

Fall 2009

Development of a Distributed Artificial Neural Network for Hydrologic Modeling

Rebecca Logsdon

University of Arkansas, Fayetteville

Follow this and additional works at: <http://scholarworks.uark.edu/inquiry>



Part of the [Fresh Water Studies Commons](#), and the [Hydrology Commons](#)

Recommended Citation

Logsdon, Rebecca (2009) "Development of a Distributed Artificial Neural Network for Hydrologic Modeling," *Inquiry: The University of Arkansas Undergraduate Research Journal*: Vol. 10 , Article 11.

Available at: <http://scholarworks.uark.edu/inquiry/vol10/iss1/11>

This Article is brought to you for free and open access by ScholarWorks@UARK. It has been accepted for inclusion in Inquiry: The University of Arkansas Undergraduate Research Journal by an authorized editor of ScholarWorks@UARK. For more information, please contact scholar@uark.edu, ccmiddle@uark.edu.

DEVELOPMENT OF A DISTRIBUTED ARTIFICIAL NEURAL NETWORK FOR HYDROLOGIC MODELING

By Rebecca Logsdon

Department of Biological Engineering

Faculty Mentor: Sreekala G. Bajwa

Department of Biological Engineering

Abstract

Hydrological models are used to represent the rainfall-runoff and pollutant transport mechanisms within watersheds. Accurate representation of these dynamic and complex natural processes within a watershed is an important step in managing and protecting a watershed. Artificial neural network (ANN) models are often used in hydrologic modeling. Typical ANN models are trained to use lumped data. However, watershed characteristics used as inputs in hydrological modeling are spatially and often temporally dynamic. Therefore, a lumped model does not have the ability to represent changes in spatial dynamics of a watershed. Therefore, the purpose of this study was to develop and test a distributed ANN model for simulating the rainfall-runoff process in the L'Angeuille River Watershed located in Eastern Arkansas. The watershed was divided into nine sub-basins to account for the spatial dynamics of flow within the watershed. Inputs for the model were rainfall, average temperature, antecedent flow and curve number. Output was runoff, collected from gage-stations at Colt and Palestine representing two of the sub-basins. Daily SCS curve numbers were developed and adjusted for crop planting and harvesting dates and crop rotation practices in each sub-basin. The model had nine layers with one neuron each to represent the nine sub-basins. The layers were connected so that if one sub-basin spatially flowed into another, its output would be an input for the downstream sub-basin. The model performed well, showing R^2 values of 0.93 and 0.98 and Nash-Sutcliffe Efficiency values of 0.92 and 0.97 for the validation and test datasets.

Introduction

Watersheds and Watershed Management

Water is one of the most important natural resources. It "drives all human systems and those of most other organisms as well" (Heathcote, 1998). Watersheds are particularly important in managing water resources, as they are broadly defined as the area of land that contributes runoff to a particular point. Managing a watershed is crucial for maintaining good ecosystems and human health. Runoff is an important aspect of watershed management. Runoff is precipitation that falls onto the earth but does not infiltrate into the soil, evapotranspire through plants, or get stored. Runoff carries with it nutrients, sediments, and pollutants until it eventually reaches a body of water. Nutrients, sediments, and pollutants that do not get deposited along the way may end up in water

bodies. Simulation of runoff is an initial step in watershed management.

Hydrological Models

Hydrological modeling is a field of study that attempts to utilize mathematical and analytical models to model watersheds and predict watershed characteristics. Many hydrologic models have been developed in attempts to model different aspects of watersheds. One very common model is the Soil and Water Assessment Tool (SWAT). SWAT models are often used for modeling watersheds, but they have difficulty accounting for LULC changes other than crop rotation. This is a problem because these parameters not only vary within a watershed, but are also interrelated with one another. For example, the runoff in one section of a watershed may contribute flow into a different section of the watershed. Therefore, typical models are incapable in handling complex relationships between large amounts of data efficiently.

