Nat Env Sci 2012 3(1): 19-33

Journal of Natural & Environmental Sciences

www.academyjournals.net

ISSN 1309 7474

Original Article

Predicting *E. coli* Concentrations in Surface Waters Using GIS

Kyna E. BOREL[1], R. KARTHIKEYAN[1*], Patricia K. SMITH[1], R. SRINIVASAN[2]
[1]*Texas A&M University, Biological and Agricultural Engineering Department, 2117 TAMU, College Station, TX 77843, USA*
[2]*Spatial Sciences Laboratory, Texas Agricultural Experiment Station, 1500 Research Parkway, Suite 221E, College Station, TX 77845, USA*

Received: 04.04.2012 ***Accepted: 18.05.2012*** ***Published: 06.06.2012***

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 because inadequate water quality monitoring data are available. According to the Texas Commission on Environmental Quality (TCEQ), there are 274 bacteria impairments in Texas water bodies out of 386 impaired water bodies. Bacteria water quality data are often more sparse than other types of water quality data, which hinders the development of WPPs or TMDLs. The Spatially Explicit Load Enrichment Calculation Tool (SELECT) is an automated Geographical Information System (GIS) tool that can assess pathogen loads in watersheds using spatial factors such as land use, population density, and soil type. A spatial watershed model was developed to simulate bacteria concentrations in streams resulting from non point sources using SELECT combined with a simple rainfall-runoff model.This model was used to simulate *E. coli* concentrations in the Geronimo Creek watershed. The watershed model applies a rainfall-driven loading function to the potential *E. coli* loads calculated by the output of SELECT. The simulated runoff volumes and *E. coli* concentrations from the model were compared to actual monthly *E. coli* data collected at two sampling sites near the outlet of a subwatershed. The runoff volumes were predicted with very good agreement (Nash-sutcliffe efficiency (E_{NS}) = 0.95, Ratio of root-meansquare error to standard deviation (*RSR*) = 0.21to 0.22) for both sampling sites. The predicted *E. coli* concentrations did not agree with measured concentrations for both sites using four different methods to determine a delivery factor for 2 sites. The results indicate that the model does not include significant factors contributing to the transport of *E. coli* bacteria.

***Key words**: Bacteria modeling, Nonpoint source pollution, Pathogens, Surface runoff, Total maximum daily load, Water quality*

[*]***Corresponding Author**: R. Karthikeyan, e-mail: karthi@tamu.edu, Phone:+019798457951, Fax:+019798623442*

INTRODUCTION

When developing a Watershed Protection Plan (WPP) or a Total Maximukm 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. 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). 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.

In order to develop WPPs or TMDLs, additional data on waterborne bacteria 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, the timeframe for a substantial water quality dataset can range from a year to multiple years. The US EPA estimates the cost of water quality monitoring of all TMDLs nationally "is expected to be approximately $17 million per year" (USEPA 2001b). A considerable portion of effort in developing a TMDL is inallocating pollutant loads and identifing potential sources. This can be done with modeling, which can be less costly thanthan extensive sampling.

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; Sadeghiand 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 where best management practices (BMPs) should 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 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 waterbody. A simple model that is able to predict actual bacteria concentrations in a waterbody 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 contributed by surface runoff at a sampling site drainage area outlet.

Study Area

Located in the Guadalupe River basin, the Geronimo Creek watershed is located across Comal and Guadalupe Counties in south central Texas (Figure 1). The Geronimo Creek watershed consists of Geronimo Creek and its tributary, Alligator Creek. Alligator Creek is an intermittent stream that typically only has flow after a rainfall event. Geronimo Creek is a tributary of the Guadalupe River, which is used for recreation by local residents and tourists. The watershed is 17,868 hectares (44,152 acres) and is primarily agricultural with some urban area near the towns of Seguin and New Braunfels (Dictson 2009).

Geronimo Creek was chosen as the study site because it was listed as a bacterially impaired waterbody on the 2008 303(d) list (TCEQ 2008). A WPP for Geronimo Creek was also being developed by the Texas AgriLife Extension Service – Department of Soil and Crop Sciences through a Texas State Soil and Water Conservation Board project with the Clean Water Act 319(h) Non Point Source Grant Program. The SELECT model was used to assess the potential *E. coli* loads to develop the load allocation portion of the WPP.

AWPP project collects crucial data, such as potential sources, population densities of animals, and the areas or land uses where potential sourcescould be present, from a local stakeholder group consisting of affected owners and citizens.

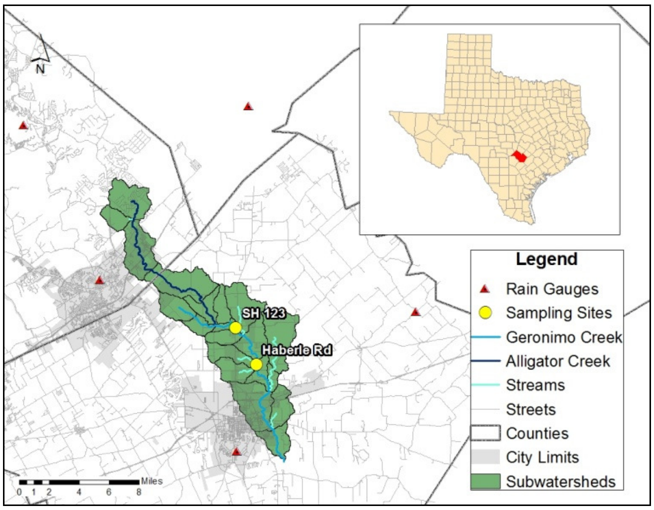

Figure 1 Geronimo Creek watershed study area with NCDC rain gauges and water quality sampling sites.

Stream Flow and *E. coli* Data

Historical and routine stream flow and *E. coli* concentration sampling data from 1996 to 2010 were obtained from the Guadalupe Brazos River Authority (GBRA). Personnel from GBRA collected surface water grab samples in sterile plastic bottles. The bottles were kept on ice and brought to the central water testing laboratory located in New Braunfels, TX. *E. coli* in the water samples were enumerated using EPA 1603 method (USEPA 2002) within four hours of sample collection.The SH 123 and Haberle Road sampling sites are both historical sites (beginning in 1996) while sampling atthe other 13 samplings sites in the watershed began in September 2008. Haberle Road samples were taken on a monthly basis beginning in September 2003 and ending in December 2010 resulting in 84 samples. 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 precipitation-induced runoff. Only 12 data points out of the 84 for the Haberle Road site samples were taken when precipitation-induced runoff occurred.

MATERIAL AND METHODS

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

$$C = (Y\,D)/(u\,R\,A) \quad (1)$$

Where $C$is the concentration of *E. coli* at sampling site (colony forming units [CFU]/mL), $Y$is the daily loading rate of *E. coli* at the sampling site (CFU), $D$is the delivery factor (dimensionless), $u$is the unit conversion factor (2.54×10^4) used to convert from (inm^2) to (mL), $R$is the daily runoff at the sampling site (in), and $A$is the grid cell area (m^2).

Equation 1 was originally intended for livestock facilities but was applied to multiple non-point sources in this study using SELECT and ArcGIS 9.X (McElroy et al. 1976). The details for obtaining each of the variables in Equation 1 are presented below.

Runoff (*R*)

Daily precipitation data wereobtained at 5 sites:Canyon Dam, Kingsbury, New Braunfels, San Marcos, and Seguin, from the National Climatic Data Center (NCDC) for 1996 to 2010. Selection of gauges was based on long-term availability and consistency of the rainfall data. The NCDC rain gauges shown in Figure 1 were utilized to develop a daily precipitation grid using inverse distance weighted (IDW) interpolation over the entire watershed area. A grid cell size of 30 meters was selected to be consistent with the spatial resolution inthe SELECT model.

