GENETIC DIVERSITY OF *ESCHERICHIA COLI* IN SOILS AND SEDIMENTS OF AN AGRICULTURAL WATERSHED AND THEIR SPATIOTEMPORAL INFLUENCES ON WATER QUALITY

by

Gregory Stuart Piorkowski

Submitted in partial fulfilment of the requirements
for the degree of Doctor of Philosophy

at

Dalhousie University
Halifax, Nova Scotia
November 2013

© Copyright by Gregory Stuart Piorkowski, 2013

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

In a series of field and watershed scale studies, the genetic diversity of *Escherichia coli* in secondary habitats (e.g. soils and sediments) of an agricultural watershed was assessed in order to examine the dynamics of *E. coli* inhabiting these matrices and to determine their contribution to waterborne populations. Using replicated field plots, persistent subpopulations of *E. coli* were observed to be significantly affected by hillslope position due to inherent differences in soil texture and moisture content. The dynamics of *E. coli* populating tile drainage effluent in a working cultivated field were monitored and it was observed that putatively naturalized *E. coli* dominated the effluent after approximately 55 days following manure amendments. The contribution of tile drainage effluents to the waterborne *E. coli* population in an adjacent stream was exponentially related to tile discharge rates, regardless of whether the effluent was populated by manure-associated or naturalized *E. coli* strains. Streambed *E. coli* populations differed according to stream geomorphological features, with strains responding to sediment texture and water velocity distributions among the features. In a temporal study of sediment *E. coli*, population turnover was observed to be affected by sediment redistribution in high-energy stream reaches and was stabilized by immigration from adjacent catchment sources in low-energy stream reaches. Reach-specific connectivity between sediment and waterborne *E. coli* populations was observed in this watershed. Reach- and catchment-scale hyporheic processes are speculated to be occurring, which may be in part influenced by strain-dependent attachment behaviour of *E. coli* strains in disjoint stream reaches influenced by different catchment sources of *E. coli*. The attachment of waterborne *E. coli* to suspended particles was observed to be associated with land use, water quality and suspended particle variables. The relationship of land use type to particle attachment reinforces the hypothesis that strain-specificity in attachment behaviour can affect the transport of *E. coli* in fluvial systems. This work provides evidence that putatively naturalized strains in cultivated fields can contribute a large part to waterborne *E. coli*, and that reach-specific hydrological factors need to be considered when relating sediment- to waterborne *E. coli* in fluvial systems.

LIST OF ABBREVIATIONS USED

AAFC	Agriculture and Agri-Food Canada
AICc	Akaike's information criterion with second-order correction
ANOVA	Analysis of variance
API	Antecedent precipitation index
BOX-PCR	Repetitive element palindromic polymerase chain reaction using BOX-A1R primers
CART	Classification and regression trees
CCA	Canonical correspondence analysis
CFU	Colony forming unit
DNA	Deoxyribonucleic acid
DO	Dissolved oxygen
DT	Drainage tile
EC	Electrical conductivity
EPEC	Enteropathogenic *E. coli*
ET	Evapotranspiration
ETEC	Enterotoxigenic *E. coli*
FIB	Fecal indicator bacteria
FST	Fecal source tracking
GC	Gene copy number
GUIDE	Generalized, unbiased, interaction detection and estimation
LASSO	Least absolute shrinkage and selection operator
LDM	Liquid dairy manure
LISST	Laser in-situ scanning and transmissometer
MARS	Multivariate adaptive regression splines
MAS	Manure-amended soil
ME	Moran's eigenvector
MST	Microbial source tracking
mtDNA	Mitochondrial deoxyribonucleic acid
NPMANOVA	Nonparametric multivariate analysis of variance
NSE	Nash-Sutcliffe efficiency
OM	Organic matter content
PAGE	Polyacrylamide gel electrophoresis
PBIAS	Percent bias
PCA	Principle components analysis
pCCA	Partial canonical correspondence analysis
PCoA	Principle coordinates analysis
PCR	Polymerase chain reaction
PSD	Particle size distribution
qPCR	Quantitative polymerase chain reaction
RCC	Rate of correct classification
REP-PCR	Repetitive element palindromic polymerase chain reaction
RSR	Root mean square error-observations standard deviation ratio
SWAT	Soil and water assessment tools
TBW	Thomas Brook Watershed

