Soil & Water Assessment Tool | SWAT

A Web-Based Interface for SWAT Modeling on the TeraGrid

L. Zhao[1], K. S. Subramanian[1], C. X. Song[1], S. Kumar[2], V. Merwade[2], C. Maringanti[3], I. Chaubey[3], M. Sayeed[1], R. S. Govindaraju[2]

1. Rosen Center for Advanced Computing, Purdue University, West Lafayette, IN 47907
2. Department of Civil Engineering, Purdue University, West Lafayette, IN 47907
3. Department of Agricultural and Biological Engineering, Purdue University, West Lafayette, IN 47907

Abstract

The Soil and Water Assessment Tool (SWAT) is a river basin scale model developed by the USDA Agricultural Research Service (ARS). It is widely used by researchers in various scientific domains to study the impact of land management practices on water quantity, sediment, and water quality in large watersheds over long periods of time. Currently most users run SWAT tests on their own desktop or laptop computers. While this is adequate for some, users who need to complete a large number of SWAT runs or calibration of the SWAT model for complex watersheds require more powerful computational and storage resources. In this paper we describe an effort to make the SWAT model easily accessible and useable to a larger class of users through a web portal interface. The SWAT portal allows users to run SWAT simulations using the distributed resources provided by the TeraGrid - a national cyberinfrastructure for high end computing funded by the National Science Foundation. This portal supports three types of SWAT simulations: regular simulation, auto calibration, and sensitivity analysis. It provides an intuitive interface for users to configure one or multiple SWAT cases, submit these runs as computation jobs to the TeraGrid, monitor job status, visualize results and download output. This TeraGrid based SWAT portal uses a shared community account to submit jobs to the TeraGrid resource, thus eliminating the need for users to know the details of TeraGrid allocation request and usage. Any user with a browser can connect to this portal over the Internet and benefit from the resources on the TeraGrid. This paper describes the design and implementation of the SWAT portal as well as several case studies by our early users and our future plan to improve the SWAT portal.

Keywords: SWAT model, Portal, Web Interface, TeraGrid, Community Account, Cyberinfrastructure

Introduction

The SWAT (Soil Water Assessment Tool) model was developed by USDA Agricultural Research Service (ARS) [Arnold 1998]. It is widely used to study the long term impacts of agricultural and land management practices on water quantity, sediment, and water quality in large complex watersheds over long periods of time [Neitsch 2002, Gassman et al. 2007]. Currently the SWAT model is typically run as an application on a personal computer. While the model is computationally efficient and easy to use for large watershed simulations, it has a few limitations: (1) those who need to calibrate the model for large watersheds using all available parameters or to perform sensitivity analysis require larger storage and more powerful computational resources. To run such computations on a personal computer, the SWAT model often needs to run for multiple days. The user may need to restart the simulation if there is a power outage during that period. Furthermore, it may take months to run a set of experiments with slightly different parameter settings using this conventional approach. (2) Even for regular simulations which take far less time, sometimes a user needs to run hundreds of cases with different configuration files and input files. To run these cases using the desktop application is error prone and time consuming.

To address the above challenges, we have developed a SWAT web interface that allows users to easily configure and run the SWAT model using the distributed resources provided by the TeraGrid - a NSF funded national cyberinfrastructure of high end computing and storage resources for researchers in the U.S. Our overarching goal is to provide a “one stop shop” for running SWAT simulations. We began by building an online SWAT simulation portal that allows users to (1) easily run long simulation cases by utilizing the TeraGrid resources, and (2) automate the process of running a large number of SWAT simulations to test multiple scenarios. Depending on the type of SWAT simulation requested, different TeraGrid resources are utilized. While it is important to provide a user friendly portal interface to run SWAT simulations on the TeraGrid, managing large amount of data efficiently for post processing and analysis are critical to scientific discovery and user productivity. Traditionally, users must download simulation results, which may be huge amount, to a local workstation in order to run post processing scripts and generate graphs or other visualization. The SWAT portal integrates simulation execution along with data management, post processing and visualization in one place, aimed at significantly increasing research productivity. The development of the SWAT portal is part of a NSF funded project named C4E4 (CyberInfrastructure for End-to-End Environmental Exploration) [Zhao 2007]. It has been used in several research groups at Purdue University and is making direct impacts.

