VARIATIONS OF INDICATOR BACTERIA IN A LARGE URBAN WATERSHED

advances in watershed management

A. M. Ibekwe, S. M. Lesch, R. M. Bold, M. B. Leddy, A. K. Graves

ABSTRACT. *The contamination of water resources by nonpoint-source fecal pollution is a major concern to human health and water quality throughout the world. The Santa Ana River (SAR) in southern California is an impaired stream with historically high fecal coliform counts. This study evaluated the presence of indicator bacteria at 13 sites in the middle Santa Ana River watershed (MSAR). The objectives of this study were to: (1) examine spatial and temporal characteristics of fecal bacteria loading during dry weather (low or baseline) flow, wet weather (storm) flow, and recessional flow (72 h after storm) along two creeks, at two wastewater treatment plant (WWTP) outlets, and at a control site in the MSAR; and (2) determine how the various sampling locations affected indicator bacteria concentrations in the watershed. Total coliform (TC), fecal coliform (FC),* E. coli, *enterococci, and total bacterial concentrations were characterized at 13 locations in the watershed over a two-year period. Analysis of covariance (ANOCOVA) was used to test each specific set of bacteria counts, site effects, water flow conditions, and the four water quality covariate effects. Our results showed that the control site (S1) and WWTP estimates always showed significantly lower indicator bacteria than the channels influenced by urban runoff and agricultural activities. The water flow effects indicated that the recessional flow transported significantly lower bacterial counts into the watershed than either the dry weather flow or the storm or wet weather flow. In addition, bacterial count estimates changed far more significantly across different sites in comparison to estimates across seasons or time. These results imply that total TC, FC,* E. coli*, and enterococci bacterial counts in the MSAR watershed were strongly influenced by spatial location effects, with contamination due to local agricultural and/or urban runoff, in contrast to elevated upstream contamination and/or discharge contamination associated with the two WWTPs. Therefore, this study has provided data for evaluation of the Santa Ana River watershed's total maximum daily load (TMDL) management plans that could also be applicable to other large watersheds with different nonpoint-source pollutants.*

Keywords. *Bacteria, Contamination,* E. coli*, Enterococci, Fecal coliform, Wastewater treatment plants.*

Fecal indicator bacteria (FIB) present at concentrations above certain thresholds are believed to be correlated with fecal pollution, thus increasing the risk of gastroenteritis. Fecal pollution may result from point and nonpoint sources (Hagedorn et al., 1999; Jagals et al., 1995). Agricultural runoff, urban storm water, and streams as well as point sources such as overflows from wastewater treatment plants (WWTPs) have been linked to increases in microbial loads to natural bodies of water (McLellan, 2004). The U.S. Environmental Protection Agency (USEPA) has proposed maximum levels for fecal bacteria in surface water for single sampling events (USEPA, 2000). FIB concentrations for the recreational water quality criterion are 61 CFU 100 mL^{-1} for enterococci and 235 CFU 100 mL^{-1} for *E. coli* for single sample maximum, and the previous USEPA-recommended FC criterion of 400 CFU 100 mL^{-1} is used in many parts of the country for national pollutant discharge elimination system (NPDES) and total maximum daily load (TMDL) purposes (USEPA, 2001). Although concentrations in water of these fecal indicator bacteria above the designated levels are presumed to indicate fecal contamination, several recent studies indicate that various FIB may grow in the environment (Ishii et al., 2006; Walk et al., 2007) or may be associated with non-fecal sources (Whitman et al., 2003).

Currently available data from southern California watersheds demonstrate that both existing and USEPA-recommended bacteria water quality criteria are routinely exceeded in the watersheds, often by one or more orders of magnitude. Measurement data and numerous literature sources have both shown that both wet and dry weather bacteria concentrations frequently exceed objectives in creeks and rivers, and that bacteria concentrations rise dramatically during wet weather periods (Izbicki et al., 2004; Rice, 2005). Some of the rivers and streams in this region of southern California are listed as impaired waterbodies due to high levels of fecal coliform bacteria. Currently, many inland surface waterbodies in the middle Santa Ana River (MSAR)

Submitted for review in October 2010 as manuscript number SW 8845; approved for publication by the Soil & Water Division of ASABE in September 2011. Presented at the 21st Century Watershed Technology Conference: Improving Water Quality and the Environment. ASABE Publication No. 701P0210cd.

The authors are **A. Mark Ibekwe,** Microbiologist, USDA-ARS U.S. Salinity Laboratory, Riverside, California; **Scott M. Lesch,** Statistician, Riverside Public Utilities, Resources Division, Riverside, California; **Richard M. Bold,** Senior Scientist, and **Menu B. Leddy,** Principal Scientist, Orange County Water District, Fountain Valley, California; and **Alexandra K. Graves,** Assistant Professor, Department of Soil Sciences, North Carolina State University, Raleigh, North Carolina. **Corresponding author:** A. Mark Ibekwe, USDA-ARS U.S. Salinity Laboratory, 450 W. Big Springs Rd., Riverside, CA 92507; phone: 951-369-4828; fax: 951-342-4964; e-mail: Mark.Ibekwe@ars.usda.gov.

watershed use water criteria and thresholds that are specified by EPA water quality criteria and are in the water quality Basin Plan for the MSAR region.

This study aimed to determine the sources of bacteria contributing to poor water quality in the MSAR. However, since there are many confounding sources of fecal contaminants to the watershed, there is a need to better understand the fate and transport of the indicator bacteria and the influences of chemical and biological variables on their fate and transport in such a large watershed. The purpose of this study was therefore to statistically quantify bacterial measurements through the use of rigorous and statistically accepted modeling methodology and to determine the degree of any long-term temporal trends at each monitoring location. The statistical methodology employed in this analysis facilitated the estimation of specific monitoring location effects and global bacterial averages for each river, in addition to the determination of temporal trends. Analysis of covariance (ANOCOVA) was specified as a plausible model for modeling each specific set of bacteria samples (Montgomery, 2001). Specifically, the models tested for the effects of body of water, month, and sampling location within a body of water on environmental variables as the covariates affecting the concentration of fecal indicator bacteria at specific locations along the rivers. The statistical methodology also facilitated the estimation of specific monitoring location effects, cyclic (seasonal) effects, and global (two-year) fecal indicator bacterial averages for each river, in addition to the determination of temporal trends. All of the above-mentioned estimates are of scientific and regulatory interest and are thus discussed in detail in the Statistical Analysis section. Finally, spatial-temporal trends in various indicator bacteria along Chino Creek and Cypress channel were correlated with other measured chemicals and biological variables to understand the fate of indicator bacteria in the MSAR watershed.

