

# REPORT

**Title** Watershed Protection Plan Development for the Geronimo Creek Watershed

**Project Number** 2010TX363B

***Principal Investigator***

Kyna McKee  
Master of Science Student  
Department of Biological and Agricultural Engineering  
Texas A&M University  
[kmckee@tamu.edu](mailto:kmckee@tamu.edu)  
Cell: (210) 289-0068  
325 Scoates Hall  
2117 TAMU  
College Station, TX 77843

***Co-Principal Investigators (faculty advisors)***

Patricia Smith  
Associate Professor  
Department of Biological and Agricultural Engineering  
Texas A&M University  
[Patti-smith@tamu.edu](mailto:Patti-smith@tamu.edu)  
Office: (979) 845-3630  
302 B Scoates Hall  
2117 TAMU  
College Station, TX 77843

R. Karthikeyan  
Assistant Professor  
Department of Biological and Agricultural Engineering  
Texas A&M University  
[karthi@tamu.edu](mailto:karthi@tamu.edu)  
Office: (979) 845-7951  
306 A Scoates Hall  
2117 TAMU  
College Station, TX 77843

## **Abstract**

When developing a Watershed Protection Plan (WPP) or a Total Maximum Daily Load (TMDL), it is often difficult to accurately assess the pollutant load for a watershed as a result of inadequate water quality monitoring data. According to the Texas Commission on Environmental Quality (TCEQ), there are 274 bacteria impairments in Texas water bodies out of 386 impaired water bodies. Data on bacteria in water bodies is often more sparse than other types of water quality data, which hinders the development of WPPs or TMDLs. To address this problem, a spatial watershed model was developed to simulate bacteria concentrations in streams resulting from nonpoint sources using the Spatially Explicit Load Enrichment Calculation Tool (SELECT) combined with a simple rainfall-runoff model. SELECT is an automated Geographical Information System (GIS) tool that can estimate potential *E. coli* loads from point and non point sources in watersheds using spatial characteristics such as land use, population density, and soil type. The watershed model applies a rainfall-driven loading function to the potential *E. coli* loads calculated by the output of SELECT. The SELECT methodology combined with this watershed model was applied to estimate *E. coli* loads in the Geronimo Creek watershed, located in central Texas. The simulated *E. coli* concentrations from the model were compared to actual monthly routine grab sample *E. coli* data collected at two sampling site near the outlet of the subwatershed. The runoff volumes were predicted with good to very good agreement for both sampling sites. Nash – Sutcliffe efficiencies range from 0.74 to 0.84 and root mean square error – observations standard deviation ratio (RSR) range from 0.51 to 0.40. The predicted *E. coli* concentrations performed unsatisfactorily for both sites and four calibration methods. The results show that the model does not include significant factors contributing to the transport of *E. coli* bacteria but can be modified to include these factors.

## **Problem and Research Objectives**

When developing a Watershed Protection Plan (WPP) or a Total Maximum Daily Load (TMDL), it is often difficult to accurately assess the pollutant load for a watershed because not enough water quality monitoring data is available. Bacteria are the most common reason for impairment of Texas water bodies. According to the Texas Commission on Environmental Quality (TCEQ), there are 274 bacteria impairments in Texas water bodies out of 386 impaired water bodies (TCEQ 2008). Bacteria water quality data is often more sparse than other types of water quality data, which hinders the development of WPPs or TMDLs.

In order to develop WPPs or TMDLs, additional bacteria water quality data must be collected which is costly and time consuming. The bacteria load analysis for a watershed cannot begin until the water quality monitoring data collection is completed. Generally, water quality data can take anywhere from a year to multiple years to collect for a substantial dataset. The U.S. EPA estimates water quality monitoring of all TMDLs nationally, “The cost of water quality monitoring to support the development of TMDLs is expected to be approximately \$17 million per year” (USEPA 2001). A considerable portion of developing a TMDL is to allocate pollutant load and to identify potential sources. This can be done with modeling which can be costly and require a significant amount of input data.

Models such as Soil and Water Assessment Tool (SWAT) and Hydrological Simulation Program- FORTRAN (HSPF) have been used for bacterial modeling (Benham, et al., 2006; Sadeghi & Arnold, 2002). Other simplistic microbial models such as, the potential non-point pollution index (PNPI) and a Spatially Explicit Delivery MODEL (SEDMOD), have been

developed to rank the potential pollution impacts of areas from nonpoint sources primarily utilizing land use and geomorphology (Fraser, et al. 1998; Munafo, et al. 2005).

SELECT is an automated Geographic Information System (GIS) tool that can assess potential *E. coli* loads in a watershed based on spatial factors such as land use, population density, and soil type (Teague, et al., 2009). SELECT is able to calculate a potential *E. coli* load and highlight areas of concern for best management practices (BMPs) to be implemented. The potential *E. coli* load in SELECT is calculated by distributing the contributing sources spatially over the entire watershed. The population densities of potential contributors are determined with stakeholder input to accurately represent the watershed, however, SELECT is a worst case scenario model and assumes that the largest amount of contribution possible from individual sources.

Current bacteria models either require extensive monitoring data within the watershed for calibration or are not able to predict actual *E. coli* concentrations in the water body. A simple model that is able to predict actual bacteria concentrations in a water body is needed in order to develop TMDLs or WWPs within the state of Texas. The objective of this study was to develop a model that would estimate the runoff volume and the *E. coli* concentration contributing from surface runoff at a sampling site drainage area outlet.