Rainfall Event Selection

The minimum rainfall needed to produce runoff was estimatedusing the NRCScurve number approach.First, the maximum soil water retention parameter (S) in inches was calculated as (Haan et al. 1994):

$$S=1000/CN - 10 \quad (2)$$

Where$CN$is the area weighted curve number for the Geronimo Creek watershed. In the NRCScurve number method rainfall must exceed$0.2\ S$before runoff will be produced. This threshhold rainfall amount was used to determine the days whenrunoff-producing precipitation and a sampling event occurred simultaneously in the watershed. The model was only run on these days to prevent overestimation of runoff and therefore *E.coli*. A customclassification of land use (Figure 2) produced using 2008 National Agriculture Imagery Program (NAIP) imagery (USDA-FSA-APFO 2008) and a prior Texas Parks and Wildlife (TPWD) classification was provided by the Texas A&M University Spatial Sciences Laboratory (SSL). The watershed curve number grid (Figure 3) was developed in ArcGIS 9.3. byoverlaying the Soil Survey Geographic (SSURGO) hydrologic soil group (USDA-NRCS 2004b) with the land use type and using an NRCS curve number lookup table (Soil Conservation Service 1986).

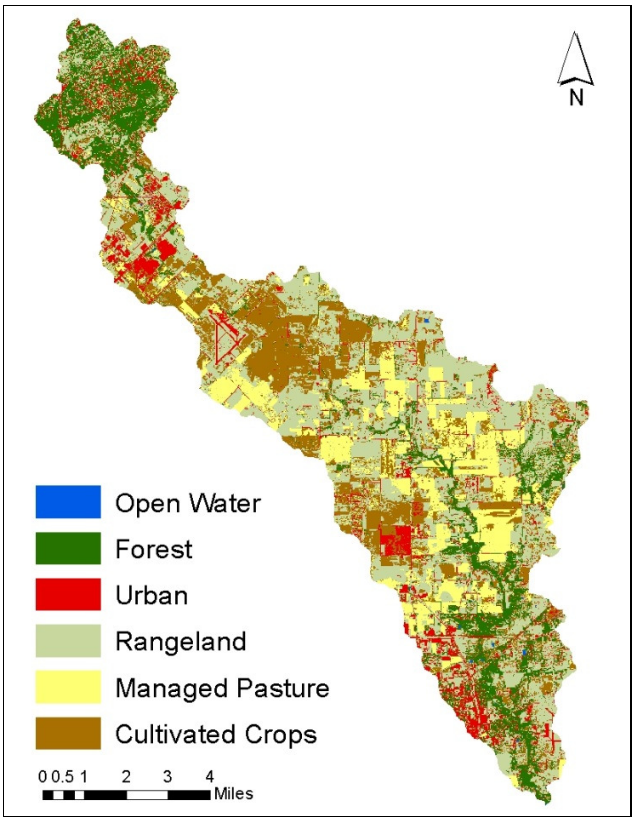

Figure 2 Geronimo Creek watershed land use classification.

The curve numbers (Table 1; Figure 3) used in the NRCS lookup table (Soil Conservation Service 1986) were determined based on the assumption of antecedent moisture condition (AMC) II. The area-weighted curve number for the Geronimo Creek Watershed was 82. The minimum rainfallneeded to produce runoff calculated using the area weighted curve number was 11 mm (0.44 in).

If the daily precipitation in one of the five rain gauges exceeded the minimum rainfall to induce runoff (0.2 S) on a day that coincided with a routine *E. coli* sampling event, a precipitation grid was generated using the ArcGIS Spatial Analysis Extension.The inverse-distance weighting (IDW)technique was used to interpolate rainfall depths across the watershed.IDW assumes that observations closer to one another are more alike than ones further apart (Zhang andSrinivasan 2009).

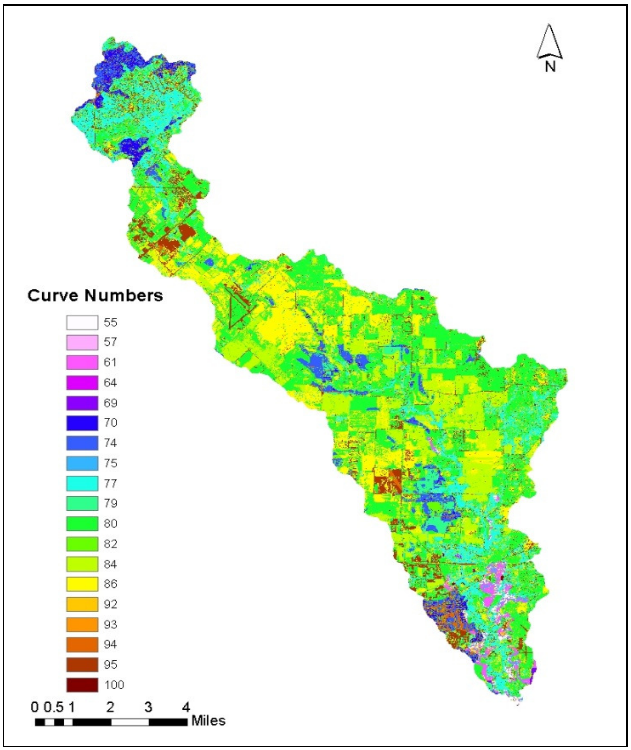

Figure 3 Geronimo Creek curve number grid**.**

Runoff Calculation

The runoff volume at a sampling site was calculated from the precipitation grid (Figure 4). An automated tool was programmed into ArcGIS to calculate a runoff grid with the input of 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 volume grid (Figure 4(a-e)) was calculated using the NRCS curve number approach with the equation(USDA-NRCS 2004a):

$$Q = [(P - 0.2S)^2/(P + 0.8S)]\, A \qquad (3)$$

where$Q$is the runoff volume (inm^2), $P$is precipitation (in), $S$is the maximum soil water retention parameter (in), and $A$is the area of a grid cell (m^2).The curve number grid is convertedinto an S grid using Equation 2. Equation 3 requires that P must exceed 0.2S before any runoff is generated; therefore, when cells in the 0.2 Sgrid (Figure 4(a)) contained negative values, they were changed to zero. (Figure 4(b)). The values in the grid of Figure 4(b) were squared, creating the numerator of Equation 3(Figure 4(c)). The grid for the denominator of Equation 3 was created by adding 0.8S to P in each cell (Figure 4(d)). The numerator (Figure 4(c)) was then divided by the denominator (Figure 4(d)), which produced the runoff grid for the entire watershed. Runoff depth was then converted to a runoff volume per grid cell by multiplying by the cell area, 900 m^2 (Figure 4(e)).

Table 1 NRCS curve number lookup table.

Land Use Type	Hydrologic Soil Group	Curve Number
Open Water	A	100
	B	100
	C	100
	D	100
Forest	A	25
	B	55
	C	70
	D	77
Urban	A	89
	B	92
	C	94
	D	95
Rangeland	A	39
	B	61
	C	74
	D	80
Managed Pasture	A	49
	B	69
	C	79
	D	84
Cultivated Crops	A	65
	B	75
	C	82
	D	86

Next the automated tool calculates the runoff volume accumulation grid for the watershed shown in Figure 4(f-g). The inputs to the tool were the runoff volume grid and a 30 × 30 meter Digital Elevation Model (DEM) (USGS 2009).The result of the flow accumulation is the total amount of runoff volume going through a specific grid cell. The runoff volume fromeach sampling site was estimated as the runoff volume (in)at the outlet of the drainage area.