Materials and Methods

Sample Sites

This study was conducted in the MSAR watershed, which covers approximately 1264 km^2 and lies largely in the southwestern corner of San Bernardino County and the northwestern corner of Riverside County (fig. 1). A small part of Los Angeles County (Pomona/Claremont area) is included. The current population of the watershed, based on 2000 census data, is approximately 1.4 million people. Land uses vary in the MSAR watershed and include urban, agriculture, and open space. Although originally developed as an agricultural area, the watershed is rapidly urbanizing. Open space areas include national forest and state park lands. The principal re–

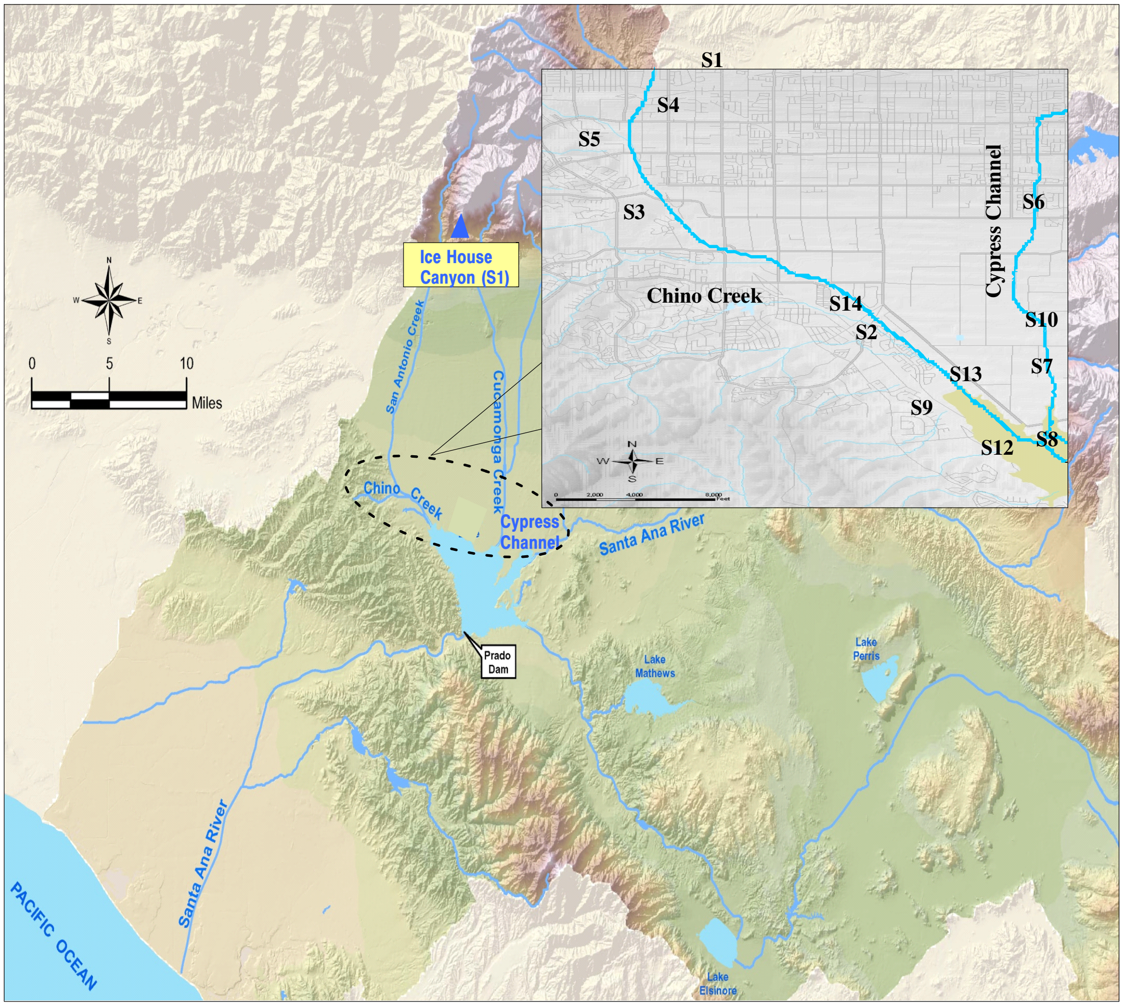

Figure 1. Sites used for the study along the MSAR watershed.

Table 1. Sampling locations and site descriptions.[a]

Zone	Sample Site	Zone Description	Geographic Positioning System (GPS) Location	Land Use	No. of Samples Collected
1	S1	Ice House Canyon (control)	N 34° 15.057 min; W 117° 37.977 min 1447 m elevation	Open space	8
3	S2	Chino Creek	N 33° 58.420 min; W 117° 41.302 min 174 m elevation	Urban runoff	8
	S3	Chino Creek	N 34° 0.246 min; W 117° 43.628 min 207 m elevation	Urban runoff	9
	S4	Chino Creek	N 30° 1.543 min; W 117° 43.652 min 222 m elevation	Urban runoff and commercial washout	9
	S5	Chino Creek	N 34° 1.144 min; W 117° 44.204 min 207 m elevation	Urban runoff	9
	S9	Chino Creek	N 33° 57.364 min; W 117° 40.788 min 163 m elevation	Urban runoff and possible agricultural runoff during storm events	7
	S12	Chino Creek	N 33° 56.941 min; W 117° 39.986 min 155 m elevation	Urban runoff and wastewater	9
4	S6	Cypress channel	N 34° 0.262 min; W 117° 39.766 min 208 m elevation	Agricultural runoff	4
	S7	Cypress channel	N 33° 58.113 min; W 117° 39.624 min 177 m elevation	Agricultural runoff	9
	S8	Cypress channel	N 33° 57.057 min; W 117° 39.555 min 160 m elevation	Agricultural runoff	9
	S10	Cypress channel	N 33° 58.109 min; W 117° 40.286 min 184 m elevation	Agricultural runoff	4
2	S13	Wastewater treatment plant	N 33° 57.840 min; W 117° 40.826 min 180 m elevation	Effluent from wastewater treatment plant	9
	S14	Wastewater treatment plant	N 33° 58.799 min; W 117° 41.655 min 184 m elevation	Effluent from wastewater treatment plant	9

[a] Sampling dates were selected based on weather conditions. Storm weather events were sampled on 28 Dec. 2004, 11 Feb. 2005, and 18 Oct. 2005 using two teams. Recessional flow samples were collected approximately 72 h after a storm event, with the exception of the storm that began on 11 Feb. 2005, which lasted several weeks. Recessional flow samples for that storm were collected on 14 March 2005. Dry weather flows samples were collected during the summer months of July and August 2005 and May and June 2006 (103 samples total). For dry and recessional flow, the Ice Canyon House (S1) location at Mount Baldy was sampled on the first day of the sample schedule; the Chino Creek and Cypress channel locations were sampled on the following day. This was to ensure the proper holding time for the bacteriological analyses.

maining agricultural area in the watershed was formerly referred to as the Chino Dairy Preserve. This area is located in the south central part of the Chino Basin subwatershed and contains approximately 250,000 cows (although this number is quickly declining as the rate of development increases) in a 50,000 ha^2 area.

