Surface Water Bacterial Fluctuation in the Upper Hoosic River Watershed

Elena Traister

February, 2005

Abstract

The Hoosic River in northwestern Massachusetts drains into the Hudson River across from the Town of Stillwater, New York. Bacterial pollution is a continuing health concern in this region. While various efforts aim to monitor the levels of indicator bacteria (such as *Escherichia coli* and fecal coliform) in freshwater systems such as this, there has been relatively little research focused on determining how bacterial concentration fluctuates over different time scales throughout a watershed. Understanding regular spatial and temporal variation exhibited by populations of indicator organisms is necessary before robust water quality sampling programs can be put into place. This study examined the variation in *E. coli* at different time scales, including seasonal, diurnal, and weather-related in the upper Hoosic River Watershed. Based on research findings, several recommendations for improving water quality monitoring methods are offered.

Introduction

In response to mounting public concern over the pollution of U.S. surface waters, the Federal Water Pollution Control Act was passed in 1972 and amended in 1977 to create what is commonly known as the Clean Water Act (CWA). The objective of the CWA was "to restore and maintain the chemical, physical, and biological integrity of the Nation's waters." Organisms such as *E. coli* and fecal coliform, which often do not cause disease directly, indicate the presence of harmful pathogens (EPA, 2002). Leaky septic systems, faulty wastewater treatment plants, and agricultural practices that allow direct animal defecation into surface water are some of the sources contributing to the spread of fecal contamination (Weiskel et al., 1996). Such patterns in land use and location of point sources would therefore be expected to influence levels of bacterial contamination throughout a watershed. Tong and Chen document "a significant relationship between land use and in-stream water quality, especially for nitrogen, phosphorus and Fecal coliform. (2002)" Surface waters with elevated fecal bacteria levels are restricted from public use, and their economic and recreational value is severely diminished (Weiskel et al., 1996).

The Hoosic River runs through the historically industrial towns of northern Berkshire County in Massachusetts and drains into New York's Hudson River. The entire Hoosic River watershed drains approximately 1865 square kilometers of Massachusetts, Vermont, and New York. According to the Massachusetts DEP, high levels of fecal coliform bacteria have been documented in parts of the upper Hoosic River Watershed, and current bacteria levels need to be determined in order to assess the status of Primary and Secondary Contact Recreational Uses (1997).

The Massachusetts Surface Water Quality Standards (314 CMR 4.00), "designate the most sensitive uses for which the various waters of the Commonwealth shall be enhanced, maintained and protected" and "prescribe the minimum water quality criteria required to sustain the designated uses..." These Standards describe both the minimum and maximum flows at which criteria should be met. Monitoring efforts often sample a particular location only once each month during the summer to evaluate whether or not the river meets the expected standard. While regulations allow for a higher bacterial concentration during wet weather, they ignore patterns in bacterial fluctuation throughout the day and over the course of the year.

Bacterial concentration is expected to increase with wet weather as evidenced by the higher cut-off value for fecal coliform bacteria suggested by the Massachusetts Department of Environmental Protection in their 1997 Water Quality Assessment of the Hudson River Basin. For less than five samples within a one month period, Fecal coliform bacteria should not exceed 400 colonies per 100 mL sample, while wet weather samples must only fall below 2000 colonies per 100 mL sample. Hunter and McDonald (1991) have documented this behavior in the UK, where they described the relationship between bacterial concentration and individual storm events. Their research also describes seasonal fluctuations, with bacterial concentration being lower during the winter months. Frenzel et al. record high variability of fecal-indicator bacteria in samples taken over a two-day period of stable stream flow and suggest that this "may have implications for testing of compliance to water-quality standards (2002)." Ignoring these potentially large seasonal and diurnal fluctuations in bacterial concentration clearly leads to a poor overall assessment of a watershed's bacterial pollution.

Rodgers et al. (2003) monitored the correlation of bacterial concentration and increased discharge caused by summer storms and noted that the strength of this correlation varied with how wet conditions were prior to each storm event. They noted that bacterial concentrations during storms closely following other storms tended to be lower that those concentrations during storms following dry weather. They suggested that storms flush out bacteria from the bed sediment and from the watershed surfaces, leaving a lower bacterial concentration available for suspension by a storm following closely thereafter. Jameison et al. (2003) and Crabill et al. (1999) documented the presence of fecal microorganisms in stream sediments and also suggested that this may be a significant source of bacterial loading to the water column during storms.

Weiskel et al. (1996) documented a similar seasonal pattern in bacterial concentration in Buttermilk Bay in Massachusetts. In this embayment, they observed higher fecal coliform (FC) concentrations in streams draining into the bay during the spring and summer. They suggested three possible mechanisms, "(1) the increased wet-weather flows associated with precipitation events in the warm months, when temperatures are always above freezing, (2) possible warm season increases in wildlife FC loading to watershed surfaces... contributing runoff to the streams, and (3) the possible effect of temperature-induced stress on indicator survival and (or) culturability in the winter samples."

Gameson and Saxon (1967) found that coliform bacteria in seawater died-off in the presence of sunlight. Fujioka et al. (1981) found that sunlight inactivated fecal coliforms and fecal streptococci in seawater, but not in freshwater from mountain streams. Boehm et al. (2002) observed diurnal fluctuations in bacterial concentrations in the surf zone of Huntington Beach in California. They were able to show that sunlight-induced bacterial die-off was responsible, at least in part, for this pattern. Inactivation of E. *coli* due to exposure to sunlight at a Lake Michigan swimming beach was also documented by Whitman et al. (2004).

Method

The Upper Hoosic River Watershed drains approximately 427 square kilometers of northwestern Massachusetts, all located within Berkshire County. This watershed varies in land-use, including historically industrial, urbanized areas, as well as farms and forested lands. The most recent Water Quality Assessment of the Hudson River Basin was prepared by the Massachusetts DEP in 1997. This assessment identified stream reaches where results from past bacterial sampling have proven inconclusive and where further study has been identified as necessary step to assess “Primary and Secondary Contact Recreational Uses” of these areas.

Twelve sampling sites were selected with consideration given to this 1997 Water Quality Assessment (see Figure 1 for the location of all sampling sites). Staff gauges were constructed and installed at ten of the sampling locations. One sampling site (Site 5) was located immediately downstream of a USGS real-time gauging station.

