a project to the modelers, stakeholders, and management planners in an appropriate manner (Campo, Bousquet, & Villanueva, 2010; Voinov & Bousquet, 2010). Tailoring a model's user interface to specifically support each population as a distinct user innately makes the model and its results transparent. Thus, everyone gains a level of understanding of the river restoration/rehabilitation project, particularly the stakeholders. This level of understanding will promote the cooperation and support of solutions to river restoration/rehabilitation (Gortmaker et al., 2004). This transparency is still limited by SWAT-DEG's dependence upon location.

Location dependency minimizes the ability for a user to apply SWAT-DEG when the user is not physically in the office. SWAT-DEG must also embrace the concept of different devices executing SWAT-DEG. The platform

eRAMS is developed to be device independent, so long as internet access and functionality is available. Hence, the user base of SWAT-DEG is increased due to maximizing access points for SWAT-DEG (La et al., 2010). Increasing access points for SWAT-DEG is beneficial only if external sources are minimized. An external source is defined as a site other than eRAMS that the user must visit in order to finish the analysis of a project via SWAT-DEG.

The server, eRAMS, is a flexible distributed-modeling platform for creating an environment conducive to effective and time efficient modeling, regardless of user's experience or knowledge. As such, eRAMS has four key features to empower the user and further the modeling capability of SWAT-DEG for river restoration/rehabilitation projects. These four features are: (1) geospatial operations including layer operations such as merging and intersecting, built in base layers such as National Hydrography Dataset (NHD), and google map layers, and ability to manage and manipulate user defined layers, (2) sharing results, inputs, and documents between any eRAMS users, (3) a single location to download data from multiple data servers such as Soil Survey Geographic Database (SSURGO), State Soil Geographic Database (STATSGO), NCDC, NHD, STOrage and RETrieval water quality database (STORET), snowpack telemetry database (SNOTEL), and (4) other models to supplement analysis and/or outputs. This is useful but all features are still inhibited by dependencies upon external sources.

Localized data downloads is one method eRAMS implements to reduce external dependencies. eRAMS provides a tool to download data from multiple data servers without leaving the modelling environment. This data can then be shared via eRAMS to anyone given permission to see the data. Thus, both transparency and team cooperation is enhanced while minimizing time spent on navigating multiple sites to obtain and share data between engineers, stakeholders, and planners. Due to the underlying modelling environment, the modeler will have extra time to run supplementary analysis. Supplementary analysis also benefits from localizing resources to one platform because eRAMS can help provide background information for each model hosted to help ensure the modeler is maximizing efficiency in choosing and running additional models.

3.2 Cloud Computing:

Service Oriented Architecture (SOA) allows for application functionality through services given by external providers (Gortmaker, Jeffrey, Janssen, Marijn, Wagenaar, René W., 2004). Service-oriented computing is the key

to deploying SWAT-DEG in a scalable manner. Service-oriented computing is defined as decomposing a software system into independent distributed components. These components are deployed on remote servers, such as CSIP, and designed to respond to eRAMS' users (Castronova et al., 2013; Huhns & Singh, 2005; M. Papazoglou, 2006). eRAMS utilizes the Infrastructure-as-a-Service (IaaS) CSIP cloud to execute models via web services, as demonstrated in figure 10.
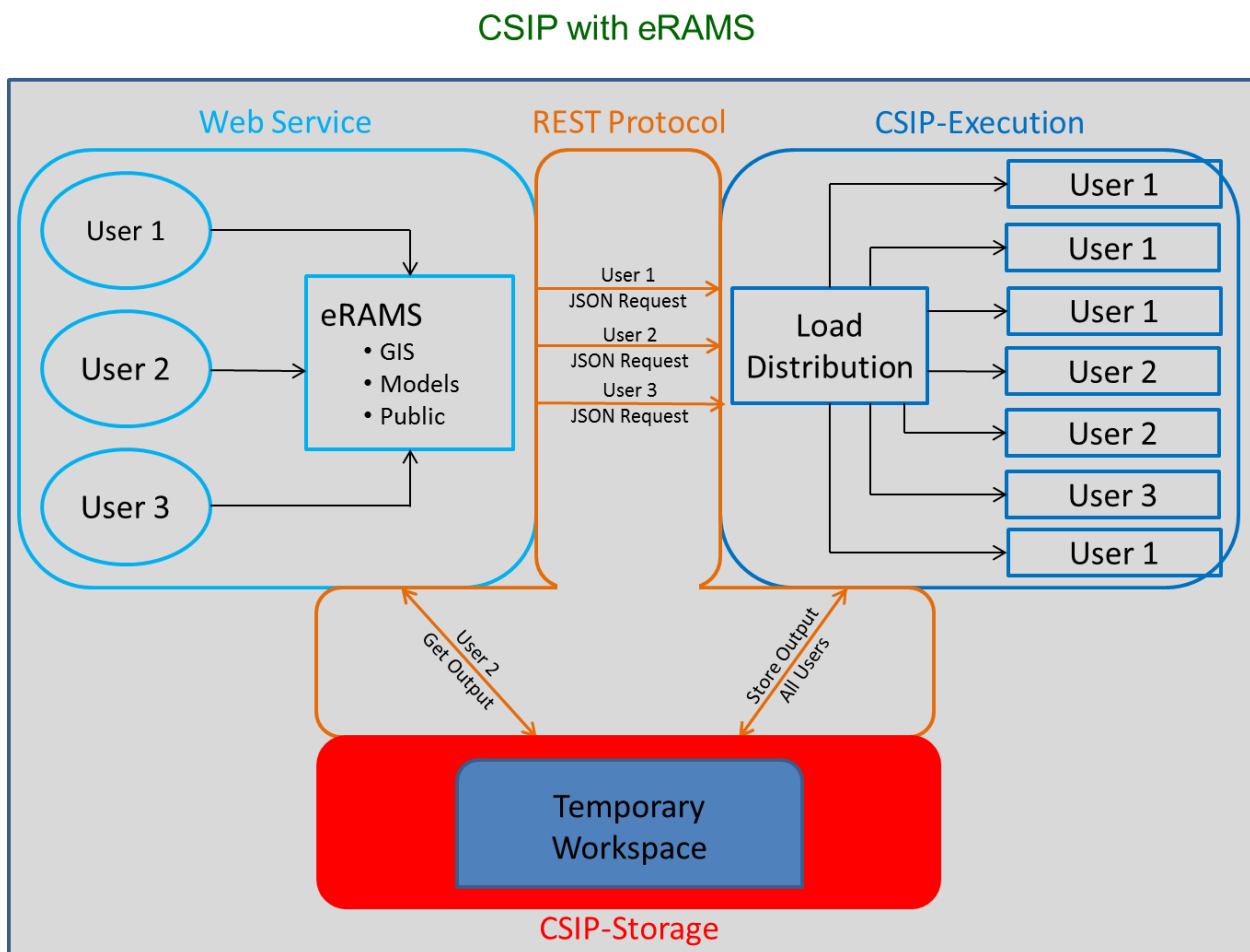
Figure 10: The Integration of CSIP with eRAMS for Executing Environmental Models within the Cloud Environment

Figure 10 represents the integration of eRAMS and CSIP to provide modeling as a service (Lloyd, W, David, O, Lyon, J, Rojas, K.W., Ascough II, J.C., Green, T.R., Carlson, J.R., 2012). The user will interact, manipulate, and visualize input and outputs of environmental models via the web platform eRAMS. The execution of an environmental model is invoked within eRAMS, however, eRAMS only acts as the gateway to executing the model within the cloud environment, CSIP. After CSIP receives the request for executing a model, CSIP utilizes a load

distributor to assist in optimizing CSIP's efficiency in minimizing time a request spends within the cloud platform, as shown in figure 10. The processes demonstrated above accomplish three primary objectives for an environmental model, like SWAT-DEG:

1. Enable multi-threading or high demand loading for environmental model.
2. Reduce the relative computational time of environmental models executed within CSIP versus a local machine.
3. Minimize consumption of local machine's resources in execution of environmental models within CSIP versus a local computer.

The benefits of CSIP are exhibited when users post high load requests, such as implementing a stochastic analysis. The benefits are revealed in the comparison of execution times of environmental models within CSIP versus a local machine, which are directly dependent upon the load of the request posted. The load of a request refers to how resource intensive and computationally expensive is the execution of the environmental model requested. A simple example is analyzing the consumption of resources when executing a 10 year simulations versus a 100 year simulation. Executing a simple simulation versus a complex simulation within CSIP will not necessarily portray the same results of running the two simulations locally. This is due to the initial overhead required to execute models within cloud infrastructure (CSIP). CSIP's initial overhead is due to utilizing Representational State Transfer (REST) as a third party between the user and CSIP.

JavaScript Objection Notation (JSON) encodes the users' input when posting a request to CSIP and is also utilized when returning output (Lloyd, W, David, O, Lyon, J, Rojas, K.W., Ascough II, J.C., Green, T.R., Carlson, J.R., 2012; Lloyd, Wes, Pallickara, Shrideep, David, Olaf, Lyon, Jim, Arabi, Mazdak Rojas, Ken., 2011). After CSIP receives a request, RESTful protocol automatically distributes the request's loadings by assigning virtual machines (VM) to the request. A virtual machine is essentially a software-based emulation of a computer, in this case, virtual machines are the emulation of a cluster of computers that generates the server of CSIP. When distributing the loadings from a request, there are multiple methods of transferring the load from the load distributor to a VM. A single request can be transferred to a VM in its entirety or a single request is segmented into multiple VMs. Minimizing time a request spends within the CSIP environment dictates which of the two previously stated options is applied to a request, given CSIP's current allocable resources. Thus, CSIP will hold requests if there are no

allocable resources available to execute new requests. CSIP utilizes REST, for the processes stated above, because of RESTful protocol's inherent properties.

REST normally relies upon Hyper Text Transfer Protocol HTTP protocol. HTTP provides four operations for machine to machine interaction, create, read, update, and delete. Thus, REST provides a lightweight alternative for creating a network application, relative to architecture styles such as Simple Object Access Protocol (SOAP) (Guan, Zhang, & Clarke, 2006).

Lightweight network application is a necessity to keep the overhead of service-oriented-modeling at a minimum. By doing so, service-based environmental modeling minimizes additional computational time due to server overhead. Thus, after diminishing the consequences of modeling within the CSIP environment, many models, of varying complexities should exist and execute within the CSIP environment. Additionally, integrating a simplistic environmental model with CSIP provides increased flexibility in scaling and modifying the original model.

3.3 Input/Output Traffic:

SWAT-DEG has two options for analysis, a stochastic or a deterministic analysis. The stochastic analysis consumes large amounts of resources and thus, displacing the load from the local machine to another machine is ideal. IaaS cloud enables users to deploy their requests, from eRAMS, to CSIP, where the load is distributed to multiple VM images based upon an algorithm (Lloyd, W., Pallickara, S., David, O., Lyon, J., Arabi, M., Rojas, K. 2013a). VMs provide SWAT-DEG the potential of reducing risk of inter-component interference to allow more efficient runs simultaneously (Lloyd, Wes, Pallickara, Shrideep, David, Olaf, Arabi, Mazdak, Rojas, Ken., 2013). In addition to reducing risk, SWAT-DEG utilizes CSIP for the following objectives:

1. Enable multiple user access, by executing SWAT-DEG in CSIP, where execution of SWAT-DEG includes both stochastic and deterministic analyses.
2. Reduce the relative computational time of running a stochastic analysis on CSIP versus on a local machine.
3. Minimize the user's local machine's resource consumption when invoking SWAT-DEG on a local computer versus on CSIP.

These objectives are dependent upon the ability to invoke and execute SWAT-DEG within CSIP in the most efficient manner, as to reduce both overhead and redundancies. The efficiency of CSIP for a given model is revealed by a ratio (Kollet & Maxwell, 2006). CSIP handles SWAT-DEG with the following workflow to ensure the continual minimization of said ratio.

1. Enter user defined inputs into SWAT-DEG's user interface (UI).

2. Transform UI inputs into JSON objects.

3. Send JSON objects, via web services, to cloud services as a platform (CSIP).

4. Allocate JSON object to a single VM.

5. Check if JSON object is requesting a stochastic or deterministic analysis.

    i. If deterministic, proceed to step 6.

    b. If stochastic, follow the subsequent steps.

        i. VM generates multiple input files (.cio, .wth, .crp) from data encoded by JSON.

        ii. Distribute input files to multiple VMs, based upon RESTful protocol. CSIP will distribute to minimize the current requests time spent within CSIP.

        iii. Proceed to step 6.

6. Execute model run.

    a. If a deterministic run is made, the results are returned via text files.

    b. If a stochastic run is made, the results are appended to one file asynchronously.

        i. CSIP then applies map reduction to reduce the outputs to manageable output files.