TSB	Tryptic soy broth
TSS	Total suspended solids
UPGMA	Unweighted pair group method with arithmetic mean
USSC	US System of Soil Classification
VSS	Volatile suspended solids
WEBs	Watershed Evaluation of Beneficial Management Practices

ACKNOWLEDGEMENTS

First and foremost I offer my sincerest gratitude to my supervisory committee for their leadership throughout this process. Particularly, I thank Greg Bezanson for his expertise and constructive criticisms, Lisbeth Truelstrup Hansen for her enthusiastic support and direction in scientific thinking and communication, Chris Yost for his guidance and feedback, and Rob Jamieson for his personal and professional mentorship. I am truly grateful for the opportunity to work under Rob and have the utmost respect for his ability to serve as a sounding board, counselor and confidant to all his students.

Next, I must acknowledge the peer support that I have received throughout my time at Dalhousie. Nick Dourado and Erin Mentink, in particular, were instrumental to the outcomes of this project. I am grateful for their technical help, as well as their friendship both in and out of the lab. My fellow graduate students were most influential throughout this process. I am lucky to have had the opportunity to share the multitude of triumphs and tribulations in grad school with Andrew Sinclair, Tristan Goulden, Jenny Hayward, Janeen McGuigan, Colin Ragush, Mark Greenwood, Evan Bridson-Pateman, Katie Campbell, Lee Hynes, Andrew Macintosh, Bopeng Zhang, Rebecca Mason and many others. I wish them the best of luck wherever life takes them.

Finally, much love goes out to my family for their support throughout this process. To my parents and siblings for understanding my need to venture into a PhD on the other side of the country, and their unwavering support despite my periodic absence from their lives. To my wife, Theresa, who was most understanding and supportive throughout the long hours, weekend work and interminable stress that this work required, and continues to surprise me by her personal resilience and her commitment to our family. To my son, Thomas, for providing me the motivation to complete this work and catalyzing my newfound form of personal growth. And to my dogs, Shelby and Clara, for their stalwart exuberance and helping me focus on the sillier things of life.

CHAPTER 1 INTRODUCTION

In agricultural watersheds, fecal microorganisms can be transported to surface water resources through point source pollution from residential septic systems and manure holding tanks, or nonpoint sources attributed to from pastured livestock and manure amended fields (Ritter et al., 2001). Direct disinfection of fecal material is often not conducted in rural agricultural watersheds due to cost and logistical constraints. Natural attenuation of fecal microbiota during their passage through porous media is more frequently relied upon, as with septic fields and land-applied manure (AAFC, 2006; Goss and Richards, 2008). The capacity of septic fields and soils to attenuate fecal microbiota is dependent on proper septic field design and maintenance, proper timing and application rates of livestock manures to cultivated fields, and suitable hydrological conditions that limit off-site transport to surface water sources (Goss and Richards, 2008). Despite best management practices being initiated to restrict its occurrence, fecal contamination of water resources from point- and nonpoint pollution sources in agricultural watersheds remains an important problem (Meals et al., 2009).

The primary issue with fecally contaminated water resources is the potential for concomitant introduction of enteric pathogens, which can then be transmitted to human populations through recreational exposure and consumption of inadequately processed drinking water or irrigated food crops (Solomon et al. 2002; Reynolds et al. 2008; Yoder et al. 2008). Water quality monitoring programs make use of fecal indicator bacteria (FIB) because of the economic and logistical constraints associated with direct pathogen detection (Field and Samadpour, 2007). *Escherichia coli* is the FIB most commonly used for monitoring freshwater resources, where it is assumed to have high fecal specificity