Six sites along Chino Creek and four sites along the Cypress channel were sampled during storms, recessional flow, and dry weather flow, along with a single control site (Ice House Canyon, site S1) and two WWTPs (S13 and S14). One site (S10) along Cypress channel was dropped after only one sample collection due to construction activities (fig. 1, table 1). Table 1 identifies the four sampling zones corresponding to the sites, sample site locations, land uses, and the number of samples acquired at each site. Some locations were not sampled every time due to weather conditions or lack of access. A number of publicly owned WWTPs discharge highly treated effluent to MSAR waterbodies, e.g., a significant portion of the flow along segments of Chino Creek is comprised mostly of treated effluent.

SAMPLE COLLECTION

All samples were collected in duplicate using sterile Nalgene sampling bottles. For sites that were deep enough to obtain a grab sample, 1 L samples were collected about 10 to 15 cm below the surface of the water. Sites with shallow flow were sampled using a narrow-diameter stainless-steel sterile sampling device. The device was washed, rinsed with 95% ethanol, and dried before use in another site. Field parameters consisting of electrical conductivity, pH, temperature, turbidity, and dissolved oxygen were taken at each sample location using standard methods (APHA, 1995). All bottles were placed in a cooler with a temperature probe, transported on ice to the laboratory, and analyzed within 6 h. Sampling events were separated into two consecutive days for dry and recessional flow, and two teams for storm flow. Usually the Ice Canyon House (S1) sample location at Mount Baldy was collected on the first day of the sample schedule; Chino Creek and Cypress channel locations were sampled on the following day. This was to ensure proper holding time for the bacteriological analyses. Sample turbidity was determined using a portable turbidity meter (model 2100P, Hach Co., Loveland, Colo.) according to the manufacturer's instructions and calibrated each day of use.

All samples were collected in storm, recessional, and dry weather flows between December 2004 and June 2006. Storm flow samples were collected when more than 25.4 mm of accumulated precipitation (i.e., rain) in the Chino Basin and at Mount Baldy were recorded at AccuWeather.com. Three separate storm weather events were sampled during the study on 28 December 2004, 11 February 2005, and 18 October 2005. Recessional flow samples were collected approximately 72 h after a storm event, except for the storm that began on 11 February 2005, which lasted several weeks.

Recessional flow samples for that storm were collected on 14 March 2005. Dry weather flows were analyzed during the summer months of July and August 2005 and May and June 2006. The mean annual streamflow from USGS gauges was 133.6 $m^3 s^{-1}$ at Chino Creek (Chino Creek at Schaefer Ave., site S3) and 96.8 $m^3 s^{-1}$ at Cypress channel (Cypress channel at Shaefer Ave., site S6). Chino Creek provides a good representation of urban runoff, with significant influence from the San Antonio Dam releases and Orange County water transfers with variable nuisance flow from household irrigation. Inputs from agriculture and dairies are considered insignificant, whereas along the Cypress channel inputs from agriculture and dairies are the dominant sources.

DETERMINING CONCENTRATIONS OF INDICATOR BACTERIA

Water samples were processed in the laboratory within 6 h of sample collection. Samples were transported in coolers maintained between 2°C to 10°C using ice packs. The membrane filtration method was used to determine the number of colony-forming units (CFU) per 100 mL sample for indicator organisms. Briefly, the methods used for TC and FC were Standard Methods 9222B and 9222D, respectively, and EPA Methods 1600 and 1603 for *Enterococcus* and *E. coli*, respectively (USEPA, 1986). TC and FC concentrations were determined by using m-Endo LES and MFC media, respectively. *Enterococcus* and *E. coli* concentrations were determined using mEI and m-TEC media, respectively. Total bacteria in the water samples was filtered and stained with water-soluble DNA-binding fluorochrome: 4',6-diamindino-2-phenylindole (DAPI). The bacteria were counted in 30 individual fields using a 100× oil emersion lens on the microscope, and the average number of bacteria (cells mL^{-1}) was calculated (Hobbie et al., 1977; Porter and Feig, 1980).

STATISTICAL ANALYSIS

Sampling sites were grouped into four zones based on site similarity and geographical location. Sites S13 and S14 were classified into a separate zone since the effluent water from each treatment plant was sampled immediately at the discharge outlet (located off the Chino Creek tributary). Likewise, site S1 was classified into its own zone, given that its location is at the base of the foothills and its land use is open space. This location (S1) represents a control site.

The primary factors of interest in this study were the geographical locations of the sample sites (i.e., site effect) and the temporal conditions surrounding the sampling location based on streamflow and the surrounding landscape. Thus, an analysis of variance (ANOVA) model using both site and surface flow classification effects represents a statistical model for analyzing the ($\log_{10}$ transformed) density of indicator bacteria. Along with TC and FC densities, four additional field parameters were also measured at each sample point: pH, salinity (EC), turbidity, and temperature of the surface water. Variations in each of these water quality parameters are known to affect bacterial concentrations. Therefore, the field parameters were measured as possible covariates in the subsequent statistical analyses.

Given the primary factors of interest and additional covariate effects, the following analysis of covariance (ANCOVA) model was initially specified as a plausible model for modeling each specific set of bacteria samples (Montgomery, 2001):

$$\begin{aligned}\ln(y_{ij}) &= \mu + \tau_i + \delta_k + \beta_1(pH_{ij}) + \beta_2(\ln[EC_{ij}]) \\ &\quad + \beta_3(\ln[tB_{ij}]) + \beta_4(T_{ij}) + \varepsilon_{ij} \end{aligned} \tag{1}$$

where y_{ij} represents the average bacteria count at the ith site during the jth sampling period, the τ_i parameters quantify the 13 distinct site effects, the δ_k parameters quantify the three (temporally dependent) water flow conditions, the four β parameters quantify the four water quality covariate effects (water pH, salinity, turbidity, and temperature), and the ε_{ij} residual errors are assumed to be normally distributed but possibly temporally and/or spatially correlated. Bacterial data were log-transformed to induce approximate normality in the residual error distribution. Additionally, the salinity and turbidity covariate readings were also log-transformed in order to reduce the influence of a few large covariate readings (i.e., both covariate distributions appeared to follow approximate lognormal distributions).

Given the spatial-temporal nature of the sampling design, four different covariance structures were examined for modeling the ε_{ij} residual error distributions. In addition to the default assumption that the errors were independently and identically distributed (IID), one temporal and two spatial covariance structures were specified and estimated using restricted maximum likelihood (REML) estimation techniques (McCulloch and Searle, 2001). An autoregressive, order 1 (AR1) model defined as:

$$\begin{aligned} Var(\varepsilon_{ij}) &= \sigma^2 \ \forall\, ij, \\ Cov(\varepsilon_{ij}, \varepsilon_{ik}) &= \rho^{|j-k|}\sigma^2 \ \forall\, i, \\ Cov(\varepsilon_{ij}, \varepsilon_{lk}) &= 0 \ \forall\, i \neq l \end{aligned} \tag{2}$$

was used to formally test for temporal correlation in the residual errors. Additionally, zone-specific compound symmetric (CS) and spatial exponential (Sp-Exp) covariance structures were estimated using REML techniques and then used to test for spatial correlation in the residual errors. The zone-specific CS structure was defined as:

$$Var(\varepsilon_{ij}) = \begin{cases} \sigma^2_{BL} & \text{for zone 1} \\ \sigma^2_{TP} & \text{for zone 2} \\ \sigma^2_{CH} & \text{for zone 3} \\ \sigma^2_{CY} & \text{for zone 4} \end{cases}$$

$$Cov(\varepsilon_{ij}, \varepsilon_{mj}) = \begin{cases} \phi_{CH} \cdot \sigma^2_{CH} \ \forall (i,m) \in \text{zone 3},\ 0 < \phi_3 < 1 \\ \phi_{CY} \cdot \sigma^2_{CY} \ \forall (i,m) \in \text{zone 4},\ 0 < \phi_4 < 1 \\ 0 \text{ otherwise} \end{cases} \tag{3}$$