7. After all runs are finished, CSIP returns the output files.

Separation of SWAT-DEG's deterministic and stochastic analysis is a necessity because the amount of input files generated via a stochastic analysis consumed too much memory within CSIP. Thus, the two types of analysis, stochastic and deterministic, became two separate services provided by CSIP. The deterministic service became the base service for any invocation of SWAT-DEG. In other words, if the user requests a stochastic service and desires 10,000 simulations, the stochastic service calls the deterministic service 10,000 times. However, calling the deterministic service only occurs when executing the model's executable. The stochastic service provides different

pre-processing and post-processing then the deterministic service. The stochastic service from eRAMS to CSIP is described within figure 11.
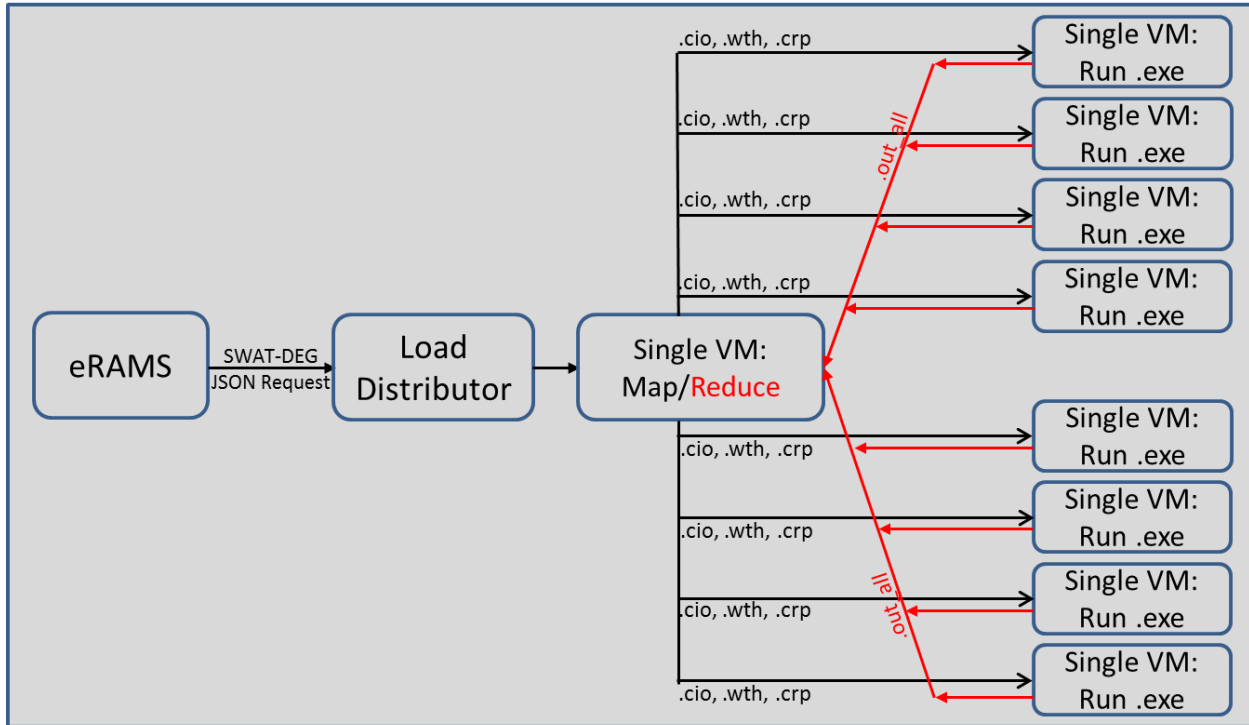
Figure 11: Integration of SWAT-DEG with CSIP. This Demonstrates the Specific Flow of Input and Output of SWAT-DEG from eRAMS to CSIP and Back to eRAMS.

The deterministic service and stochastic service are the same from eRAMS to the load distributor of figure 5. Both the deterministic and stochastic service shares the transference of a user request from eRAMS to the load distributor to a single VM. Within this VM, the JSON contains information for whether a deterministic or stochastic service is to be executed.

A deterministic approach requires three input files: (1) a crop look-up file (SwatDegProgram.crp), (2) a climate data file (SwatDegProgram.wth or climateData.csv), and (3) a parameter file (SwatDegProgram.cio). In the VM, CSIP searches for two records, SwatDegProgram.crp and a JSON object. The .crp file is a static file that is indefinitely stored within CSIP. SwatDegProgram.wth and SwatDegProgram.cio are not static, but are generated before the JSON request is sent. These two files are then attached to the JSON object and thus, the VM can execute SWAT-

DEG without any file manipulation or generation. The process of uploading these three files, which will be known as a set from here on out, is trivial but a stochastic analysis requires uploading a set in far larger magnitudes.

Uploading large magnitudes of information from eRAMS to CSIP via REST is not only inefficient but also impractical. To alleviate overwhelming amounts of memory consumption in the creation and post of such a large JSON request, the JSON object will not contain input files. Instead, the JSON object will contain parameters that define a singular, deterministic SWAT-DEG analysis.

The JSON parameters contains information to populate a Java Hashmap ("Java Hashmap," 2014), provide information to manipulate the climate data, and provide information in regards to the parameters undergoing the stochastic analysis. The information provided in the JSON request will generate input files needed to run a deterministic run for each run requested by the user for a stochastic analysis. Thus, if the user wants to simulate 10,000 runs for a stochastic analysis, the VM will generate 10,000 sets from the JSON request. When every set is generated, sets will be distributed to multiple VMs to obtain higher efficiency. Higher efficiency refers to the amount of time past starting with a user request and ending with CSIP returning the output via RESTful protocol. When a VM receives a set or multiple sets, the VM invokes a deterministic service that consumes a single set. Multiple VMs will asynchronously undergo the same process until the stochastic request is fulfilled.

Due to each VM receiving different sets at different times, the VM will receive output from SWAT-DEG's executable at different times. Storing these output files until the stochastic analysis concludes is not practical because of memory limits. Thus, to maximize CSIP's memory and to ensure the independency of each VM, the stochastic service utilizes an asynchronous post-processing approach. After a VM uses the deterministic service to call SWAT-DEG, the VM then appends the outputs returned to a single file. The single file is stored into temporary memory by CSIP and is constantly accessed by multiple VMs. Thus, CSIP's memory is preserved by an asynchronous post-process of SWAT-DEG's executable's output files. After the stochastic analysis is done within every VM, the appended output file undergoes map reduction. map reduction is the process of decreasing large amounts of data through mathematical processes. This file can then be used by SWAT-DEG's user interface for pre-determined graphical visualizations.

3.4 Future SWAT-DEG User Interface Suggestions

SWAT-DEG's User Interface enables the user to begin defining an area of interest, entering in data for the model, running the model, and ending with graphing the model. However, these features can be enhanced to further the modelling experience of the modeler and stakeholder. Improvement of SWAT-DEG's UI requires adding GIS-like functionality, beta testing & audience targeting, and graphing capabilities.

GIS functionality will improve SWAT-DEG's user-friendliness by minimizing modeler's navigation time on sites other than eRAMS.com (eRAMS) and providing initial values or insight into needed parameters to run SWAT-DEG. There are three key ARCGIS functionalities that would improve SWAT-DEG's UI: (1) watershed layers, (2) Data extractions/manipulation, (3) and climate data queries. Each of these functionalities would be done through eRAMS' source code to ensure eRAMS independence from third party licenses.

SWAT-DEG is limited to one GIS watershed layer and deletes the previous watershed layer if a new one is defined. User defines these watershed layers via US databases (i.e. HUC 12, HUC 10, and HUC 8) or by literally drawing the boundary on the eRAMS' map. To improve SWAT-DEG's current watershed allocation, two new features should be added. The first feature is to allow the user to upload a GIS watershed layer and to manipulate that layer within eRAMS map interface. This feature will allow users more flexibility in defining their area of interest and, if needed, to use their own methods of manipulating GIS layers. The second feature is to enable multiple watershed layers to exist within a single SWAT-DEG project. Multiple GIS watershed layers will support analyzing the influence of changes in the land use, such as urbanization. These GIS watershed layers will also retain the latest climate station query, the watershed area, and the affiliated data corresponding to each GIS watershed layer within a given project.

Defining a GIS watershed layer also provides eRAMS an area of interest for which additional queries can proceed. Queries could occur for growing season lengths, land covers, soil texture, curve number, and time of concentration for the current area of interest. These values will be automatically extracted and inputted when the user defines the watershed boundary. Such extractions are possible by permitting eRAMS to automatically request data from established online databases such as Soil Survey Geographic Database (SSURGO), State Soil Geographic Database (STATSGO), NCDC, and NHD..

The watershed boundary also enables for the user to query climate stations within a given buffer. Currently, the user is able to quickly query climate stations without the need to navigate external sites. Thus, the potential learning curve required to efficiently download data is reduced. Furthering this concept is a dialog box that appears when the mouse is hovering over climate stations on eRAMS map interface. The dialog box will display the climate stations name and two methods of downloading the data. The first method is to click on a hyperlink that leads to an external website and the second method is to provide the user a universal tool for downloading. The universal tool will download the data required by the user without the user ever accessing an external site. Instead, eRAMS will be navigating the external websites or local databases to retrieve the specified data. Subsequently, the data will be automatically restructured to fit the format required by the model SWAT-DEG, which eliminates human error in downloading and restructuring the input data. More features, such as the ones described above, should be added to SWAT-DEG but require beta testing to determine what the features are.

Beta testing for SWAT-DEG must occur in order to guarantee SWAT-DEG's future in modeling. The purpose of beta testing is not necessarily to debug and document errors or bugs within the eRAMS platform and/or SWAT-DEG's UI, but rather to cater SWAT-DEG to the modeler, stakeholder, and researcher. Beta testing should be distributed to three main audiences, stakeholders, modelers, and academia. To cater to a particular audience, SWAT-DEG should contain unique interfaces for the modeler, stakeholder, and researcher. Thus, each interface is tailored to maximize the current user's experience and modeling capability.

Stakeholders will use the model to ensure that past, present, and future river restoration/rehabilitation solutions are necessary and appropriate for the given situation. However, to ensure the transparency between stakeholders and the engineers, stakeholders must be able to easily navigate and understand how to run and interpret the output of the model. To assist stakeholders, both eRAMS and SWAT-DEG can disable features, automatically populate values, and provide documentation to help the stakeholder learn the processes being analyzed with SWAT-DEG. Documentation could take the form of theoretical documents and descriptions of SWAT-DEG parameters. The description should include a basic description of the parameter, a range of values in which the parameter typically falls, and documentation that further describes the parameter's influence in river restoration/rehabilitation.

The modelers will want a more advanced version of SWAT-DEG and full access to eRAMS' features. An advanced view of SWAT-DEG would include higher potential to modifying parameters, modifying watershed disaggregation,

and full access to GIS features or operations. As a result, modeler's retain full flexibility of SWAT-DEG without deterring other users from also modeling with SWAT-DEG. SWAT-DEG's advanced view will provide the modeler full flexibility in manipulating the model and will also provide access to all features within eRAMS to ensure flexibility and efficiency when conducting project analysis. These same features hold true for academia, however academia would require different options within the graphical interface.

The graphical interface should be changed in three regards, different interfaces for different audiences, default graphs, and stochastic graphing capabilities. The graphing interface should be as simplistic as possible for the stakeholders. In addition to simplicity, general guidelines to interpreting the graphical visualizations should be displayed to help the stakeholder understand and support the engineer's decisions. The graphical interfaces, for modeler's and an academic researcher, would differ in the statistical options available and the graphical types displayed in the output tab. In addition, each view point's graphical interface should contain a default effective discharge graph. This graph will be a summary of all the results from SWAT-DEG, but modelers and academia will be able to enhance the summary graph via stochastic analysis.

Stochastic analysis must also undergo beta testing to ensure that the graph types chosen are meaningful and significant to the audience utilizing SWAT-DEG. However, the stochastic analysis should also be used to enhance the effective discharge graph to help minimize the impact of uncertainty within river restoration concepts

4.1 Introduction

The following chapter is a step through guide to executing SWAT-DEG from the beginning to the end. The guide will provide pictures to provide guidance and visualizations to each of the steps. Details of specific features or processes are detailed in chapter 2 and chapter 3 of this thesis. In addition, this guide will only provide steps and minimal details about a given feature or process.

4.2 Creating a new project

The first step to starting a SWAT-DEG project is substantiating a SWAT-DEG project. This is done by navigating to eRAMS "My Account" page, which is shown in the top left and highlighted by a blue oval in figure 12. On this page, click "Create Project," which is highlighted by a red oval in figure 12.
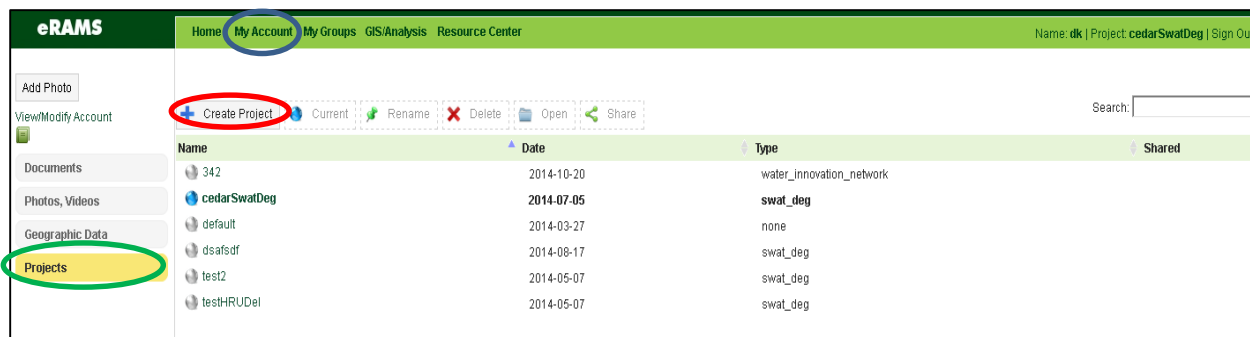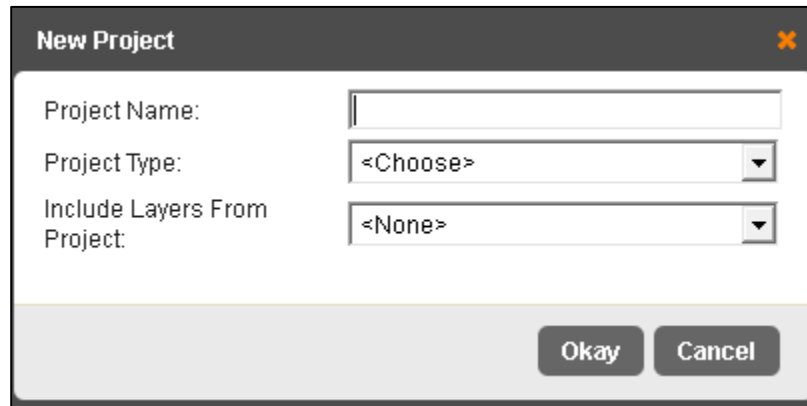
Figure 12: "My Account" Page of eRAMS
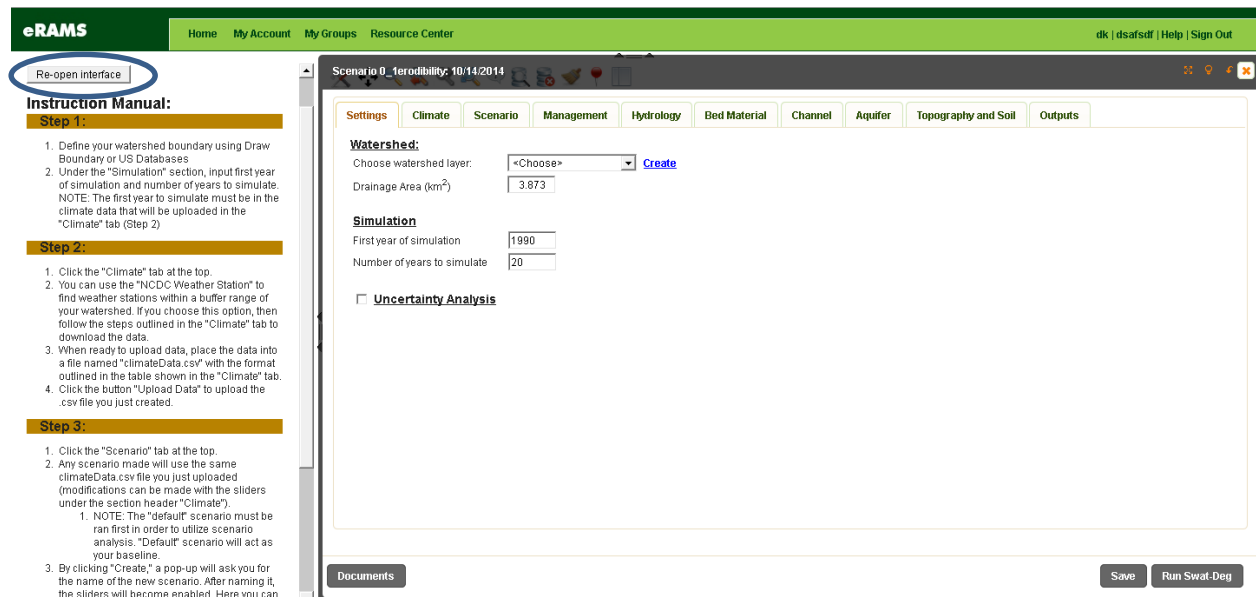
Clicking "Create Project" pops up a dialogue box, as shown in figure 13.