The overall objective of this research project was to develop a conceptual model in ArcGIS 9.X utilizing the potential *E. coli* load estimated by SELECT to simulate *E. coli* concentrations occurring in Geronimo Creek. It was presumed that precipitation is the main driving factor for the transport of *E. coli* bacteria from sources to the stream. Also the affects of temperature were negligible, since in Texas watersheds the monthly normal daily mean temperatures do not vary from month to month by more than 10 °F.

- (1) To apply SELECT to Geronimo Creek watershed using stakeholder inputs concerning the *E. coli* sources and the population densities.
- (2) Another sub objective was to develop an automated rainfall-runoff model in ArcGIS 9.X utilizing rain gauges located in and around the Geronimo Creek watershed and to estimate the *E. coli* concentrations in the creek.

## Methodology

*E. coli* concentrations were calculated using a modified delivery factor originally developed by McElroy et al. (1976) for pollutant loading from livestock facilities:

$$C = (Y * D) / (a * R * A) \quad (1)$$

where

*C* = concentration of *E. coli* at sampling site (CFU/mL)

*Y* = daily loading rate of *E. coli* at sampling site (CFU)

*a* = unit conversion factor ( $2.54 \times 10^4$ ) – to convert from in•m<sup>2</sup> to mL

*R* = daily runoff at sampling site (in)

*A* = grid cell area (m<sup>2</sup>) – 900 m<sup>2</sup>

*D* = delivery factor (dimensionless)

The equation was intended for livestock facilities but was applied to multiple non-point sources calculated using SELECT and ArcGIS 9.X. The variable concentration of pollutant in runoff (*C*) was calculated using the equation above to determine the concentration of *E. coli* in Geronimo Creek. The loading rate (*Y*) was calculated in SELECT for livestock, wildlife, and domestic sources. McElroy et al. (1976) acknowledged that the quantity of pollutants discharged depends mostly on runoff volume.

Runoff (R)

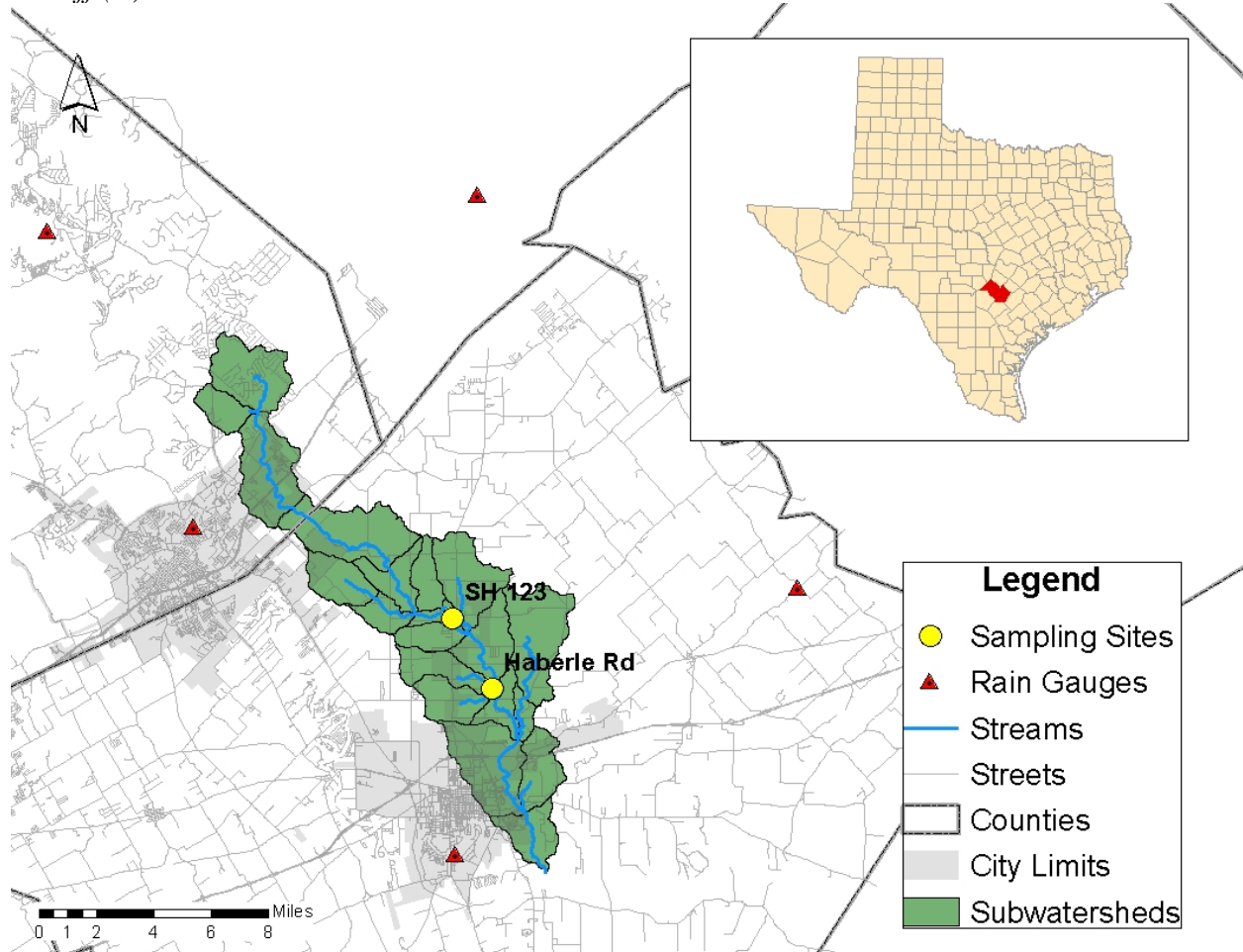


Figure 1. Geronimo Creek Watershed Study Area With Rain Gauges and Sampling Sites

Daily precipitation data was obtained at 5 sites, Canyon Dam, Kingsbury, New Braunfels, San Marcos, and Seguin, from the National Climatic Data Center (NCDC) for 1996 to 2010. The NCDC rain gauges shown in figure 1 were utilized to develop a daily precipitation grid.