Artificial Neural Networks

Artificial Neural Networks (ANNs) were designed to process and transfer information similarly to the neurons in a human brain. Broadly, a neural network is given a variety of inputs and corresponding outputs (Figure 1). These inputs enter into a hidden layer or layers that contain neurons. As the inputs pass through the hidden layer, weights and biases are added to the data. When the weighted data goes through a neuron, it is processed with a non-linear function in an attempt to relate the input data to the target data. Simply put, ANNs have the ability to relate input and output variables in complex systems

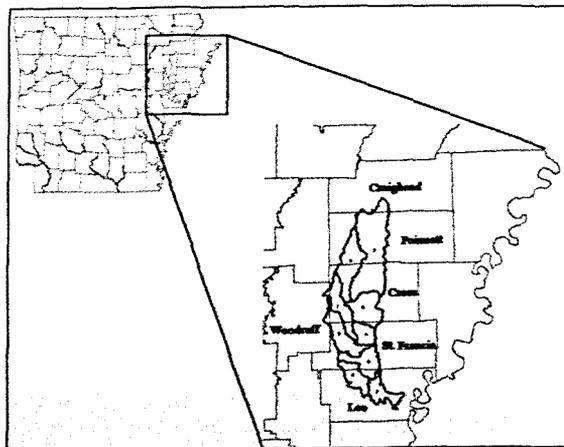


Figure 1. Location of LRW in Arkansas and counties encompassed.

(Dawson et al., 2001).

Artificial Neural Networks are relatively new to hydrologic modeling, but have the ability to handle multiple data inputs and relate them in non-linear spatial ways (Dawson et al., 2001). ANNs also have the capability to account for dynamic changes in a watershed, such as changes in land use and land cover. This property is especially important for watershed management, because increasing human population leads to a rapidly changing landscape. Typically, ANNs used in hydrologic models are feed-forward, back-propagation networks with one hidden layer of neurons. Input and target data along with network parameters are entered. Data flows forward through the network, where the network compares the computed output to the known target by calculating an error (usually mean squared error). If the error goal is not met, the network keeps re-running the data, changing the weights and biases until a given network parameter is met. The problem with this typical use of ANNs is that it does not have the ability to spatially relate the input parameters.

In this research, however, a pre-defined network in MatLab® was not used to model the LRW. Instead, a custom ANN with a specific architecture was defined in order to better capture the spatial dynamics of the flow within the watershed.

Significance of Research

Being able to accurately and efficiently model aspects of a watershed, particularly runoff, is very important in monitoring and controlling non-point source pollution within the watershed. Unlike point-source pollution, non-point source pollution is difficult to pinpoint and quantify. It is carried through runoff and sediment flow in and out of watersheds. Because of the Clean Water Act (1972) and its regulations, it is important to be able to quantify pollutant and sediment transport in a given watershed. Water health and quality is a good indication of ecosystem health and health of the human population. Water is the most essential resource for human survival. It is needed for drinking, for growing food, and for cleansing purposes. A lack of clean water leads to many waterborne diseases and even death. Being able to quantify, monitor, and even predict runoff and pollutant loads in runoff is a great step towards conserving and managing watersheds and water resources.

Objectives

In this study, an ANN model was developed to simulate and predict the watershed scale rainfall-flow process using historical flow data from USGS gage stations. Other objectives of this study were to perform a sensitivity analysis on input variables and evaluate the performance of the ANN model.

Methods

Determining Watershed for Case Study

L'Anguille River Watershed (LRW) is located in Eastern Arkansas, United States and encompasses six counties (Figure 1). LRW was divided into nine sub-basins to account for the

spatial dynamics of flow (Figure 1).

The watershed is mostly agricultural land (rice, soybean, and cotton), followed by forest and urban areas (Figure 3). LRW was chosen as a case study because, due to its large agricultural production, it has some major pollution problems.

Under section 303(d) of the Clean Water Act, states are

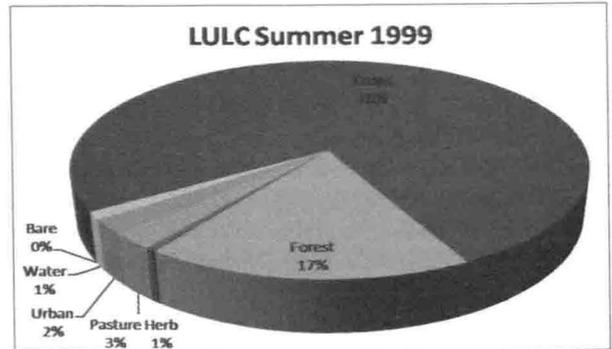


Figure 2. LULC for LRW for summer 1999.

required to develop a list of impaired waters that are too polluted or degraded to meet water quality standards set by that state (USEPA, 2009). The states are then required to establish