and exhibit similar environmental fate and transport to fecal pathogens (Tallon et al. 2005). However, the environmental behaviour of *E. coli* has been found to be complex and difficult to predict. Decay rates and subsequent persistence of *E. coli* in soil and sediment matrices have been reported to be a function of texture, soil moisture content, nutrient status, organic matter content, ambient temperature, and antagonistic biological influences (Cools et al., 2001; Craig et al., 2004; Zaleski et al., 2005; Unc and Goss, 2006; Ishii and Sadowsky, 2008; Haller et al., 2009; Brennan et al., 2010; Garzio-Hadzick et al., 2010; Ishii et al., 2010). Surbeck et al. (2010) argue that such complexity arises from *E. coli* being a biological entity that responds to ambient nutrient availability, resource competition with other bacteria, and predation.

The complex response of *E. coli* to environmental systems can be attributed in large part to differential gene expression among strain types that enables certain strains to achieve greater environmental adaptation than others (Foppen et al., 2010; van Elsas et al., 2011). Whereas most *E. coli* inhabiting fecal material have limited capacity for survival in environmental media, some strains may be better suited for long-term environmental persistence and may even become part of the natural, authochthonous environmental microbial community; a transition referred to as "naturalization" (Ishii and Sadowsky, 2008; Byappanahalli et al., 2012). Indeed, several studies have reported on the perennial persistence of *E. coli* genotypes in forest soils, beach sands and freshwater sediment (Byappanahalli et al., 2003; Ishii et al., 2006; Ishii et al., 2007; Halliday and Gast, 2011; Byappanahalli et al., 2012). In cultivated fields, *E. coli* has been detected in agricultural drainage waters well outside of the range of decay expected in manure-amended soils (Brennan et al., 2010; Vanderzaag et al., 2010; Esseilli et al., 2011). In

laboratory and mesocosm studies, putatively naturalized *E. coli* strains have demonstrated longer persistence in soil and sediment matrices than fecal strains (Topp et al., 2003; Anderson et al., 2005). The transport of "naturalized" strains to water resources and their subsequent detection in water monitoring programs may skew assessements of human health risk since these strains would not be indicative of fecal inputs and their associated pathogens (Ishii and Sadowsky, 2008; Van Elsas et al., 2011).

In addition to being used as a generic indicator of microbiological water quality, *E. coli* is a target organism for library-dependent microbial source tracking (MST) programs. Library-dependent MST requires isolating target organisms from a fecal source, determining phenotypic or genotypic characteristics of the isolates, compiling source libraries of the isolates, and comparing waterborne isolates to the source libraries (Stoeckel and Harwood, 2007). Thus, this method utilizes the genotypic and phenotypic variability of *E. coli* strains isolated from different host species to assign waterborne *E. coli* to a fecal source. However, "naturalized" *E. coli* strains can confound library-dependent MST programs by increasing the complexity of the waterborne *E. coli* populations and subsequently reducing the capacity for fecal source characterization under reasonable sampling efforts (Santo Domingo et al., 2007). For example, Kon et al. (2009) determined that environmentally adapated *E. coli* strains contributed up to 23% of the waterborne *E. coli* population, making a strong case for including environmental isolates in MST investigations.

Despite the complexity of its response to environmental matrices, *E. coli* remains the primary FIB for water quality monitoring programs and for simulating microbial dynamics in deterministic watershed models, such as SWAT and WATFLOOD (Dorner

et al., 2005, Kim et al., 2010). Process-based water quality models have most effectively modeled *E. coli* concentrations by recognizing that it can be derived from both: (i) catchment sources transported via surface and subsurface runoff fluxes; and (ii) channel stores, or the reservoir of bacteria associated with stream bed sediments (Rodgers et al., 2003; Jamieson et al., 2005; Wilkinson et al., 2011; Pandey et al., 2012). The potential for *E. coli* to become naturalized in soils and sediments, and be subsequently transported to the water column can affect agricultural water quality monitoring and modeling programs. Stream systems in agricultural watersheds can demonstrate chronically elevated *E. coli* concentrations above water quality guidelines (Jamieson et al., 2003; Muirhead et al., 2004; Sinclair et al., 2008). Water useage consequently would be limited in these watersheds, and perhaps unnecessarily so should "naturalized" *E. coli* be contributing the larger proportion of the waterborne population.