The minimum rainfall to induce runoff was calculated using the SCS curve number approach by using the average area weighted curve number for the Geronimo Creek watershed. The watershed curve number grid was developed in ArcGIS 9.X. by intersecting the Soil Survey Geographic (SSURGO) hydrologic soil group with the land use type and using an NRCS lookup table. The area weighted curve number for the Geronimo Creek Watershed was calculated as 82. The minimum rainfall to induce runoff calculated using the area weighted curve number was 0.44 inches.

Runoff precipitation was assumed to occur in the watershed if one of the five rain gauges measured precipitation greater than the minimum rainfall to induce runoff. A precipitation grid was developed in ArcGIS 9.X. for each day with runoff precipitation occurring on the same day as when routine *E. coli* samples were taken from the Geronimo Creek sampling sites using the ArcGIS Spatial Analyst Extension. The interpolation method used will be inverse distance weighted (IDW). Inverse distance weighting assumes that observations closer to one another are

more alike than ones farther apart (Zhang & Srinivasan, 2009).

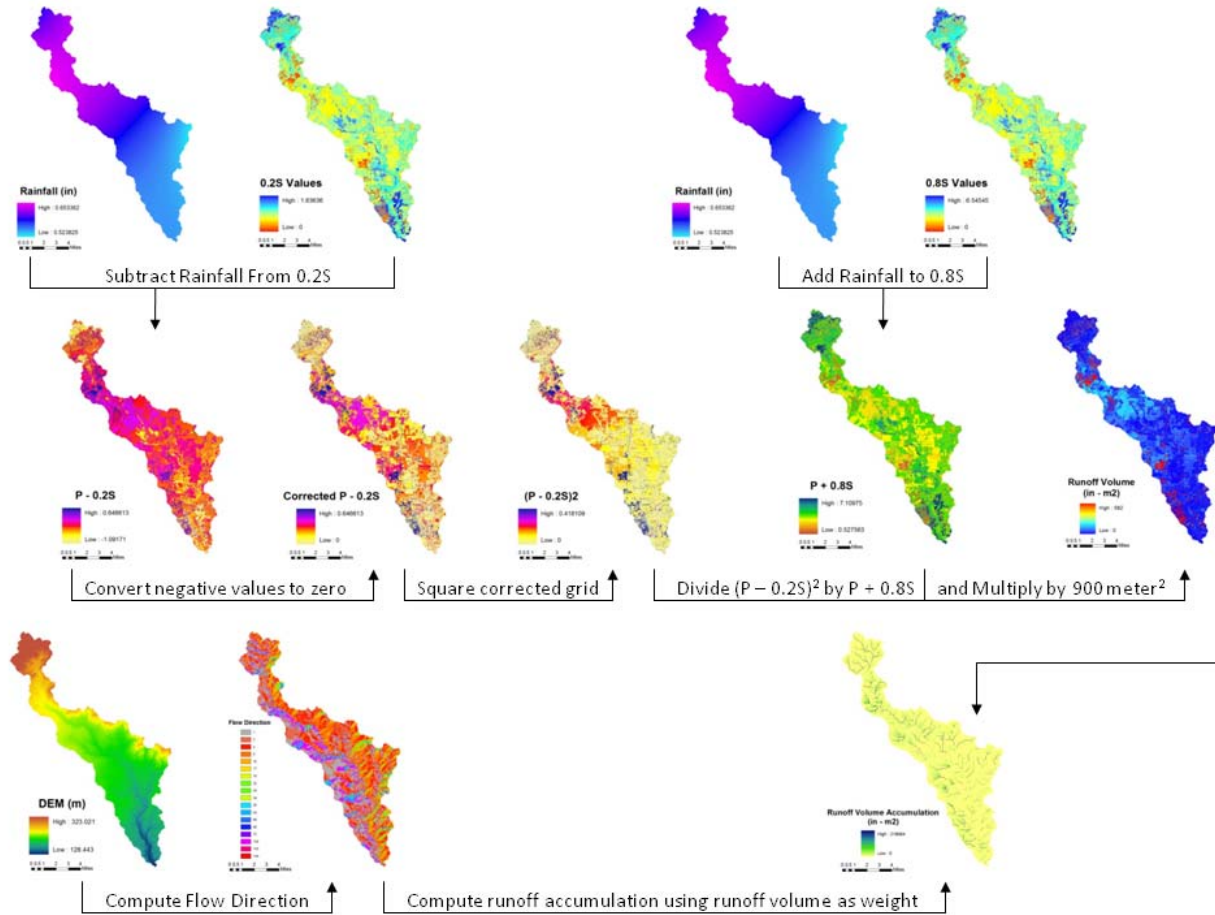


Figure 2. Flow chart illustrating the calculation of accumulated runoff volume

The runoff volume at a sampling site was then calculated from the precipitation grid (Figure 2). An automated tool was programmed into ArcGIS to calculate a runoff grid with the inputs being a rain gauge shapefile with the measured amounts of rainfall for each rain gauge as fields in the attribute table and an S grid calculated from the curve number grid. The runoff grid was calculated using the SCS curve number approach with the equation  $Q = (P - 0.2S)^2 / (P + 0.8S)$  where  $Q$  is the runoff depth in inches,  $P$  is the precipitation and  $S$  is the maximum soil water retention parameter derived from the curve number. The runoff equation requires that  $P$  must exceed  $0.2S$  before any runoff is generated. The average runoff is first calculated for the entire watershed on a 30 meter grid cell basis. The curve number grid is calculated into an  $S$  grid using the equation,  $S = (1000/CN) - 10$  where  $CN$  is the curve number. For the results of  $(P - 0.2S)$ , the negative values were given a value of zero so that runoff was not calculated for cells with  $P$  less than  $0.2S$ . After the runoff depth was calculated, the runoff depth was then converted to a runoff volume per grid cell by multiplying by the cell area which was 900 square meters creating a runoff volume grid.

An additional part of the Arc GIS 9.X. tool was used to automatically calculate with flow accumulation grid for the watershed. The inputs to the tool were the previously generated runoff volume grid and a Digital Elevation Model (DEM) over the watershed area provided by the Texas A&M University SSL which had a 30 meter grid cell size. The result of the flow

accumulation would be the total amount of runoff volume going through a specific grid. The runoff volume at a sampling site is estimated by identifying the runoff volume value grid cell at the sampling site drainage area outlet.