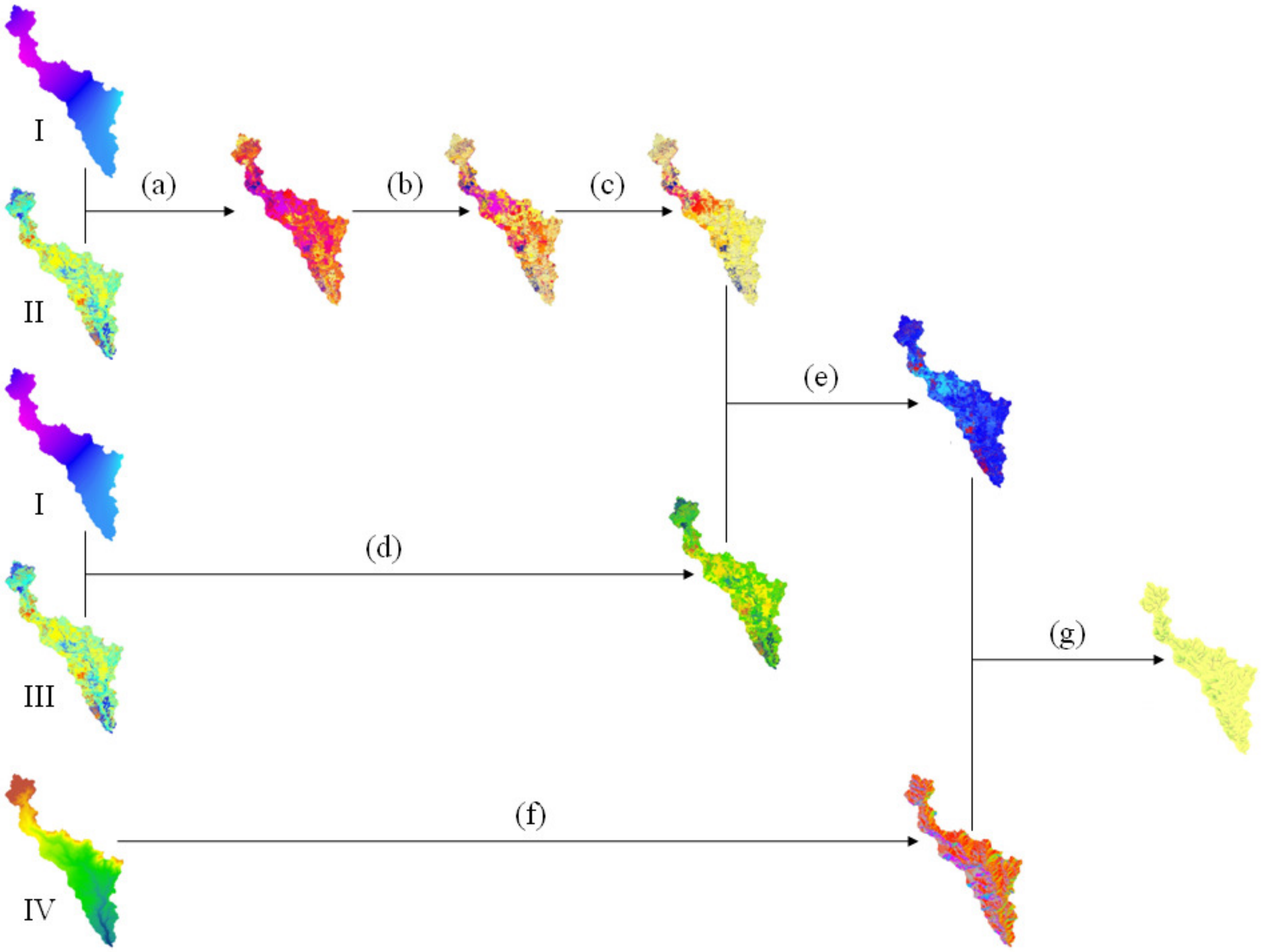

Figure 4 Flow chart illustrating the calculation of accumulated runoff volume. I – rainfall grid, II – 0.2S grid, III – 0.8S grid, IV –DEM,(a) Subtract 0.2*S* from rainfall grid, (b) Convert negative values to zero, (c) Square corrected grid, (d) Add 0.8*S* to rainfall grid, (e) Divide grid created from (c) by grid created from (d) then multiply result by 900 m^2 to create a runoff volume grid, (f) Compute flow direction from DEM grid, (g) Compute flow accumulation from flow direction grid using the runoff volume grid as the accumulation weight.

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

Potential *E. coli* loads for Geronimo Creek were predicted using SELECT and stakeholder input for stocking rates and potentialsources. A custom land use classification (Figure 2) provided by the Texas A&M University SSL was used to distribute sourceson land use types that were determined to be suitable for a specific source.

Twenty-one subwatersheds were delineated using the SWAT model. The stream channel was determined with the SWAT model using the DEM. The fecal production rates used in the model,based on EPA guidance, are offecal coliform (USEPA 2001a). These rates were converted from fecal coliform to *E. coli* using aratioof 0.63 fecal coliform to *E. coli*based on USEPA's regulatory standardsin recreational waters. The regulatory standard for fecal coliform is200 organisms per 100 mL and is 126 organisms per 100 mL for *E. coli* (USEPA 2003).

In the Geronimo Creek watershed, relevant livestock sources were goats, horses, and cattle. Wildlife sources were deer and feral hogs. Domestic sources were dogs and on-site wastewater treatment systems (OWTS).For livestock and wildlife, the number of animals wasestimated usinganimal densities and stakeholder input. The animals were distributed evenly across suitable habitats and a fecal production rate was then applied per animal. For instance, stakeholders recommendedstocking rates forcattle of 8 and 4 hectares (20 and 10 acres) per animal should be applied to Comal and Guadalupe Counties, respectively, and should be applied to the land use types of rangeland, forest, and managed pasture.Thedensity ofhorses was estimatedto be 53 hectares

(132 acres) per animal with a population of 124 horses distributed overrangeland areas. Since goats were typically raised on goat farms, stakeholders estimatedthat 200 goats of the total watershed population of 750 animals should be distributed evenly in the watershed on rangeland, forest, and managed pasture. The remaining animals were concentrated to specific watersheds thatcontained known goat farms witha specificnumber of animals.

White-tailed deer had a population density of 4 hectares (10 acres) per animal (Lockwood 2005). The suitable habitat determined for deer were forest and rangeland with at least 8 hectares (20 acres) of contiguous terrain available. Feral hogs had a population density of 11 hectares (26 acres) per animal and were only distributed on suitable habitat (forest, rangeland, managed pasture, and cultivated crops) 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.

Household data from the 2000 US Census (USCB 2000) was used to calculate a dog density of 1 dog per household. The potential *E. coli* load for OWTSs was calculated by Espey Consultants (Ling and McFarland 2011) using spatially distributed point data of each household collected from 911 address data. Households within Certificate of Convenience and Necessity (CCN) areas, which were on city sewer lines, were removed to elimate households being serviced by a wastewater treatment facility (WWTF). A failure rate for the OWTS was determined using SSURGO soil limitation classes and the age of the treatment system to calculate the percentage of *E. coli* contributed to the watershed due to septic failure. A fecal production rate was applied to each household for dogs and OWTS.

SELECT divided the watershed into a raster grid with a 30 × 30 meter cell size thus the potential load was calculated over the entire watershed at this resolution. The individual raster files for each source were added together spatially to create a total load raster file (Figure 5) for the watershed.The total load raster file (Figure 5) estimated the potential *E. coli* load for the entire watershed based on a worst-case scenario and assumed the entire load calculated reaches the water body.

Another part of the automated tool programmed in ArcGIS 9.3. was to calculate the *E. coli* load actually reaching a specific grid cell in the watershed (Figure 6).The inputs to the tool were the runoff grid, the total load raster from SELECT, and the DEM. The first step to estimate the *E. coli* load reaching the sampling site was to only consider the *E. coli* load grid cells that had runoff generated. Therunoff volume grid generated from each precipitation event was converted to a Boolean runoff grid, where the grid cells with values greater than 0 precipitation were converted to 1 (Figure 6a). A SELECT runoff grid was estimated for each runoff event by multiplying the SELECT grid by the Boolean runoff grid, so that the cells with no runoff had a contributing *E. coli* load of zero (Figure 6b). The load accumulation was calculated using the SELECT runoff grid as an input weight and the DEM shown in Figure 6(c-d). The output of flow accumulation would then represent the amount of *E. coli* load that would flow through each cell considering the upslope cells (Figure 6d). The flowaccumulation at a sampling site would then estimate the *E. coli* load reaching that site.