In recent years, web-based simulation interfaces such as the nanoHUB, LEAD, and web-based GIS and decision supporting systems have become a popular platform that brings scientific applications to a broad user community [nanoHub, LEAD, Watergen]. We envision that the SWAT portal will similarly improve and broaden the use of SWAT and the TeraGrid for hydrological research. The SWAT portal helps eliminating technological barriers related to advanced knowledge of TeraGrid systems. Any user with a browser can connect to this SWAT portal using the internet and benefit from TeraGrid resources.

In the following sections, we first present the overall system design and workflow. We then describe in detail the user interface, job management, data access and post processing components. Two use cases are discussed next which demonstrate the benefit of the portal. The final section discusses future work and concludes the paper.

Overview of SWAT Portal Design

In this section we give a brief overview of the SWAT model and the work flow associated with a typical run. We then describe the setup of the SWAT portal and how it submits SWAT jobs to TeraGrid resources.

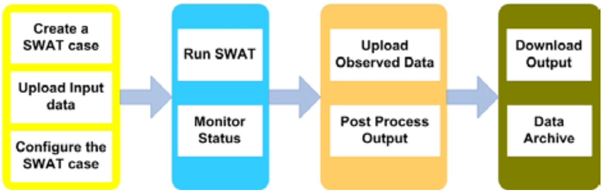


Figure 1: SWAT portal workflow

SWAT Model Workflow

A typical SWAT simulation run from the portal consists of four steps as shown in Figure 1. A user first creates a new simulation case and uploads the corresponding input file. The user then specifies the type of the simulation. The portal supports three types of SWAT simulations: regular simulation, auto calibration, and sensitivity analysis. A user can run multiple simulations at the same time. In the second step, the portal submits the SWAT simulation(s) to appropriate TeraGrid computation resources depending on the type of the simulation. A shared community account is used to submit jobs, thus eliminating the need for users to know how to obtain TeraGrid allocations and how to configure TeraGrid systems. The user can track the status of submitted runs from the portal. Once the job is completed, the user may process the output to generate plots that may be included in future publications. In the final step, the user can download the output data of interest from the portal. Old data will automatically be archived and can be accessed later when needed.

System Architecture

The setup of the SWAT portal matches the workflow described in the previous section (Figure 2). Each workflow step is a functional unit with a corresponding user interface provided by the portal. Behind the scene, the portal uses a MySQL database to manage the information about the users and their simulations. To make the portal scalable for a large number of users, the simulation jobs are submitted to a remote TeraGrid computation resource using the Globus middleware [Globus].

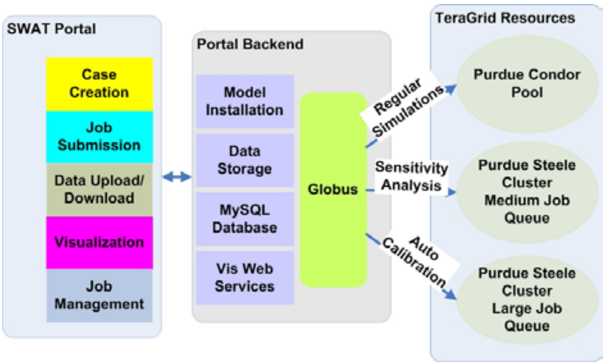


Figure 2: SWAT portal architecture

There are 11 resource providers on TeraGrid. We chose to use the Condor pool and Steele Linux cluster at Purdue University [Condor, Steele]. The Condor pool consists of over 20,000 processors of mixed architecture types and configurations. The Steele cluster consists of 893 8-core Dell 1950 systems with various combinations of 16-32 GB RAM and Gigabit Ethernet and Infiniband. These two resources are selected because (1) the vast amount of Condor nodes makes it readily available to run normal SWAT simulations. Most of the time the user can access the cycles through Condor within minutes; and (2) the Steele cluster provides several queues with different maximum wall clock limits. It is the only TeraGrid Linux cluster that allows a maximum wall clock limit of 720 hours which is long enough to run auto calibration cases.