*Potential E. coli Load (Y) Estimation using SELECT*

Potential *E. coli* loads for Geronimo Creek were predicted using SELECT and input from stakeholders for stocking rates and possible sources. A custom land use classification (Figure 4) was provided by the Texas A&M University Spatial Sciences Laboratory (SSL) using 2008 National Agriculture Imagery Program (NAIP) imagery and a prior Texas Parks and Wildlife (TPWD) Classification.

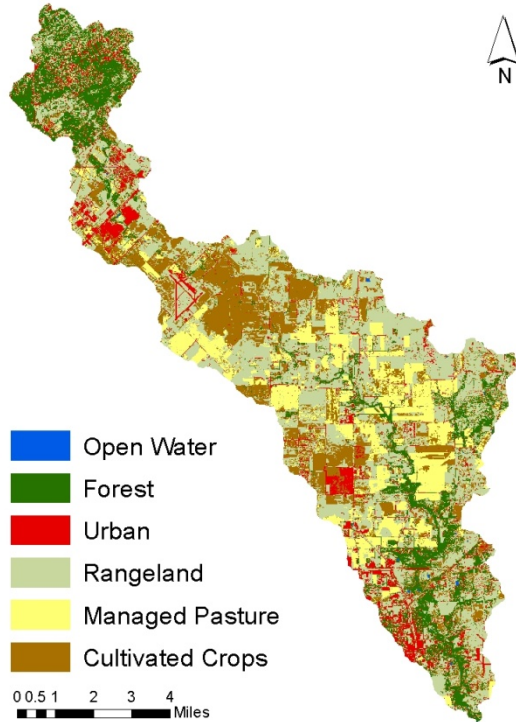


Figure 3. Geronimo Creek watershed land use classification

The SWAT model was used to delineate 21 subwatersheds as well as the watershed stream channel. In the Geronimo Creek watershed, it was determined that livestock sources for the watershed are goats, horses, and cattle. Wildlife sources are deer and feral hogs. Domestic sources consist of dogs and on-site wastewater treatment systems (OWTSs). A conversion of 0.63 fecal coliform to *E. coli* was used in the model. The conversion factor of 0.63 was decided using the USEPA’s regulatory standards for fecal coliform and *E. coli* in recreational waters. The regulatory standard for fecal coliform was 200 organisms per 100 mL and is 126 organisms per 100 mL for *E. coli* (USEPA, 2003). The conversion factor was determined by taking the ratio of these two regulatory standards.

For livestock and wildlife, the number of animals is estimated with animal densities and stakeholder input. For cattle, the stakeholders determined a stocking rate of 20 and 10 acres per animal should be applied to Comal and Guadalupe Counties respectively with a suitable habitat of rangeland, forest, and managed pasture land use types. A density for horses was determined to be 132 acres per animal over the entire watershed with a total watershed population of 124

horses with a suitable habitat of rangeland. The animals are distributed evenly across suitable habitats and a fecal production rate is then applied per animal. Due to goats being raised on goat farms, 200 goats out of the total watershed population of 750 animals were distributed evenly in the watershed on rangeland, forest, and managed pasture land use types. The remaining animals were concentrated to specific watersheds which contained known goat farms for a specified number of animals. The potential *E. coli* load for the subwatersheds containing goats was calculated per subwatershed by multiplying the number of animals per subwatershed by the fecal production rate per animal. White-tailed deer had a population density of 10 acres per animal (Lockwood, 2005). The suitable habitat determined for deer were forest and rangeland with at least 20 acres of contiguous terrain available. Feral hogs had a population density of 26 acres per animal and were only distributed on suitable habitat within 100 meters of the main stem of Geronimo Creek which is perennial. Feral hogs were not distributed around Alligator Creek because it is an intermittent creek and is an unsuitable habitat for feral hogs. The suitable habitats for feral hogs as determined by stakeholders were forest, rangeland, managed pasture, and cultivated crops.