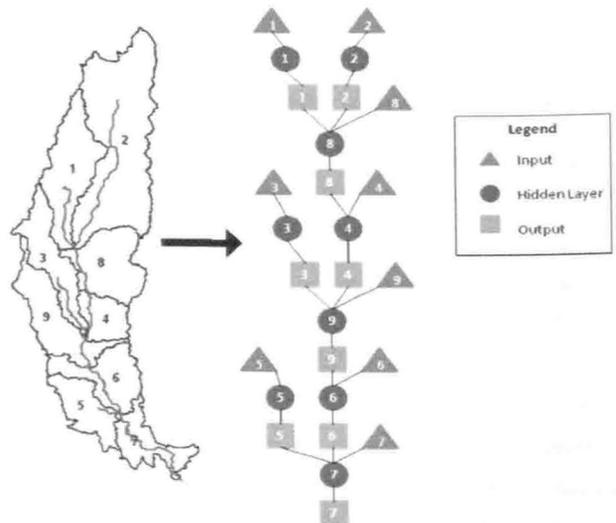


Figure 3. Simplified network architecture showing how layers were connected to account for spatial dynamics of water flow within LRW.

rankings for the impaired water bodies listed and develop Total Maximum Daily Loads (TMDLs) for the pollutant that is causing the water quality problems. Since 1995, there have been seven TMDL reports on the L'Anguille River, five for turbidity and two for fecal coliforms (USEPA, 2009). In 2008 the river had twelve of its reaches totaling over 98 miles designated as impaired (Class 5) by the Arkansas Department of Environmental Quality (ADEQ, 2008). Agriculture was the source of the pollutants and problems in all known cases (ADEQ, 2008). Five of the twelve reaches designated as

impaired in 2008 were classified as 5a streams meaning they are “truly impaired” and TMDLs need to be developed for the given parameter.

Determining Input Data for ANN Model

Runoff is precipitation that does not evapotranspire back into the atmosphere, infiltrate into the groundwater, or get stored in the soil. Therefore, to precisely quantify the amount of runoff entering a water body, it would be ideal to have exact values for these four variables. Current technology, however, does not allow for exact quantification of these parameters. Thus, there is a need to create hydrologic models that can take what data is currently available and mathematically relate the data so as to estimate runoff. Inputs chosen for this model were precipitation, average temperature, Soil Conservation Service Curve Number (SCS-CN), and antecedent stream flow where available. Precipitation and average temperature data for each sub-basin were collected from the nearest weather station; data used were from January 1, 1995-December 31, 2004. Antecedent stream flow data were collected from two different USGS gage stations on L'Anguille River at Colt and Palestine for the same years. The daily SCS-CN had to be developed based on LULC and hydrologic soil type data.

Development of SCS-CN

The SCS-CN provides a way to quantify and estimate the amount of runoff that an area of land generates, based on the LULC and hydrologic soil type of that land. Since the precipitation data were from January 1, 1995-December 31, 2004, daily SCS-CN's needed to be developed for this time period for each sub-basin. LULC data for LRW were available for spring, summer, and fall of 1999 from the University of Arkansas' Center for Advanced Spatial Technology database, so it was assumed to be the base LULC of the 1995 data. Soils data were available for LRW from the U.S. National Resource Conservation Service. Using ESRI's ArcGIS program, and specifically ArcMap, the soils and LULC data were dissolved (based on hydrologic soil group and cover name, respectively) and intersected for each sub-basin. Then, the SCS-CN was calculated for each soil-LULC complex based on NEH curve number tables (USDA, 2008). The area weighted CN was then calculated for each sub-basin for the spring, summer, and fall datasets.

Next, the CNs were adjusted in order to account for crop planting and harvesting dates (USDA, 1997) because the CN changes based on whether or not the crop is actually in the ground. The result after this adjustment was a daily CN for one year, assumed to be year 1995 (the beginning of the precipitation, temperature, and gage station data). Since crop rotation is a significant management practice in agriculture, the first year's daily CNs had to be adjusted according to common crop rotation practices. These crop rotation practices were determined based on a focus

group survey of University of Arkansas Cooperative Extension Service agents (Hill et. al, 2003). The data produced after crop rotation adjustment included 10 years (Jan. 1, 1995- Dec. 31, 2004) of daily, area-weighted CNs for each sub-basin.

Determining Network Outputs and Target Data

Since the purpose of this study was to predict the flow in the L'Anguille River, naturally the target output for the model was discharge. However, there are only two USGS gage stations along the entire reach of the L'Anguille River (at the outlets of sub-basins eight and nine). Therefore, only sub-basins eight and nine were connected to target data and used to monitor and evaluate network performance.