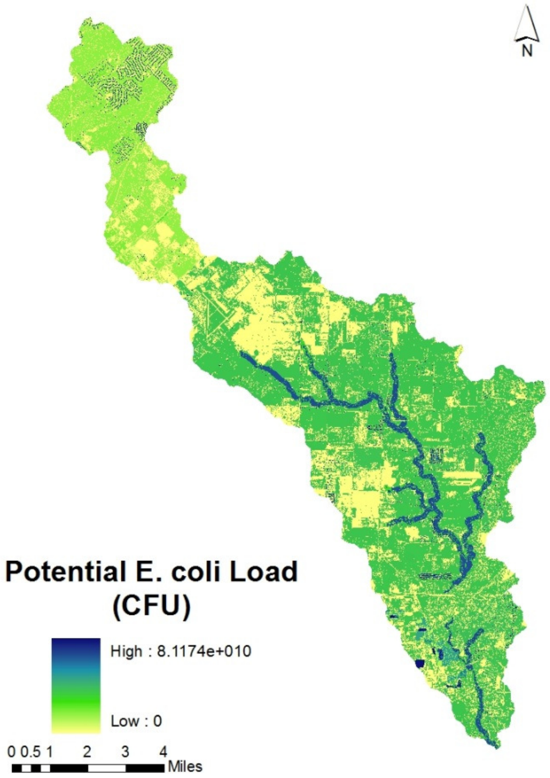

Figure 5 Total potential *E. coli* load calculated using SELECT for the Geronimo Creek watershed.

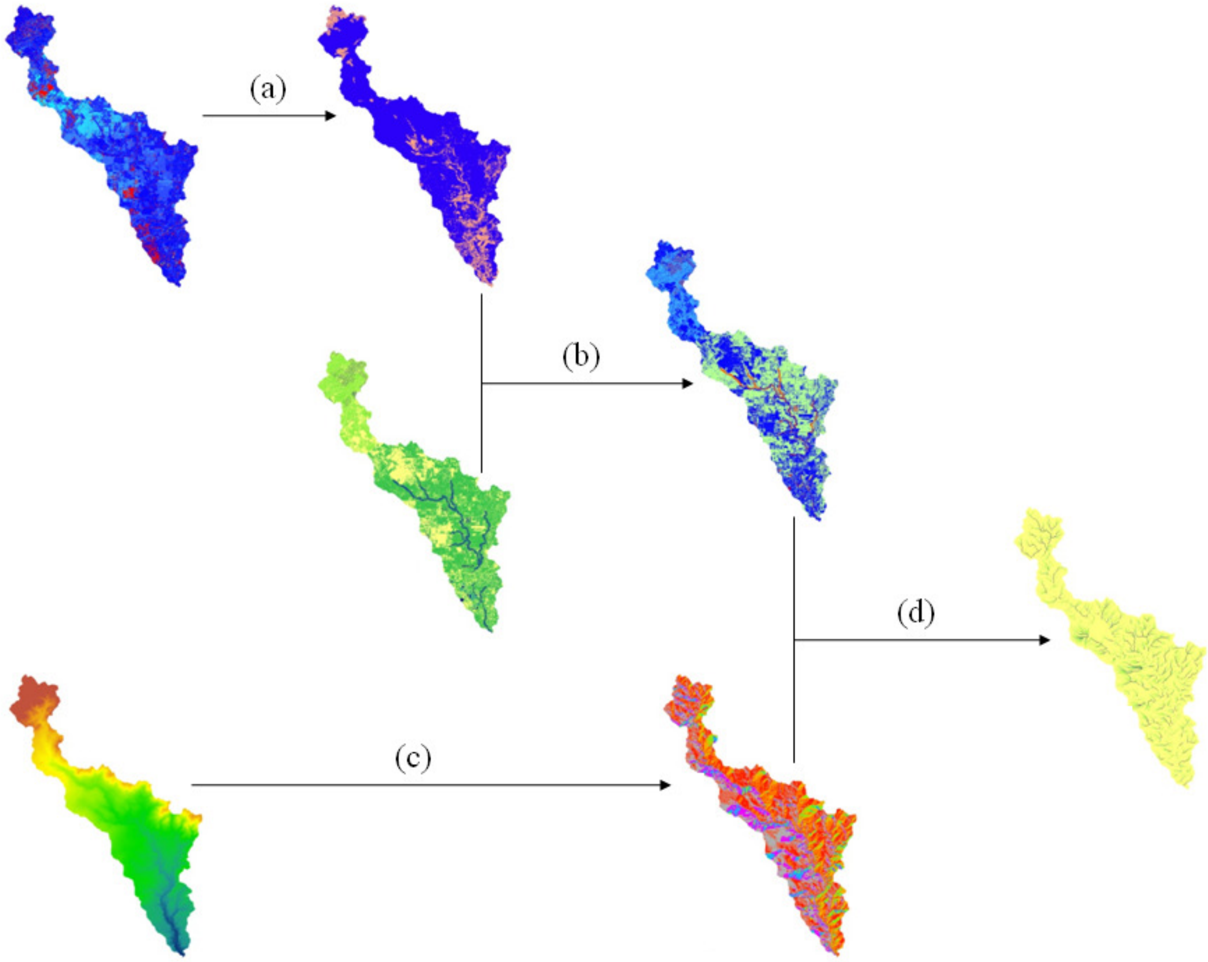

Figure 6 Flow chart illustrating the calculation of the contributing *E. coli* load. (a) Convert runoff values greater than zero to one, (b) Multiply converted runoff and SELECT load to compute contributing load, (c) Compute flow direction, (d) Compute flow accumulation using flow direction with contributing load as accumulation weight

Calculation of Observed and Predicted Runoff Volumes

Two methods for calculating flow duration curves (FDCs) were used at eachsampling site, Haberle Road and SH 123, one method using instantaneous monthly samples and the other using SWAT simulated daily flow rates. SWAT-simulated FDCswere developed using simulated daily flowrates from 1998 to 2009. The observed FDCs were developed using instantaneous observed flowrates from sampled 1996 to 2003. The SH123 dataset had a break in sampling from 2008 to 2009 The Haberle Road dataset had a break in sampling from 2003 to 2009.

To account for the volume of water present in the stream before the runoff event, three methods were used to add base flow to the predicted runoff volume calculated from the FDCdeveloped using instantaneous flows: (1) adding the 75% exceedence flow calculated from the FDC developed using instantaneous flows, (2) adding the 75% exceedence flow calculated from the FDC using SWAT simulated flows, and (3) adding the maximum instaneous flow or 100% exceedence flow developed using instantaneous flows. The 75% exceedence flow using instantaneous flows was 0.06 m^3/s (2.2 cfs)for SH 123 and 0.23 m^3/s (8.1 cfs) for Haberle Road. The 75% exceedence flow using SWAT simulated flows was0.32 m^3/s (11.35 cfs) and 0.40 m^3/s (14.29 cfs) for SH 123 and Haberle Road, respectively. The 100% exceedence flow using instantaneous flow rates for SH 123 and Haberle Road were 0.028 m^3/s (1.0 cfs) and 0.096 m^3/s (3.4 cfs), respectively.Observed and stream flows and base flows (cfs)were converted to a runoff volume (m^3) by multiplying by the watershed response time using a combination of methods (Table 2): SWAT calculated time of concentration (t_C), SCS method lag time (t_L) calculated using the t_C, SCS method calculated $t_C$and SCS method calculated t_L using SCS method calculated t_c. SWAT calculatedt_C for each hydrological response unit (HRU). All HRUs in a subbasin were averaged to obtain subbasint_C. Thet_Cs for the subbasins containing the main channels of Geronimo and Alligator Creeks were added together starting at the upper portion of the watershed and ending ateither the SH 123 or Haberle Road sampling sites. The SWAT calculated t_C was converted to a t_L by using a method developed by the SCS (Haan et al. 1994):

$$t_L = 0.6t_C \qquad (4)$$

where$t_L$is the lag time (hrs) and $t_C$is the time of concentration (hrs).

The SCS t_L equation based on natural watersheds (Haan et al. 1994) is:

$$t_L = (L^{0.8}\,(S+1)^{0.7})/(1900Y^{0.5}) \qquad (5)$$

where$L$is the hydraulic length of the sampling site drainage area (ft), $S$is the average maximum soil water retention parameter (in), and $Y$is the average land slope of the sampling site drainage area (%). The hydraulic length of the SH 123 subbasin was determined by measuring the longest distance along the SWAT delineated stream channel to the drainage area outlet. The stream length 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 was located downstream of SH 123, there was a log jam located at SH 123 which may haveinhibited flow from upstream of SH 123 to

Haberle Road. Therefore, two hydraulic lengths were used forHaberle Road, one for the entire upstream portion from Alligator Creek, and one from SH 123.