Samples were collected from all twelve sites approximately every fourteen days beginning March 28, 2004 and ending December 5, 2004. Intensive sampling over 24-hour periods and following storms was carried out during the summer at Sites 3 and 5. Sampling schedules are outlined in Tables 1-3. At the time each sample was collected, the sample site, time of day, water level (if a gauge was present), and water temperature were recorded. Recent weather conditions were determined from online weather records. Duplicate samples were collected with a 15% frequency. Field blanks consisted of autoclaved nano-pure water dispensed at the sampling site and otherwise transported and processed identically to other samples. Each sample was analyzed for *E. coli* using Colilert and the Defined Substrate Technology® developed by Idexx Laboratories. Defined Substrate Technology® simultaneously quantifies E. *coli* and total coliform density using the nutrient indicators MUG and ONPG, which are metabolized by E. *coli* and coliform enzymes, respectively. A byproduct of E. *coli* metabolism of MUG fluoresces under UV light, while a byproduct of coliform metabolism of ONPG is yellow in color. Samples were distributed into separate wells using Quanti-Tray/2000 and the Quanti-Tray® Sealer, and then the number of bacteria present in the original sample was estimated using the standard MPN method. While both E. *coli* and total coliform data were obtained, only E. *coli* data were analyzed. Quality control samples with a known range of E. coli density were analyzed successfully on March 3rd and July 15th of 2004.

To calculate percent land use/land cover in the subwatersheds above each sampling location, data layers for land use (1999) and drainage sub-basins were downloaded from the MassGIS website. For each of the twelve sampling sites, the drainage sub-basin polygons comprising the watershed above each site were merged. The land use data layer was subsequently clipped by each of these twelve sub-watersheds, and percent cover represented by agriculture, forest, open land, wetland, and urban land use classes was calculated for each sub-watershed (see Table 4 for land cover groupings).

The average relative standard deviation among all duplicate samples (N=66) was 19%, which compares favorably with previous results in this lab using this method. Two positive blanks were analyzed when samples were collected in autoclaved, previously used bottles. Subsequent samples were collected in disposable bottles, and there were no further positive blanks.

Six samples exceeded the maximum discernible *E. coli* density by saturating the sample tray.

Number of Sites	*Samples Collected*	*Total Trips*	*Frequency*
12	(15 Total) -- 1 per site -- 15% Duplication Frequency -- 1 Blank	19	Approximately every 14 days

Table 1. Ten month, biweekly watershed-wide sampling plan

Number of Sites	*Samples Collected During Each Storm*	*Total Storms*
2 (Sample Sites Number 3 & 5)	-- 1 Blank -- 15% Duplication Frequency -- <u>*Rising Hydrograph Limb*</u> One sample every two hours from each site -- <u>*Falling Hydrograph Limb*</u> One sample approximately every four hours from each site	5 (4 storms were sampled at both sites, and a fifth was sampled only at Site 3)

Table 2. Storm sampling plan

Number of Sites	*Samples per Day (over 24 hours)*	*Total Days*
2 (Sample Sites Number 3 & 5)	(29 Total) -- 1 sample every two hours from each site -- 15% Duplication Rate -- 1 Blank	5

Table 3. 24-hour sampling plan

<u>**CODE**</u>	<u>**ABBREV**</u>	<u>**CATEGORY**</u>	<u>**DEFINITION**</u>	<u>**Grouping**</u>
1	AC	Cropland	Intensive agriculture	Agriculture
2	AP	Pasture	Extensive agriculture	Agriculture
3	F	Forest	Forest	Forest
4	FW	Wetland	Nonforested freshwater wetland	Water/Wetland
5	M	Mining	Sand; gravel & rock	Urban
6	O	Open Land	Abandoned agriculture; power lines; areas of no vegetation	Open Land
7	RP	Participation Recreation	Golf; tennis; Playgrounds; skiing	Open Land
8	RS	Spectator Recreation	Stadiums; racetracks; Fairgrounds; drive-ins	Urban
9	RW	Water Based Recreation	Beaches; marinas; Swimming pools	Water/Wetland
10	R0	Residential	Multi-family	Residential
11	R1	Residential	Smaller than 1/4 acre lots	Residential
12	R2	Residential	1/4 - 1/2 acre lots	Residential
13	R3	Residential	Larger than 1/2 acre lots	Residential

15	UC	Commercial	General urban; shopping center	Urban
16	UI	Industrial	Light & heavy industry	Urban
17	UO	Urban Open	Parks; cemeteries; public & institutional greenspace; also vacant undeveloped land	Open Land
18	UT	Transportation	Airports; docks; divided highway; freight; storage; railroads	Urban
19	UW	Waste Disposal	Landfills; sewage lagoons	Urban
20	W	Water	Fresh water; coastal embayment	Water/Wetland
24	PL	Power lines (part of #6)		Open Land
26	RG	Golf (part of #7)		Open Land
31	UP	Urban public (part of #17)		Open Land
32	TF	Transportation facilities (part of #18)		Urban
34	CM	Cemeteries (part of #17)		Open Land
35	OR	Orchard (part of #21)		Agriculture
36	N	Nursery (part of #21)		Agriculture