Constructing Network Architecture

One objective of the project was to create an ANN that could account for the spatial dynamics of flow within the watershed. Instead of using a pre-defined network in MatLab®, a custom, distributed ANN with a unique architecture was defined in order to better capture the spatial dynamics of the runoff within the watershed. By custom defining the network, the architecture was arranged in such a way that the output of one sub-basin was an input into another sub-basin if the first sub-basin's flow entered into the second sub-basin.

The network created contained three or four inputs for each sub-basin (rainfall, average temperature, SCS-CN, and antecedent stream flow for sub-basins 8 and 9), nine layers with one neuron representing each sub-basin, and target output for sub-basins eight and nine. The network layers were

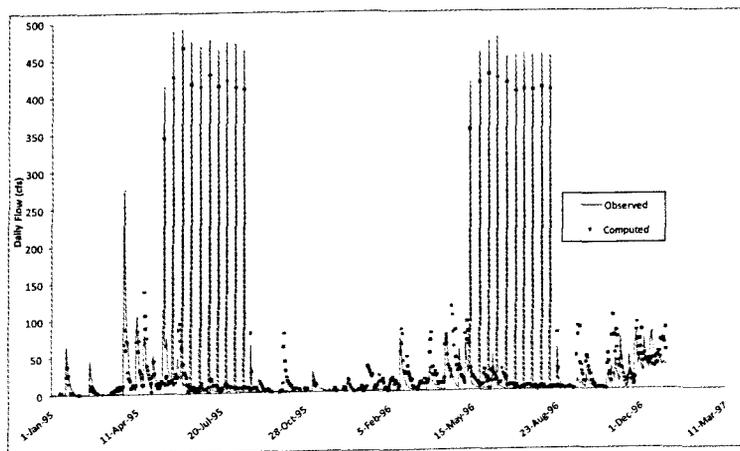


Figure 4. Hydrologic time series for validation set for sub-basin 8.

connected in such a way as to account for the spatial dynamics of water flow between the sub-basins within LRW (Figure 4).

For network training, the Levenberg-Marquardt algorithm was used as the training algorithm and the performance of the network was measured by the Mean Squared Error (MSE). The data sets were divided into training, testing, and validation datasets. Because the training of the network requires the most

data, 60% (1997-2002) of the dataset was used for training, whereas 20% (1995-1996) was used for testing and 20% (2002-2004) for validation.

Optimization of Network Parameters

Because only one neuron was used in each layer, it was not necessary to optimize the number of neurons. Thus, optimization was performed only on the training parameters. Since the training function chosen was the Levenberg-Marquardt algorithm, the only option for optimization of the network was the learning rate. A trial and error procedure was followed by varying learning rate at different increments. The optimized learning rate was identified as the one that resulted in the lowest MSE.

Sensitivity Analysis

The sensitivity of the model to each input was determined, by examining the weights of the network inputs. It was determined that the model was “sensitive” to all inputs of the model (rainfall, average temperature, SCS-CN, antecedent flow).

Results

The custom defined neural network was run, using the optimized learning rate of 0.15. The simulated flow closely followed the actual flow (Figures 5 and 6). The model was evaluated at the outputs of sub-basins eight and nine using three different criteria: (1) correlation coefficients between

Table 1. Calculated results of neural network model performance.

		Correlation Coefficient	R-Square Value	Nash-Sutcliffe Efficiency
Sub-basin 8	Training	0.82	0.68	0.68
	Testing	0.74	0.55	0.46
	Validation	0.96	0.93	0.92
Sub-basin 9	Training	0.98	0.96	0.96
	Testing	0.96	0.93	0.93
	Validation	0.99	0.98	0.97

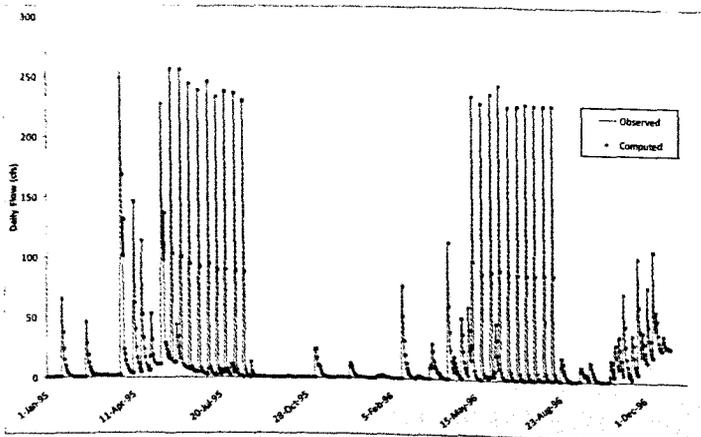


Figure 5. Hydrologic time series for validation set for sub-basin 9.