Figure 13: Pop-up Dialogue Box for Generating a New Project

Enter in the project name, which will be logged in the "projects" directory under the "My Account" page of eRAMS. This directory is highlighted in figure 12 by a green oval and has a yellow fill for the button. Clicking the drop down menu for "Project Type" generates a list of projects that are accessible by the current user's account permissions. Some accounts will have access to different projects depending on the status of the user's account in eRAMS. The last drop down menu "Include Layers from Project" is optional. It is to allow the transfer of layers from one project to another. For example, if the user already digitized an area of interest layer, the user can transfer that from project to project. Once the user clicks, "okay" on the "New Project" dialogue box, figure 14 is shown.

Figure 14: Full Representation of User Experience when Generating a SWAT-DEG Project

Here the SWAT-DEG interface is in front of eRAMS GIS interface. This is typical of many projects within eRAMS but this possible detriment to the accessibility of the user interface is limited by two features of eRAMS. In eRAMS, each dialogue box comes with the ability to dock and close the dialogue box. The docking option docks SWAT-DEG's user interface to the bottom of the screen and re-opening SWAT-DEG's user interface is done by clicking the "Re-open interface" button on the far upper-left hand corner, highlighted by a blue oval. This enables the user to interact with the GIS interface, shown in figure 15, while simultaneously viewing the brief instruction guide, on the far left, and viewing the parameters of the SWAT-DEG interface.
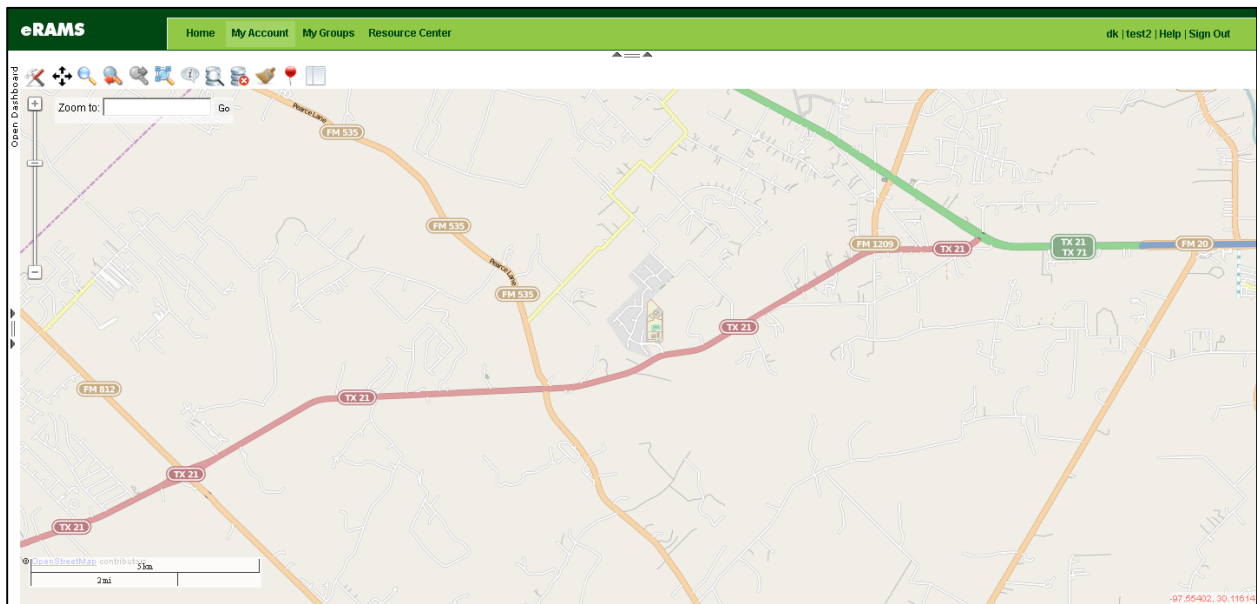
Figure 15: GIS User Interface Provided by eRAMS

The SWAT-DEG interface, figure 16, and eRAMS GIS interface, figure 15 are the key interfaces for the user.
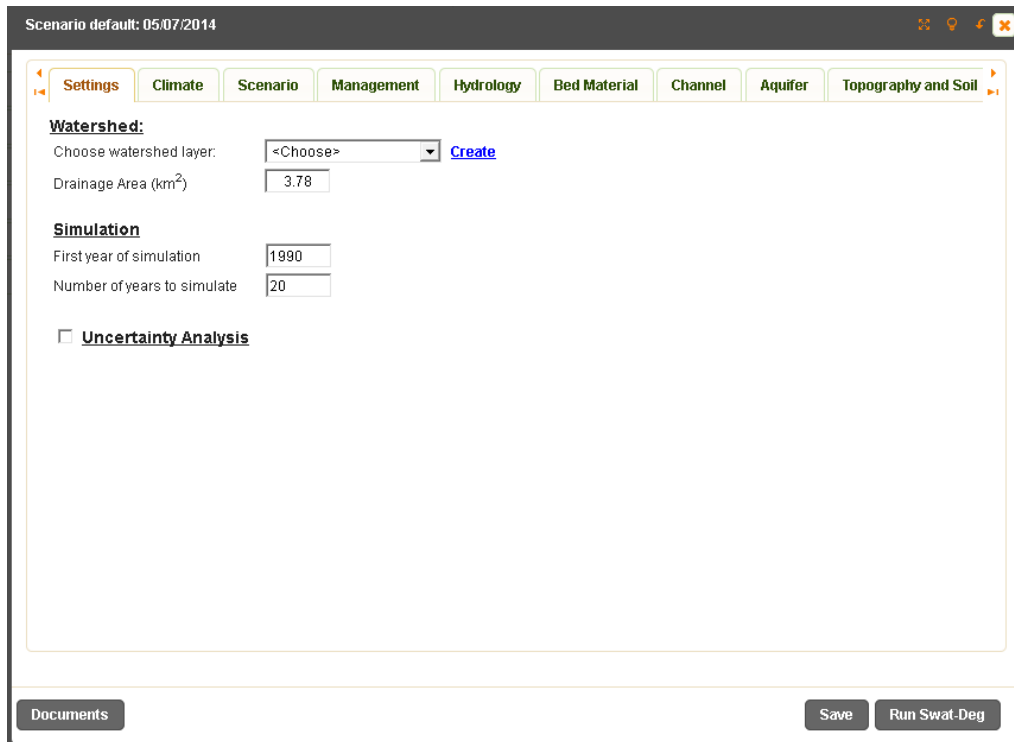
Figure 16: SWAT-DEG Main Interface

Within the main SWAT-DEG user interface, the user can click on any tab to begin inputting the necessary data for the current analysis. Despite this capability, the tabs are in an order that is deemed as the natural progression for maximizing user experience in executing a SWAT-DEG simulation.

4.3 Settings Tab

The first tab is the settings tab, which contains a combination of logistical parameters and SWAT-DEG parameters, shown in figure 13. The drainage area is a SWAT-DEG parameter whereas the rest are for model logistics or type of model run.

Figure 17: Settings Tab of SWAT-DEG's User Interface

First the "Watershed" section requires the user to define the area of interest for the current SWAT-DEG analysis.

When the user clicks the "Create" button, which is highlighted in blue, a dialogue box is displayed.

Figure 18: User Interface for HRU Generation and for Defining the Watershed

Figure 18 is the user interface for defining HRU and for defining the watershed or area of interest. Click New Folder

or select the folder that all HRU and watershed definition GIS layers are to be stored. As shown in figure 19, the

folder starting with "HRU Data: 1" is the folder that will contain the rest of the analysis and will be the basis for

querying any future layers for the current project.
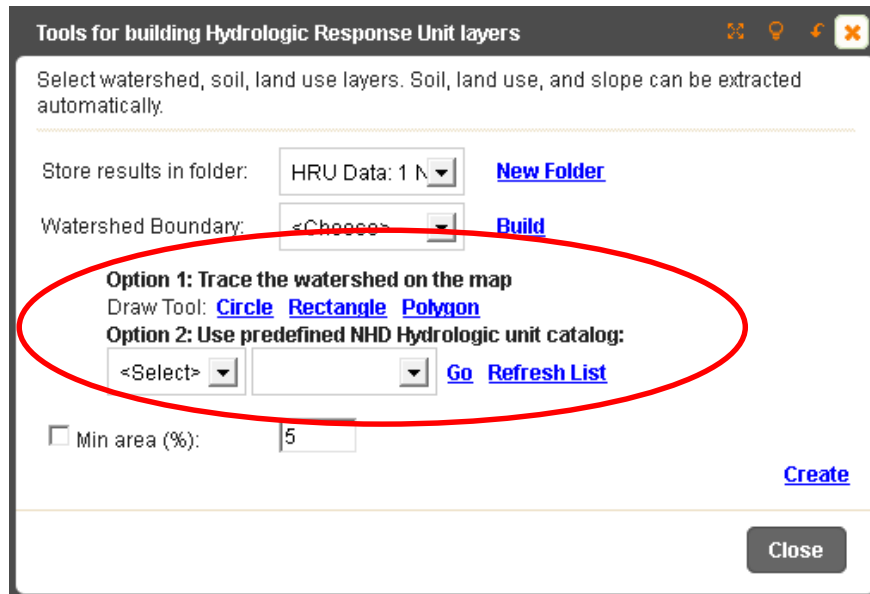
Figure 19: Defining Watershed Boundary

If the user created a new folder to store all future layers, then they will have to define a new watershed boundary by clicking "Build" on far right hand side of the row containing "Watershed Boundary." If the user chose an existing folder, then SWAT-DEG will populate the rest of the drop down menus with possible layer choices.

After clicking "Build" for the watershed boundary, the parameters highlighted by a red circle are visible. Here the user has the choice of three options, circle, rectangle, polygon for interactive methods to define the watershed boundary on eRAMS GIS user interface. The second option is to use NHD's Hydrologic Unit Code (HUC). After choosing the folder and defining the watershed boundary, the next step is building the land use layer in eRAMS.

If the user defined a new folder, then the user will also have to build a new land use layer for eRAMS. Otherwise, the user can click the select box to define the land use layer with a previously created layer.

To create a new land use layer in eRAMS, click the "Build" button in the "Land use" row. After doing so, the parameters highlighted in the blue oval are visible. Here the user clicks the drop down menu, which is shown as NASS in figure 20.
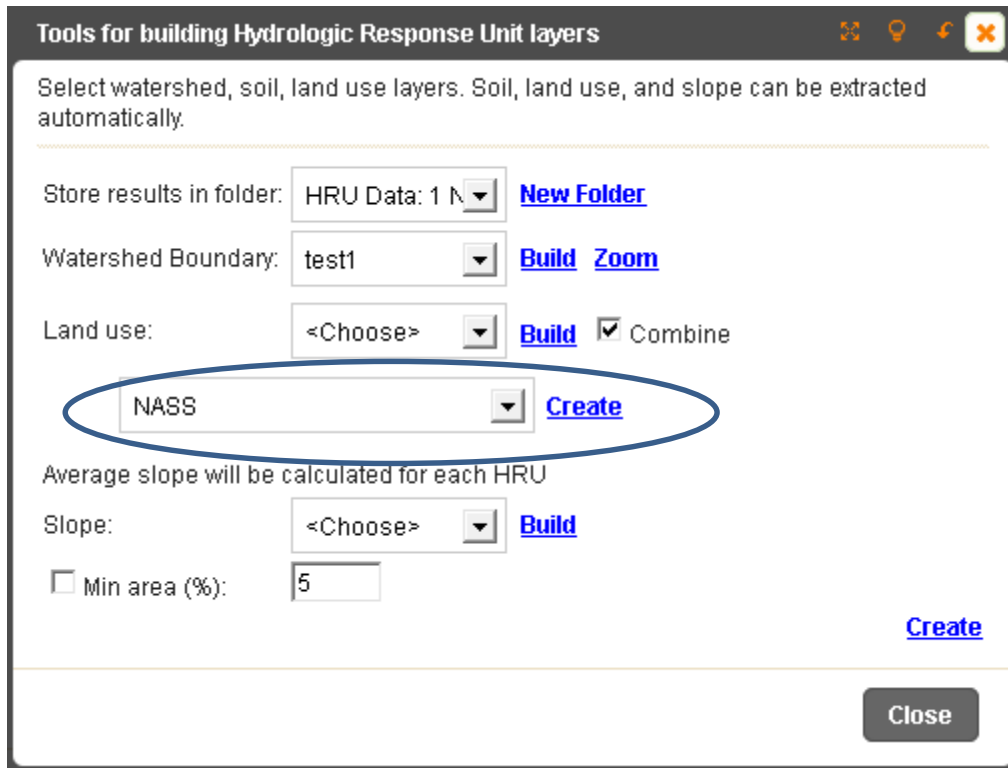
Figure 20: Define Land Use Layer in eRAMS

The only option currently offered by eRAMS is NASS. After selecting NASS, click the "Create" button on the right.