The research presented herein was directed at elucidating the genetic diversity of environmental *E. coli* in a chronically contaminated agricultural watershed. Over two growing seasons and an overwinter period, *E. coli* was isolated from streambed sediments, soils, and tile drainage effluent and molecularly characterized. The intent was to assess the potential for putatively naturalized strains to exist in streambed sediments and agricultural soils, to determine the spatial and temporal stability of environmental *E. coli* populations, and to evaluate the relative influence of the environmentally persistent subpopulations on waterborne *E. coli*.

Chapters 2 and 3 relate to *E. coli* populations, in terms of the presence and abundance of particular strains, in soil systems. Chapter 2 investigates the differences in *E. coli* decay and population structure according to manure application rates and hillslope

position, with associated differences in soil texture and moisture. *E. coli* populations were simultaneously compared to library-independent fecal source tracking markers. Relevant to the theme of the dissertation (i.e. *E. coli* genetic diversity), it was hypothesized that persistent *E. coli* strains would be different between drier upslope and moister toe-slope soil types and that higher nutrient availability under greater application rates would lead to different population structures given possible strain-specific responses to soil and nutrient variables.

Chapter 3 presents the results of a mensurative study that characterized *E. coli* genotypes occurring in manure-amended soils and tile drainage effluent in a working cultivated field in the Thomas Brook Watershed (TBW). Chronically persistent *E. coli* have been identified in the tile effluents of this field, and it is hypothesized that naturalized strains may be populating the tile effluents. Surface soils and tile drainage effluents were monitored for two growing seasons and an overwinter period, with solid and liquid dairy manure being applied in successive years during the spring period (May and June, respectively). The intent of this study was to determine: (i) the temporal influence of manure amendments on *E. coli* populating the tile effluents; (ii) the potential for putatively naturalized *E. coli* to exist in the tile drainage plots; and (iii) the influence of tile drainage effluents on waterborne *E. coli* in an adjacent stream.

Chapters 4, 5 and 6 relate to the genetic diversity of *E. coli* inhabiting the bed sediments and water column of the TBW stream network. Chapter 4 presents the results of a study examining spatial patterns in sediment-borne *E. coli* as a function of stream morphological features. *E. coli* has been shown to be influenced by sediment texture and organic matter content; variables that are sorted by differential hydraulic behavior in

fluvial systems into spatially distinct morphological features. It was hypothesized that *E. coli* populations would demonstrate differences in concentration and/or population structure among the reaches owing to spatially explicity differences in sediment properties. The stability of the observed spatial patterns was assessed through studying three separate monitoring locations over two different time periods.

Chapter 5 reports on the temporal stability of sediment *E. coli* populations and their contribution to waterborne populations in three stream reaches differing in their hydrological characteristics, sediment properties and adjacent upland *E. coli* sources. The temporal stability of *E. coli* genotypes in the streambed was assessed to determine if putatively naturalized *E. coli* inhabit stream sediments. Population-level similarity was correlated with environmental variabiles to determine the factors that drive change in sediment *E. coli* populations. Finally, the contribution of sediment *E. coli* strain types to waterborne *E. coli* was assessed, and the environmental variables that correlated with the contribution of sediment strain types to the waterborne population were determined.

Chapter 6 reports on the application of statistical models for predicting waterborne *E. coli* particle attachment, an important parameter for modeling *E. coli* transport in fluvial systems. Four monitoring sites differing in their receiving catchment area and land uses, as well as water quality and suspended particle properties were investigated. The objective of this study was to test the utility of recursive-based nonlinear regression models for predicting *E. coli* particle attachment, considering attachment is driven by strain-specific responses to environmental conditions and particle properties. It was hypothesized that the recursive nature of the model types used would better accommodate dynamic *E. coli* behaviour in comparison to linear models.