The SWAT model developed by USDA is currently only available for the MS Windows platform. In order to run SWAT on the TeraGrid Linux systems, we first ported the source code of SWAT 2005 to Linux, using the Intel FORTRAN 90 compiler. This executable is then used by the portal to run SWAT simulations on the Steele and Condor. For the backend system, we installed the Linux version of SWAT executable in a TeraGrid “community software area” on Steele. There are also several shell scripts invoked by the portal to create user directories, launch SWAT on a compute node, and archive the output. In order to support multiple users from the portal, separate directories are created for each user to hold his/her model input, model output, and observed data for post processing.

SWAT Portal Implementation

In this section we describe the design and implementation of the main components of the SWAT portal.

User Interface

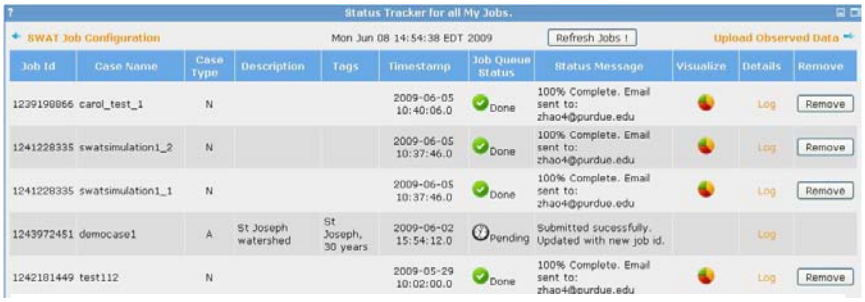
The SWAT portal interface is implemented using the Gridsphere portal development framework [GS]. The Gridsphere framework provides an open source portlet API that is JSR 168 compliant [JSR], a simple architecture for portlet integration, and a tag library for user interface design. Each function unit of the portal is implemented as a *portlet* which dynamically generates the user interface and invokes the services on the backend.

SWAT Execution

As the first step, the user needs to upload an input data archive using the data upload interface. On the job configuration page, the user specifies a set of configuration settings including the type of the simulation, the name of the experiment, description, keywords, and an email address to send notification when the simulation completes. The user can then click on the submit button to send the simulation job to the TeraGrid. Internally, the portal uses Java CoG kit and GRAM API to interact with the TeraGrid resource [Cog, GRAM]. A shared community account and GSI authentication is used when submitting jobs to the GRAM server. There are three types of jobs and they are submitted to different computation resources based on their characteristics: normal simulations are dispatched to the Purdue Condor Pool; sensitivity analysis jobs are sent to the medium size PBS job queue of the Steele cluster; auto calibrations go to the large size PBS job queue of the Steele cluster. All of these operations are transparent to the user.

Job Management

The job management component allows users to track the status of their simulation runs. The jobs are listed in a table which can be sorted based on various attributes, making it easy to find the one of interest. There are five possible job states: *submitting*, *pending*, *active*, *done*, and *error*. The portal uses a custom implementation of GRAM JobListener to get real time update on the status of a submitted job. It also provides links to the log files for debugging purpose. Users can also delete unwanted cases from this interface.