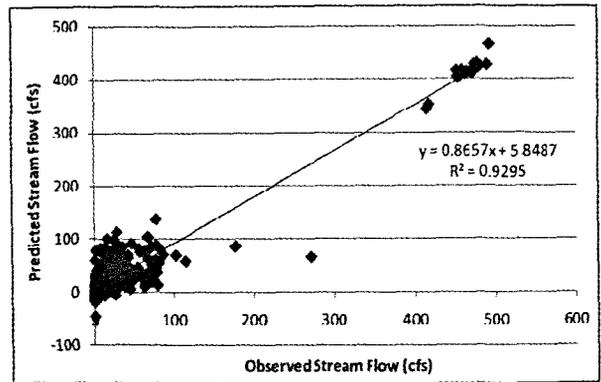


Figure 6. Linear regression of observed stream flows versus computed stream flows for sub-basin 8.

computed and observed results, (2) R-square values between the computed and observed results, and (3) Nash-Sutcliffe Efficiency coefficient.

Linear regression was performed using Excel© (Figures 7 and 8). Both the correlation coefficient and R-square value

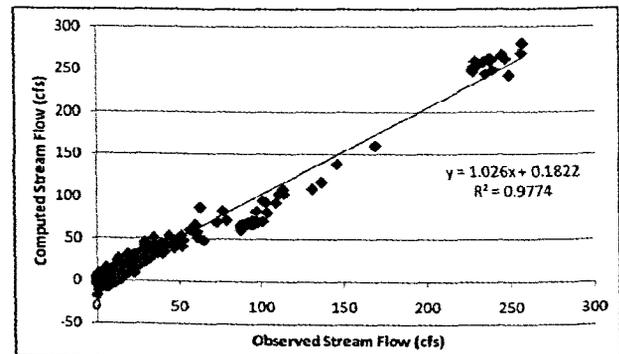


Figure 7. Linear regression of observed stream flows versus computed stream flows for sub-basin 9.

were calculated for sub-basins 8 and 9 for the training, testing, and validation datasets (Table 1).

The Nash-Sutcliffe Efficiency value is often used in evaluating hydrologic models because it is “insensitive to additive and proportional differences between model simulations and observations” (Harmel et. al, 2007). The Nash-Sutcliffe Efficiency value was calculated for sub-basins 8 and 9 for the training, testing, and validation datasets using equation 1 (Table 1).

The performance of the model was very good in sub-basin 9, but poor in sub-basin 8, particularly for training and test data. This may be because the model optimized itself for sub-basin 9 and not sub-basin 8.

Summary and Conclusions

The goal of this project was to develop a spatially distributed, custom ANN to model the flow within a watershed. The model was trained using historical daily rainfall, average temperature, SCS-CN, and stream flow data. The model was trained with 6 years of data

and tested and validated using 2 years of data. The results of the model show that the model was able to simulate the stream flow at sub-basin 9 very well, with all R2 values >0.93 and the Nash-Sutcliffe Efficiency values all being greater than 0.93 as well. Sub-basin 8 results were not quite as good as sub-basin 9; however, the results are acceptable with training and testing R2 values >.55 and validation R2 = 0.93. Also, the Nash-Sutcliffe

$$NSE = 1 - \frac{\sum_{i=1}^n (obs_i - comp_i)^2}{\sum_{i=1}^n [(obs_i - \overline{obs})]^2}$$

Efficiency values were all positive and greater than 0.46. The better performance of the model at sub-basin 9 could be due to the fact that it is spatially further down from sub-basin 8, and therefore had more previous inputs.

Table 1. Calculated results of neural network model performance.