For dogs, the 2000 census data was used to calculate the contribution by using a dog density of 1 dog per household. The potential *E. coli* load for OWTSSs was calculated by Espey Consultants. For OWTSSs, spatially distributed point data of each household was collected from 911 address data and households within Certificate of Convenience and Necessity (CCN) areas were removed to not include households being serviced by a wastewater treatment facility. A failure rate was determined for the OWTSSs using SSURGO soil limitation classes and the age of the system to calculate the percentage of *E. coli* contributing to the watershed due to septic failure. A fecal production rate was then applied to each household for dogs and OWTSSs. Since SELECT divides the watershed into a raster grid with a 30 meter cell size, the potential load is calculated over the entire watershed at a 30 meter cell size. The individual raster files for each source are then added together spatially to create a total load raster (Figure 5) for the watershed that is divided into 30 meter grid cells.

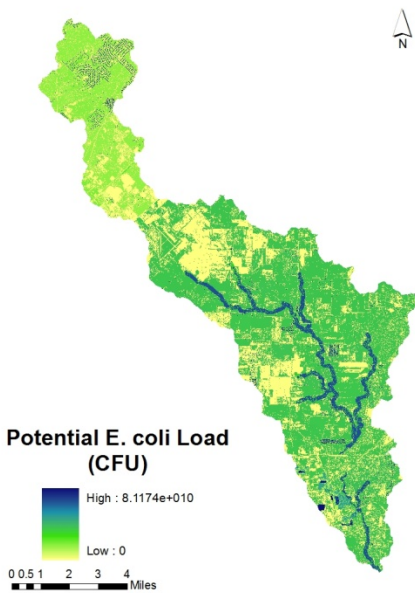


Figure 4. Total potential *E. coli* load calculated using SELECT for the Geronimo Creek watershed

The total load raster (Figure 4) estimates the potential *E. coli* load for the entire watershed based on a worst case scenario assuming the entire load calculated reaches the water body. Another part of the tool programmed in ArcGIS 9.X. was to calculate the *E. coli* load actually reaching a specific grid cell in the watershed.

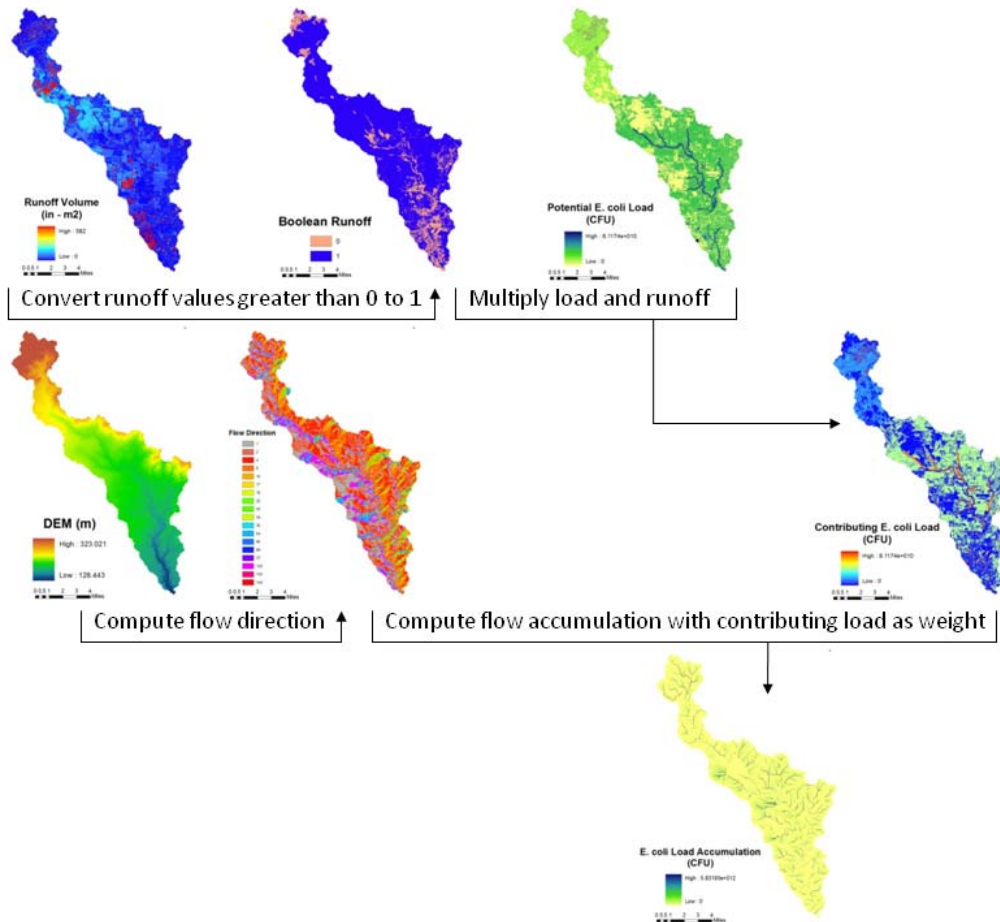


Figure 5. Flow chart illustrating the calculation of the contributing *E. coli* load

The inputs to the tool were the previously calculated runoff grid, the total load raster which was an output from running SELECT, and the DEM. The first step to estimating the *E. coli* load reaching the sampling site was to only consider the *E. coli* load grid cells that have runoff generated. A runoff SELECT grid was estimated for each runoff event in which the cells with no runoff generated had a contributing *E. coli* load of zero. The flow accumulation was calculated using the runoff SELECT grid as an input weight and the DEM. The output of flow accumulation would then represent the amount of *E. coli* load that would flow through each cell considering the upslope cells. The flow accumulation at a sampling site would then estimate the *E. coli* load reaching that site.