CHAPTER 2 EFFECT OF HILLSLOPE POSITION AND MANURE APPLICATION RATES ON THE PERSISTENCE OF FECAL SOURCE TRACKING INDICATORS IN AN AGRICULTURAL SOIL

2.1 Introduction

Non-point source water pollution from manure-amended soils represents an important factor in global water quality impairment (Ritter, 2001). Land application of untreated livestock manure is a common practice in Canadian agricultural watersheds as a means of waste disposal and as a soil amendment to introduce plant nutrients, mainly N and P (AAFC, 2006). Traditional water quality monitoring programs in agricultural watersheds use fecal indicator bacteria (FIB), such as *E. coli* and enterococci, to assess water quality and trigger potential restrictions on water uses. However, FIB offer no information regarding fecal sources, and do not allow for targeted watershed management activities to mitigate fecal pollution (Field and Samadpour, 2007). Recent advances in fecal source tracking (FST) methodologies have facilitated the determination of source contribution arising from point and non-point source water pollution, and can be adapted into integrated watershed management programs for improving agricultural water quality (Smith and Perdeck, 2004). FST studies can be conducted through library-independent and library-dependent strategies, which have been summarized elsewhere (Field and Samadpour, 2007; Stoeckel and Harwood, 2007). Library-independent markers are better suited for routine monitoring of larger geographic areas, whereas library-dependent markers are more suited for targeted sampling campaigns in small geographic areas to answer specific hypotheses.

Since non-point source water pollution from manure-amended fields is an important driver of water quality impairment in agricultural watersheds, understanding the fate and transport of FST indicators in soils is important for selecting appropriate tools for water quality monitoring programs. A fundamental assumption underpinning the use of library-independent FST markers is that the marker demonstrates consistent and predictable die-off in environmental matrices (Harwood, 2007). Most studies on the persistence of library-independent FST markers have focused on their fate in water systems, or in laboratory-based soil studies (Rogers et al., 2011; Tambalo et al., 2012). The degree to which the assumption of predictable decay holds true in agricultural soils under field settings is not clear. This information is important for water quality monitoring since manure-amended soils can either provide a source of manure-borne pathogens (Rogers et al., 2011), or confound specific hypotheses about source contributions (e.g. feedlot source vs. diffuse manure-amended soil) (Roslev and Bukh, 2011). Greater knowledge of FST marker decay would allow for a more informed approach as to how long runoff from manure-amended fields could be monitored using FST markers, or how long a manure-amended field would confound a monitoring program focusing on a specific livestock industry.

For library-dependent indicators, it is assumed that the target organism demonstrates a similar clonal composition between primary (i.e. host feces) and secondary habitats (e.g. liquid manure holding tanks, manure-amended soils) (Harwood et al., 2007). However, the clonal composition of *E. coli* shifts between primary and secondary habitats, invalidating this assumption in terms of building source-specific libraries from fecal matter alone (Gordon, 2002). Environmentally adapted strains of *E.*

coli often are distinct from host strains (Whittam, 1989; Byappanahalli et al., 2006), lending further support for probing secondary habitats for library-dependent MST indicators. Library-dependent approaches can and should still be used for water quality monitoring, but source libraries should be generated from all habitats, including soils, because of potential confounding effects (Whitman et al., 2006; Kon et al., 2007; Byappanahalli et al., 2012). The extent to which environmentally persistent *E. coli* strains respond to environmental conditions in field settings remains unknown, but Topp et al. (2003) observed differential persistence of individual *E. coli* strains in various soil textures in laboratory experiments. Agricultural practices, such as manure application rates, may also influence the population structure of persistent *E. coli* as Ishii et al. (2010) found that nutrient supply and soil moisture differentially affected persistent *E. coli* strains in laboratory incubations. Should manure nutrients provide a selective advantage for some strains, then manure application rates would affect population structure even though higher rates of fecal loading have not been found to affect the inactivation rate of total *E. coli* concentrations (Jamieson et al., 2002).