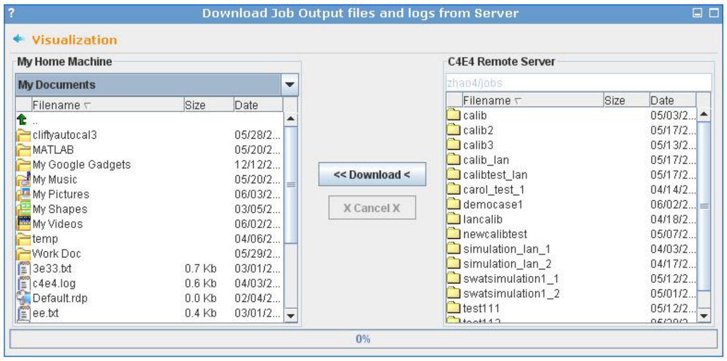
Mon Jun 08 14:54:38 EDT 2009

Job Id	Case Name	Case Type	Description	Tags	Timestamp	Job Queue Status	Status Message	Visualize	Details	Remove
1239198866	carol_test_1	N			2009-06-05 10:40:06.0	✓ Done	100% Complete. Email sent to: zhao4@purdue.edu	●	Log	Remove
1241228335	swatsimulation1_2	N			2009-06-05 10:37:46.0	✓ Done	100% Complete. Email sent to: zhao4@purdue.edu	●	Log	Remove
1241228335	swatsimulation1_1	N			2009-06-05 10:37:46.0	✓ Done	100% Complete. Email sent to: zhao4@purdue.edu	●	Log	Remove
1243972451	democase1	A	St Joseph watershed	St Joseph, 30 years	2009-06-02 15:54:12.0	Pending	Submitted sucessfully. Updated with new job id.		Log	
1242181449	test112	N			2009-05-29 10:02:00.0	✓ Done	100% Complete. Email sent to: zhao4@purdue.edu	●	Log	Remove

Figure 3: Job management interface

Visualization

SWAT output files can be huge in volume. As a result, users find it difficult to download the data and extract the variables for a particular sub-basin/reach/HRU. The visualization component addresses this need and provides interactive web–based plotting services. It asynchronously invokes a visualization web service which parses and plots selected variables in STD, SUB, RCH, and HRU files. Four types of plots can be generated using gnuplot [gnuplot]: a simulation plot on a specific variable, a comparison plot using the observed data, a multi-variable plot of two different variables using Y1 and Y2 axes, and finally, an all-in-one plot. For each plot generated, the user can download both the plot and the raw data from the portal.



Filename	Size	Date
..		
cliftyautocal3		05/28/2...
MATLAB		05/20/2...
My Google Gadgets		12/12/2...
My Music		05/20/2...
My Pictures		06/03/2...
My Shapes		03/05/2...
My Videos		06/02/2...
temp		04/06/2...
Work Doc		05/29/2...
3e33.txt	0.7 Kb	03/01/2...
c4e4.log	0.6 Kb	04/03/2...
Default.rdp	0.0 Kb	02/04/2...
ee.txt	0.4 Kb	03/01/2...

Filename	Size	Date
calib		05/03/2...
calib2		05/17/2...
calib3		05/13/2...
calib_lan		05/17/2...
calibtest_lan		05/17/2...
carol_test_1		04/14/2...
democase1		06/02/2...
lancalib		04/18/2...
newcalibtest		05/07/2...
simulation_lan_1		04/03/2...
simulation_lan_2		04/17/2...
swatsimulation1_1		05/12/2...
swatsimulation1_2		05/01/2...
test111		05/12/2...

Figure 4: Data access interface

Data Access

The portal provides easy-to-use interfaces for uploading input/observed data and downloading output data. Each interface is implemented as a portlet with an embedded Java applet. The data access applet is a client-side Java component that supports uploading and downloading files and folders to any web server. The applet is designed for cross-browser support. Java Server Pages deployed on the server side act as the backend and handle the file streams over HTTP and HTTPS. All file operations are processed in the context of the user account and the access is restricted and secure.

Use Cases

In this section we describe two usage cases involving real users. In both cases the capabilities provided by the SWAT portal significantly increased the researchers' productivity. It enabled large scale studies which are not practical using the conventional approach.

Effect of non-linear optimization technique on stability of auto-calibration

Shuffled Complex Evolution algorithm (SCE-UA; Duan et al. 1992) is a widely used global optimization technique in watershed modeling (Eckhardt and Arnold, 2001; van Griensven and Bauwens 2003; Kannan et al., 2008). Probabilistic approach followed in SCE-UA technique (evaluation of objective function at randomly selected parameter set from specified parameter space) raises question about the stability of the autocalibration results, particularly when large parameters are used. For example, a modeler cannot confirm if the autocalibration routine would give same results if it is repeated multiple times using the same inputs. Investigation of this issue is hindered by the high computational demand of the SCE-UA implementation in hydrologic models. In this study, SCE-UA implementation in SWAT auto-calibration is evaluated by using the SWAT portal. Such an investigation would be impossible using a personal computer.