This will execute eRAMS' server to begin downloading the needed data from the NASS database. The layer is

uploaded to an eRAMS directory, specified by the folder chosen for beginning the HRU creation and the defining of

the watershed. After which, the downloaded layer is clipped to the boundary defined when the user generated a

watershed boundary.

If the user defined a new folder, then the user must also generate a new soils layer in eRAMS. Otherwise the user

can click the desired layer from the drop down menu.

To create a new soil layer in eRAMS, click the "Build" button that is within the same row as the label "Soil." Once

"Build" is clicked, the parameters highlighted in figure 21 by the blue oval are visible.
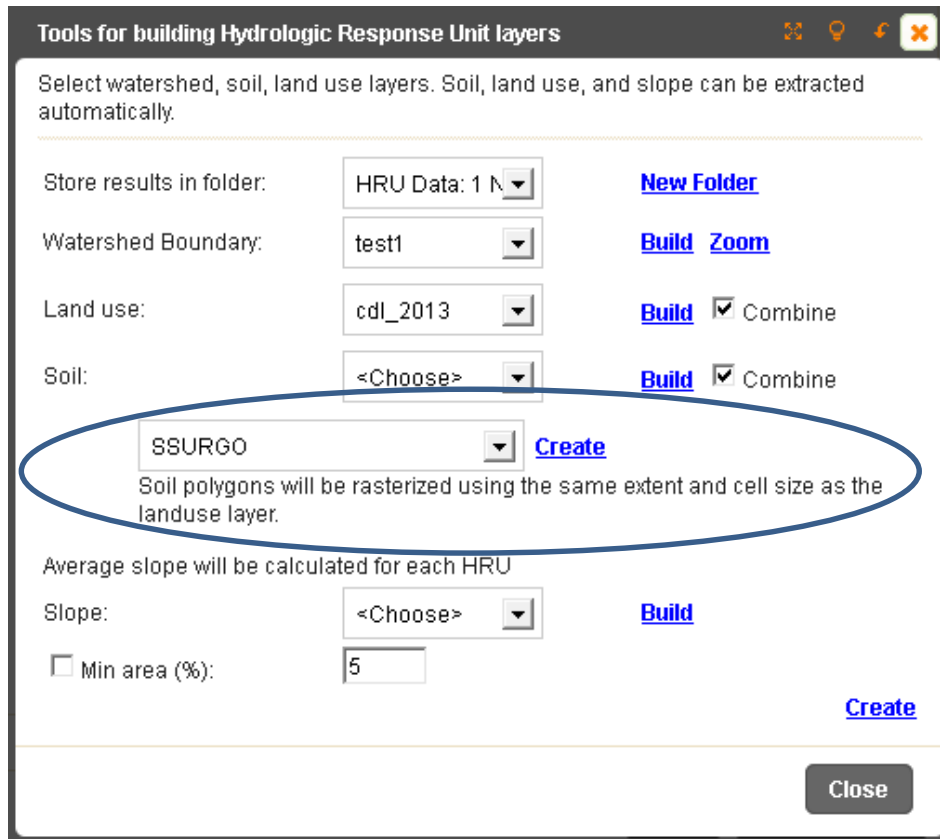
Figure 21: Define Soil Layer in eRAMS

Soil contains two options as the source for downloading the needed soil data. The choices are SSURGO or STATSGO. Once, the user has chosen the downloading source, click the "Create" button. The create button will execute the download from the corresponding source and store the layer into the folder specified at the beginning of the HRU creation and watershed definition process. Once the layer is downloaded, the layer is then clipped to the boundary of the watershed layer.

Once the soil and land use layers are populated, the last layer to generate is the slope layer. If the user defined a new folder, then the user must also generate a new slope layer in eRAMS. Otherwise the user can click the desired layer from the drop down menu.

To create a new slope layer in eRAMS, click the "Build" button that is within the same row as the label "Slope." Once "Build" is clicked, the parameters highlighted in figure 22 by the blue oval are visible.
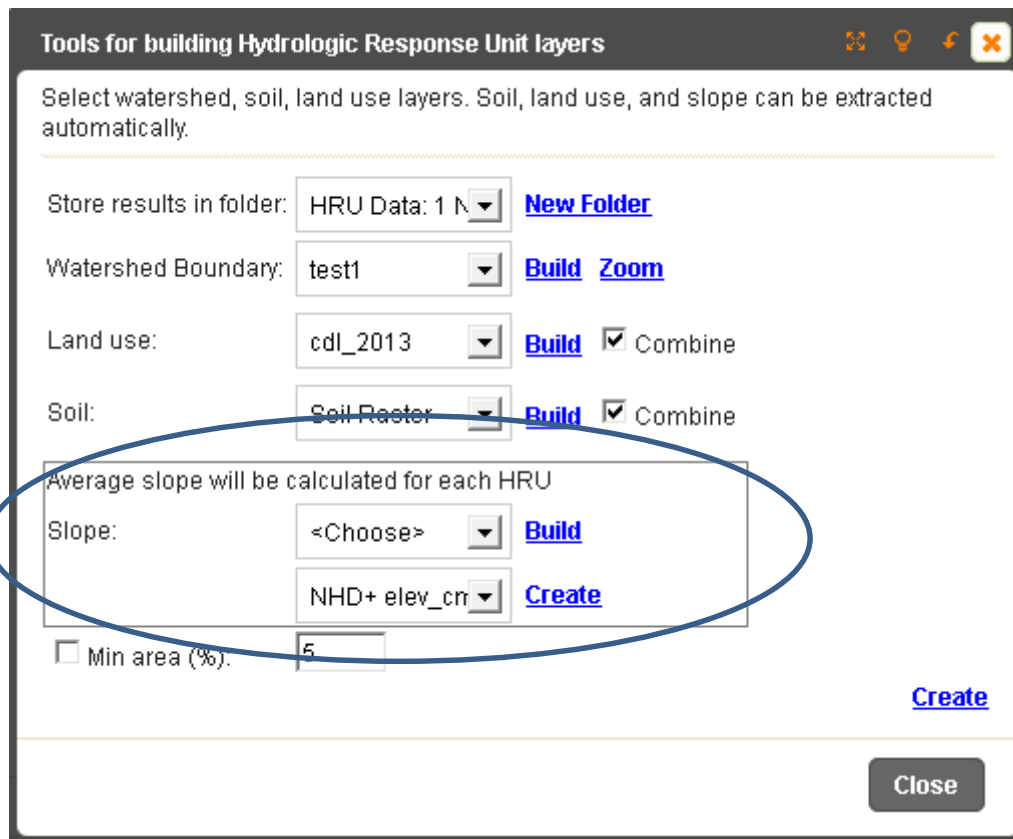
Figure 22: Define the soil layer in eRAMS

Slope only contains one source for downloading the needed data, this source is NHD. The "Create" button will

execute the download from the corresponding source and store the layer into the folder specified at the beginning of

the HRU creation and watershed definition process. Once the layer is downloaded, the layer is then clipped to the

boundary of the watershed layer.

After defining all three layers for the HRU generation, the user must decide what the minimum threshold is for a

given HRU's percent area. This threshold enables larger HRUs to absorb the HRUs smaller than the percent area

threshold.

After generating the HRUs and defining the watershed, the user then decides whether to implement a deterministic

or stochastic analysis. The basic look of a deterministic analysis is shown in section 3.6 but a stochastic analysis has

a slightly different look, as shown in figure 23.

Figure 23: Stochastic User Interface for Hydrology Tab

The difference between the user interface of a deterministic run versus a stochastic run is the ability to apply a

distribution to most of the input parameters of SWAT-DEG. Figure 23 shows the hydrology tab when the

"Uncertainty Analysis" checkbox is checked in figure 16. The user can change the distribution applied to a given

parameter by selecting a new distribution from the drop-down menu. The user can also define the distribution via the

parameters on the right hand side of the distribution. These parameters will automatically change dependent upon

the distribution chosen and the greyed values within each box shows the default parameter assigned to each

distribution parameter. Thus, if the user does not define the distribution parameter, these values are used.

4.4 Climate Tab

Figure 24 shows the parameters available under the climate tab of SWAT-DEG's user interface. Here the user

uploads the necessary climate data for their analysis.

Figure 24: Climate Tab User Interface

The user can upload data immediately or use the "NCDC Weather Stations" section to find the necessary data. The "NCDC Weather Stations" shows all weather stations within a certain radius of the defined watershed boundary described in section 3.3. Once the user chooses a station, the user has multiple options to download the data.

The first option is to download the data via the actual data host's website and the other option is to use eRAMS data download user interfaces. These interfaces are spoken about in further detail in section 2.4.

After downloading the data, the user must ensure that the data fits the requirements of SWAT-DEG. These requirements were minimized by enabling SWAT-DEG to regenerate the climate data for each run on the fly. Thus, the regeneration enables SWAT-DEG to format the data in the exact format FORTRAN needs. However, SWAT-DEG does require the data to fit a much simpler format for uploading. This format can be seen by clicking the "here" button in the top left corner of the climate tab user interface and is shown in figure 25.

ClimateData.csv Format
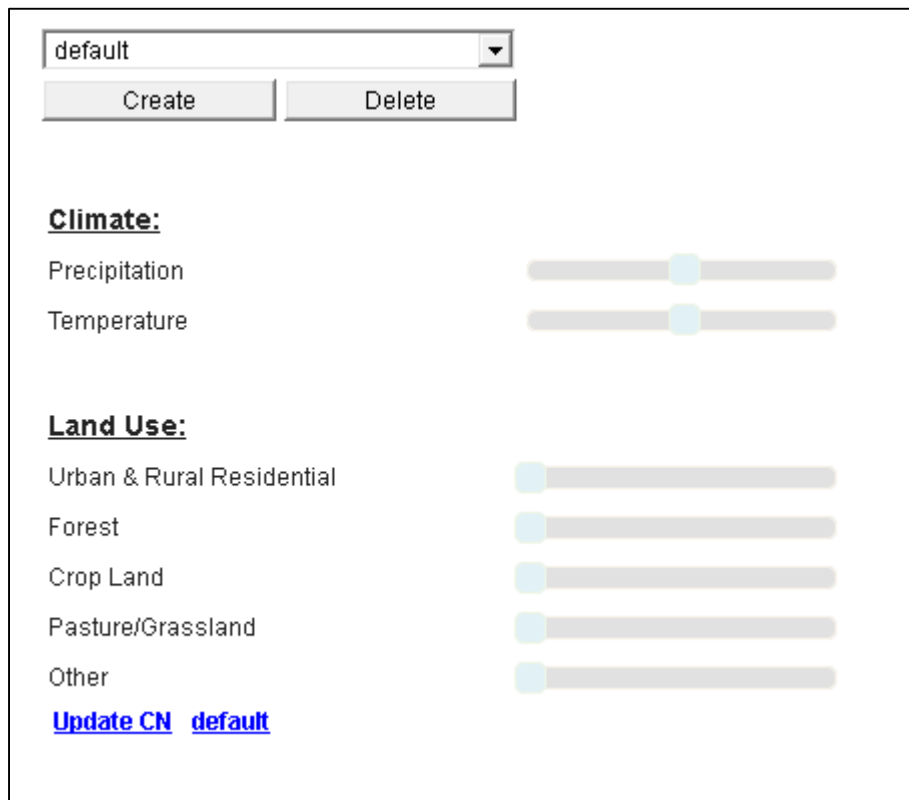
1. Place the data (using Excel) in their respective columns as demonstrated below
2. Save the Excel file as a ".csv"
3. NOTE: exclude the first two rows in your Excel File

Year	Julien Day	Precip	Avg Temp	Max Temp	Min Temp
		mm	(^{0}C)	(^{0}C)	(^{0}C)
1974	1	0	-3.8	5	-8.3
1974	2	1	6.8	8.9	4.4
1974	3	0	2.5	10	-3.4
1974	4	0.5	1.1	5	-1.2
1974	5	0	16.2	24.4	10.4

Figure 25: Required format for SWAT-DEG Climate Data Uploading

4.5 Scenario Analysis

The scenario tab is to enable the user to generate multiple scenarios for a given SWAT-DEG analysis. Figure 26 shows the parameters of the scenario analysis and is broken into three sections: (1) the creation or deletion of a scenario, (2) the climate change modifiers, and (3) the land use modifiers.
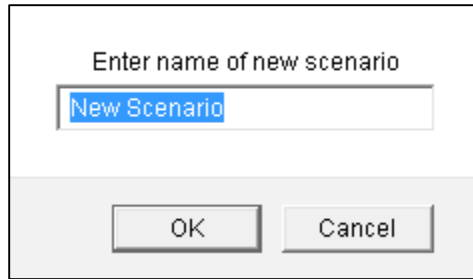

Figure 26: Scenario Tab User Interface

When first clicking on this tab, after generating a new SWAT-DEG project, the select box will show the scenario as "default," as shown in figure 26. This scenario does not allow for any modifiers, in climate change or land use to change. The reason is because the "default" scenario is supposed to be the baseline for which comparisons can be made from. To change the modifiers requires generating a new scenario, which is shown in figure 27.