Table 4. Land-Use classification grouping

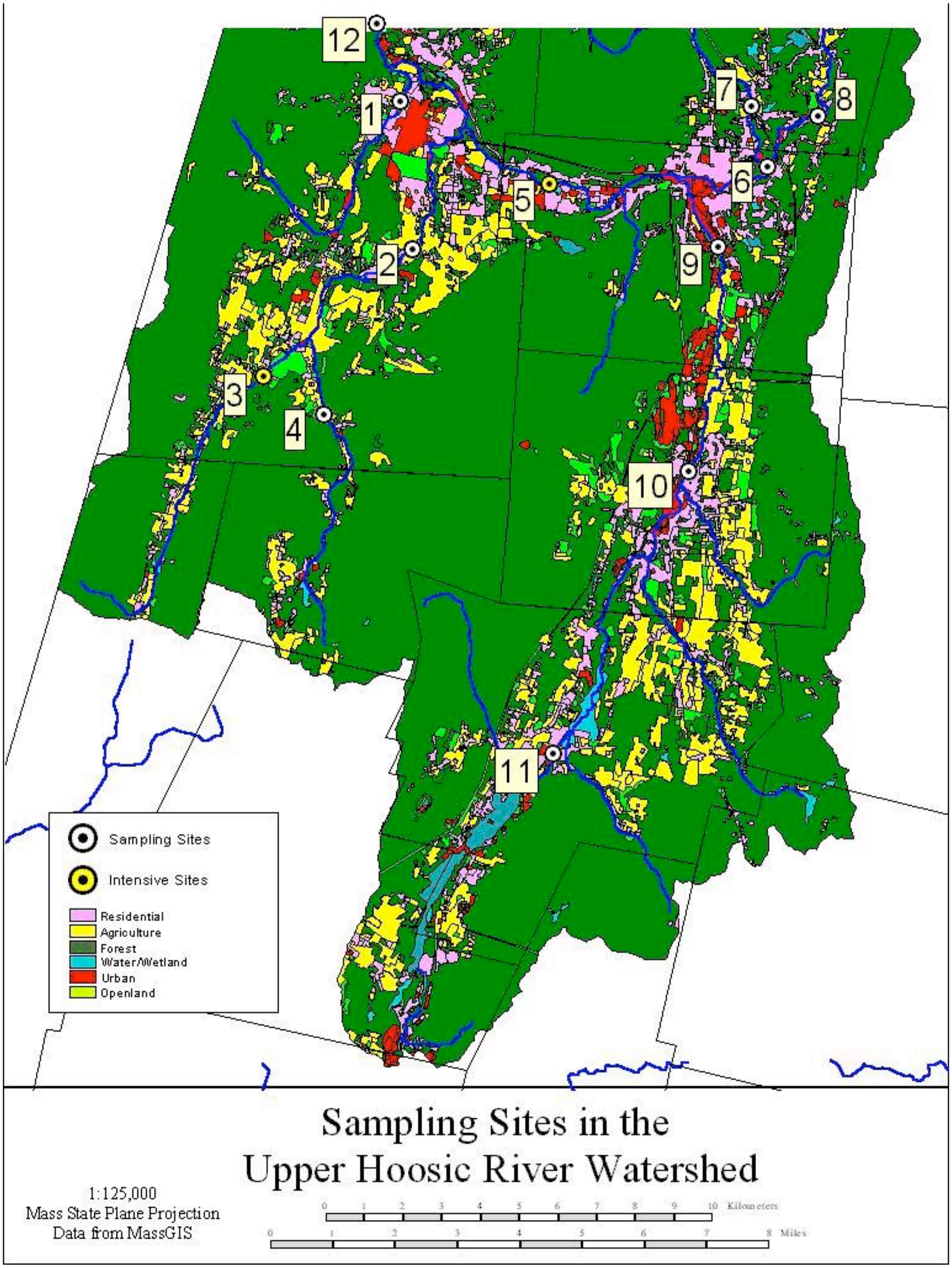

Figure 1. Sampling sites and land use in the upper Hoosic River watershed

Data Analysis

Ten-Month Watershed Sampling Data

All statistical analyses performed for this study were carried out using MiniTab. To determine which grouping of variables best predicted *E. coli* density at the twelve sites sampled throughout the Hoosic River Watershed over ten months, I used best subset regression analysis. Best subset regression evaluates models constructed with all possible combinations of predictor variables and is used in deciding which variables to include in a regression model. The following variables were used in best subsets regression analysis to predict ln(*E. coli* density): percent land cover in agriculture, forest, and urban land use classes, discharge (determined from the real-time USGS gage at Site 5), temperature, whether the sample tray was saturated (exceeding the maximum discernible density), and whether it had rained during the twenty-four hours prior to the sample being collected. Site indicator variables were not included at this step. No further transformations were necessary, as the trends between predictor variables and ln(*E. coli*) appeared reasonably linear. The subset meeting the Cp selection criteria (Cp value less than the number of variables plus one) included all variables.

I ran a regression using the variables identified by best subsets regression, which showed all predictors except percent agriculture to be significant at the 95% confidence level. Site indicator variables were added one by one to the model, and the Site 10 indicator was the only significant site predictor at the 95% confidence level. The final regression was able to predict ln(*E. coli* density) with an adjusted R^2 value of 66.4%. Variables included in this model, as well as their coefficients and p values are shown in Table 6. Residuals appeared to be normally distributed and did not exhibit any trends when plotted alongside observation order or fitted value.

Variable	Coefficient	p Value
% Urban Cover	65.14	0.000
% Forest Cover	-6.209	0.000
Saturated	1.4130	0.000
Rain Previous 24 Hours	1.2940	0.004
Temperature	0.11711	0.000
Discharge	0.0007491	0.000
Site 10	1.8658	0.013

Table 6. Regression variables, coefficients, and p values for a model of the 10-month sampling data. The model had an adjusted R^2 value of 66%.

24-Hour Trends

To determine which grouping of variables best predicted *E. coli* density at the two sites sampled over 24 hours on five different days, I used best subset regression analysis. The following variables were used in best subsets regression analysis to predict ln(*E. coli* density): indicator variables for each separate day, an indicator variable distinguishing one site from the other, an indicator variable for one outlying data point, curves representing sunlight exposure at each site, temperature, water level at Site 3 (water level was recorded from the staff gage installed at Site 3) and discharge at Site 5 (discharge was recorded from real-time online data from the USGS gauge at Site 5). No further transformations were necessary, as the trends between predictor variables and ln(*E. coli* density) appeared reasonably linear. A sun exposure curve was generated to increase sinusoidally starting at 5:30 am (sunrise), reach a maximum at 12:30 pm, and then return to zero at 8:30 pm (sunset). This curve is equal to zero at all points between 8:30 pm and 5:30 am. Sun exposure curves were generated in the same manner for both sites but were included as separate variables in order to tease out the difference significance this predictor had on each site. Therefore, for each data point collected at Site 3, the sun curve for Site 5 would have a value of zero, and visa versa. The first two subsets that met Cp selection criteria (Cp value less than the number of variables plus 1) included nine variables. Among these two subsets, all variables were included.

I created a general linear model using the all of the variables that were included in the best subsets regression discussed above. Date was included as a random factor because this variable constituted only a sample of all possible levels associated with the effect associated with date. Each level of this random factor, rather than contributing a fixed amount towards the expected value of ln(E. *coli*), has an effect that is treated as a sample from a normal distribution in the regression model. Factors that were not significant at the 95% confidence level were removed from the model in order of p value (largest first). An interaction between the sun exposure curve at Site 5 and the date was included to help explain the observation that the magnitude of "sunlight-induced die-off" appeared to vary somewhat among sampling days. The final model showed all predictors, except for water level and sun exposure at Site 3, and discharge at Site 5, to be significant at the 95% confidence level. This regression was able to predict ln(*E. coli* density) with an adjusted R^2 value of 80%. Date accounted for 14% of the overall variability of ln(*E. coli* density) in the model. Variables included in this model, as well as their coefficients and p values are shown in Table 7. Residuals appeared to be normally distributed and did not exhibit any trends when plotted versus observation order or fitted value.