Table 2 Times (lag and concentration) used to estimate runoff volumes in the Geronimo Creek watershed.

Sampling Site	Method	Time (hrs)
SH 123	SCS t_L from SCS t_c starting from Alligator Creek	7.2
	SCS t_c starting from Alligator Creek	12.0
	SWAT t_c starting from Alligator Creek	6.9
Haberle Road	SCS t_L from SCS t_C starting from Alligator Creek	9.2
	SCS t_c starting from Alligator Creek	7.8
	SCS t_L from SCS t_C starting from SH 123	2.9
	SCS $t_c$from SH 123	0.97
	SCS t_L from SWAT t_c from SH 123	0.58

Delivery Factor (*D*)

The delivery factor was calculated from Equation 1 using observed *E. coli* concentration data provided by the Guadalupe Blanco River Authority (GBRA) taken at both the SH 123 and Haberle Road sampling sites. The delivery factor represents all factors influencing movement of the potential *E. coli* load into the waterbodywith the exception of surface runoff. Two delivery factors were calculated, one using the observed runoff volume, and the other from runoff predicted by the the ArcGIS tool, developed in this study.

The mean of the delivery factors calculated for each rainfall event was determined using both anarithmetic mean and a geometric mean to obtain a site specific delivery ratio. This resulted in the calculation of eight different delivery factors (Table 3), four for each watershed.

Table 3 Delivery factors used for *E. coli* concentration calculation.

Sampling Site	Average	Observed Runoff Volume	Predicted Runoff Volume
SH 123	Mean	0.752	0.942
	Geomean	0.015	0.015
Haberle Road	Mean	0.480	0.316
	Geomean	0.065	0.059

Outlier Testing

The model assumes that runoff transporting the *E. coli* to the stream was generated by rainfall occurring on the day the sample was taken. An ideal sample would be one that was taken on a day where the preceding day had either no rainfall or not enough rainfall to induce runoff. Eleven out of the 17samples fit these criteria and were considered ideal samples. Five of the remaining samples had runoff generated from one or more subbasins the day preceding sampling, but the cumulative rainfall of all subbasins was less for thepreceding day than the day sampled. Of these five, there were two cases where runoff was generated from multiplesubbasins, but the cumulative rainfall for the day the sample was taken was at least 4.5 times greater than rainfall.. It was decided that these 16 samples would not violate the model assumptions because a majority of the rainfall impacting runoff occurred on the day of sampling. One sampling day, March 3, 2005, at the Haberle Road site did notmeet these criteria. The cumulative rainfall on all subbasins for the preceding day was higher (55 mm, 2.01 in) than the day the sample was taken (29 mm, 1.13 inches). Precipitation grids were computed for both days, and the grid generated from the precipitation onMarch 3, 2005 had lower minimum, maximum, and mean precipitation values for grid cells over the entire watershed than the precipitation grid generated for March 2, 2005.

The Dixon-Thompson test was applied to test thehypothesis that the March 3, 2005 sampling point was an outlier. The Dixon-Thompson test is suitable for sample sizes as small as three and can test for both lowest and highest outliers (McCuen 2003). However, the March 3, 2005 data point was the second largest observed runoff volume, so the sample with the highest runoff volume was not included in this test,reducing the sample size to 11. The equation for the Dixon-Thompson High Outlier Test Statistic for sample sizes of 11 is (McCuen 2003):

$$R_{DT} = (X_{11} - X_9)/(X_{11} - X_2) \qquad (9)$$

where$X_n$are the data are ranked from smallest to largest and the subscript indicates the rank of the value from smallest to largest.

Assuming a normal population, the test statistic (R_{DT} = 0.801; Equation 9) was larger than all of the critical values (R_C) at 5% (0.570), 2.5% (0.617), and 1% (0.670) for the sample size of 11. Therefore, the March 3, 2005runoff volume was considered an outlier by the Dixon-Thompson test. The point March 3, 2005 was removed from the Haberle Road data set. The largest runoff volume was not removed from the dataset because there was not a substantial amount of rainfall the day before the sample was taken as there was with the March 3, 2005 sampling day.

Statistics

The accuracy of the model was evaluated using the Nash-Sutcliffe efficiency (E_{NS}), root mean square error (*RMSE*), and RMSE-observations standard deviation ratio (*RSR*). Legates and McCabe (1999) recommend including at least one relative error measure (E_{NS} or R^2) and at least one absolute error measure (*RMSE* or mean absolute error) for a complete assessment of model performance.According to Nash and Sutcliffe (1970) the E_{NS} value is an index of agreement or disagreement between observed and predicted values. E_{NS} is computed as (Nash and Suttcliffe 1970):

$$E_{NS} = 1 - \left[\sum_{i=1}^{n}(O_i - P_i)^2 / \sum_{i=1}^{n}(O_i - \bar{O})^2\right] \qquad (6)$$

where$O_i$is an observed value, $P_i$is the corresponding predicted values, $\bar{O}$ is the mean of the observed values, and $n$is the number of data points.

The E_{NS} value ranges from -∞ 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_{NS} = 0.75 to 1), good (E_{NS}= 0.5 to 0.74), fair (E_{NS} = 0.25 to 0.49), poor (E_{NS} = 0 to 0.24) and unsatisfactory (E_{NS}< 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). *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:

$$RMSE = \sqrt{\sum_{i=1}^{n}(O_i - P_i)^2 / n} \qquad (7)$$

RSR is a model evaluation statistic that standardizes *RMSE*with the observed data standard deviation (Moriasi, et al. 2007):

$$RSR = \left[\sqrt{\sum_{i=1}^{n}(O_i - P_i)^2}\right] / \left[\sqrt{\sum_{i=1}^{n}(O_i - \bar{O})^2}\right] \qquad (8)$$

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 efficiences are classified by Moriasi et al. (2007) as very good (*RSR* = 0.00 to 0.50), good (*RSR* = 0.51 to 0.60), satisfactory (0.61 to 0.70), and unsatifactory (*RSR*> 0.70).

Moriasi et al. (2007) states that the model evaluation guidelines for both E_{NS} and *RSR* values 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, evalution time step, model calibration procedure, and project scope and magnitude (Moriasi et al. 2007).

RESULTS AND DISCUSSION

The runoff volumes were simulated for both the SH 123 and Haberle Road sampling sites and compared to the instantaneous observed flows converted to a volume. For both SH 123 and Haberle Road, *E. coli* concentrations were simulated using Equation 1 with both predicted and instantaneous observed runoff volumes.

Runoff Volume

SH 123

The model predicted runoff volume at the SH 123 sampling site with accuracies ranging from very good to unsatisfactory (Table 4). Adding baseflow using the 75% exceedence flow from the SWAT generated FDC resulted in an unsatisfactory agreement for both the $E_{NS}$and *RSR* values no matter which method was used to determine the travel time. Baseflow added using the 75% exceedence flow from the FDC developed using instantaneous flows resulted in a very good agreement for both $E_{NS}$and *RSR* valuesfor all travel time methods. The $E_{NS}$and *RSR* values calculated from volumes calculated by adding the 100% exceedence flow from an FDC developed using instantaneous flows resulted in satisfactory to very good agreement. Overall the best fit was found when using the lag time based on the SCS method time of concentration and baseflows calculated as the 75% exceedence on the FDC developed from instaneous flows.

Runoff volumes at the SH 123 site were underestimated for three events and overestimated for two events. Runoff volumes measured at SH 123 were primarily from the fall and winter seasons with only one sample collected in the spring. Therunoff volumes for SH 123 followed the general trend of the observed runoff volumes and were in close agreement for the methods using instantaneous flow, both at 75% exceedence and 100% exceedence.

Table 4 Model performances for predicting runoff volume at SH 123, n= 5, using 3 methods for determining watershed travel time and 3methods for determining baseflow.