Likewise, the zone-specific Sp-Exp structure was defined as:

$$Var(\varepsilon_{ij}) = \begin{cases} \sigma^2_{BL} \text{ for zone 1} \\ \sigma^2_{TP} \text{ for zone 2} \\ \sigma^2_{CH} \text{ for zone 3} \\ \sigma^2_{CY} \text{ for zone 4} \end{cases}$$

$$Cov(\varepsilon_{ij}, \varepsilon_{mj}) = \begin{cases} \sigma^2_{CH} \ \exp(-d_{im} / \alpha_{CH}) \ \forall(i,m) \in \text{zone 3} \\ \sigma^2_{CY} \ \exp(-d_{im} / \alpha_{CY}) \ \forall(i,m) \in \text{zone 4} \\ 0 \text{ otherwise} \end{cases} \quad (4)$$

where d_{im} represents the physical distance between the ith and mth sites within zones 3 or 4, respectively. The covariance structures defined in equations 3 and 4 were both used to model and test for positively correlated residuals within zones 3 and/or 4, in addition to non-constant variance effects across the four zones. Note also that both covariance structures assume that the errors associated with the two treatment plant sites (S13 and S14) are independent (since the effluent output from each plant was sampled independently and both sets of effluent were collected before reaching Chino Creek).

After estimating equation 1 under each of the four residual error assumptions, likelihood ratio tests were used to test for temporal and spatial correlation in the residual errors (Schabenberger and Gotway, 2005). This is discussed with examples in the next section; the residual errors associated with all five bacterial count ANOCOVA models exhibited statistically significant spatial correlation. Thus, each ANOCOVA model was re-estimated as a mixed linear model using REML estimation techniques (McCulloch and Searle, 2001). Additionally, adjusted F and t-tests for all parameter estimates and/or contrasts of interest were computed using the Kenward-Roger adjustment technique (Kenward and Roger, 1997). All of the statistical modeling and analysis for this study was performed using the GLM and MIXED procedures in the SAS STAT software package, version 9 (SAS, 2001).

RESULTS

The basic univariate summary statistics for the FIB measurements of interest; i.e., TC, FC, *E. coli*, and enterococci counts, are shown in figure 2. The statistics shown summarize the $\log_{10}$ transformed counts for each of the response variables. TC counts showed the highest concentrations and the greatest variability. Examination of each site throughout the watershed indicated that bacterial concentrations along Chino Creek and Cypress channel routinely exceeded the applicable water quality objectives for FC (single sample of 400 CFU per 100 mL) and USEPA's recommended water quality criteria for *E. coli* (235 CFU per 100 mL) except at the control site (S1) and the WWTPs (fig. 2). As shown in figures 3a through 3d, the univariate summary statistics during the three seasons showed total coliform to be consistently higher than the EPA-recommended level for a single sample of 400 CFU per 100 mL (fig. 3a). The same trends were observed for fecal coliform (fig. 3b), *E. coli* (fig. 3c), and *Enterococcus* (fig. 3d) during the storm season for most sites. Total bacterial counts did not significant differ between sites and seasons (fig 3e).

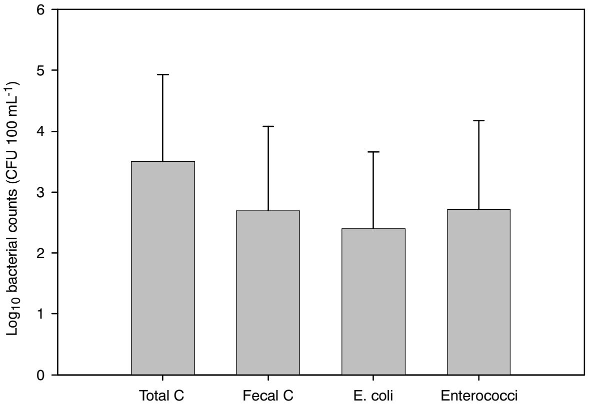

Figure 2. Univariate summary statistics for bacterial counts using Shipiro-Wilk test for residual normality for total coliform, fecal coliform, *E. coli*, enterococci, and total bacteria.

IDENTIFICATION OF SUITABLE COVARIANCE STRUCTURES FOR INDICATOR BACTERIA

The identification of suitable covariance structures for indicator bacteria was determined using an ANOCOVA model. The summary statistics for the $\log_{10}$ transformed bacteria concentration models are shown in table 2. All five models (TC, FC, *E. coli*, enterococci, and total bacteria) were fitted using the 99 sample observations that had a complete set of associated covariate readings (four samples were missing salinity covariate readings and thus were excluded from the ANOCOVA modeling analyses). The Shapiro-Wilk (SW) test scores for residual normality are also displayed in table 2, along with the associated p-values (Shapiro and Wilk, 1965). The R^2 values for all five ordinary ANOCOVA models ranged from 0.790 to 0.847, suggesting that equation 1 could consistently explain about 80% or more of the variation in the log bacteria count data.

A preliminary analysis of the ANOCOVA model residuals suggested that the normality assumption was reasonably well satisfied in all five models. However, the residual variance estimates were noticeably different across the four sampling zones in each model, particularly for the FC, *E. coli*, and *Enterococcus* models. Perhaps not surprisingly, one (*Enterococcus*) of these three residual error distributions was not significant according to the Shapiro-Wilk test at the 0.05 level of significance (table 2). Additionally, although none of the residual error distributions exhibited any noticeable temporal correlation structure, all five distributions exhibited some degree of spatial correlation structure.