Variable	Coefficient	p Value
Date		0.000
Site 3 Indicator 0	0.8240	0.000
Outlier 0	-1.1861	0.000
Site 5 Sun Exposure	-0.5248	0.000
Temperature	-0.15095	0.000
Date*Site 5 Sun Exposure		0.025
6-08-2004	-0.3783	0.047
6-11-2004	0.0212	0.909
6-15-2004	-0.2943	0.120
6-20-2004	0.5123	0.007

Table 7. Model variables (including values of categorical predictors), coefficients, and p values for 24-hour sampling. The model had an adjusted R^2 value of 80%.

Storm Sampling

I used best subset regression analysis to determine which grouping of variables best predicted *E. coli* density at Site 3 and Site 5 in wet weather. Best subset analysis was performed separately for each site. The following variables were used to predict ln(*E. coli* density) at Site 3: the generated sun exposure curves for each site, water level (read from the staff gage), temperature, whether the hydrograph was rising at the time the sample was taken, and an indicator variable for each storm. The subset with the lowest Cp value (7.0) and highest adjusted R^2 value (86%) was selected. This subset included all of the above variables, except whether the hydrograph was rising at the time the sample was taken.

I created a general linear model using all of the variables identified by the subset discussed above. Storm number was included as a random factor. Only storm number and water level were significant predictors at the 95% confidence level. The final model, including only storm number and water level to predict ln(*E. coli* density) had an adjusted R^2 value of 83.75%. Storm number accounted for 38.0% of the overall variability of ln(*E. coli* density) in the model. Variables included in this model, as well as their coefficients and p values are shown in Table 8. Residuals appeared to be normally distributed and did not exhibit any trends when plotted alongside observation order or fitted value.

The same variables were used in the best subsets regression carried out for Site 5, using discharge at Site 5 (as obtained from the USGS real-time online data) in place of "water level." The subset with the lowest Cp value (4.6) and highest adjusted R^2 value (42.9%) was selected. This subset included all of the above variables, except the sun exposure curve and the indicator variable for storm 3.

I created a general linear model using all of the variables identified by the subset discussed above. Storm number was included as a random factor. Sun exposure and storm number were not significant predictors of ln(*E. coli* density) at the 95% confidence level. The final model, including temperature, discharge, and whether the hydrograph was rising, as well as an interaction between discharge and whether the hydrograph was rising, had an adjusted R^2 value of 45%. Variables included in this model, as well as their coefficients and p values are shown in Table 9. Residuals appeared to be normally distributed and did not exhibit any trends when plotted alongside observation order or fitted value.

Variable	Coefficient	P Value
Storm		0.000
Water Level	15.165	0.002

Table 8. Model variables, p values, and coefficients for fixed effects for storm sampling at Site 3. The model had an adjusted R^2 value of 86%.

Variable	Coefficient	P Value
Temperature	-0.24452	0.000
Rising Hydrograph 0	-1.4680	0.002
Discharge	0.006264	0.045
Discharge*Rising Hydrograph 0	0.008586	0.007

Table 9. Model variables, coefficients, and p values for storm sampling at Site 5. The model had an adjusted R^2 value of 45%.

Results

Table 5 shows percent land use/land cover for each of the twelve sub-watersheds drained at each of the twelve sampling sites.

Site	Agriculture	Forest	Water/Wetland	Urban	Open/Park
1	8.6%	82.3%	0.1%	0.5%	2.6%
2	12.5%	79.8%	0.3%	0.2%	3.9%
3	13.4%	80.6%	0.4%	0.0%	2.6%
4	5.4%	91.3%	0.6%	0.3%	1.0%
5	10.3%	73.1%	2.1%	2.2%	4.0%
6	5.7%	80.2%	2.3%	0.9%	2.0%
7	3.9%	89.6%	0.8%	1.0%	1.6%
8	5.9%	80.3%	3.1%	0.4%	1.5%
9	12.6%	70.7%	2.3%	2.5%	4.6%
10	13.5%	74.4%	3.0%	0.7%	2.7%
11	9.3%	75.8%	5.4%	1.5%	3.0%
12	10.9%	74.6%	1.5%	1.6%	4.1%

Table 5. Percent land cover statistics for sub-basins drained at each sampling location

Figure 2 shows ln(*E. coli* density) of all samples collected throughout the upper Hoosic River watershed over ten months at each of the twelve sampling sites. *E. coli* density appears to increase following wet weather. When wet weather samples (no rain reported for the 24 hours prior to sampling) are removed from the graph, it is also more apparent that density increases from spring to summer, and then decreases again in the fall (Figure 3). Site 4, draining the sub-basin with the greatest percent forest cover, consistently exhibits relatively low *E. coli* density, while Site 9, draining the sub-basin with the least percent forest cover, consistently exhibits relatively high *E. coli* density.

Figures 4 and 5 show ln(E. coli density) of samples collected concurrently at Sites 3 and 5 over twenty-four hours on five different days. There is a noticeable decline of E. coli during the day at Site 5. Site 5 is located on the main stem of the Hoosic River at the downstream end of several miles of flood control structures that include earthen berms and floodwalls, which greatly increases exposure of the channel to sunlight. Site 3 is located on the smaller Green River, which is significantly more shaded than the main stem of the Hoosic River. Figure 6 shows diurnal temperature fluctuation at Sites 3 and 5 recorded when samples were collected for *E. coli* analysis. Stream temperature tends to be lower at Site 3, which is likely due to the increased shade at this site. However the magnitude of the temperature variation is quite similar at the two sites. This is likely due to the fact that although the main stem of the Hoosic River is exposed to more sunlight than the Green River, there is also more water to heat, and so the size of the stream dampens the diurnal temperature variation.

Storm sampling data for Site 5 and Site 3 are shown in Figures 7 and 8, respectively. In general, *E. coli* density appears to increase and decrease with the hydrograph.