		Correlation Coefficient	R-Square Value	Nash-Sutcliffe Efficiency
Sub-basin 8	Training	0.82	0.68	0.68
	Testing	0.74	0.55	0.46
	Validation	0.96	0.93	0.92
Sub-basin 9	Training	0.98	0.96	0.96
	Testing	0.96	0.93	0.93
	Validation	0.99	0.98	0.97

In conclusion, this study has shown that a spatially distributed ANN is very capable at accurately simulating the stream flow of a river in a watershed. A lumped, SWAT model that was developed for LRW using total monthly flows instead of daily flows had R2 values between measured and predicted total flows of 0.84 and 0.87 for calibration and validation periods, respectively (Srivastava, et. al, 2005). The Nash-Sutcliffe Efficiency values were also 0.86 and 0.91 for calibration and validation periods, respectively. For both sub-basin 8 and sub-basin 9 validation data sets, the R2 and Nash-Sutcliffe Efficiency values were higher than the SWAT model. This shows the capabilities and possibilities that a distributed artificial neural network has in modeling stream flow.

Acknowledgements

The authors thank the Arkansas Department of Higher Education for funding of this project through a State Undergraduate Research Fellowship. Thanks are also due to Dr. Findlay Edwards and Dr. Marty Matlock for their guidance and support, and to Mansoor Leh for his help on ANNs and ArcGIS.

References

Arkansas Department of Environmental Quality. 2008. List of Impaired Water Bodies (303(d) List).

Dawson, C. W., and R. L. Wilby. 2001. "Hydrological Modelling Using Artificial Neural Networks." *Progress in*

Physical Geography 25: 80-108.

Harmel, R. D and P. K. Smith. 2007. Consideration of Measurement Uncertainty in the Evaluation of Goodness-of-Fit in Hydrologic and Water Quality Modeling. *Journal of Hydrology*. 337, 3-4.

Heathcote, Isobel W. 1998. *Integrated Watershed Management Principles and Practices*. New York: John Wiley & Sons.

Hill, J., M. Popp, and P. Manning. 2003. Focus Group Survey Results: Typical Arkansas Crop Producer Production and Marketing Practices. Publication no. 971. Arkansas Agricultural Experiment Station.

Srivastava, Vibhava, Sreekala Bajwa, and Indrajeet Chaubey. 2005. "Hydrological Modeling of the L'Anguille River Watershed using Soil and Water Assessment Tool." 2005 ASAE Annual International Meeting. Available at: <http://asae.frymulti.com/azdez.asp?search=1&JID=5&AID=20207&CID=tf12005&v=&i=&T=1>. Accessed 20 February 2009.

USDA.-NASS. 1997. Usual Planting and Harvesting Dates for U.S. Field Crops. Agricultural Handbook 628.

USDA. 2008. National Resources Conservation Service. *National Engineering Handbook*.

USEPA. 2009. Impaired Waters and Total Maximum Daily Loads. Available at: <http://www.epa.gov/owow/tmdl/>. Accessed 25 February 2009.

Mentor Comments:

Professor Sreekala Bajwa reflects on the innovative quality of Rebecca Logsdon's work and the reasons why it merits publication.

Rebecca Logsdon's research merits publication in the Inquiry journal for a number of reasons. Her work is quite innovative as there are no published articles on the distributed artificial neural network model for representing watershed scale hydrological processes in large watersheds. When Rebecca started this project, her background in ecological engineering, particularly in hydrology, geographic information systems (GIS), statistics, etc was very minimal. She taught herself many of the basic concepts, and learned the relevant information quickly. The quality of the work she has done is excellent. The fact that her research is accepted for presentation at the World Congress on Computers in Agriculture and Natural Resources is a testimony to its quality and relevance in today's world.

Artificial neural networks (ANN) models are usually fast, accurate and easy to implement. They have been used in hydrologic modeling for simulating rainfall-runoff, groundwater movement as well as nutrient and pollutant transport. The practical use of ANN models in hydrology was fairly limited in the past due to the lump nature of these models. Watershed scale hydrologic processes are highly distributed processes as the soil, topography, vegetation

and weather vary spatially and temporally in a watershed. Therefore, Rebecca focused on developing a watershed scale distributed ANN model to represent the spatial and temporal dynamics of rainfall runoff. Such a model has the potential to broaden its application to flood forecasting, water quality modeling, water planning, understanding the impact of urban development, etc. Therefore, her research is an important step towards utilizing innovative modeling tools that are faster and easier to run towards protecting the environment.

Rebecca's work is high quality and fairly complex for an undergraduate research project. She devoted an enormous amount of time to complete this work. This work has served as a platform for her success in attracting an NSF fellowship and admission to graduate school at Purdue University.