The REML -2 log-likelihood (-2LL) scores for each of the five ANOCOVA models estimated using all four residual error covariance structures are shown in table 3. Since the IID error structure represents a special (i.e., nested) case of the autoregressive order 1 (AR1), compound symmetric (CS), and spatial exponential (Sp-Exp) covariance structures, formal likelihood ratio tests could be computed by directly differencing the -2LL scores (Schabenberger and Gotway, 2005). The results for these 15 tests are shown in table 3, along with Akaike's information criterion (AIC) (Akaike, 1974). With respect to Akaike's criterion, smaller AIC scores imply a more parsimonious (i.e., better fit) mixed linear model. The results shown in table 3 are consistent on three points. First, the AR1 covariance structures never produced statistically significant likelihood ratio test scores; these results im–

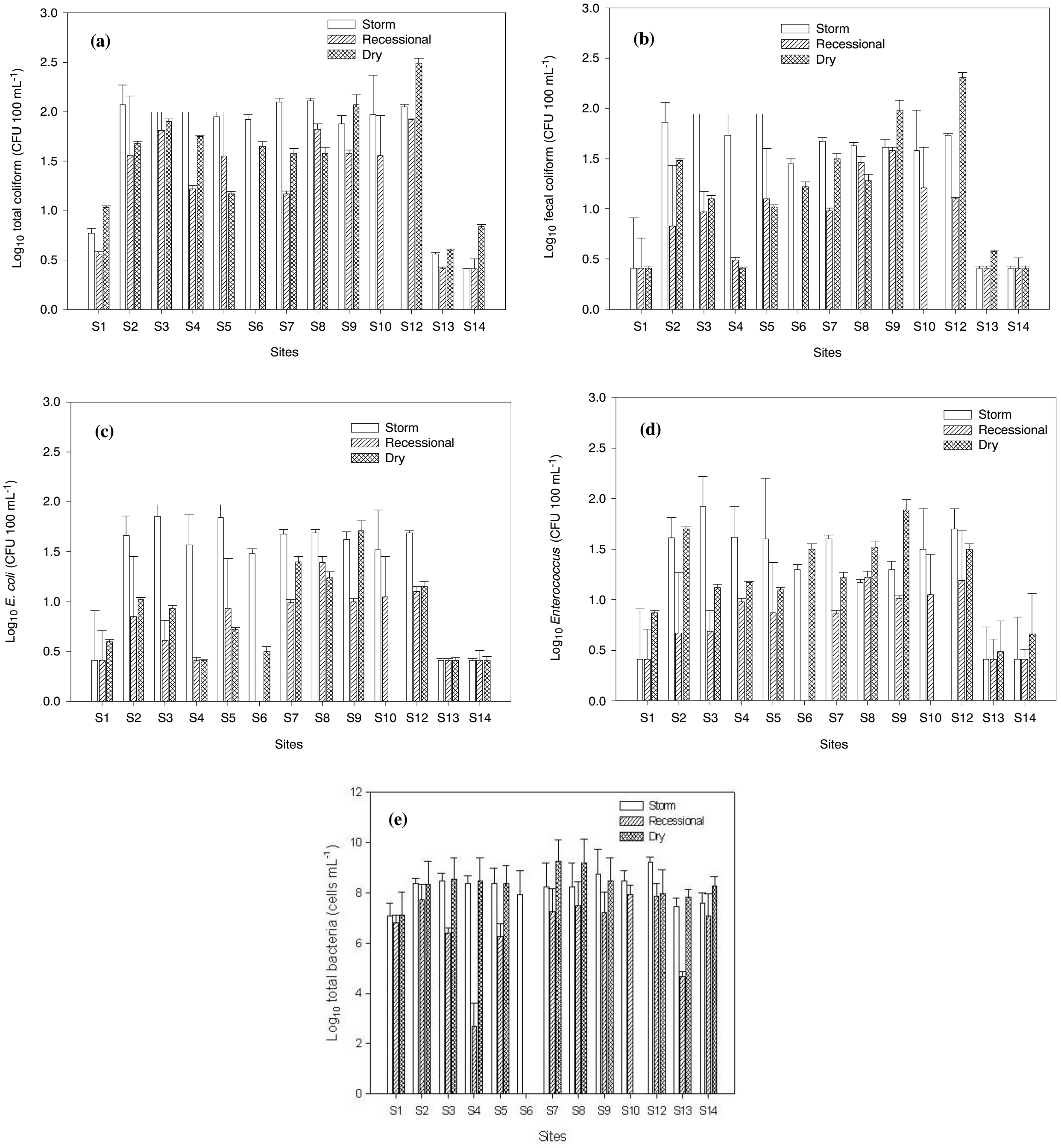

Figure 3. Bacterial counts during storm, recessional, and dry weather flow: (a) total coliform, (b) fecal coliform, (c) *E. coli*, (d) enterococci, and (e) total bacteria.

ply that the ANOCOVA residual errors did not exhibit any temporal, autoregressive correlation structure in any of the five models. Second, the CS and Sp-Exp covariance structures always produced statistically significant likelihood ratio test scores, confirming that the residual errors associated with sites within Chino Creek and Cypress channel were positively correlated in all five models and that the variances changed across zones. Third, the ANOCOVA models fit using the Sp-Exp structure always produces the smallest -2LL and AIC scores. This latter result implies that a heterogeneous variance, spatially correlated residual error assumption should be adopted (in place of the IID error assumption) when estimating each ANOCOVA model.

The REML estimated and Sp-Exp covariance parameter values for Chino Creek and Cypress channel and the two heterogeneous variance components for the control site and the

Table 2. Ordinary ANOCOVA model summary statistics and Shapiro-Wilk test at the 0.05 level of significance.

Statistic	Total Coliform	Fecal Coliform	*E. coli*	Enterococci	Total Bacteria
R^2	0.84	0.79	0.85	0.79	0.81
MSE	2.09	2.53	1.57	2.46	0.79
CV (%)	17.30	25.50	22.40	24.70	4.80
SW test	0.99	0.99	0.97	0.99	0.99
SW p-value	0.79	0.004	0.04	0.59	0.41

Table 3. REML -2LL scores for various residual covariance structures.

Response Variable	Covariance Structure	-2LL	Difference	DF	Chi-square p-value	AIC Score
Total coliform	IID	334.6				336.6
	AR1	334.5	0.1	1	0.752	338.6
	CS	315.5	19.1	5	0.002	327.5
	Sp-Exp	305.2	29.4	5	<0.001	317.2
Fecal coliform	IID	350.0				352.0
	AR1	349.9	0.1	1	0.752	354.0
	CS	316.9	33.1	5	<0.001	328.9
	Sp-Exp	301.4	48.6	5	<0.001	313.4
E. coli	IID	311.9				313.9
	AR1	311.8	0.1	1	0.752	315.9
	CS	273.6	38.3	5	<0.001	285.6
	Sp-Exp	265.5	46.4	5	<0.001	277.5
Enterococci	IID	347.9				349.9
	AR1	347.5	0.4	1	0.527	351.5
	CS	297.7	50.2	5	<0.001	309.7
	Sp-Exp	295.5	52.4	5	<0.001	307.5
Total bacteria	IID	254.1				256.1
	AR1	252.1	2.0	1	0.157	256.1
	CS	229.7	24.4	5	<0.001	241.7
	Sp-Exp	224.3	29.8	5	<0.001	236.3

WWTPs for the indicator bacteria mixed linear ANOCOVA model were analyzed (data not shown). In comparison to the ordinary ANOCOVA model mean square error (MSE) estimate of 1.57, the variance estimates associated with the control (S1) and WWTP (S13 and S14) data were noticeably lower (0.74 and 0.35, respectively), while Chino Creek and Cypress channel were clearly higher (2.77 and 2.37). These results confirm that bacteria concentrations collected along either channel tended to be much more variable in comparison to the control (S1) or WWTP (S13 and S14) sampling locations. The predicted correlation coefficients (data not shown) agreed reasonably well with their observed counterparts (calculated using the residual errors derived from the IID model). In general, the covariance modeling results for all five mixed linear ANOCOVA models tended to be fairly similar. More specifically, the channel variance components were always larger than the corresponding control (S1) or WWTP variance components. Additionally, the estimated Sp-Exp correlation structures always agreed reasonably well with the observed, within-zone correlation coefficients, particularly given the small sample sizes analyzed in this study.