#### Calculation of Observed Runoff Volume

The observed instantaneous stream flow taken during the time the *E. coli* grab sample was sampled was converted to a runoff volume. The base flow was removed from the stream flow by subtracting the 100% exceedence flow. Flow duration curves were developed for the sampling sites SH 123 and Haberle Road using SWAT simulated flow rates ranging from 1998



to 2009. The 100% exceedence flow for the Haberle Road sampling site was determined as 1.89 cfs and 1.0 cfs for the SH 123 site.

The stream flow was converted to a runoff volume by multiplying by the lag time calculated for each sampling site using the SCS lag equation based on natural watersheds  $t_L = L^{0.8}(S + 1)^{0.7} / 1900Y^{0.5}$  where L is the hydraulic length of the sampling site drainage area in feet, S is the average maximum soil water retention parameter calculated from the curve number grid for the sampling site drainage area, and Y is the average land slope of the sampling site drainage area in percentage.

The SH 123 sampling site L parameter was determined by measuring the longest length for SWAT delineated stream channel to the drainage area outlet. The stream length measuring 78926 feet included the entire length of Alligator Creek and the length of Geronimo Creek from its confluence with Alligator Creek to the drainage area outlet. Although the Haberle Road sampling site is located downstream of SH 123, the flow accumulation showed due to slope that the Haberle Road site had about 1/5 of the area contributing of 1033 pixels compared to the SH 123 site's 5035 pixels. To remedy this difference, the hydraulic length for the Haberle Road sampling site was determined by measuring the longest stream length from the site determined with the flow accumulation grid which was 4738 feet. The lag time for the SH 123 site was 7.18 hours and 0.78 hours for the Haberle Road sampling site.

#### *Delivery Factor (D)*

The delivery factor is back calculated from equation 1 using observed *E. coli* concentration data. All factors influencing the processes affecting the runoff of the potential *E. coli* load into the creek are meant to be included in the delivery factor with the exception of runoff. Two separate delivery factors were calculated, one using the observed runoff volume converted from the observed stream flow. The other delivery factor is calculated from the simulated runoff volume.

A delivery ratio was calculated for all data points both using the observed and simulated runoff volume for each site separately. For each site, the average and the geomean was calculated for the separate delivery ratios. This resulted in the calculation of eight different delivery ratios to be applied to the data. For both sites, an observed and simulated delivery ratio was calculated with each type applying both an average and geomean.

#### *Calibration*

We obtained historical and routine stream flow and *E. coli* concentration sampling data ranging from 1996 to 2010 from the Guadalupe Brazos River Authority (GBRA). The SH 123 and Haberle Road sampling sites were both historical sites while the other 13 samplings sites in the watershed began sampling in September 2008. 84 Haberle Road samples were taken on a monthly basis beginning in September 2003 and ending in December 2010. For the SH 123 sampling site, monthly sampling began in October 1996 and ended in August 2003, but then resumed on September 2008 until August 2010. Out of the 105 data points taken at the SH 123 sampling site only 5 coincided with runoff precipitation. Only 12 data points out of the 84 for the Haberle Road site samples were taken when runoff precipitation occurred. The model was calibrated for both the Haberle Road site and the SH 123 site separately.

## Statistics

The accuracy of the model was evaluated using the Nash-Sutcliffe efficiency ( $E$ ), root mean square error ( $RMSE$ ), and RMSE-observations standard deviation ratio ( $RSR$ ). According to Nash and Sutcliffe (1970) the  $E$  value is an index of agreement or disagreement between observed and predicted values. The  $E$  value evaluates how consistently the predicted values agree with the observed values by applying linear regression analysis (Nash and Sutcliffe, 1970).  $E$  is computed with the equation:  $E = 1 - [\sum_{i=1}^n (O_i - P_i)^2 / \sum_{i=1}^n (O_i - \bar{O})^2]$  where  $O_i$  is observed values,  $P_i$  is predicted values, and  $\bar{O}$  is the mean of the observed values (Nash and Sutcliffe, 1970). The  $E$  value ranges from negative infinity to 1, where negative values are considered a biased model and values between 0 and +1 are considered an unbiased model (McCuen, et al., 2006). Model efficiencies were classified similar to Moriasi et al. (2007) and Parajuli et al. (2009) as very good ( $E = 0.75 - 1$ ), good ( $E = 0.5 - 0.74$ ), fair ( $E = 0.25 - 0.49$ ), poor ( $E = 0 - 0.24$ ) and unsatisfactory ( $E < 0.0$ ).

$RMSE$  is an error index used in model evaluation and is valuable because the error is indicated in the units of the constituent of interest (Moriasi, et al., 2007). Legates and McCabe (1999) recommend including at least one relative error measure ( $E$  or  $R^2$ ) and at least one absolute error measure ( $RMSE$  or mean absolute error) for a complete assessment of model performance.  $RSME$  values close to 0 indicate a perfect fit but values half the standard deviation are still considered low (Singh, et al., 2004). The equation for  $RMSE$  is:

$RMSE = \sqrt{\sum_{i=1}^n (O_i - P_i)^2 / n}$  where  $O_i$  is observed values,  $P_i$  is predicted values, and  $n$  is the number of samples.