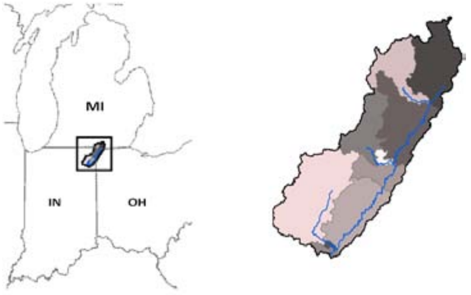


Figure 5: St. Joseph River Watershed and its sub-watersheds

SWAT is used to create watershed model for St. Joseph River Watershed (SJRW) in Northern Indiana (Figure 5). SJRW (total area: 2800 km2) is divided into 10 subwatersheds and 97 hydrologic response units (HRUs). The SJRW model is calibrated for 7 years of daily streamflow data (1993-1999) at watershed outlet using SCE-UA algorithm through auto-calibration routine in Arc-SWAT (version 1.0.5). Fourteen model parameters are included in model calibration based on sensitivity analysis results and available literature (see Kumar and Merwade, 2009 for detail). Calibration results are validated using 4 additional years of streamflow data (2000-2003) and results are found to be satisfactory (Table 1, Figure 6). The calibration runs are repeated 50 times using the same set of inputs, and results are compared with respect to model performance during calibration and validation, and the range of values associated with each parameter. All 50 simulations were submitted in parallel through the SWAT portal and completed within 72 hours using the Steele cluster at Purdue.

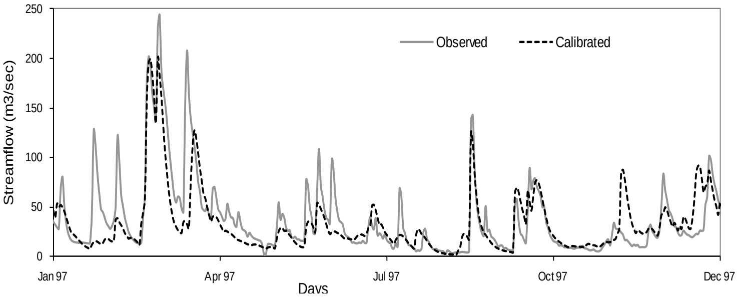


Figure 6: streamflow hydrograph (Group2) for a calibration year 1997

Results from 50 calibration runs are divided into three groups such that the results are the same in each group. Group 1 includes the calibration runs from 1-30; Group 2 includes runs from 31-40; and Group 3 includes runs from 41-50. The calibration and validation results from each group are presented in Table1. Results from Group 1 is slightly poor compared to Groups 2 and 3, and also fewer good parameter sets (<20) were obtained in Group 1 compared to the other two groups, which produced more than 1700 good parameter sets, hence Group 1 results are not included in parameter uncertainty analysis . Uncertainty range associated with selected parameters for Groups 2 and 3 are shown in Figure 7. First row of parameters in Figure 7 show similar uncertainty range (except for *Alpha_bf*) in Groups 2 and 3, while the second row of parameters show substantially dissimilar uncertainty range in Groups 2 and 3. Kumar and Merwade (2009) has classified first row of parameters as significant parameters and second row of parameters as insignificant parameters.

Table 1: Model calibration and validation results for daily streamflow output

Model	Calibration (1993-1999)		Validation (2000-2003)	
	R2NS	Mbias (%)	R2NS	Mbias (%)
Group1	0.54	0.4	0.55	21.8
Group2	0.58	-2.9	0.57	18.3
Group3	0.57	-8.8	0.57	11.0

Fifty auto-calibration runs performed in this study using same set of input resulted into 3 different groups of optimized parameters. Model results in terms of final model output (streamflow) are not significantly different among different calibration runs. Probabilistic nature of optimization technique does introduce sources of uncertainty in autocalibration results; however, uncertainty introduced by the model structure (significant vs. insignificant parameters) seems to play a larger role. The issue of model structure can be investigated by using simpler model (fewer model parameters) or including only significant parameters in model calibration.