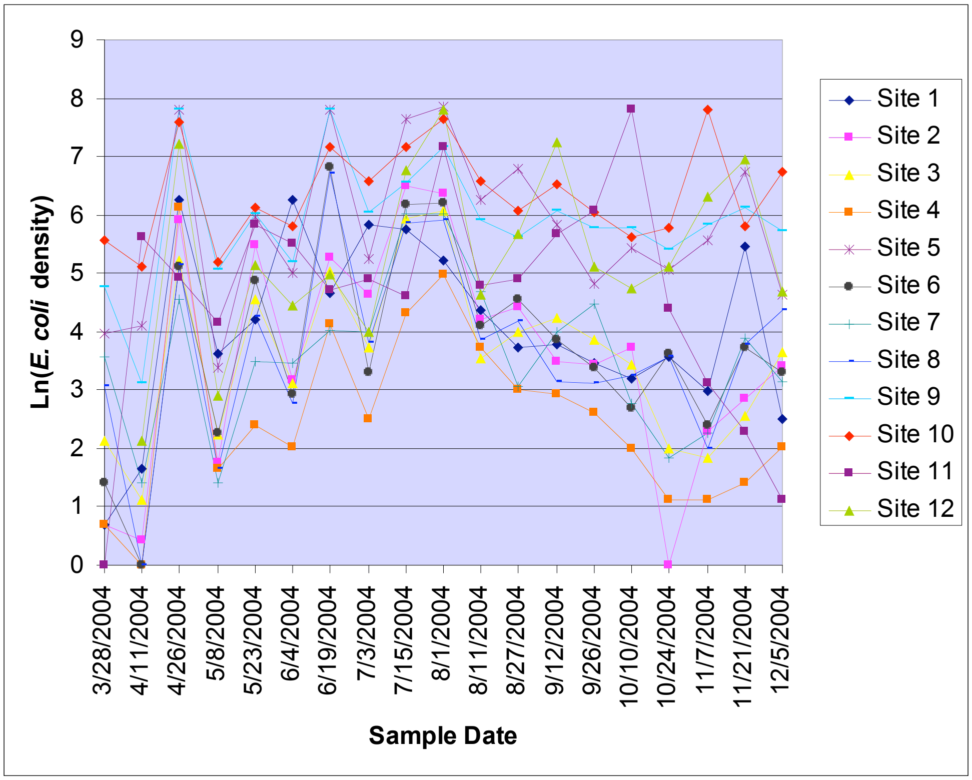

Figure 2. Samples collected at twelve sites over ten months

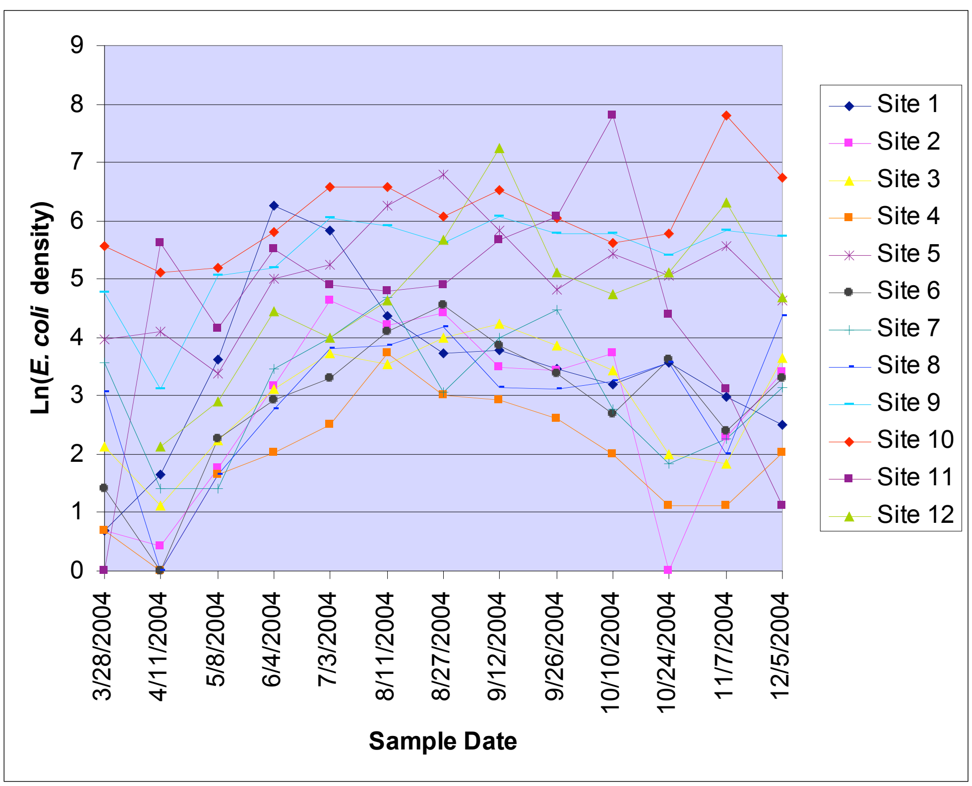

Figure 3. Dry weather samples collected at twelve sites over ten months

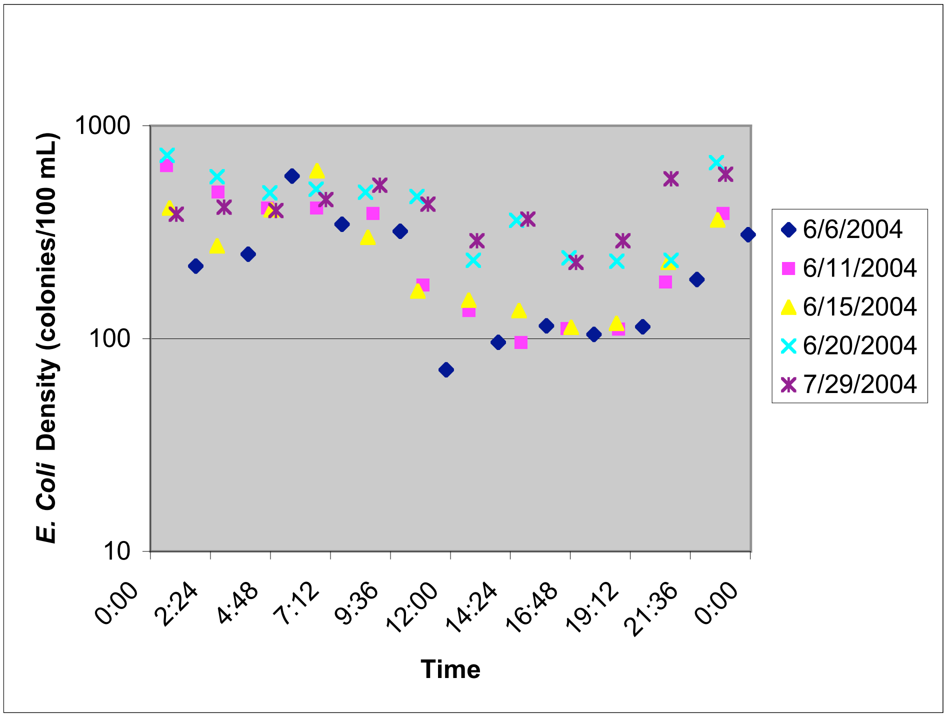

Figure 4. Diurnal *E. coli* fluctuation at Site 5 as sampled on five different days