$RSR$  is a model evaluation statistic that standardizes  $RMSE$  with the observed data standard deviation (Moriasi, et al., 2007). Moriasi et al. (2007) developed  $RSR$  to fill the need of an error index with additional information provided for using  $RSME$  with the standard deviation recommended by Legates and McCabe (1999). The equation for  $RSR$  is:

$RSR = [\sqrt{\sum_{i=1}^n (O_i - P_i)^2}] / \sqrt{\sum_{i=1}^n (O_i - \bar{O})^2}$  where  $O_i$  is observed values,  $P_i$  is predicted

values, and  $\bar{O}$  is the mean of observed values (Moriasi, et al., 2007). The value of  $RSR$  ranges from 0, which is the optimal value and indicates a perfect model, to a large positive value (Moriasi, et al., 2007). Model efficiencies are classified by Moriasi et al. (2007) as very good ( $RSR = 0.00 - 0.50$ ), good ( $RSR = 0.51 - 0.60$ ), satisfactory ( $0.61 - 0.70$ ), and unsatisfactory ( $RSR > 0.70$ ).

Moriasi et al. (2007) states that the model evaluation guidelines for both  $E$  and  $RSR$  values given apply to a continuous, long-term simulation for a monthly time step. The guidelines should be adjusted based on a multitude of factors including quality and quantity of measured data, single-event simulation, evaluation time step, model calibration procedure, and project scope and magnitude (Moriasi, et al., 2007). Moriasi et al. (2007) continues to say that when a complete measured time series does not exist, such as when only a few grab samples per year are available, that the data may not be sufficient for analysis using the recommended statistics.

## Principal Findings

The runoff volumes and the *E. coli* concentrations were simulated for both the Haberle Road and SH 123 sampling sites.

### *Runoff Volume*

The model was able to predict the runoff volume at the Haberle Road sampling site outlet with good agreement and at the SH 123 sampling site outlet with very good agreement. For the Haberle Road sampling site, both the *E* and *RSR* values (Table 1) had a good performance rating. The *RMSE* value is considered low because it less than half of the observed standard deviation as shown in Table 1. The SH 123 sampling site had a very good performance rating for both *E* and *RSR* values. The *RMSE* value for the SH 123 sampling station was considered low as well.

Table 1. Runoff volume model performance

<b>Sampling Site</b>	<b>Statistic</b>	<b>Value</b>
Haberle Road	E	0.74
	RSR	0.51
	RMSE	984
	Observed Average	1329
	Observed Standard Deviation	2015
SH 123	E	0.84
	RSR	0.40
	RMSE	1494
	Observed Average	3764
	Observed Standard Deviation	4128

For the Haberle Road sampling site, simulated runoff volumes were mostly underestimated with the exception of one point. This point may have been overestimated because it was taken during the driest season in a year (in August) whereas; the other points were taken in wetter months. The dataset does not include any data points taken in the fall months (October and November); September is not considered a fall month because the weather is still similar to the summer weather for this region. The dataset also has a gap for the 2009 year where no data points collected had contributing runoff occurring at the same time,

The runoff volumes for the SH 123 site were all underestimated for the five data points. The data was not taken continuously and there is therefore a gap between the years 2002 and 2010 with no data taken in 2001, where no data was collected where runoff occurred at the same time. The SH 123 site data points only include the fall and winter seasons with only one data point taken in the spring. This may skew the data some because the points do not include the summer season which is typically the driest season for the region.

### *E. coli Concentrations*

For both the Haberle Road and SH 123 sampling sites, the model predicted *E. coli* concentrations with unsatisfactory agreement (Table 2) for all four methods of delivery factor calibration for both *E* and *RSR* values. The *RMSE* values for both sites using all four methods, were higher than the observed standard deviations and observed averages (Table 2) indicating an unsatisfactory agreement between the observed and predicted *E. coli* concentrations. The delivery factor estimated from the geomean of simulated runoff volumes performed the best for both the Haberle Road and SH 123 sampling sites. The Haberle Road site consistently performed better than the SH 123 site with the *E* and *RSR* values of -0.67 and 1 (Table 2) for the

Haberle Road site and value of -101.21 and 10.11(Table 2) for the SH 123 site, respectively, estimated using the geomean simulated delivery factor for both. The delivery factor estimated using simulated runoff volumes versus observed runoff volumes was able to better predict the *E. coli* concentrations because simulated runoff volumes were consistently under predicted for both the Haberle Road and SH 123 sampling sites. The runoff volume was meant to dilute the *E. coli* load to an *E. coli* concentration entering the stream.

Table 2. *E. coli* concentration model performance.

Sampling Site	Statistic	Simulated Delivery Factor		Observed Delivery Factor	
		Geomean	Average	Geomean	Average
Haberle Road	E	-0.67	-72	-1155	-54189
	RSR	1	9	34	233
	RMSE	21	138	646	4421
	Observed Average	12	12	12	12
	Observed Standard Deviation	20	20	20	20
SH 123	E	-101.21	-408.58	-54143	-84609620
	RSR	10.11	20.24	233	9198
	RMSE	8.00	16.01	184	7275
	Observed Average	1.79	1.79	1.79	1.79
	Observed Standard Deviation	0.88	0.88	0.88	0.88

The observed *E. coli* concentrations had values ranging from 0.46 to 57 CFU/mL. The *E. coli* concentrations predicted using the delivery factor estimated from the geomean of the simulated runoff volumes was the method that had the closest range of concentrations (0.54 to 42.32 CFU/mL) to the observed concentration range. The method predicting *E. coli* concentrations using the delivery factor estimated from the average of the observed runoff volumes performed the poorest and grossly over predicted with a range of concentrations from 187 to 14739 CFU/mL. The *E. coli* concentrations predicted using the delivery factor estimated from the average of the simulated runoff volumes had a closer range of 5 to 445 CFU/mL than the concentrations predicted using the geomean of the observed runoff volumes with a range of 27 to 2162 CFU/mL.