CLIMATE AND WATER QUALITY TRENDS IN THE MSAR USING ANOCOVA

In the MSAR, the water temperature rises in the summer months to about 30°C to 32°C on the average; the upward trend begins in April, the maximum occurs in July and August, and the temperature decreases to an average of about 27°C between November and April. The mean annual rainfall for the MSAR watershed is about 460 mm per year, predominantly between December and April, during wet years and $\leq$250 mm during dry years. In the two years of this study, rainfall was slightly heavier during the first year than the second. ANOCOVA results for the water quality variables for the two creeks, control, and WWTPs suggested potentially important differences among the variables that may influence microbial dynamics. The primary parameter F and t-tests for the estimated mixed linear ANOCOVA model are shown in table 4. All of the p-values shown in table 4 have been computed using the Kenward-Roger adjustment technique. As shown, the site effect was significantly below the 0.001 level in four of the five ANOCOVA models and significantly below the 0.05 level in the total bacteria model. In contrast, the water flow effect was significant in only two of the five models: TC ($p = 0.030$) and total bacteria ($p = 0.001$). This is significant because it shows the influence of the dominant or larger populations, such as TC or total bacteria, rather than the subpopulations such as FC, *E. coli*, and enterococci, whose populations are a small subset of the total population.

As will be discussed in more detail subsequently, the log bacteria count estimates tended to change far more significantly across different sites (data not shown) in comparison to the differences in the average estimates across the three water flow conditions. The measured log turbidity levels were significantly related to the log bacteria count data in three of the five ANOCOVA models: *E. coli*, enterococci, and total bacteria. In each model, higher turbidity levels corresponded with higher bacterial counts. In contrast, the log salinity levels (ln[ECe]) were never significantly related to the bacteria count levels in any of the five models. Similarly, the water pH levels were only found to be significantly related to the TC count data. In this ANOCOVA model, higher pH levels corresponded with lower bacterial counts. Finally, the water temperature was found to be significantly related to the log bacteria count levels in two of the five ANOCOVA models: *E .coli* and total bacteria. However, the temperature effect was not consistent across these two models; higher water temperatures resulted in higher total bacterial counts but lower *E. coli* counts.

The effects of individual site, water flow, and zone average on log bacterial count estimates derived from each estimated ANOCOVA model are shown in table 5. These estimates and tests fully quantify how the various sampling locations and water flow conditions affected TC, FC, *E. coli*, enterococci, and total bacterial count, respectively. Table 6 shows the t-scores (and associated p-values) for five prespecified zone-average contrast tests of interest. Based on the

Table 4. Climate and water quality trends estimated with mixed linear ANOCOVA models. All p-values were computed using the Kenward-Roger adjustment technique.

		Response Variable				
Factor or Covariate	Statistic	TC	FC	*E. coli*	Enterococci	Total bacteria.
Site	F score	8.56	8.53	9.73	4.72	2.13
	p-value	<0.001	<0.001	<0.001	<0.001	0.034
Water flow	F score	3.93	0.39	2.83	0.93	8.28
	p-value	0.030	0.686	0.085	0.413	0.001
ln[turbidity]	t score	1.03	0.40	3.35	3.23	7.90
	p-value	0.305	0.687	0.001	0.002	<0.001
ln[ECe]	t score	0.06	-0.20	-0.35	-0.07	-1.55
	p-value	0.952	0.839	0.727	0.942	0.128
pH	t score	-3.48	-1.78	-1.48	0.52	-1.10
	p-value	<0.001	0.081	0.145	0.607	0.275
Temperature	t score	-0.49	-0.65	-2.41	0.61	3.46
	p-value	0.626	0.523	0.019	0.544	0.001

Table 5. Average seasonal water flow effect estimates of response variables.

Sites	Total Coliform	Fecal Coliform	*E. coli*	Enterococci	Total Bacteria
Recessional flow	3.37 ±0.14	2.63 ±0.17	2.18 ±0.14	2.64 ±0.17	7.81 ±0.10
Dry weather flow	3.75 ±0.23	2.83 ±0.23	2.71 ±0.20	2.83 ±0.23	8.12 ±0.13
Storm flow	3.89 ±0.16	2.68 ±0.19	2.32 ±0.15	2.82 ±0.18	8.26 ±0.11

Table 6. Pre-specified zone-average contrast of response variables.

Sites[a]	Total Coliform (t-test)	Fecal Coliform (t-test)	*E. coli* (t-test)	Enterococci (t-test)	Total Bacteria (t-test)
SI vs. WWTPs	1.38 (0.17)	-61 (0.55)	-1.82 (0.07)	0.72 (0.48)	1.57 (0.13)
S1 vs. CCK	-3.86 (<0.001)	-4.39 (<0.001)	-4.72 (<0.001)	-2.19 (0.03)	1.08 (0.28)
S1 vs. CCH	-3.91(<0.001)	-4.96 (<0.001)	-4.43 (<0.001)	-2.21 (0.03)	0.49 (0.63)
WWTPs vs. CCK	-11.31 (<0.001)	-6.18 (<0.001)	-5.48 (<0.001)	-5.11 (<0.001)	-1.39 (0.17)
CCK vs. CCH	0.71 (0.49)	0.30 (0.77)	1.09 (0.29)	0.18 (0.86)	-1.19 (0.26)

[a] S1 indicates ice house (control), WWTPs = waste water treatment plants (S13 and S14), CCK = Chino Creek (S2, S3, S4, S5, S9, and S12), and CCH = Cypress channel (S6, S7, S8, and S10).

significant F score for recessional flow in table 4 (0.03), recessional water flow can be judged to have significantly lower effects on transport of TC than either dry weather flow or storm flow. The zone-average contrast tests (table 6) confirm that the control site (S1) estimate is significantly lower than the Chino Creek and Cypress channel estimates, and the WWTP estimate is significantly lower than the Chino Creek and Cypress channel estimates (all p-values < 0.001). In contrast, the S1 and WWTP estimates cannot be judged to be significantly different. Likewise, the two average tributary estimates (Chino Creek and Cypress channel) are also not significantly different from each other.