		Baseflow Methods		
		75% Exceedence		**100% Exceedence**
Travel Time	**Statistic**	**SWAT**	**Instantaneous**	**Instantaneous**
SCS t_L and SCS t_c calculated from Alligator Creek	E_{NS}	-1.83	0.95	0.83
	RSR	1.68	0.22	0.41
	RMSE (m^3)	6217	813	1506
	Observed Mean (m^3)	4496	4496	4496
	Observed Std. Dev. (m^3)	4128	4128	4128
SCS t_c calculated from Alligator Creek	E_{NS}	-1.11	0.78	0.63
	RSR	1.45	0.47	0.61
	RMSE(m^3)	8931	2902	4488
	Observed Mean (m^3)	7495	7495	7495
	Observed Std. Dev. (m^3)	6882	6882	6882
SWAT t_c calculated from Alligator Creek	E_{NS}	-1.95	0.94	0.84
	RSR	1.72	0.25	0.40
	RMSE (m^3)	6072	872	1405
	Observed Mean (m^3)	4307	4307	4307
	Observed Std. Dev. (m^3)	3955	3955	3955

Haberle Road

The model predicted runoff volumes at the Haberle Road sampling site with accuracies ranging from very good to unsatisfactory (Table 5). The travel times originating from Alligator Creek calculated using both the SCS and SWAT methods had poor to unsatisfactory agreement for $E_{NS}$and *RSR* values This implies that there was something happening in the watershed to prevent stream flow from above SH 123 to reach Haberle Road. A log jam that occurred during the simulation period at SH 123 may have caused poolingat that site and prohibitedrunofffrom travelingdownstream. Accuracies improved significantly when travel times were calculated beginning from SH123.

The most accurate model for Haberle Road occurred when the lag time based on the SWAT t_C calculated from SH 123was used regardless of baseflow method selected. As with the SH 123 model, the overall best performance occurred when the 75% exceedence flow was from the FDC created using instantaneous flows was used to estimate baseflow.

Simulated runoff volumes were underestimated at Haberle Road with the exception of one eventduring the driest month (August).All other samples 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 2009, an extremely dry year in whichno samples collected had contributing runoff occurring at the same time.

The runoff volumes estimated with the method that performed the best were used to calculate *E. coli* concentrations.

Table 5 Model performance for estimating runoff volume at Haberle Road n = 11, using 5 methods for determining watershed travel time and 3methods for determining baseflow.

Time	**Statistic**	**75% Exceedence**		**100% Exceedence**
		SWAT	**Instantaneous**	**Instantaneous**
SCS t_L calculated from Alligator Creek	E_{NS}	0.09	-0.05	-0.25
	RSR	0.95	1.03	1.12
	RMSE(m^3)	21473	23145	25205
	Observed Mean (m^3)	17368	17368	17368
	Observed Std. Dev. (m^3)	23669	23669	23669
SWAT tc calculated from Alligator Creek	E_{NS}	0.12	-0.03	-0.22
	RSR	0.94	1.01	1.11
	RMSE (m^3)	18184	19613	21383
	Observed Mean (m^3)	14881	14881	14881
	Observed Std. Dev. (m^3)	20279	20279	20279
SCS tL calculated from SH 123	E_{NS}	0.35	0.23	0.05
	RSR	0.81	0.88	0.97
	RMSE (m^3)	5685	6188	6857
	Observed Mean (m^3)	5423	5423	5423
	Observed Std. Dev. (m^3)	7390	7390	7390
SWAT tc calculated from SH 123	E_{NS}	0.82	0.77	0.66
	RSR	0.42	0.47	0.59
	RMSE (m^3)	1011	1133	1401
	Observed Mean (m^3)	1837	1837	1837
	Observed Std. Dev. (m^3)	2504	2504	2504
SWAT tc SCS tL calculated from SH 123	E_{NS}	0.92	0.95	0.89
	RSR	0.28	0.21	0.33
	RMSE (m^3)	401	317	476
	Observed Mean (m^3)	1100	1100	1100
	Observed Std. Dev. (m^3)	1499	1499	1499

Comparison to Other Studies

Performance statistics for event-based runoff simulation, after base flow correction, by the spatial watershed model developed from this study (Tables 4 and 5)were comparable to those from studies using continuous process models, such as SWAT and HSPF for streamflow prediction. Coffey et al. (2010) were able to validate daily flows using a calibrated SWAT model from January 2004 to February 2005 with very good accuracy (E_{NS}= 0.78) in Irish catchments. Parajuli et al. (2009) were able to calibrate the SWAT model in the Upper Wakarusa watershed for mean daily flow of a subwatershed with very good agreement (E_{NS} = 0.83) and validate in two subwatershedswithin the calibrated subwatershed with very good agreement (E_{NS} = 0.83 and E_{NS}= 0.76). SWAT was also used in simulation of a watershed with multiple karst features such as multiple springs, sinkholes, and losing streams (Baffaut and Benson 2009), validating daily stream flow values from 2001 to 2007 with E_{NS}values ranging from 0.24 to 0.56 for five stations. Chin et al. (2009) predicted daily and monthly averaged flow for an experimental watershed from 1996 to 2002 using both SWAT and HSPF. SWAT was able to predict monthly and daily flows with very good (E_{NS} = 0.88) for monthly flows to good (E_{NS} = 0.65) agreement for dailyflows. HSPF predicted both monthly and daily flows with very good agreement (E_{NS}= 0.89 for monthly and E_{NS}= 0.87 for daily). Paul et al. (2004) simulated daily mean flow using HSPF in the Salado Creek watershed for a calibration period from 1991 to 1993 with good agreement (E_{NS} = 0.55). These studies were chosen because they also predicted *E. coli* or fecal coliform bacteria that generally had very good to good agreement using the E_{NS}values.

E. coli Concentrations

Based on the E_{NS} and the *RSR*, the model predicted *E. coli* concentrations for all four methods of delivery factor calibration with unsatisfactory agreement at both the Haberle Road and SH 123 sampling sites (Table 6 and Table 7). The *RMSE* values at both sites using all four methods were higher than both the observed standard deviations and observed meansanother indication of unsatisfactory agreement between the observed and predicted *E. coli* concentrations(Table 6).

Table 6 Model performance for *E. coli* concentrations at SH 123.

Delivery Factor Calculation	Statistic	Simulated Delivery Factor		Observed Delivery Factor	
		Geomean	Average	Geomean	Average
Observed Flow	E_{NS}	-90	-441687	-44	-281054
	RSR	10	665	7	530
	RMSE (m^3)	8	526	5	419
	Observed Mean (m^3)	1.8	1.8	1.8	1.8
	Observed Std. Dev. (m^3)	0.9	0.9	0.9	0.9
Simulated Flow	E_{NS}	-142	-656256	-71	-417698
	RSR	12	810	8	646
	RMSE (m^3)	9	641	7	511
	Observed Mean (m^3)	1.8	1.8	1.8	1.8
	Observed Std. Dev. (m^3)	0.9	0.9	0.9	0.9

Table 7 Model performance for estimating *E. coli* concentrations at Haberle Road.

Delivery Factor Calculation	Statistic	Simulated Delivery Factor		Observed Delivery Factor	
		Geomean	Average	Geomean	Average
Observed Flow	E_{NS}	-56	-1641	-67	-3800
	RSR	8	41	8	62
	RMSE(m^3)	121	653	133	994
	Observed Mean (m^3)	8	8	8	8
	Observed Std. Dev. (m^3)	17	17	17	17
Simulated Flow	E_{NS}	-4	-154	-5	-362
	RSR	2	12	2	19
	RMSE (m^3)	36	200	40	307
	Observed Mean (m^3)	8	8	8	8
	Observed Std. Dev. (m^3)	17	17	17	17

The delivery factor estimated from the geometric mean of all of the delivery factors back calculated using simulated runoff volumes performed the best for Haberle Road while the geometric mean of the delivery factors back calculated using observed runoff volume performed the best for SH 123.

The Haberle Road site consistently performed better than the SH 123 site.