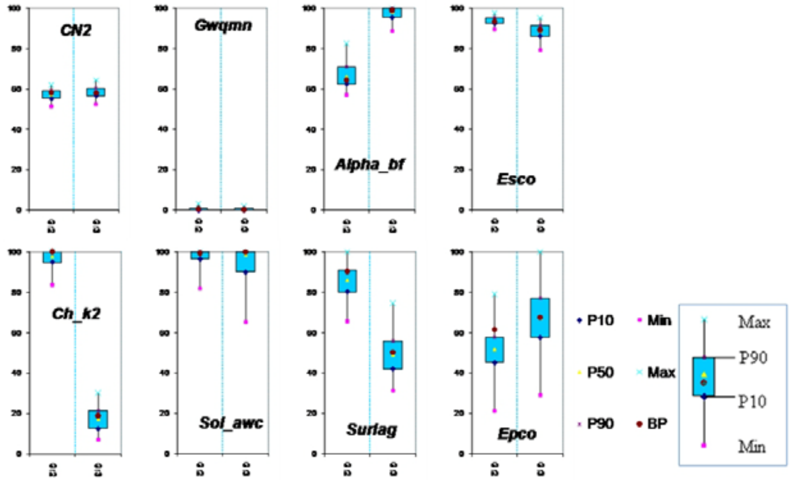


Figure 7: Uncertainty plots for good parameters sets from Group2 (Gr2) and Group3 (Gr3). Y axis represents normalized value of parameter uncertainty range (P10, P50 and P90 are 10th, 50th, and 90th percentile, Min: Minimum, and Max: Maximum of good parameter sets. BP: Best Parameter set)

Clifty creek watershed initiative project

Clifty creek watershed, located in south eastern Indiana, has an area of 522 km^2. The land use in the watershed is predominately agricultural with corn and soybean occupying 84% of the total area. Close to 4% of the watershed is urban and 10% forest. The water quality in Columbus city located at the downstream is of concern in the watershed which has been degrading mostly due to the Nonpoint Source (NPS) pollutants, such as surface runoff, nutrients, sediment, and pesticide that is a result of intensive fertilizer and pesticide application in the agricultural regions to achieve better yields. The goal of this project is to implement Best Management Practices (BMPs) in the Clifty creek watershed to reduce the NPS pollutants from the watershed, as well as to implement urban BMPs such as rain barrel in the watershed.

The SWAT model was developed using the ArcSWAT interface available in ArcGIS 9.2. The watershed is delineated from a 30 m Digital Elevation Model (DEM) obtained from United States Geological Survey (USGS). The watershed was then divided into subwatershed based on the user defined outlets in the watershed for which detailed outputs were required. These outlets also consist of observed stream flow and water quality monitoring stations located in the watershed. The subwatershed were further divided into Hydrologic Response Units (HRUs), based on common land use and soil, that form the unit at which the SWAT model performs the calculations. The land use data was obtained in a gridded form from National Land Cover Database (NLCD) 2001 and the soil data was obtained from STATSGO.

The watershed model was calibrated using the SWAT web portal for stream flow. Table 2 provides the details about various parameters that were used for calibration, which were observed to be sensitive in literature that used the SWAT model to simulate stream flow. The auto-calibration method available in the SWAT model uses a shuffled complex algorithm to perform the optimization of the objective function (which was sum squared errors in the simulation) by obtaining optimal parameter values. The total execution time of the model calibration was 12 hours using the Steele cluster compared to 41 hours when running on a personal computer. The optimal parameter values obtained are shown is Table 2.

Table 2. Stream flow calibration of the SWAT model for Clifty Creek watershed

Parameter	Usle_P	Slsubbsn	Slope	Esco	Ch_K2	Timp	Surlag	Cn2	Usle_C	Epco	Ch_N	Smfmx
Low	0.1	-25	0	0	-25	0.01	0	-10	0.001	0	0.01	0
high	1	25	0.6	1	25	1	10	10	0.5	1	0.5	10
optimal	0.247	7.466	0.10	0.829	-6.153	0.08	8.627	-5.874	0.165	0.081	0.245	4.573

Future Work

As future work, we plan to develop services that dynamically submit SWAT jobs to other TeraGrid sites primarily based on the availability of the computation resources. This will significantly reduce the average waiting time of jobs before they actually run on a cluster. We would also like to work closely with the SWAT user community to get feedbacks on the portal design and add new features based on their needs. For example, we are currently working with a research group in the Agricultural and Biological Engineering department at Purdue University to implement an interface that enables batch configuration and submission of a large number of jobs with slightly different parameter settings. Finally we are also in the process of making the portal available for instruction use.