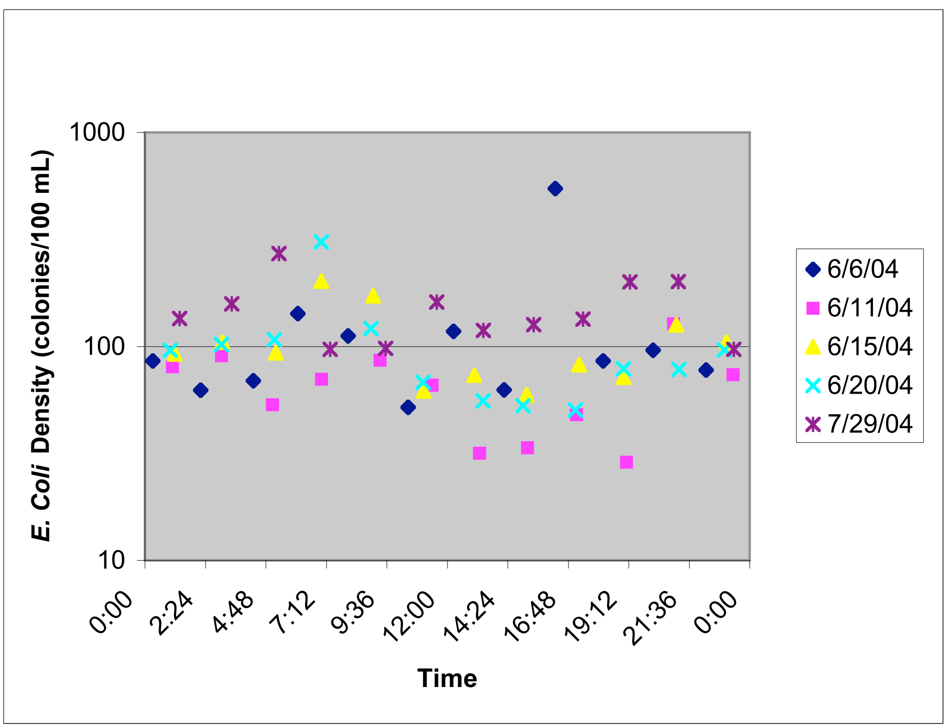

Figure 5. Diurnal *E. coli* fluctuation at Site 3 as sampled on five different days