DISCUSSION

Microorganisms in densities above certain levels in water can cause adverse effects, including death in humans and wildlife as a result of exposure. Adverse health effects in humans can be grouped into gastrointestinal, respiratory, eye, ear, nose, skin, etc. For bodies of water used for non-contact recreational purposes, such as the one in this study, high densities of bacteria can result in immediate closure for public use at points where a contaminated stream empties into a body of water used for contact recreation. As seen in figure 2, high densities of indicator bacteria resulted in inclusion of Chino Creek and Cypress channel in the 303(d) list of bodies of water with bacterial counts above the 1995 water quality control plan for the Santa Ana River basin (Basin Plan) for beneficial uses (Rice, 2005). The Basin Plan currently specifies fecal coliform log mean concentrations of less than 200 organisms per 100 mL based on five or more samples per 30-day period, and not more than 10% of the samples can exceed 400 organisms per 100 mL for any 30-day period. The TMDL for these waterbodies established compliance targets for both fecal coliform and *E. coli*. The target for fecal coliform is a five-sample/30-day logarithmic mean of less than 180 organisms per 100 mL, and not more than 10% of the samples can exceed 360 organisms per 100 mL for any 30-day period. For *E. coli*, the target is a five-sample/30-day logarithmic mean of less than 113 organisms per 100 mL, and not more than 10% of the samples can exceed 212 organisms per 100 mL for any 30-day period. According to our data, the rate of non-compliance was much higher during wet weather and dry weather flow conditions than during recessional flow on a single sampling event.

The elevated FIB counts along Chino Creek and Cypress channel coincide with mostly nonpoint sources of fecal contamination. Most sampling sites along Chino Creek and Cypress channel were in violation of local and USEPA water quality standard for FIB counts. All of these sites are situated near known human point sources or agricultural operations (Ibekwe et al., 2007; Ibekwe and Lyon, 2008) (fig. 1). For example, the values of FIB along Cypress channel and Chino Creek were strongly influenced by land use. According to land use data obtained from the Southern California Association of Governments (SCAG), 26%, 16%, 2%, and 57%, respectively, were categorized as agriculture, commercial/industrial, natural/vacant, and residential, with a total of 2004 ha of land covering the reach, while 4%, 30%, 14%, and 52% in Chino Creek were categorized as agriculture, commercial/industrial, natural/vacant, and residential with a total land area of 16319 ha. From the land use data, pre-specific zone-average contrast of the response variables showed that there were no significant differences in FC, TC, *E. coli*, enterococci, and total bacterial counts between Chino Creek and Cypress channel. However, significant differences in FC, TC, *E. coli*, and enterococci were observed between the WWTPs and Chino Creek or Cypress channel, indicating that the treatment processes at the WWTPs removed some of the FIB. Our results indicate a potential health risk if treatment failure occurs, or if treatment fails to remove potential pathogenic bacteria.

In our study, TC and total bacterial counts were significantly different among the dry, storm, and recessional flows. *E. coli* was only significant at the 0.1% level. Sinclair et al. (2009) showed storm flow loads were higher than base flow loads of *E. coli* in a rural watershed. Their finding was consistent with the findings of Jamieson et al. (2003), Jamieson et al. (2004), and Gentry et al. (2006) for the same watershed and with other rural watershed investigations (Nagels et al., 2002; Oliver et al., 2003; Reeves et al., 2004). This is in agreement with our study on TC and weakly so with *E. coli*. The loading of microbial contaminants from urban runoff and agricultural activities along Chino Creek and Cypress channel represents a steady-state source of pollution that would be less influenced by other events. Lewis et al. (2005) measured an average event storm flow TC loading rate of 6.10×10^9 CFU ha^{-1}. The high numbers of FIB in the creeks may indicate the ability of these bacteria to survive and grow outside their intestinal habitat, and *E. coli* in particular has been shown to persist in the environment long after it has been introduced (Anderson et al., 2005; Desmarais et al., 2002; Fujioka et al., 1999; Wheeler-Alm et al., 2003; Whitman et al., 2003; Shanks et al., 2006; Haack et al., 2009). Therefore, environmental persistence or regrowth during summer months could confound the interpretation of baseline dry weather flow dynamics.

CONCLUSION

We have shown in this study that urban and agricultural activities contributed about the same levels of indicator bacteria to the MSAR watershed. Here we have separated indicator bacteria to TC, FC, *E. coli*, and enterococci in order to verify compliance with regulatory requirements and potential health risk. The human health risk from agricultural animal feces is usually assumed to be less than the risk from human feces, in part because viruses, which are the most common cause of human illnesses from exposure to fecal contamination in water, are highly host-specific. Our results are consistent with previous studies that showed increases in fecal bacteria concentrations along gradients of coastal urbanization (DiDonato et al., 2009). These authors noted the work of Karn and Harada (2001) with data collected from various agencies in South Asia documenting elevated coliform concentrations with increased urbanization. This study has provided useful information that could be used in the evaluation of watershed management plans for mixed watersheds.

ACKNOWLEDGEMENTS

We wish to express our appreciation to Pamela Watt, Greg Woodside, Nira Yamachika, Gary Hackney, and Bill Rice for their technical assistance and generous support during this study. The study was made possible in part by the financial support of a Proposition 13 Grant from the State Water Resources Control Board Nonpoint-Source Pollution Control Programs, the Inland Empire Utilities Agency, and the Orange County Water District. This research was also supported in part by the 206 Manure and Byproduct Utilization Project of the USDA-ARS.

REFERENCES

Akaike, H. 1974. A new look at the statistical model identification. *IEEE Trans. Auto. Control* 19(6): 716-723.

Anderson, K. L., J. E. Whitlock, and V. J. Harwood. 2005. Persistence and differential survival of fecal indicator bacteria in subtropical waters and sediments. *Appl. Environ. Microbiol.* 71(6): 3041-3048.

APHA. 1995. *Standard Methods for the Examination of Water and Wastewater.* 19th ed. Washington, D.C.: American Public Health Association.

Desmarais, T. R., H. M. Solo-Gabriele, and C. J. Palmer. 2002. Influence of soil on fecal indicator organisms in a tidally influenced subtropical environment. *Appl. Environ. Microbiol.* 68(3): 1165-1172.

DiDonato, G. T. J. R. Stewart, D. M. Sanger, B. J. Robinson, B. C. Thompson, A. F. Holland, and R. F. Van Dolah. 2009. Effects of changing land use on the microbial water quality of tidal creeks. *Marine Poll. Bull.* 58(1): 97-106.

Fujioka, R., C. Sian-Denton, M. Borja, J. Castro, and K. Morphew. 1999. Soil: The environmental source of *Escherichia coli* and enterococci in Guam's streams. *J. Appl. Microbiol.* 85(supp. 1): 83S-89S.

Gentry, R. W., J. McCarthy, A. Layton, L. D. McKay, D. Williams, S. R. Koirala and G. S. Sayler. 2006. *Escherichia coli* loading at or near base flow in a mixed-use watershed. *J. Environ. Qual.* 35(6): 2244-2249.

Haack, S. K., J. W. Duris, L. R. Fogarty, D. W. Kolpin, M. J. Focazio, E. T. Furlong, and M. T. Meyer. 2009. Comparing wastewater chemicals, indicator bacteria concentrations, and bacterial pathogen genes as fecal pollution indicators. *J. Environ. Qual.* 38(1): 248-258.