The prediction of *E. coli* concentrations for the SH 123 sampling site was poorer than the prediction for the Haberle Road sampling site. The SH 123 sampling site followed similar trends as the Haberle Road sampling site. As with the Haberle Road sampling site, for the SH 123 sampling site, the delivery factors computed using the simulated runoff volumes performed better than the delivery factors computed using the observed runoff volumes. The delivery factors computed using the geomean instead of the average of the respective runoff volumes performed better as well for the SH 123 sampling site. The range for the observed *E. coli* concentrations was from 1.12 to 3.2 CFU/mL. The *E. coli* concentrations predicted using the delivery factor estimated using the geomean of simulated runoff volumes had the closest range from 0.26 to 19 CFU/mL of predicted concentrations to the observed concentrations. Since the SH 123 site runoff volume was predicted more accurately than the Haberle Road site runoff

volume, the *E. coli* concentration data should be better predicted as well. Since this is not the case, other factors influencing the transport of *E. coli* bacteria are not accounted for in the model.

### Significance

In ungauged watersheds, historical bacteria data is sparsely available. It is expensive to collect more monitoring data. The USEPA estimates a cost of approximately seventeen million dollars a year for water quality monitoring to support the development of all national TMDL projects (USEPA, 2001). Current bacteria models require extensive monitoring data within the watershed for calibration or they cannot predict actual *E. coli* concentrations in the water body. A simple model that predicts actual bacteria concentrations in a water body is needed in order to develop TMDLs or WWP within the state of Texas.

### References Cited

- Benham, B., Baffault, C., Zeckoski, R., Mankin, K., Pachepsky, Y., Sedeghi, A., et al. (2006). Modeling Bacteria Fate and Transport in Watersheds to Support TMDLs. *Transactions of the ASABE*. 49(4) , 987-1002.
- Fraser, R. H., Barten, P. K., & Pinney, D. A. (1998). Predicting stream pathogen loading from livestock using a geographical information system-based delivery model. *Journal of Environmental Quality*. 27 , 935-945.
- Legates, D. R., & McCabe, G. J. (1999). Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation. *Water Resources Research*. 35(1) , 233-241.
- Lockwood, M. (2005). *White-tailed Deer Population Trends*. Austin, Texas: Texas Parks and Wildlife Department.
- McCuen, R. H., Knight, Z., & Cutter, A. G. (2006). Evaluation of the Nash-Sutcliffe efficiency index. *Journal of Hydrologic Engineering*. 11(6) , 597-602.
- McElroy, A. D., Chiu, S. Y., Nebgen, J. W., Aleti, A., & Bennett, F. W. (1976). *Loading functions for assessment of water pollution from nonpoint sources, EPA-600/2-76-151*. Washington D.C.: USEPA Office of Research and Development.
- Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., & Veith, T. L. (2007). Model evaluation guidelines for systematic quantifications of accuracy in watershed simulations. *Transactions of the ASABE*. 50(3) , 885-900.
- Munafò, M., Cecchi, G., Baiocco, F., & Mancini, L. (2005). River pollution from non-point sources: a new simplified method of assessment. *Journal of Environmental Management*. 77 , 93-98.
- Nash, J. E., & Sutcliffe, I. V. (1970). River flow forecasting through conceptual models part 1 - A discussion of principles. *Journal of Hydrology*. 10 , 282-290.
- Parajuli, P. B., Mankin, K. R., & Barnes, P. L. (2009). Source specific fecal bacteria modeling using soil and water assessment tool model. *Bioscience Technology*. 100 , 953-963.
- Sadeghi, A. M., & Arnold, J. G. (2002). A SWAT/microbial sub-model for predicting pathogen loadings in surface and groundwater at watershed and basin scales. *Total Maximum Daily Load (TMDL) Environmental Regulations*. St. Joseph, MI: ASAE.
- Singh, J., Knapp, H. V., & Demissie, M. (2004). *Hydrologic modeling of the Iroquois River watershed using HSPF and SWAT. ISWS CR 2004-08*. Champaign, Ill: Illinois State Water Survey.

- TCEQ. (2008, March 19). *2008 Texas Water Quality Inventory and 303(d) List*. Retrieved August 4, 2010, from Texas Commission on Environmental Quality: <http://www.tceq.state.tx.us/compliance/monitoring/water/quality/data/08twqi/twqi08.html>
- Teague, A., Karthikeyan, R., Babbar-Sebens, M., Srinivasan, R., & Persyn, R. A. (2009). Spatially Explicit Load Enrichment Calculation Tool to Identify Potential *E. coli* Sources in Watersheds. *Transactions of the ASABE*. 52(4) , 1109-1120.
- USEPA. (2003). *Bacterial water quality standards for recreational waters (freshwater and marine waters) status report, EPA 823-R-03-008*. Washington, D.C.: U.S. Environmental Protection Agency, Office of Water.
- USEPA. (2001). *The National Costs of the Total Maximum Daily Load Program (Draft Report), EPA 841-D-01-003*. Washington D.C.: U.S. Environmental Protection Agency, Office of Water.
- Zhang, X., & Srinivasan, R. (2009). GIS-based spatial precipitation estimation: a comparison of geostatistical approaches. *Journal of the American Water Resources Association*. 45(4) , 894-906.