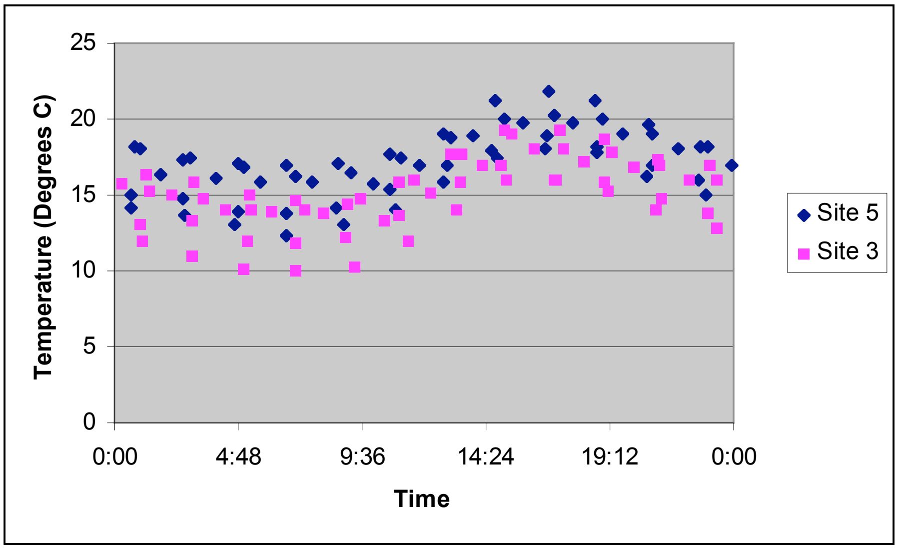

Figure 6. Diurnal temperature fluctuation at Site 3 and Site5

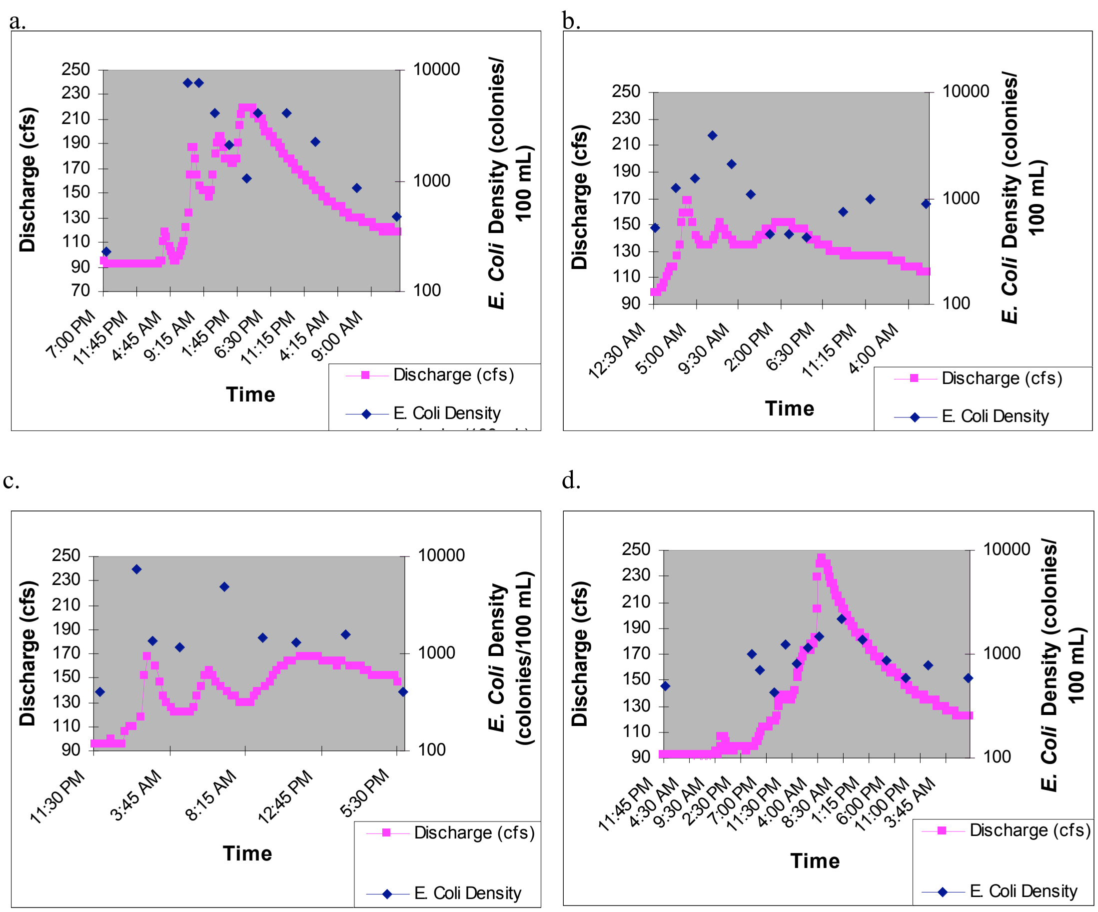

Figure 7. *E. coli* and water level at Site 5 during a storm on (a) 6/25-6/27/2004, (b) 6/29-6/30/2004, (c) 7/1-7/2/2004, and (d) 7/4-7/7/2004.

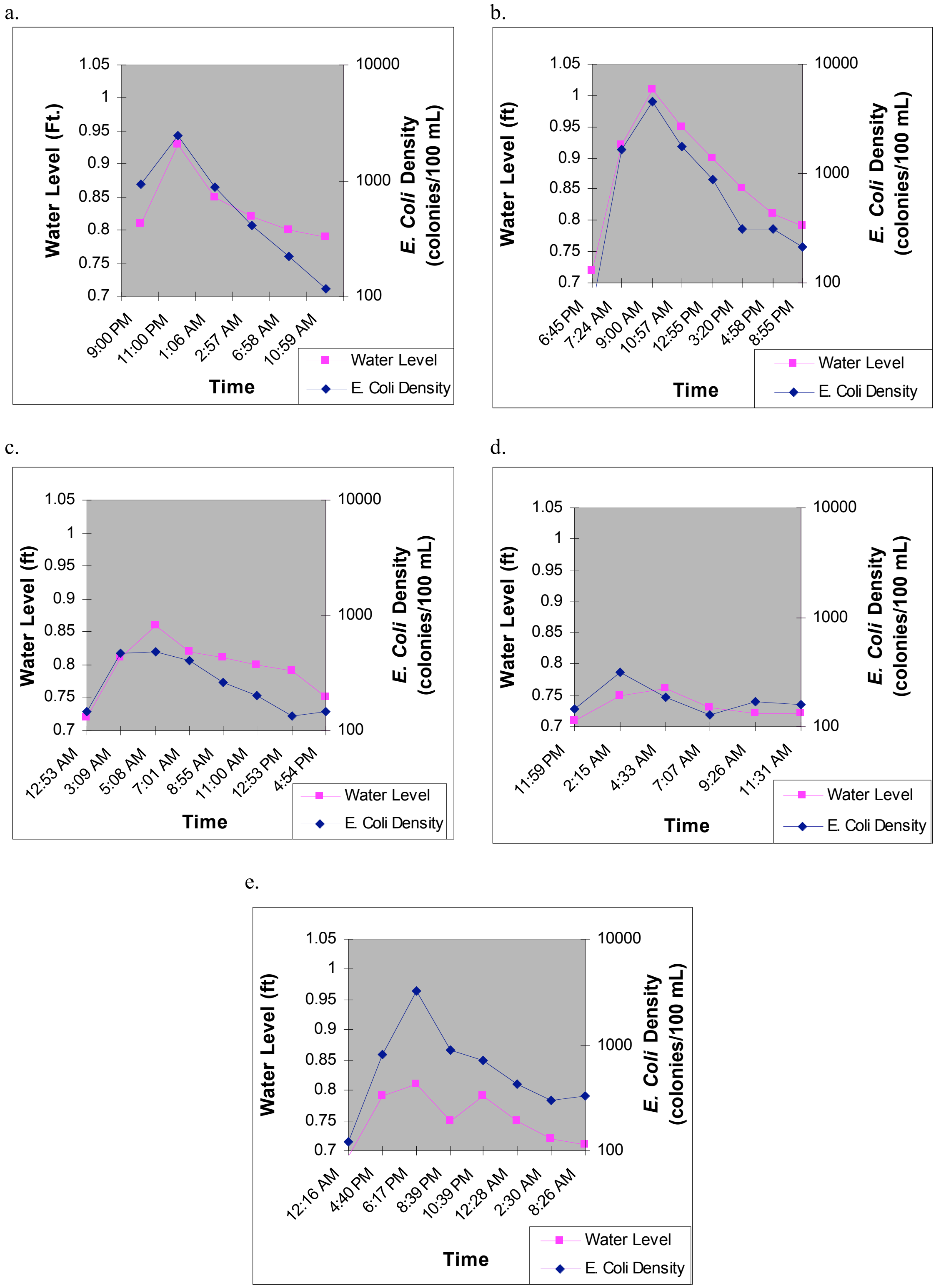

Figure 8. *E. coli* and Water Level at Site 3 During a Storm on (a) 6/22-6/23/2004, (b) 6/25-6/27/2004, (c) 6/29-6/30/2004, (d) 7/1-7/2/2004, and (e) 7/4-7/7/2004.

Discussion

10-month Watershed Sampling

The negative relationship between *E. coli* density and percent forest cover is evidenced by the negative regression coefficient and by the tendency for sites draining the most forested watersheds to have lower *E. coli* density. Lower *E. coli* density might be expected in streams draining areas with higher percentages of forest cover due to two possible mechanisms: 1) in forested areas (with little impervious cover), there is less run-off and basin surface contamination is not so quickly washed into the stream, and 2) forested areas have less development and less risk of containing leaky septic systems. This rationale could also explain the positive relationship between *E. coli* density and percent urban cover.

Higher discharge is associated with higher *E. coli* levels throughout the Hoosic River watershed over the ten month sampling period. Recent precipitation events also appear to be associated with higher *E. coli* density, as signified by the positive coefficient of this variable in the regression model.