Conclusion

In this paper we describe our design and implementation of a web portal that makes it easy to run different types of SWAT simulations using TeraGrid resources. The portal integrates a comprehensive set of services for end-to-end scientific exploration, including data upload/download, simulation composition, execution, status tracking, and visualization. The main contribution of this work is that it enables users to run long running watershed calibration cases as well as a large number of SWAT simulations using TeraGrid resources, thus significantly reducing the total amount of time required. We believe this web interface addresses an important demand in the SWAT community and will prove to be a convenient and efficient tool for research and educational users.

Acknowledgements

This research is sponsored in part by the National Science Foundation under grant DBI-0619086 and TeraGrid Resource Partners grant OCI-0503992.

References

Arnold J.G., R. Srinivasan, R.S. Muttiah, and J.R. Williams, 1998. Large area hydrologic modeling and assessment part I: model development. *Journal of the American Water Resource Association*, 34(1), 73-89.

Condor: RCAC – BoilerGrid. Available at: http://www.rcac.purdue.edu/userinfo/resources/condorpool/. Accessed 2 June 2009.

Cog: Java Cog Kit. Available at: http://wiki.cogkit.org/wiki/Main_Page. Accessed 2 June 2009.

Duan, Q., S. Sorooshian and V. Gupta, 1992. and efficient global optimization for conceptual rainfall-runoff models. *Water Resources Research*, 28 (4), 1015-1031.

Eckhardt, K. and Arnold J.G. 2001. Automatic calibration of a distributed catchment model. *Journal of Hydrology. 251: 103-109*

Gassman, P.W., M. R. Reyes, C.H. Green, and J.G. Arnold, 2007. The Soil and Water Assessment Tool: Historical Development, Applications, and Future Research Directions. *Transaction of the ASABE*, 50(4), 1211-1250

Globus: Globus Toolkit. Available at: http://www.globus.org/toolkit/. Accessed 2 June 2009.

gnuplot: gnuplot homepage. Available at: http://www.gnuplot.info/. Accessed 2 June 2009.

GRAM: GRAM-Globus. Available at: http://dev.globus.org/wiki/GRAM. Accessed 2 June 2009.

GS: The GridSphere Portal Framework. Available at: http://www.gridsphere.org. Accessed 2 June 2009.

JSR: JSR 168 Portlet Specification. Available at: http://jcp.org/en/jsr/detail?id=168. Accessed 2 June 2009.

Kannan N., C. Santhi, and J.G.Arnold, 2008. Development of an automated procedure for estimation of the spatial variation of runoff in large river basin. *Journal of Hydrology*, 359, 1-15.

Kumar S., and V. Merwade, 2009. Impact of watershed subdivision and soil data resolution on SWAT model calibration and parameter uncertainty. *Journal of the American Water Resource Association*, in Press.

Zhao, L., Song. C. X., et al. 2007. Interweaving Data and Computation for End-to-End Environmental Exploration on the TeraGrid. Proceedings of the TeraGrid'07 Conference, Madison, WI.

LEAD: LEAD Portal. Available at https://portal.leadproject.org/gridsphere/gridsphere. Accessed 2 June 2009.

nanoHub: nanoHUB.org – Simulation, Education, and Community for Nanotechnology. Available at http://nanohub.org/. Accessed 2 June 2009.

Neitsch, S. L., Arnold, J. G., Kiniry, J. R., Williams, J. R., King, K. W. 2002. Soil Water Assessment Tool Theoretical Documentation: 2002. Available at: http://www.brc.tamus.edu/swat/downloads/doc/swat2000theory.pdf. Accessed 2 June 2009.

Steele: RCAC–Steele. Available at: http://www.rcac.purdue.edu/userinfo/resources/steele/: Accessed 2 June 2009.

Van Griensven, A., and W.Bauwens, 2003. Multiobjective autocalibration for semidistributed water quality models. *Water Resources Research*, 39(12), 1348, doi:10.1029/2003WR002284.

Watergen: Watergen Web Home. Available at http://cobweb.ecn.purdue.edu/~watergen/. Accessed 2 June 2009.