Figure 27: Generate New Scenario Dialogue Box

After generating a new scenario, the climate and land use modifier sliders are enabled for editing, as shown in figure 28.
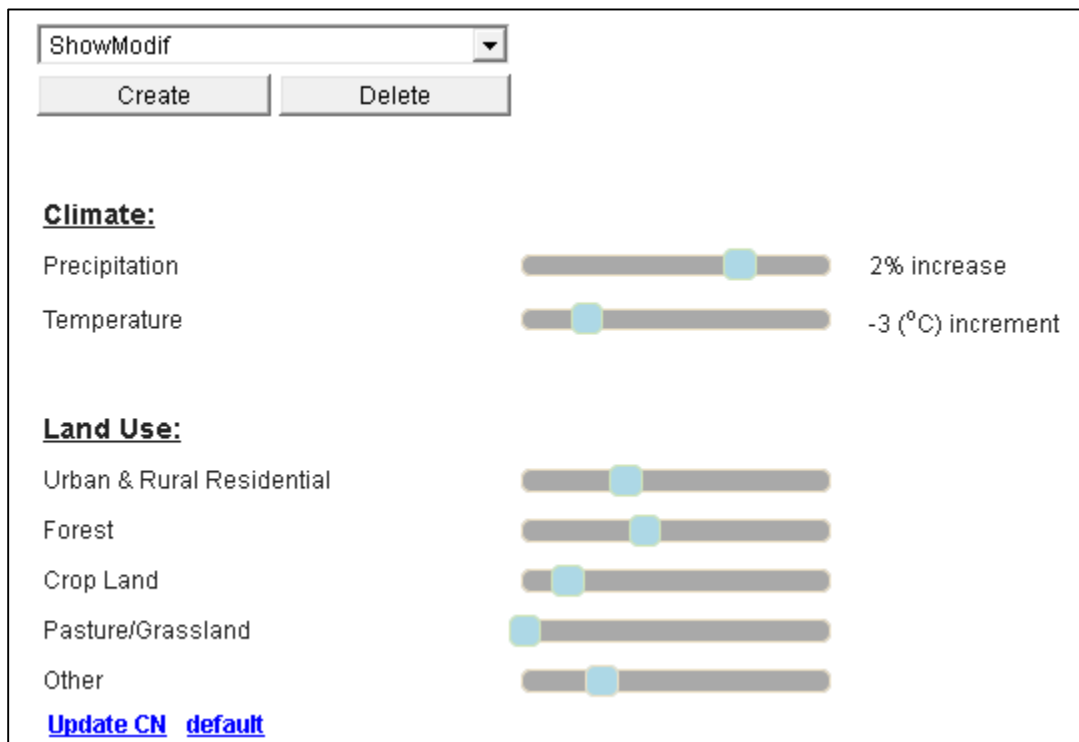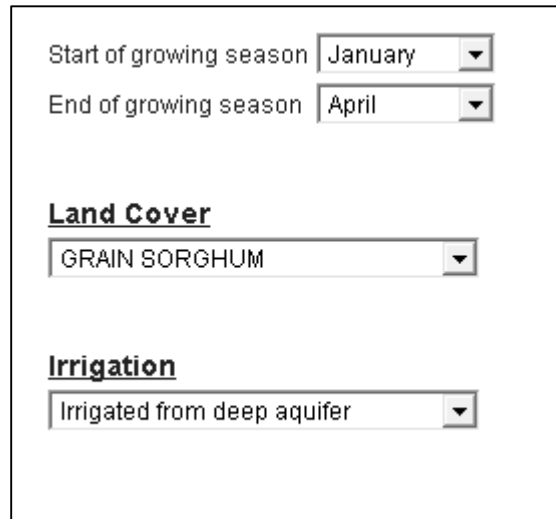

Figure 28: Scenario Tab User Interface with Modifiers Enabled

Here the user has specified that they want a 2% increase for precipitation and a -3 increment change for temperature. These two modifiers are applied to the whole time period specified in the settings tab. The land use modifiers only change the percent area of the HRUs and must always sum up to 100%.

4.6 Parameters

The parameters are spread throughout six tabs of SWAT-DEG's user interface, excluding the tabs previously

described. The first parameter tab is the management tab, shown in figure 29
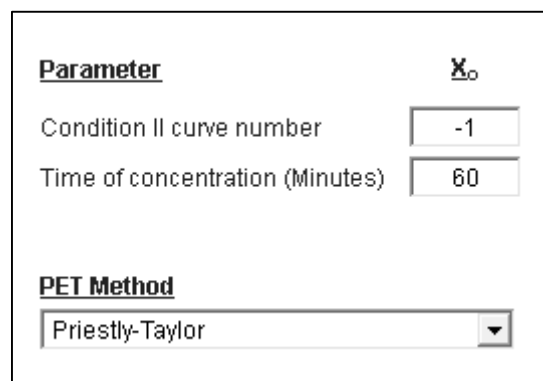

Figure 29: Management Tab User Interface

The management tab contains the parameters relating to the management and logistics of growing crops. After

entering in the necessary inputs, the next parameter tab is the hydrology tab, as shown in figure 30.

Figure 30: Hydrology Tab User Interface

The hydrology tab contains the parameters relating to the hydrological conditions of the watershed. After entering in

the necessary inputs, the next parameter tab is the bed material tab, as shown in figure 31.

Figure 31: Bed Material Tab User Interface

The bed material tab contains parameters relating to the bed of the channel undergoing analysis. The bed material

tab has a unique feature in that the "Bed Transport Equation" drop-down menu contains multiple choices, with each

option corresponding to a different equation defining the bed transport equation, as shown in figure 32 and figure

33.

Figure 32: Bagnold Velocity Power Function Equation for Bed Transport Equation

Figure 33: The Modified Simone & Li Equation for Bed Transport Equation

After entering in the necessary inputs, the next parameter tab is the channel tab, as shown in figure 34.

Figure 34: Channel Tab User Interface

The channel tab contains the parameters relating to the physical parameters of the channel undergoing analysis.

After entering in the necessary inputs, the next parameter tab is the aquifer tab, as shown in figure 35.

Figure 35: Aquifer Tab User Interface

The aquifer tab contains the parameters relating to the aquifer. After entering in the necessary inputs, the next

parameter tab is the topography and soil tab, as shown in figure 36.
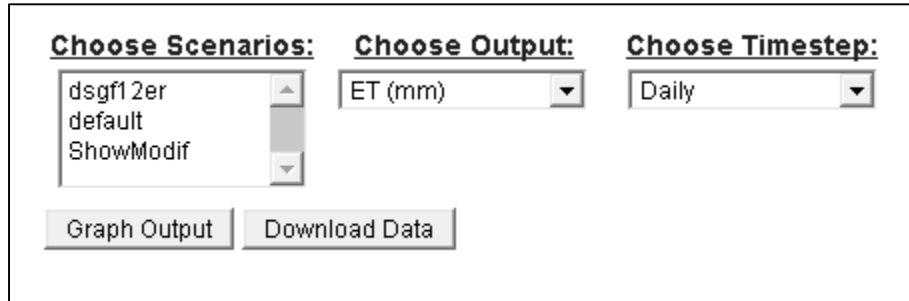
Figure 36: Topography and Soil Tab User Interface

The topography and soil tab contains the parameters relating to the topography and soil of the immediate area of the

channel. After entering in the necessary inputs, the user is able to click the "Run SWAT-DEG" button in the bottom

right hand corner of SWAT-DEG's main interface, shown in figure 16.

4.7 Outputs Tab

The last tab is for visualizing the outputs of the SWAT-DEG simulations. The layout of the output tab is shown

below in figure 37.

Figure 37: Output Tab User Interface

Here the user chooses the scenario to graph, the output parameter, the timestep, and if a monthly or yearly timestep

is chosen, statistics can also be applied. Once these inputs are chosen, click the "Graph Output" button. This will

generate a graph as shown in figure 38.
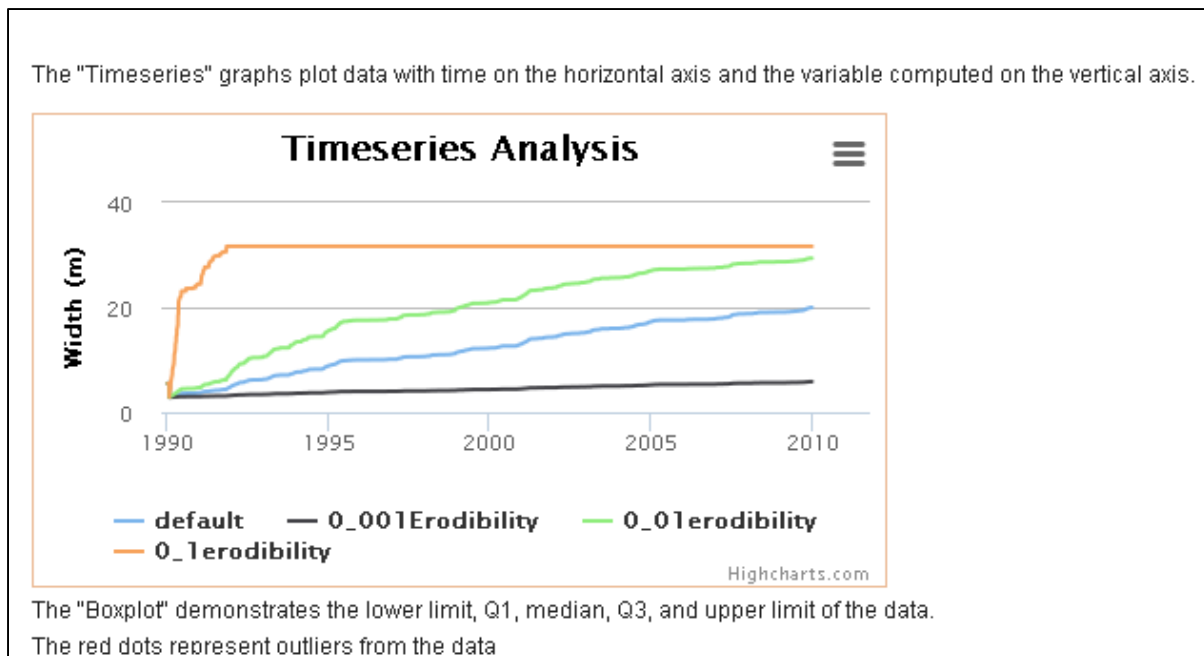
Figure 38: SWAT-DEG's Implementation of a JS HighChart Graph

Click the bars in the top right hand corner of the graph to download or print the graph for sharing or documentation

purposes. Figure 39 shows the options in regards to printing or downloading a graph.

Print chart

Download PNG image
Download JPEG image
Download PDF document
Download SVG vector image
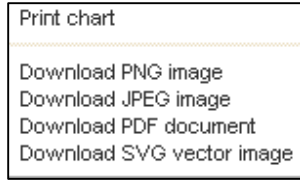
Figure 39: Options for Printing or Downloading a Graph within SWAT-DEG's User Interface

Chapter 5: Case Study Application – Cedar Creek

5.1 Introduction

The following chapter is intended to serve as a case study example of the newly added features of SWAT-DEG. The features demonstrated are deterministic runs, graphical interface, and scenario analysis within the web service, cloud environment. Parameters are typical parameters used for SWAT-DEG analysis of Cedar Creek, as seen in Allen, Peter, Arnold, Jeff & Gary, (2012).

5.2 Inputs

5.1.1 Area of Interest

Cedar Creek's bottom and sides are channelized from near the headwaters at Kite Road (approximately 32 38.975N and 97 2.134W) to Robinson Road (32 39.277N and 97 0895W). Allen, Peter, Arnold, Jeff, Stinchcomb, (2012). The engineers of AECOM, wanted to focus upon the channel stability assessment from Robinson Road to (Bardin Road (32 39.515N and 96 59.749W), which is approximately 9396 feet of stream channel. The parameters and their associated values utilized in this SWAT-DEG analysis are shown in table 3 below.

Table 3: Parameters Used in SWAT-DEG Analysis of Cedar Creek

Parameter	Value
Drainage Area (km^2)	3.873
Start Year	1990
Simulation Period (yrs)	20
Precipitation Modifier (%)	0, 3, 5
Temperature Modifier (%)	0, 3, 5
Start of Growing Season	April
End of Growing Season	October
Land Cover	Pasture

Irrigation	Irrigated from source outside of basin
Condition II Curve Number	80
Time of Concentration (min)	30
PET-Method	Priestly-Taylor
Loading Coefficient	0.5
Bed Transport Equation	Modified Simon & Li
Transport Coefficient	2123
Velocity Export Coefficient	3.3
Depth Coefficient	0.468
Slope Coefficient	0.613
Width (m)	5.5
Depth (m)	0.71
Slope (m/m)	0.007
Length (m)	1.15
Manning's N	0.04
Bottom Conductivity (mm/h)	0.9
Erodibility	0.005
Cover Factor	0.01
Width-Depth Ratio	4
Equilibrium Slope (m/m)	0.008
Median Sediment Size (mm)	4
Shear	2.5
Fraction of Day Force is Applied	0.08
Specific Yield – Shallow (mm)	0.03
Groundwater Alpha Factor	0.06
Revap Coefficient	0.1
Percolation Coefficient	0.01

Shallow Aquifer Storage (mm)	3
Shallow Aquifer Flow (mm)	0
Deep Aquifer Flow (mm)	300
Lattitude	31.6
Soil Texture	Silty Clay
Soil Profile Depth (mm)	1.5
Soil Water – Fraction	0.95
Snow Water Equivalent (mm)	0
Land Surface Slope (m/m)	0.008
Surface Slope length (m)	25
USLE Soil Erodibility Factor	0.3
USLE Soil Cover Factor	0.2
USLE Management Practice Factor	0.8

5.3 Case Study Scenarios

5.1.1 Scenario Modifications

Scenario modifications came in the form of climate change and land use changes. Specifically, the climate change simulation involves a 3% and 5% increase in precipitation and an additional increment of 3 and 5 for temperature. For land use, SWAT-DEG simulates an increase of 20% and 30% urbanization. These scenario modifications are to represent the current reality of many rivers and watersheds throughout the world.