The observed *E. coli* concentrations had values ranging from 1.12 to 3.2 CFU/mL for the SH 123 sampling site. The method of predicting *E. coli* concentrations that had the closest range of concentrations (0.01 to 9.78 CFU/mL) to the observed concentration range was estimated using a the geometric mean the delivery factors back calculated using the observed runoff volumes and with the concentration calculated using the observed runoff volumes. *E. coli* concentrations predicted using the delivery factor derived from the observed runoff volumes produced better results than the concentrations predicted using the delivery factor back calculated from the simulated runoff volumesPredictions of*E. coli* concentrations at the SH 123 sampling site were poorer than predictionsat the Haberle Road sampling site. One explanation may be the additional data available for calibration at the Haberle Road site.

For Haberle Road the delivery factors back calculated using the simulated runoff volumes performed better than the delivery factors computed using the observed runoff volumes. Concentrations calculated using the delivery factor derived using simulated runoff volumesperformed better than concentrations calculated using the delivery factor back calculated using observed runoff volumes. The range for the observed *E. coli* concentrations was from 0.46 to 57 CFU/mL. The *E. coli* concentrations predicted using the geometric mean of delivery factorscalculated with simulated runoff volumes and concentrations calculated using the simulated runoff volumes had the closest range from 0.17 to 96 CFU/mL of predicted concentrations to the observed concentrations.

Comparison to Other Studies

SWAT and HSPF have been used to predict *E. coli*concentrations in watersheds with mixed success. Coffey et al. (2010) used SWAT toto predict *E. coli* concentrations for Irish catchments that were compared to monthly grab samples over the period September 2005 to September 2006 resulting in 11 observed samples. The predicted *E. coli* concentrations were in good agreement with the observed concentrations having anE_{NS}value of 0.59. Parajuli et al. (2009) estimated fecal coliform bacteria concentrations using the SWAT model with unsatisfactory to fair agreement for calibration, validation, and verification watersheds. The calibration watershed was in poor agreement with anE_{NS}value of 0.20 and the validation watershed had anE_{NS}value of 0.31 which resulted in a fair agreement. The verification watershed had an unsatisfactory agreement with an E_{NS}value of -2.Baffaut and Benson (2009) used the SWAT model to predict fecal coliform bacteria concentrations calibrated and validated against monthly or bi-monthly grab sample concentrations for the James River Basin, a karst watershed. The SWAT model was calibrated for four different sampling sites resulting inE_{NS}values ranging from -6 to 0.11 (Baffaut and Benson 2009). Chin et al. (2009) predicted fecal coliform bacteria concentrations using both the SWAT and HSPF models for an experimental watershed. SWAT performed better than HSPF with anE_{NS}value of 0.73 compared to anE_{NS}value of 0.33 for HSPF. Paul et al. (2004) did not calibrate the HSPF model due to a lack of observed fecal coliform bacteria data. HSPF was able to simulate in-stream fecal coliform concentrations with good agreement butwas unableto captureextreme concentrations (Paul et al. 2004).

Uncertainty

Uncertainty is an important issue regarding water quality modeling because models are increasingly used to guide decisions regarding water resource policy, management, and regulation (Beck 1987; Sharpley et al. 2002; Harmel et al. 2006; Parajuli et al. 2009). Uncertainty in measured water quality data is introduced during streamflow measurement, sample collection, sample preservation/storage, and laboratory analysis (Harmel et al. 2006). Modeling bacteria transport may have one of the highest probable errors and the least confidence compared with modeling surface hydrology, sediment, and nutrients (Novotny 2003; Parajuli et al. 2009; Coffey et al. 2010). One source of potential uncertainty in these model results is the GIS data input (Parajuli et al. 2009). In this study the best available data as input into the model including stakeholder input for land use and contributing bacteria source animal numbers and distribution in the watershed. The other GIS input, including the DEM, soils, and climate data, were the best available data. Harmel et al. (2006) determined the cumulative probable uncertainty for streamflow data ranging from3-42% for best case to worst case scenerios, though uncertainties in measuring in-stream concentrations and loads for bacteria are likely greater. Uncertainty and variability surrounds bacteria modeling and can lead to large discrepancies in model results (Coffey et al. 2010).

Potential Causes for Inaccurate E. coli Bacteria Modeling

The assumption that the SELECT *E. coli* load is a constant is the most significant reason for the large discrepancy between the observed and simulated *E. coli* concentrations. The SELECT *E. coli* load does not account for buildup of *E. coli* over multiple days because it is a snapshot of one day. The *E. coli* load generated using SELECT is based on the data collected from the stakeholders regarding the densities of the contributing sources as well as the distribution of those sources for 2010. The densities collected, especially regarding livestock stocking rates and livestock distribution on land use types, varied greatly seasonally and from year to year. The land use and household data determining the distribution of sources were also valid for 2010. However, most of the *E.coli*data for the SH 123 sampling site were collected between 1998 and 2002 and between 2004 and 2010 for theHaberle Road sampling site The *E. coli* loads generated using SELECT for 2010 may have been more accurate for the Haberle Road sampling site than SH 123 simply because less change would have occurred between the earliest sampling date (2004) 2010, and the date of the source data determination (2010). One solution to improve the model would be to run SELECT for different years or seasons with land use and contributing source densities varying from year to year or season to season. Another reason for the discrepancies between the observed and predicted *E. coli* concentrations was the model didnot account for point sources contributing *E. coli* directly to the stream. Direct deposition of fecal material by livestock and wildlife was not considered in the model because *E. coli*was considered to be contributing to the stream only through surface runoff. In reality, *E. coli*could still be contributed to the stream during low flow conditions. The *E. coli* concentrations for the Haberle and SH 123 sampling sites ranged from 44 to 330 CFU/100 mL and 0 to 438 CFU/100 mL, respectively for both sites for low flow conditions.

The model also didnot account for bacteria die-off and re-growth occurring in the stream, soil, and in the fecal material itself. During a rainfall event, sediment located in the stream containing bacteria can be stirred up and further contribute to the *E. coli* concentration occuring in the sample. Coffey et al. (2010) elaborated that there are unknown spatial and temporal sources of contamination contributing bacteria and the ability to accurately account for all of these factors is debateable.

The model was unable to accurately predict the *E. coli* concentrations occuring in stream. One potential reason for the model inaccuracies for predicting *E. coli* concentrations is applying the *E. coli* load resulting from SELECT as a constant. This may be overcome by varying the SELECT *E. coli* loads for different years and/or seasons. Direct deposition was not considered in the model although *E. coli*maycontribute into the stream during periods of low flow conditions. Bacteria die-off and re-growth occuring in the stream, soil,and fecal material werenot considered in the model,though such inclusion could potentially increase the accuracy of the model at predicting *E. coli* concentrations. There are other unknown factors that contribute to *E. coli* bacteria contamination in streams, which make predicting *E. coli* concentrations with a model difficult. Although the model did not accurately predict *E. coli* concentrations, it can be modified in multiple ways to increase the model accuracy by varying the *E. coli* concentration yearly and seasonally, accounting for direct deposition, and accounting for die-off and regrowth.

CONCLUSIONS

A watershed model was developed in ArcGIS to estimate the volume of water from runoff and the *E. coli* concentrations contributing at a sampling site. Two sampling sites for the Geronimo Creek watershed were chosen, although there was a lack of observed hydrologic and water quality data coinciding with runoff events. Observed streamflow was converted into a volume by multiplying by the sampling site outlet traveltime. A base flow volume was added to the predicted runoff volume by multiplying base flow by the sampling site outlet travel time. A model calibration using four different methods was applied using a delivery factor for the predicted *E. coli* concentrations.

The model results for the runoff volume were in very good agreement (E_{NS} = 0.95, RSR = 0.21) for the Haberle Road sampling site and in very good agreement (E_{NS} = 0.95, RSR = 0.22) for the SH 123 sampling site. The *RMSE* values were less than half of the standard deviation showing a good agreement between the observed and predicted runoff volumes. The *E. coli* concentration results were in unsatisfactory agreement for both samping sites using all methods. The concentrations calculated with the geometric mean delivery factor performed the best for both sites. The Haberle Road sampling site performed consistently better for all methods than the SH 123 sampling site. More temporal data points are needed to calibrate and validate the model properly to show if it is able to consistently and accurately predict runoff volumes and *E.coli* concentrations.