Judging from the regression coefficients, the relationship between *E. coli* density and temperature is positive, which is evident in the tendency for *E. coli* density to increase and decrease with temperature over the course of ten months. Higher temperatures would be expected to lead to faster metabolism and increased survival of bacterial organisms, in comparison with very low, winter temperatures.

Site 10 was identified as a significant predictor of *E. coli* density, and this site was associated with higher *E. coli* levels than predicted by the rest of the model (signified by the positive regression coefficient). This site drained a sub-basin with a comparatively low percent forest cover and high agricultural cover, and generally exhibited the highest E. coli density, likely being nearby an upstream point source of bacterial contamination.

24-Hour Sampling

The positive coefficient for the zero value of the indicator variable designating Site 3 signifies the tendency for E. coli density to be lower at this site than at Site 5. Site 3 was chosen as a more pristine reference site, draining a sub-watershed with a high percent forest cover and with less significant stream channel modification in comparison with Site 5, so this trend was not unexpected.

Temperature has a negative coefficient in this model, which suggests an inverse relationship between temperature and ln(*E. coli* density) in short-term *E. coli* trends. The relationship between temperature and *E. coli* density cannot easily be separated from the response of *E. coli* density to exposure to sunlight, since temperature increases upon exposure to sunlight.

It was interesting to find that a diurnal cycle of *E. coli* density was not as obvious at Site 3 as it was at Site 5, and the curve generated to represent sun exposure was not a significant predictor of ln(*E. coli* density) at Site 3. As discussed above, the magnitude of diurnal temperature cycling was similar between Sites 3 and 5, but Site 5 is located downstream of several miles of river where stream cover has been significantly reduced. These findings are in line with research suggesting that diurnal *E. coli* cycling in surface water is due to exposure to sunlight (Boehm et al., 2002).

The significance of the interaction between sun exposure at Site 5 and date as a model predictor suggests that diurnal bacterial die-off upon exposure to sunlight is somewhat variable day to day.

Storm Sampling

Trends in *E. coli* density during storm flow were associated with the hydrograph in several ways. Higher water level/discharge was associated with higher E. coli density. At Site 5, samples taken while the hydrograph was rising tended to be higher than those taken while the hydrograph was falling and, furthermore, at this site there also appeared to be an interaction between discharge and whether the hydrograph was rising. The tendency for *E. coli* density to be higher while the hydrograph is rising may be due to the "first flush effect", where bacterial contamination is washed off of the surface of the watershed during the early part of the storm. Draining a more "pristine" watershed, Site 3 may not receive as great of an *E. coli* load from overland flow, which would explain the insignificance of the "rising hydrograph" factor at the 95% confidence level at this site.

The negative relationship between temperature and *E. coli* density at Site 5 may also reflect "sunlight-induced die-off," as *E. coli* levels decreased in the daylight even during storm flow. Whether bacteria are responding to increases in temperature or direct sunlight exposure is difficult to discern, however this effect is again only apparent in at Site 5.

It was also interesting to note that the model fit for Site 3 was able to explain 84% of the overall variability in ln(*E. coli* density), while the model fit for Site 5 (draining a larger and more urbanized watershed) was only able to explain 45% of the overall variability. This suggests, not unexpectedly, that there are more complex factors influencing levels of bacterial contamination in larger, more populous watersheds.

Implications for Water Quality Sampling

In light of the regular fluctuations in *E. coli* density that occur over space and time in this small northeastern watershed, it is clear that EPA water quality monitoring guidance that distinguishes only between dry and wet weather samples does not yield an estimate of the range of *E. coli* density occurring in a stream. Furthermore, without reducing the expected variability that would exist between samples due to diurnal and seasonal cycling, whether or not a body of water falls under the EPA recommended *E. coli* density cut-off values would be expected to be different depending on when samples were collected.

This study suggests that the lowest *E. coli* densities would likely be found in the winter during dry weather in streams draining watersheds with relatively high percent forest cover. This study did not investigate whether diurnal *E. coli* fluctuations also occur during the winter. The highest *E. coli* densities, on the other hand, should be found during the summer during wet weather in streams draining watersheds with relatively low percent forest cover before dawn.

Water quality monitoring programs may achieve a more comprehensive evaluation of bacterial contamination in a watershed by analyzing *E. coli* density at times when bacterial concentration should be lowest and times when it should be highest. This study clarifies that "representative" samples should be collected not only under wet or dry weather, but also at the same time of year and at the same time of day. By reducing the variability between samples, it would be possible to determine more specifically when and under what conditions a body of water is safe for human contact and which bodies of water consistently fail to meet EPA criteria for bacterial contamination.

References

Boehm, A. B., Grant, S. B., Kim, J. H., Mowbray, S. L., McGee, C. D., Clark, C. D., Foley, D. M., and Wellman, D. E. *Environmental Science and Technology* **36**, 3885 (2002)

Crabill, C., Donald, R., Snelling, J., Foust, R., and Southam G. *Water Research* **33** (9) 2163 (1999).

Environmental Protection Agency, Implementation Guidance for Ambient Water Quality Criteria for Bacteria, May 2002 Draft

Frenzel, SA, and Couvillion, *CS Journal of the American Water Resources Association* **38** (1) 265 (2002).

Fujioka, Roger S., Hashimoto, Harlan H., Siwak, Edward B., and Young, Reginald H. *Applied and Environmental Microbiology* **41** (3) 690 (1981)

Gameson, A. L., and Saxon J. R. *Water Research* **1,** 279 (1967)

Hudson River Basin 1997 Water Quality Assessment Report, 2000. Report #11/12/13-AC-1. Prepared by: Laurie E. Kennedy and Mollie J. Weinstein. 98 pages.

Hunter, Colin and McDonald, Adrian *Water Resources* **25**, 447 (1991).

Jameison, R.C., Gordon, R.J., Tattrie, S.C., and Stratton, G.W. *Water Quality Research Journal of Canada* 38, (1) 33 (2003)

Massachusetts Surface Water Quality Standards, 314 CMR 4.00.

Rodgers, P., Soulsby, C., Hunter, C., and Petry, J. *The Science of the Total Environment* **289**, 314 (2003)

Tong, STY, and Chen WL. *Journal of Environmental Management* **66** (4) 377 (2002)

Weiskel, Peter K., Howes, Brian L., and Heufelder, George R. *Environmental Science and Technology* **30**, 1872 (1996)

Whitman, Richard L., Nevers, Meredith B., Korinek, Ginger C., and Byappanahalli, Muruleedhara N. *Applied and Environmental Microbiology* **70** (7) 4276 (2004)