Hagedorn, C., S. L. Robinson, J. R. Filtz, S. M. Grubbs, T. A. Angier, and R. B. Reneau. 1999. Determining sources of fecal pollution in a rural Virginia watershed with antibiotic resistance patterns in fecal streptococci. *Appl. Environ. Microbiol.* 65(12): 5522-5531.

Hobbie, J. E., R. J. Daley, and S. Jasper. 1977. Use of nuclepore filters for counting bacteria by epifluorescence microscopy. *Appl. Environ. Microbiol.* 33(5): 1225-1228.

Ibekwe, A. M., and S. R. Lyon. 2008. Microbiological evaluation of fecal bacterial composition from surface water through aquifer sand material. *J. Water and Health* 6(3): 411-421

Ibekwe, A. M., S. R. Lyon, M. Leddy, and M. Jacobson-Meyer. 2007. Impact of plant density and microbial composition on water quality from a free water surface constructed wetlands. *J. Appl. Microbiol.* 102(4): 921-936.

Ishii, S., W. B. Ksoll, R. E. Hicks, and M. J. Sadowsky. 2006. Presence and growth of naturalized *Escherichia coli* in temperate soils from Lake Superior watersheds. *Appl. Environ. Microbiol.* 72(1): 612-621.

Izbicki, J. A., M. I. Pimentel, and M. B. Leddy. 2004. Microbial and dissolved organic carbon characterization of stormflow in the Santa Ana River at Imperial Highway, Southern California, 1999-2002. USGS Scientific Investigation Report 2004-5116. Reston, Va.: U.S. Geological Survey.

Jagals, P., W. O. K. Grabow, and J. C. de Villiers. 1995. Evaluation of indicators for assessment of human and animal faecal pollution of surface runoff. *Water Sci. Tech.* 31(5): 235-241.

Jamieson, R., R. Gordon, S. Tattrie, and G. Stratton. 2003. Sources and persistence of fecal coliform bacteria in a rural watershed. *Water Qual. Res. J. Canada* 38(1): 33-47.

Jamieson, R., R. Gordon, D. Joy, and H. Lee. 2004. Assessing microbial pollution of rural surface waters: A review of current watershed-scale modeling approaches. *Agric. Water Mgmt.* 70(1): 1-17.

Karn, S. K., and H. Harada. 2001. Surface water pollution in three urban territories of Nepal, India, and Bangladesh. *Environ. Mgmt.* 28(4): 483-496.

Kenward, M. G., and J. H. Roger. 997. Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics* 53(3): 983-997.

Lewis, D. L., E. R. Atwill, M. S. Lennox, L. Hou, B. Karle, and K. W. Tate. 2005. Linking on-farm dairy management practices to storm-flow fecal coliform loading for California coastal watersheds. *Environ. Mon. Assess.* 107(1-3): 407-425.

McCulloch, C. E., and S. R. Searle. 2001. *Generalized, Linear, and Mixed Models.* New York, N.Y.: John Wiley.

McLellan, S. L. 2004. Genetic diversity of *Escherichia coli* isolated from urban rivers and beach water. *Appl. Environ. Microbiol.* 70(8): 4658-4665.

Montgomery, D. C. 2001. *Design and Analysis of Experiments.* 5th ed. New York, N.Y.: John Wiley.

Nagels, J. W., R. J. Davies-Colley, A. M. Donnison, and R. W. Muirhead. 2002. Faecal contamination over flood events in a pastoral agricultural stream in New Zealand. *Water Sci. Tech.* 45(12): 45-52.

Oliver, D. M., C. D. Clegg, P. M. Haygarth, and A. L. Heathwaite. 2003. Determining hydrological pathways for the transfer of potential pathogens from grassland soils to surface waters. In *Proc. 7th Intl. Specialised IWA Conf.,* Diffuse Pollution and Basin Mgmt., 3B - Agriculture, 36-41. M. Bruen, ed. London, U.K.: International Water Association.

Porter, K. G., and Y. S. Feig. 1980. The use of DAPI for identifying and counting aquatic microflora. *Limnol. Oceanogr.* 25(5): 943-948.

Reeves, R. L., S. B. Grant, R. B. Mrse, C. M. Copil Oancea, B. F. Sanders, and A. B. Boehm. 2004. Scaling and management of fecal indicator bacteria in runoff from a coastal urban watershed in southern California. *Environ. Sci. Tech.* 38(9): 2637-2648.

Rice, W. 2005. Staff report on bacterial indicator total maximum daily loads in the middle Santa Ana River watershed. Santa Ana, Cal.: Santa Ana Regional Water Control Board.

SAS. 2001. *SAS/STAT User's Guide.* Version 9. Cary, N.C.: SAS Institute, Inc.

Schabenberger, O., and Gotway, C. A. 2005. *Statistical Methods for Spatial Data Analysis.* Boca Raton, Fla.: Chapman and Hall/CRC.

Shanks, O. C., C. Nietch, M. Simonich, M. Younger, D. Reynolds, and K. G. Field. 2006. Basin-wide analysis of the dynamics of fecal contamination and fecal source identification in Tillamook Bay, Oregon. *Appl. Environ. Microbiol.* 72(8): 5537-5546.

Shapiro, S. S., and M. B. Wilk. 1965. An analysis of variance test for normality (complete samples). *Biometrika* 52(3-4): 591-611.

Sinclair, A., D. Hebb, R. Jamieson, R. Gordon, K. Benedict, K. G. W. Stratton, and A. Madani. 2009. Growing season surface water loading of fecal indicator organisms within a rural watershed. *Water Research* 43(5): 1199-1206.

USEPA. 1986. Ambient water quality criteria for bacteria: 1986. EPA 440/5-84-002. Washington, D.C.: U.S. Environmental Protection Agency, Office of Water.

USEPA. 1998. *Giardia: Human health criteria*. EPA 823-R-002. Washington, D.C.: U.S. Environmental Protection Agency.

USEPA. 2000. Improved enumeration methods for recreational water quality indicators, *Enterococci* and *Escherichia coli*, EPA-821-97-004. Washington, D.C.: USEPA.

USEPA. 2001. Protocol for developing pathogen TMDLs. EPA 841-R-00-002. Washington, D.C.: USEPA.

Walk, S. T., E. W. Alm, L. M. Calhoun, J. M. Mladonicky, and T. S. Whittam. 2007. Genetic diversity and population structure of *Escherichia coli* isolated from freshwater beaches. *Environ. Microbiol.* 9(9): 2274-2288.

Wheeler-Alm, E., J. Burke, and A. Spain. 2003. Fecal indicator bacteria are abundant in wet sand at freshwater beaches. *Water Res.* 37(16): 3978-3982.

Whitman, R. L., D. A. Shively, H. Pawlik, M. B. Nevers, and M. N. Byappanahalli. 2003. Occurrence of *Escherichia coli* and enterococci in *Cladophora* (Chlorophyta) in nearshore water and beach sand of Lake Michigan. *Appl. Environ. Microbiol.* 69(8): 4714-4719.