5.4 Results/Discussion

5.1.1 Climate Scenarios

The climate scenarios showed a slight increase in scouring and thus an increase in the depth of the channel as compared to the baseline default scenario, shown in figure 40.
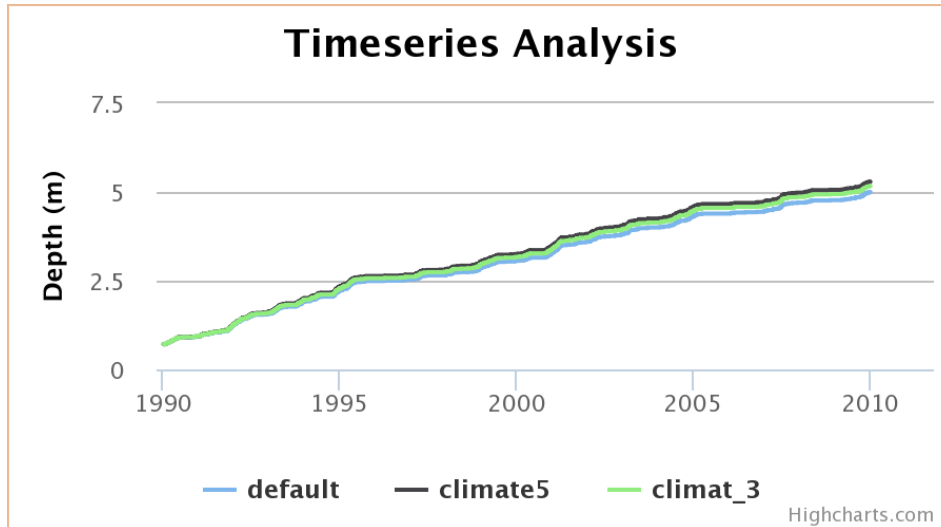
Figure 40: The Increasing Depth and thus, scouring of Cedar Creek Over Time

This increase is expected, as larger flows in a more frequent occurrence cause scouring and a continual and gradual decrease in slope, as shown in figure 41.
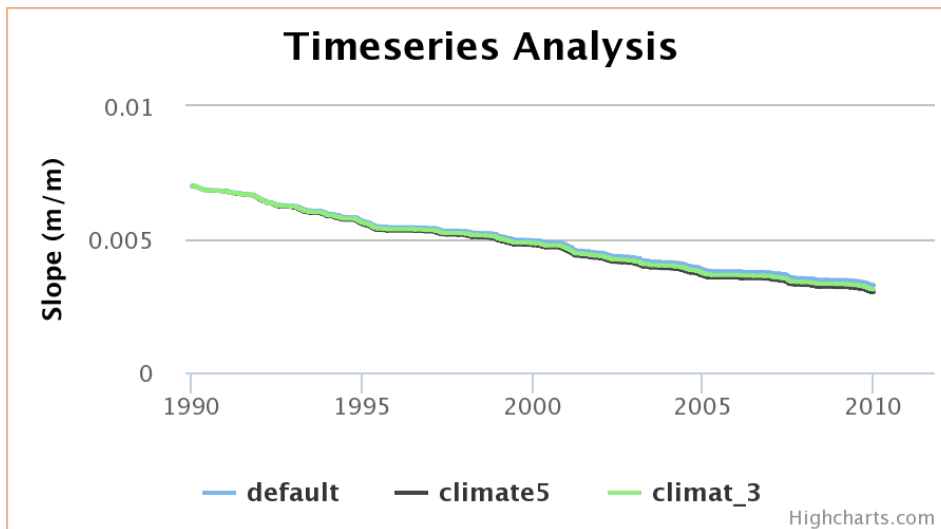
Figure 41: The Degradation of the Channel shown via the Slope

The channel will continue to degrade until quasi-equilibrium is reached. Quasi-equilibrium requires the channel to not only degrade in slope or depth, but also in width. Figure 42 shows the increase in width; though not as pronounced, there is scouring of the embankments.
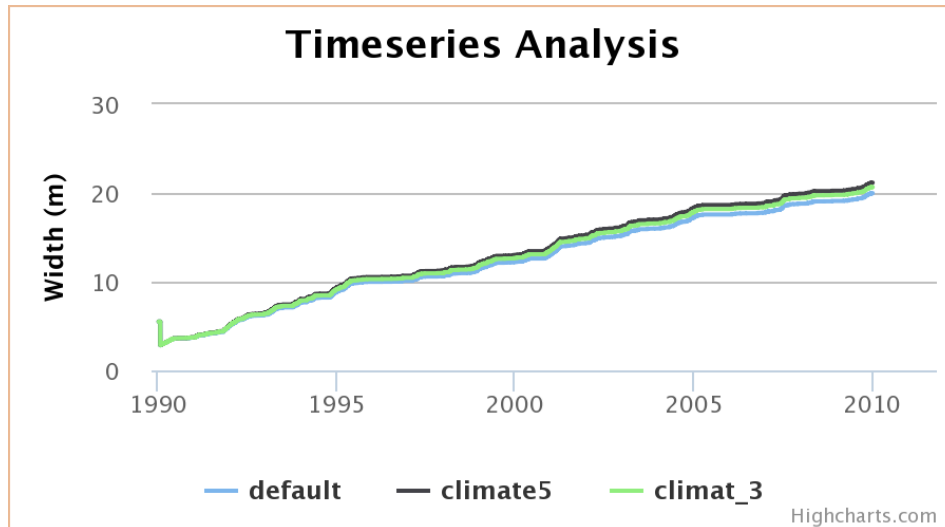
Figure 42: Widening of the Channel Due to Scouring of the Bankments

This scouring can cause bank failures, which are highly significant when a developmental area is built within a floodplain.

5.2.1 Land Use Scenarios

As shown via the climate change scenarios, the same results were seen within the land-use scenarios. Depth and slope both decreased due to scouring of the channel bed, as shown in Figures 42 and figure 43, respectively.
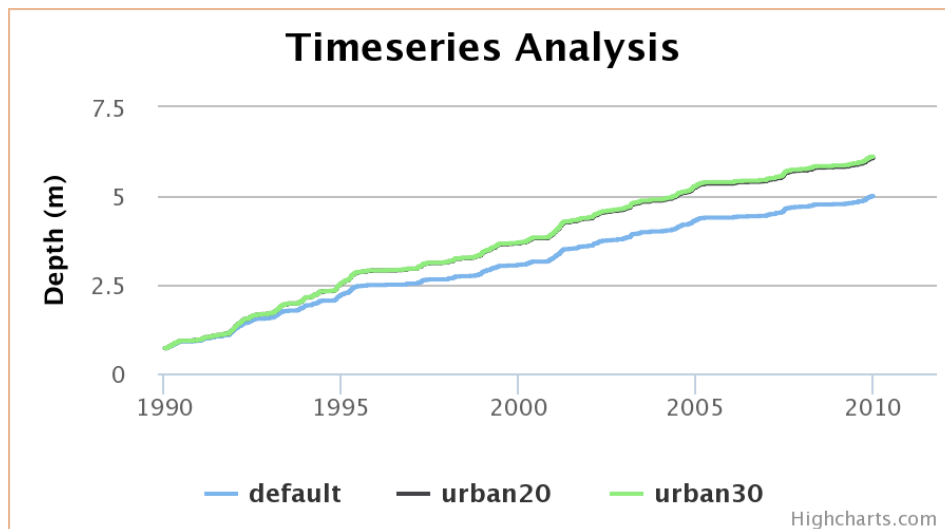
Figure 43: Depth Decreasing Due to Scouring of the Bed
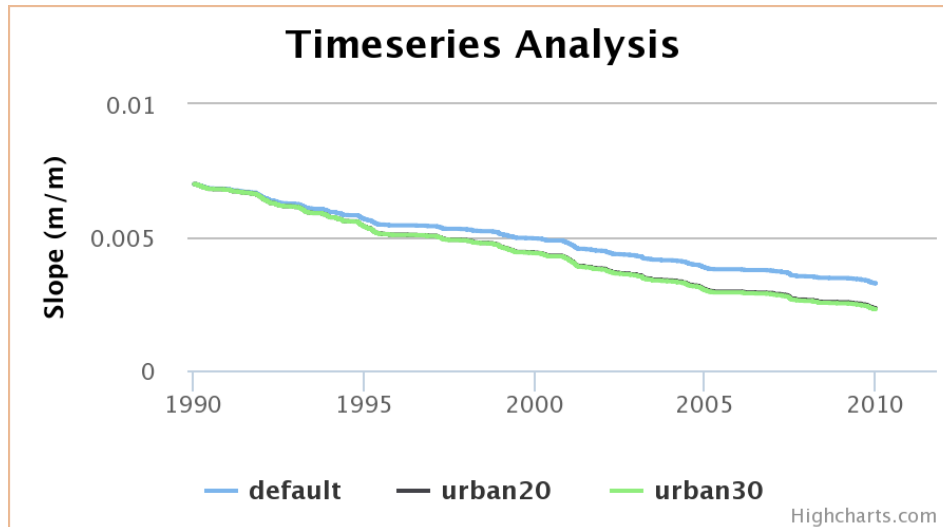
Figure 44: Slope Decreasing Due to Gradual Scouring of the Channel Bed

The depth and slope have larger rates than the default scenario, but still follow the relative pattern observed in the default time series. Thus, the rapid increase of urbanization also increased the rate at which the channel degrades until quasi-equilibrium is achieved. In addition, concluding that the quasi-equilibrium has changed in accordance to the changes in land use is also possible. If the latter holds true, then the river and the systems downstream of the scouring will see larger effects than if the channel were to reach quasi-equilibrium naturally. This trend is also shown in the timeseries of width, seen in figure 44.
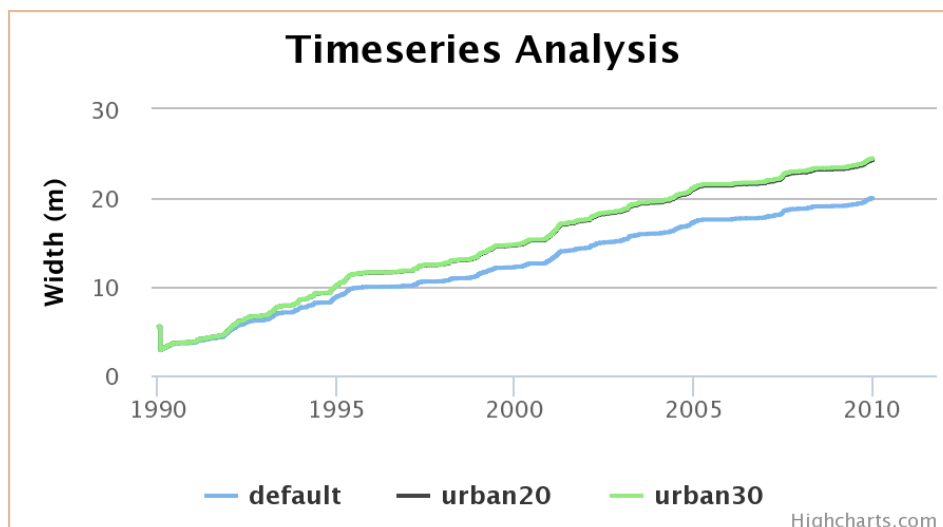
Figure 45: Widening of the Channel Due to Scouring of the Bankments

58

5.3.1 Comparison of Scenarios

The degradation of the channel stream is inevitable, as the default scenarios all show gradual scouring of the channel bed and widening of the channel. Cedar Creek is undergoing rapid urbanization and thus, this gradual scouring is occurring at higher rates. In addition, climate change could potentially worsen the situation. Both of these scenarios are shown in figures 40-45. Interestingly, rapid urbanization has a more drastic effect on width, depth, and slope. This is highly evident when observing the depth and slope caused by climate change versus those caused by urbanization, shown in figure 40-41, and 43-44. Implementing solutions to minimize impacts from rapid urbanization upon the river system would provide the best assistance for reaching a state of quasi-equilibrium with fewer downstream consequences.

5.4.1 Scalability Testing

Cloud infrastructure, via CSIP, provided mixed results in scaling SWAT-DEG. The mixed results are not due to CSIP infrastructure directly, instead, it is due to which service, deterministic or stochastic executes. The scalability tests were done utilizing Amazon's cloud infrastructure with a 4-core c3 xlarge CPU and a 4-core m2.2 xlarge

Case Study 1:

CPU. The specifications of each VM are specified below:

M2.2xlarge

Intel(R) Xeon® CPU E5-2665 0 @ 2.40GHz

850 GB disk

34.2 GB ram

4 cores, 3.25 ECUs per core

Advertised network I/O performance: 500 Mbps, moderate

Case Study 2:

C3.xlarge

Intel(R) Xeon® CPU E5-2680 v2 @ 2.80GHz

59

40 GB SSD

7.5 GB ram

4 cores, 3.5 ECUs per core

Advertised network I/O performance: 500 Mbps, moderate

The c3 xlarge CPU was used to test SWAT-DEG's deterministic runs, resulting in a large reduction in total time, total time being the time required to finish the entire workload. The entire workload starts with the initial REST request and ends with the return of outputs via REST. Figure 46 demonstrates the improvement from running SWAT-DEG on multiple VMs instead of locally with a highly limited amount of VMs.
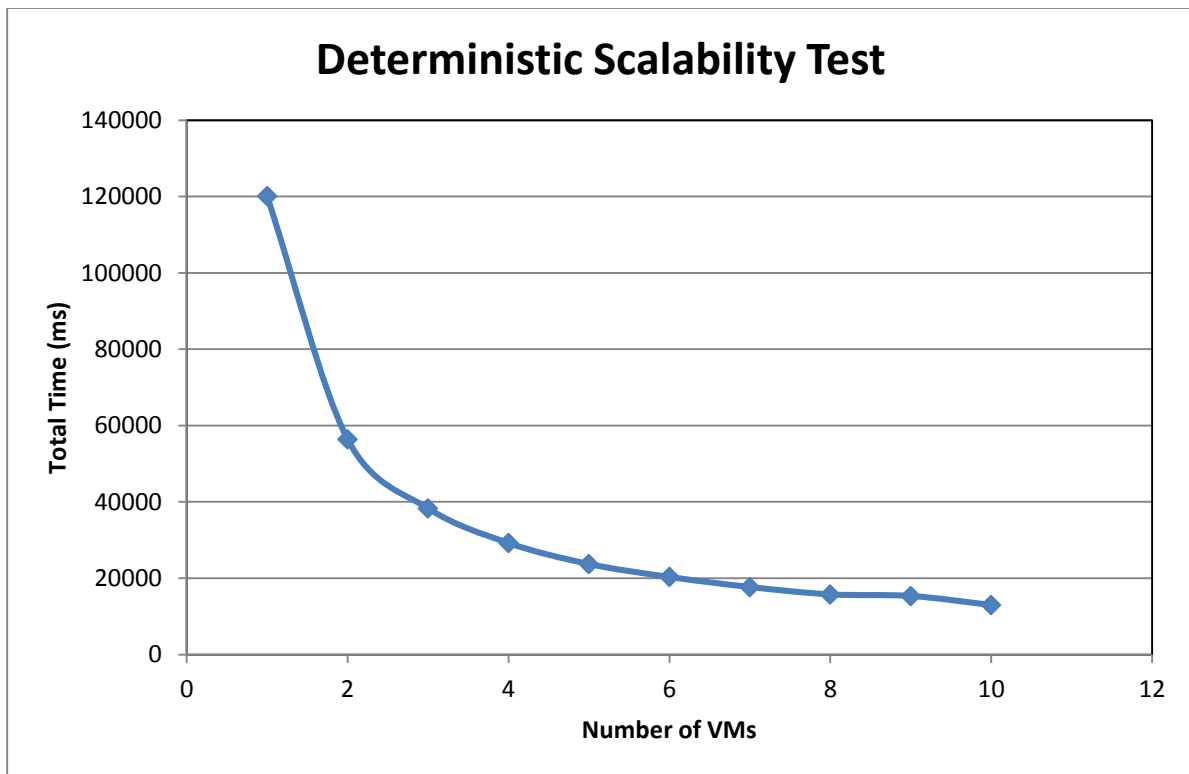
Figure 46: The deterministic scalability test results. The test simulated a constant 1000 SWAT-DEG deterministic runs while scaling the number of VMs and number of simultaneous runs. Total time represents the total amount of time it takes for a user to receive results from birth of the request

The x-axis represents the number of VMs that were allocated for executing SWAT-DEG's deterministic service. In addition, the numbers of simultaneous runs were proportionately scaled with the number of VMs. During this

60

process CSIP parallelizes multiple deterministic services as the number of VMs increased, while holding the number of runs, 1000, constant. As shown in figure 46, as the number of VMs or the number of parallel deterministic services executed are increased, the total time decreases. This demonstrates the usefulness of applying third party resources, via CSIP, as the platform to executing models with multiple users firing requests or with a single user firing multiple requests. The stochastic services did not show the same results as the deterministic services despite designing the stochastic service to call the deterministic service.

The deterministic service has two phases, whereas the stochastic service has three phases. The deterministic service's phases are executing the request via SWAT-DEG's executable and then executing a map reduction to the data for graphing and memory purposes. The stochastic service includes the two phases of the deterministic service but adds an additional phase of input or output generation. The input generation consists of generating sets of input files that will run multiple deterministic services asynchronously and the output phase consists of appending the results of map reduction to a single file. The additional phase was also tested within CSIP environment to see the effect of scalability on a stochastic analysis, via the Monte Carlo Method (Ronald Christensen, Wesley Johnson, Adam Branscum, 2011).

The results showed slight improvement when the number of simultaneous users was small, but as the number of simultaneous users scaled up the total time did not scale down as predicted. Figures 47-49 demonstrate the results of the scalability testing.
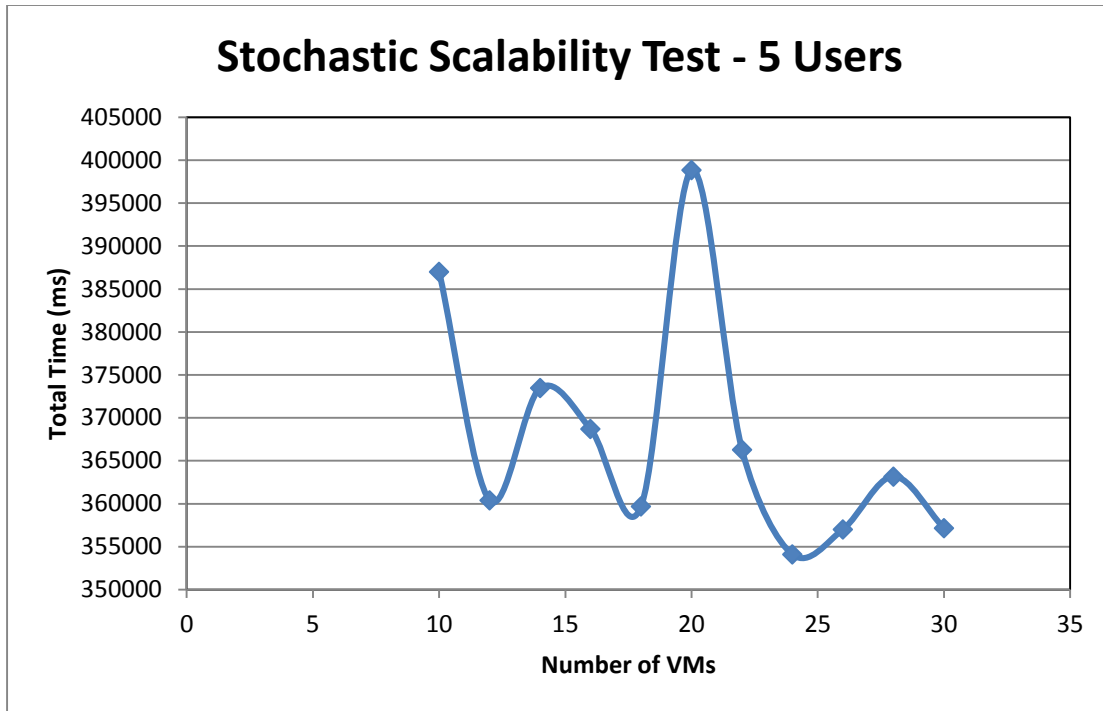
Figure 47: The stochastic scalability test with 5 simultaneous users. The stochastic test kept a constant 100 runs and 5 simultaneous users while varying the number of VMs allocated to the single stochastic request. The x-axis demonstrates the number of VMs and the y-axis represents the time a single stochastic request spends within the CSIP environment.

Figure 47 demonstrates a slight downward trend as the number of VMs is increased. This is the predicted outcome when sending a stochastic request to the CSIP environment. The time difference between increasing the number of VMs is significant in this test case because the relative time to finishing a stochastic request of 100 simulations is also small. Thus, the ratio between a small time differences for increasing the number of VMs relative to the small run time will propagate as the run time increases for a given request.
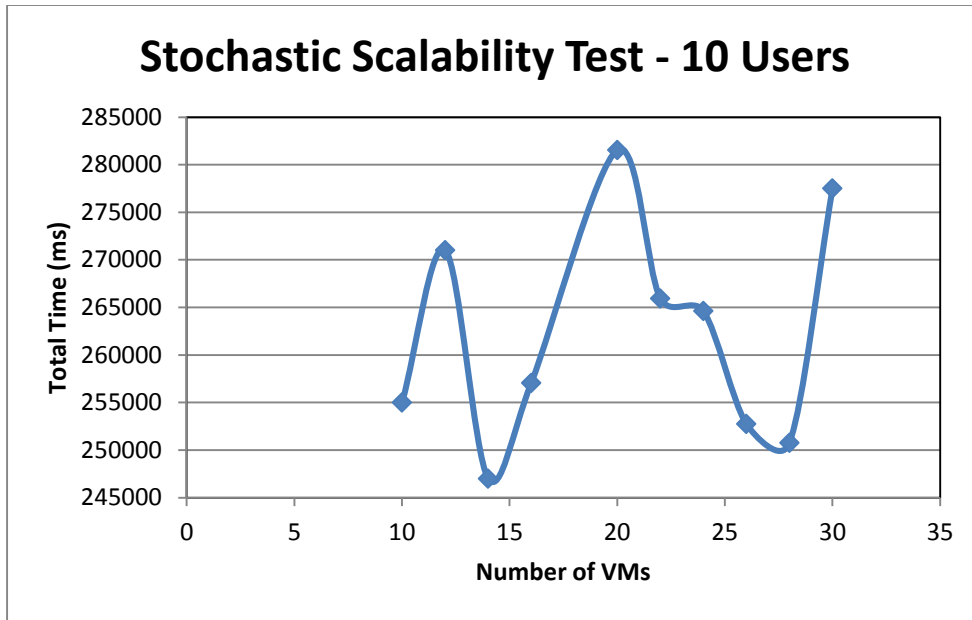
Figure 48: The stochastic scalability test with 10 simultaneous users. The stochastic test kept a constant 100 runs and 10 simultaneous users while varying the number of VMs allocated to the single stochastic request. The x-axis demonstrates the number of VMs and the y-axis represents the time a single stochastic request spends within the CSIP environment.
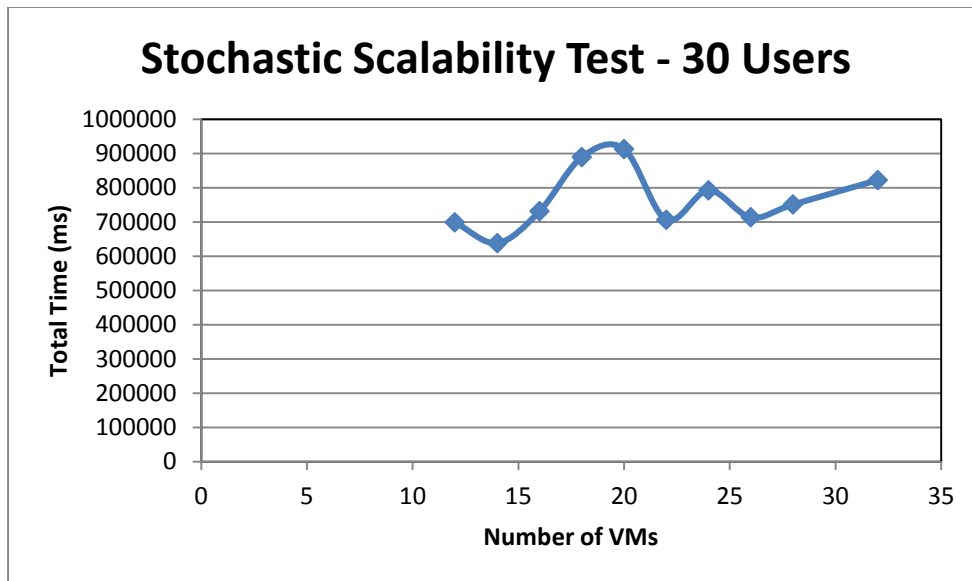
Figure 49: The stochastic scalability test with 10 simultaneous users. The stochastic test kept a constant 100 runs and 10 simultaneous users while varying the number of VMs allocated to the single stochastic request.

Figures 48 and 49 do not show a downward trend as expected, but this is not due to CSIP's infrastructure. This is most likely due to a bottleneck when implementing increasing amounts of simultaneous users for a specified amount

of VMs. The bottleneck is not due to increasing amount of simultaneous users, instead, it is due to SWAT-DEG's map reduction phase. The initial request for SWAT-DEG runs, regardless of deterministic or stochastic, utilizes a multiple VM allocation, whereas the map reduction is allocated to a single core or a single VM. Thus, the time a given stochastic or deterministic request spends in the CSIP environment is approximately 50% in the execution of a SWAT-DEG run and 50% in the map reduction phase. When allocating a given number of cores and a given number of parallel SWAT-DEG requests, all the VMs allocated to the request are only used 50% of the time. This bottleneck is skewing or hiding the actual benefit CSIP's environment provides for a stochastic analysis.

The limited benefit provided by CSIP within figure 47-49 is a demonstration of Amdahl's law (Sun & Chen, 2010). Amdahl's law refers to the maximum expected improvement due to the improvement of a specific part of a system. In SWAT-DEG's case, the specific part improved is the execution of SWAT-DEG's executable within CSIP's environment. The reduction phase is designed within a cloud framework, however bottlenecks were found in the design of the reduction phase. When testing the maximum speedup due to the improvement of executing, SWAT-DEG is limited by the time needed for the reduction phase. For example, if the reduction of a given SWAT-DEG request takes 30 minutes, the total time of the same request cannot be less than 30 minutes.

The bottleneck provided by the reduction phase minimizes the effectiveness of the VMs allocated to the stochastic SWAT-DEG request. The fact that VMs within the request spend approximately 50% of the time idle is contrary to CSIP's goal. CSIP's goal is to maximize the running time of VMs by minimizing the idle time of the physical infrastructure. SWAT-DEG has a few options to effectively realize CSIP's purpose. The first option is to parallelize the reduction phase, similar to how SWAT-DEG's execution can utilize multiple VMs. The second option is to stagger or increment the times when the reduction phase is executed. These two options do not increase the speed of executing a SWAT-DEG model, but rather decrease the total time by maximizing the efficiency of the VM activity for a given SWAT-DEG request.

5.5 Conclusions

SWAT-DEG within the web-service and cloud environment has shown that it can provide analysis of the stability of channels. Moreover, the additional features of scenarios, via climate change and land use change, gives greater

insight into the problems associated with maintenance of channels during urbanization. SWAT-DEG's purpose is to provide management assistance to project managers during major urbanization around rivers; SWAT-DEG's scenarios have vital implications due to its ability to not only analyze the current situation, but also future conditions resulting from current construction plans or floodplain development.

There are still stipulations that the data entered into SWAT-DEG dictates the validity and accuracy of the outputs. In addition, the case study previously shown is not a full demonstration of every feature. SWAT-DEG's stochastic analysis and HRU generation are additional features that will be further explored.