ACKNOWLEDGEMENTS

Financial assistance was provided to KynaBorelby USGS Water Resources Grant through the Texas Water Resources Institute. This project was funded by Texas State Soil and Water Conservation Board (TSSWCB) through 319(h) non point source management program.

REFERENCES

Baffaut C, Benson VW, 2009. Modeling Flow and Pollutant Transport in A Karst Watershed with SWAT. Transactions of the ASABE 52 (2): 469-479

Beck MB, 1987. Water Quality Modeling: A Review of the Analysis of Uncertainty. Water Resources Research 23 (8): 1393-1442

Benham BL, Baffaut C, Zeckoski RW, Mankin KR, Pachepsky YA, Sadeghi AM, Brannan KM, Soupir ML, Habersack MJ, 2006. Modeling Bacteria Fate and Transport in Watersheds to Support TMDLs. Transactions of the ASABE 49 (4): 987-1002

Chin DA, Sakura LD, Bosch DD, Gay PA, 2009. Watershed-Scalefate and Transport of Bacteria. Transactions of the ASABE 52 (1): 145-154

Coffey R, Cummins E, Bhreathnach N, Flaherty VO, Cormican M, 2010. Development of A Pathogen Transport Model for Irish Catchments Using SWAT. Agricultural Water Management 97:101-111

Dictson N, 2009. Geronimo and Alligator Creeks Watershed Partnership Information Sheet. The Geronimo and Alligator Creeks watershed partnership. http://www.geronimocreek.org/GeronimoInfoSheet.pdf.Accessed 13 August 2010

Fraser RH, Barten PK, Pinney DA, 1998. Predicting Stream Pathogen Loading from Livestock Using A Geographical Information System-Based Delivery Model. Journal of Environmental Quality 27: 935-945

Haan C, Barfield B, Hayes J, 1994. Design Hydrology and Sedimentology for Small Catchments.Academic Press, Inc, San Diego, CA, USA

Harmel RD, Cooper RJ, Slade RM, Haney RL, Arnold JG, 2006. Cumulative Uncertainty in Measured Streamflow and Water Quality Data for Small Watersheds. Transactions of the ASABE 49 (3): 689-701

Legates DR, McCabe GJ, 1999. Evaluating the Use of "Goodness-of-Fit" Measures in Hydrologic and Hydroclimatic Model Validation. Water Resources Research 35 (1): 233-241

Ling W, McFarland M, 2011. Geronimo and Alligator Creeks Watershed Protection Plan. The Geronimo and Alligator Creeks watershed partnership.http://www.geronimocreek.org/GeronimoWPPworkingDraft.pdf. Accessed 12 March 2012

Lockwood M, 2005. White-tailed Deer Population Trends. Texas Parks and Wildlife Department, Austin, TX, USA

McCuen RH, 2003. Modeling Hydrologic Change: Statistical Methods.CRC, Boca Raton, FL, USA

McCuen RH, Knight Z, Cutter AG, 2006. Evaluation of the Nash-Sutcliffe Efficiency Index. Journal of Hydrologic Engineering 11 (6): 597-602

McElroy AD, Chiu SY, Nebgen JW, Aleti A, Bennett FW, 1976. Loading Functions for Assessment of Water Pollution from Nonpoint Sources. US Environmental Protection Agency. EPA-600/2-76-151. USEPA Office of Resources and Development, Washington, DC, USA

Moriasi DN, Arnold JG, Van Liew MW, Bingner RL, Harmel RD, Veith TL, 2007. Model Evaluation Guidelines for Systematic Quantifications of Accuracy in Watershed Simulations. Transactions of the ASABE 50 (3): 885-900

Munafo 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 JE, Sutcliffe IV, 1970. River Flow Forecasting Trhough Conceptual Models Part 1 - A Discussion of Principles. Journal of Hydrology 10: 282-290

Novotny V, 2003. Water Quality, Diffuse Pollution and Watershed Management.John Wiley and Sons Inc, Hoboken, NJ, USA

Parajuli PB, Mankin KR, Barnes PL, 2009. Source Specific Fecal Bacteria Modeling Using Soila and Water Assessment Tool Model. Bioresource Technology 100: 953-963

Paul S, Haan PK, Matlock MD, Mukhtar S, Pillai SD, 2004. Analysis of the HSPF Water Quality Parameter Unvertainty in Predicting Peak In-stream Feacal Coliform Concentrations. Transactions of the ASAE 47 (1): 69-78

Sadeghi AM, Arnold JG, 2002. A SWAT/Microbial Sub-model for Predicting Pathogent Loadings in Surface and Groundwater at Watershed and Basin Scales. Total Maximum Daily Load (TMDL) Environmental Regulations.ASAE, St. Joseph, MI, USA

Sharpley AN, Kleinman PJ, McDowell RW, Gitau M, Bryant RB, 2002. Modeling phosphorus transport in agricultural watersheds: processes and possibilities. Journal of Soil and Water Conservation 57 (6): 425-439

Singh J, Knapp HV, Demissie M, 2004. Hydrologic Modeling of the Iroquois River Watershed Using HSPF and SWAT. ISWS CR 2004-08. Illinois State Water Survey,Champaign, IL, USA

Soil Conservation Service, 1986. Urban Hydrology for Small Watersheds, Technical Release No. 55. Soil Conservation Service, US Department of Agriculture, Washington, DC, USA

TCEQ, 2008. 2008 Texas Water Quality Inventory and 303(d) List. From Texas Commission on Environmental Quality. http://www.tceq.state.tx.us/compliance/monitoring/water/quality/data/08twqi/twqt08.html.Accessed4 August 2010

Teague A, Karthikeyan R, Babbar-Sebens M, Srinivasan R, Persyn RA, 2009. Spatially Explicit Load Enrichment Calculation Tool to Identify Potential *E. coli* Sources in Watersheds. Transactions of the ASABE 52 (4): 1109-1120

USCB, 2000.Census 2000 TIGER / Line_ Files.US CensusBureau, Washington, D.C.http://www.census.gov/geo/www/tiger/tiger2k/tgr2000.html. Accessed March 2010

USDA-FSA-APFO, 2008. National Agriculture Imagery Program (NAIP). United States Department of Agriculture Farm Service Agency Aerial Photography Field Office.http://datagateway.nrcs.usda.gov. Accessed May 2009

USDA-NRCS, 2004a.Part 630 National Engineering Handbook Chapter 10 Estimation of Direct Runoff from Storm Rainfall.United States Department of Agriculture National Resource Conservation Service, Washington, D.C. http://policy.nrcs.usda.gov/OpenNonWebContent.aspx?content=17752.wwb.Accessed 8 May 2012

USDA-NRCS, 2004b. Soil Survey Geographic (SSURGO) Database.
United States Department of Agriculture National Resource
Conservation Service, Washington, D.C. http://soils.usda.gov/
survey/geography/ssurgo/. Accessed March 2010

US Environmental Protection Agency (USEPA), 2002. Method 1603: Escherichia coli (*E. coli*) in Water by Membrane Filtration Using Modified Membrane-thermotolerant Escherichia coli agar (modified m TEC). Publication EPA-821-R-02-023.Washington, D.C. USEPA Office of Water, Office of Science and Technology

USEPA, 2001a. Protocol for Developing Pathogen TMDLs: Source Assessment, 5-1-5-18 1st ed. EPA841-R-00-002US Environmental Protection Agency, Office of Water,Washington, DC, USA

USEPA, 2001b. The National Costs of the Total Maximum Daily Load Program (draft report), EPA 841-D-01-003. US Environmental Protection Agency, Office of Water,Washington, DC, USA

USEPA, 2003. Bacterial Water Quality Standards for Recreational Waters (freshwater and marine waters) Status Report, EPA 823-R-03-008. US Environmental Protection Agency, Office of Water, Washington, DC, USA

USGS, 2009.1-Arc Second National Elevation Dataset, US Geological Survey, Sioux Falls, SD http://seamless.usgs.gov/website/seamless/viewer.htm. Accessed March 2009

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