REFERENCES

Aldrawiesh, K., Al-Ajlan, a., Al-Saawy, Y., & Bajahzar, a. (2009). A comparative study between computer programming languages for developing distributed systems in web environment. *Proceedings of the 2nd International Conference on Interaction Sciences Information Technology, Culture and Human - ICIS '09*, 457–461. doi:10.1145/1655925.1656009

Allen, P., Arnold, J., & Skipwith, W. (2002). Erodibility of Urban Bedrock and Alluvial Channels, North Texas. *JAWRA Journal of the …*, *38*(5). Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/j.1752-1688.2002.tb04360.x/abstract

Allen, P., Arnold, J., & Skipwith, W. (2008). Prediction of channel degradation rates in urbanizing watersheds. *Hydrological Sciences Journal*, (March 2013), 37–41. Retrieved from http://www.tandfonline.com/doi/abs/10.1623/hysj.53.5.1013

Allen, P.M., Arnold, J.G., Stinchcomb, G. (2012). *Cedar Creek Channel Assessment For The City of Grand Prairie, Texas. Report to AECOME Engineers for the City of Grand Prairie, Deparment of Engineering* (p. 112). City of Grand Prairie, Texas.

An Approach for Using Load Duration Curves in the Development of TMDLs. (n.d.). Retrieved June 23, 2013, from http://water.epa.gov/lawsregs/lawsguidance/cwa/tmdl/techsupp.cfm.

Argent, R., Grayson, R., & Ewing, S. (1999). Integrated models for environmental management: issues of process and design. *Environment International*, (617), 693–699. Retrieved from http://www.sciencedirect.com/science/article/pii/S0160412099000525

Argent, R. M., Perraud, J.-M., Rahman, J. M., Grayson, R. B., & Podger, G. M. (2009). A new approach to water quality modelling and environmental decision support systems. *Environmental Modelling & Software*, *24*(7), 809–818. doi:10.1016/j.envsoft.2008.12.010

Arnold, J., Allen, P., Muttiah, R., & Bernhardt, G. (1995). Automated base flow separation and recession analysis techniques. *Ground Water*. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/j.1745-6584.1995.tb00046.x/full

Arnold, J. G., & Fohrer, N. (2005). SWAT2000: current capabilities and research opportunities in applied watershed modelling. *Hydrological Processes*, *19*(3), 563–572. doi:10.1002/hyp.5611

Arnold, J., & Srinivasan, R. (1999). Continental scale simulation of the hydrologic balance1. *… of the American Water …*, *35*(5). Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/j.1752-1688.1999.tb04192.x/abstract

AS, H. (2014a). Demo Graph. Retrieved October 31, 2014, from http://www.highcharts.com/demo/line-basic/dark-unica

AS, H. (2014b). JS HighCharts. Retrieved October 31, 2014, from http://www.highcharts.com/about

Atkins, D. (2003). *Revolutionizing science and engineering through cyberinfrastructure: Report of the National Science Foundation blue-ribbon advisory panel on cyberinfrastructure*. Retrieved from https://arizona.openrepository.com/arizona/handle/10150/106224

Beran, B., & Piasecki, M. (2009). Engineering new paths to water data. *Computers & Geosciences*, *35*(4), 753–760. doi:10.1016/j.cageo.2008.02.017

Brown, C., & Hollis, J. (1996). SWAT—A Semi-empirical Model to Predict Concentrations of Pesticides Entering Surface Waters from Agricultural Land. *Pesticide Science*, 41–50. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/(SICI)1096-9063(199605)47:1<41::AID-PS385>3.0.CO;2-N/abstract

Brown, L., & Barnwell, T. (1987). *The enhanced stream water quality models QUAL2E and QUAL2E-UNCAS: documentation and user manual*. Retrieved from http://medina.cee.duke.edu/CE245/sect1-1.pdf

Campo, P. C., Bousquet, F., & Villanueva, T. R. (2010). Modelling with stakeholders within a development project. *Environmental Modelling & Software*, *25*(11), 1302–1321. doi:10.1016/j.envsoft.2010.01.005

Castronova, A. M., Goodall, J. L., & Elag, M. M. (2013). Models as web services using the Open Geospatial Consortium (OGC) Web Processing Service (WPS) standard. *Environmental Modelling & Software*, *41*, 72–83. doi:10.1016/j.envsoft.2012.11.010

Chow, V. Te. (1959). *Open-channel Hydraulics* (Print.). New York: McGraw-Hill.

Clarke, S. J., Bruce-Burgess, L., & Wharton, G. (2003). Linking form and function: towards an eco-hydromorphic approach to sustainable river restoration. *Aquatic Conservation: Marine and Freshwater Ecosystems*, *13*(5), 439–450. doi:10.1002/aqc.591

Conan, C., de Marsily, G., Bouraoui, F., & Bidoglio, G. (2003). A long-term hydrological modelling of the Upper Guadiana river basin (Spain). *Physics and Chemistry of the Earth, Parts A/B/C*, *28*(4-5), 193–200. doi:10.1016/S1474-7065(03)00025-1

Dustdar, S., & Schreiner, W. (2005). A survey on web services composition. *International Journal of Web and Grid Services*, *1*(1), 1–30. Retrieved from http://inderscience.metapress.com/index/473epwhagp6hbeba.pdf

Gassman, P., Reyes, M., Green, C., & Arnold, J. (2007). *The soil and water assessment tool: historical development, applications, and future research directions* (Vol. 50, pp. 1211–1250). Retrieved from http://publications.iowa.gov/archive/00005419/01/paper_12744.pdf

Gassman, P., Williams, J., & Benson, V. (2005). *Historical development and applications of the EPIC and APEX models*. Retrieved from http://ageconsearch.umn.edu/bitstream/18372/1/wp050397.pdf

Goodall, J., Horsburgh, J., Whiteaker, T., Maidment, D., & Zaslavsky, I. (2008). A first approach to web services for the National Water Information System. *Environmental Modelling & Software*, *23*(4), 404–411. doi:10.1016/j.envsoft.2007.01.005

Gortmaker, J., Janssen, M., & Wagenaar, R. W. (2004). The advantages of web service orchestration in perspective. *Proceedings of the 6th International Conference on Electronic Commerce - ICEC '04*, 506. doi:10.1145/1052220.1052284

Granell, C., Díaz, L., & Gould, M. (2010). Service-oriented applications for environmental models: Reusable geospatial services. *Environmental Modelling & Software*, *25*(2), 182–198. doi:10.1016/j.envsoft.2009.08.005

Guan, Q., Zhang, T., & Clarke, K. (2006). GeoComputation in the grid computing age. *Web and Wireless Geographical Information …*, 237–246. Retrieved from http://link.springer.com/chapter/10.1007/11935148_22

Huhns, M., & Singh, M. (2005). Service-oriented computing: Key concepts and principles. *Internet Computing, IEEE*, (February), 2–8. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1407782

Izaurralde, R. C., Williams, J. R., McGill, W. B., Rosenberg, N. J., & Jakas, M. C. Q. (2006). Simulate Soil C dynamics with EPIC: Model description and testing against long-term data. *Ecological Modelling*, *192*(3-4), 362–384. doi:10.1016/j.ecolmodel.2005.07.010

Jain, S. K. (2010). Simulation of Runoff and Sediment Yield for a Himalayan Watershed Using SWAT Model. *Journal of Water Resource and Protection*, *02*(03), 267–281. doi:10.4236/jwarp.2010.23031

Java Hashmap. (2014). Retrieved from http://docs.oracle.com/javase/7/docs/api/java/util/HashMap.html

Jha, S., Western, A., Rutherfurd, I., & Grayson, R. (2005). Testing uncertainty in a model of stream bank erosion. *MODSIM 2005 International …*, 2714–2720. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Testing+Uncertainty+In+A+Model+Of+Stream+Bank+Erosion#2

Knisel, W. (1980). CREAMS: A field-scale model for chemicals, runoff and erosion from agricultural management systems. *USDA Conservation Research Report*. Retrieved from http://www.cabdirect.org/abstracts/19801956438.html

Kollet, S. J., & Maxwell, R. M. (2006). Integrated surface–groundwater flow modeling: A free-surface overland flow boundary condition in a parallel groundwater flow model. *Advances in Water Resources*, *29*(7), 945–958. doi:10.1016/j.advwatres.2005.08.006

Kondolf, G. M., Anderson, S., Lave, R., Pagano, L., Merenlender, a., & Bernhardt, E. S. (2007). Two Decades of River Restoration in California: What Can We Learn? *Restoration Ecology*, *15*(3), 516–523. doi:10.1111/j.1526-100X.2007.00247.x

La, H., Oh, S., & Kim, S. (2010). Methods to utilizing cloud computing in developing mobile internet device (MID) applications. *Proceedings of the 4th International Conference …*, (Mid), 1. doi:10.1145/2108616.2108654

Lettenmaier, D., Wood, A., & Palmer, R. (1999). Water resources implications of global warming: A US regional perspective. *Climatic Change*, 537–579. Retrieved from http://link.springer.com/article/10.1023/A:1005448007910

Li, Q., Yu, X., Xin, Z., & Sun, Y. (2012). Modeling the effects of climate change and human activities on the hydrological processes in a semiarid watershed of loess plateau. *Journal of Hydrologic Engineering*, (April), 401–412. doi:10.1061/(ASCE)HE.1943-5584

Lloyd, W., David, O., Lyon, J., Rojas, K. W., Ascough II, J. C., Green, T. R., & Carlson, J. R. (2012). *The Cloud Services Innovation Platform–Enabling Service-Based Environmental Modelling Using Infrastructure-as-a-Service Cloud Computing*. *Meeting …*. Retrieved from http://www.iemss.org/iemss2012/proceedings/D3_1_0905_David_et_al.pdf

Lloyd, W., Pallickara, S., David, O., Lyon, J., Arabi, M., & Rojas, K. (2011). Migration of Multi-tier Applications to Infrastructure-as-a-Service Clouds: An Investigation Using Kernel-Based Virtual Machines. *Grid Computing (GRID), 2011 IEEE/ACM 12th International Conference on Grid Computing*, 137–144. doi:10.1109/Grid.2011.26

Lloyd, W., Pallickara, S., David, O., Lyon, J., Arabi, M., & Rojas, K. (2013). Performance implications of multi-tier application deployments on Infrastructure-as-a-Service clouds: Towards performance modeling. *Future Generation Computer Systems*, *29*(5), 1254–1264. doi:10.1016/j.future.2012.12.007

M. Papazoglou, D. G. (2006). Service-oriented computing. *Communications of the ACM*, *46*(10), 25–28. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1607964

McIntosh, B. S., Jeffrey, P., Lemon, M., & Winder, N. (2005). On the design of computer-based models for integrated environmental science. *Environmental Management*, *35*(6), 741–52. doi:10.1007/s00267-004-0172-0

Mineter, M. J., Jarvis, C. H., & Dowers, S. (2003). From stand-alone programs towards grid-aware services and components: a case study in agricultural modelling with interpolated climate data. *Environmental Modelling & Software*, *18*(4), 379–391. doi:10.1016/S1364-8152(02)00073-7

Neumann, L., Western, A., & Argent RM. (2010). The sensitivity of simulated flow and water quality response to spatial heterogeneity on a hillslope in the Tarrawarra catchment, Australia. *Hydrological Processes*, *86*(October 2009), 76–86. doi:10.1002/hyp

Palmer, M., Allan, J. D., Meyer, J., & Bernhardt, E. S. (2007). River Restoration in the Twenty-First Century: Data and Experiential Knowledge to Inform Future Efforts. *Restoration Ecology*, *15*(3), 472–481. doi:10.1111/j.1526-100X.2007.00243.x

Peter Allen, Jeff Arnold, Stephanie Capello, D. C. (2010). Channel Erosion Estimate for Urbanizing Watersheds: Submerged Jet testing and SWAT-DEG. In *21st Century Watershed Technology: Improving Water Quality and Environment Conference Proceedings*. Universidad EARTH, Costa Rica. doi:10.13031/2013.29391

Poff, N., Brinson, M., & Jr, J. D. (2002). Aquatic ecosystems and global climate change. *… on Global Climate Change, …*, (January). Retrieved from http://www.sysecol2.ethz.ch/AR4_Ch04/Ch4-GreyLit/Po59.pdf

R. A. Leonard, W.G. Knisel, D. A. S. (1987). GLEAMS: Groundwater Loading Effects of Agricultural Management Systems. *Transactions of the ASAE*, *30*, 1403–1418. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:GLEAMS:+Groundwater+Loading+Effects+of+Agricultural+Management+Systems#0

Ronald Christensen, Wesley Johnson, Adam Branscum, T. H. (2011). *Bayesian Ideas and Data Analysis*. (J. Z. Bradley Carlin, Julian Faraway, Martin Tanner, Ed.) (p. Ch 3.4). Boca Raton: CRC Press.

S.L. Neitsch, J.G. Arnold, J.R. Kiniry, R. Srinivasan, J. R. W. (2005). *Soil and Water Assessment Tool Theoretical Documentation*.

Simon, A., & Rinaldi, M. (2006). Disturbance, stream incision, and channel evolution: The roles of excess transport capacity and boundary materials in controlling channel response. *Geomorphology*, *79*(3-4), 361–383. doi:10.1016/j.geomorph.2006.06.037

Srinivasan R., Ramanarayanan T, Arnold J, B. T. (1998). Large area hydrologic modeling and assessment part II: Model application1. *JAWRA Journal of …*, *34*(1), 91–101. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/j.1752-1688.1998.tb05962.x/abstract

Sun, X.-H., & Chen, Y. (2010). Reevaluating Amdahl's law in the multicore era. *Journal of Parallel and Distributed Computing*, *70*(2), 183–188. doi:10.1016/j.jpdc.2009.05.002

Van Griensven, a., Ndomba, P., Yalew, S., & Kilonzo, F. (2012). Critical review of the application of SWAT in the upper Nile Basin countries. *Hydrology and Earth System Sciences Discussions*, *9*(3), 3761–3788. doi:10.5194/hessd-9-3761-2012

Vijay P. Singh, D. K. F. (Ed.). (2002). *Mathematical Models of Large Watershed Hydrology* (p. 4). Highlands Ranch: Water Resources Publications, LLC.

Voinov, A., & Bousquet, F. (2010). Modelling with stakeholders☆. *Environmental Modelling & Software*, *25*(11), 1268–1281. doi:10.1016/j.envsoft.2010.03.007

Yang, J. (2011). Convergence and uncertainty analyses in Monte-Carlo based sensitivity analysis. *Environmental Modelling & Software*, *26*(4), 444–457. doi:10.1016/j.envsoft.2010.10.007