The present study was conducted along a soil toposequence (i.e. hillslope) to determine whether finer textured soils at lower slope positions, and subsequent higher moisture conditions, combined with differences in manure application rate would affect the decay rate constants of host-specific *Bacteroidales* and mitochondrial DNA (mtDNA) markers and *E. coli* population structure under field conditions. This information is important to determine whether fecal indicator decay rates and contributions in a watershed monitoring program should be assessed as a function of watershed soil properties and manure management protocols.

2.2 Materials and Methods

2.2.1 Site Description and Sampling Design

The study was conducted during June through August 2011 at Agriculture and Agri-Food Canada's (AAFC) Atlantic Food and Horticultural Research Centre in Kentville, Nova Scotia. During these months, the cumulative precipitation was 301 mm and the average daily temperature ranged between 15.0˚C and 19.3˚C. Two 180 m^2 blocks were established at upslope and toe-slope positions to represent differences in soil texture and moisture content. The upslope block was classified as a Typic Haplorthod, and the toe-slope block was classified as an Aquic Haplorthod, reflecting saturation of the soil profile at various points throughout the year. All soils within the AAFC's Horticultural Research Centre are classified as Haplorthods, and the two blocks chosen represented the greatest variation in soil texture and moisture characteristics available. Chemical and textural properties of each soil are presented in Table 2.1. Within each block, seven 4 m^2 plots were demarcated, with three plots receiving liquid dairy manure (LDM) at a rate of 12.5 L/m^2 (125 kL/ha), three plots receiving LDM at a rate of 25 L/ m^2 (250 kL/ha), and one control plot with no LDM addition. The lower application rate was chosen based on the rates used in a concurrent study on a working cultivated field (Piorkowski, unpublished) in order to compare the results from the plot to field scale. The rates used were analogous to the application rate on the working field, and no calculation of nutrient demand was performed for the blocks in this study. Each plot was separated from adjacent plots by a 2 m distance. The LDM was incorporated into the soil by rototilling to a depth of 10 cm the day following block preparation and manure addition. Incorporation was conducted to reduce microbial transport through runoff or leaching by increasing soil contact and

disturbing soil macropores, respectively (Quinton et al., 2003; Meals and Braun, 2006). Liquid dairy manure is commonly applied to croplands (96.8% of farms) in Atlantic Canada, and is often incorporated into the soil (52.3% of farms) within 2 days of application (AAFC, 2006).

Table 2.1. Soil properties of the two experimental blocks established at AAFC's Horticultural Research Centre in Kentville, Nova Scotia.

Horizon*	Depth (cm)	Texture**	pH	CEC*** (meq/ 100g)	Base Saturation (%)	Organic Matter (%)	N (%)	P_2O_5 (kg/ha)
Block 1 – Aquic Haplorthod (toe-slope position)								
Ap	0-28	L	6.3	9.1	77.0	2.8	0.13	150
Bsg	28-71	SL	5.3	8.4	78.1	1.4	0.04	16
Cg	71+	LS	5.7	9.8	95.9	1.3	0.05	52
Block 2 – Typic Haplorthod (upslope position)								
Ap	0-25	SL	5.9	8.9	51.7	3.3	0.16	159
Bs	25-61	SL	5.4	5.1	38.9	1.5	0.06	112
BC	61-73	L						
C	73+	L	5.1	6.4	43	1.3	0.04	140

*A, B, C refer to the dominant soil horizons. Horizon suffixes are: p = affected by anthropogenic disturbance (i.e. plowing); s = illuvial sesquioxide concentration; and g = gleying/mottling.
**L = loam, SL = Silty loam, LS = loamy sand, as defined by the USDA–NRCS Official Soil Series Descriptions (https://soilseries.sc.egov.usda.gov/osdname.asp)
***CEC = Cation exchange capacity

Following incorporation, the blocks were hand seeded with a forage mixture consisting of creeping red fescue (*Festuca rubra*), perennial ryegrass (*Lolium perenne*), and white clover (*Trifolium repens*). Fecal indicator concentrations were obtained from manure samples collected during the time of application. Duplicate topsoil samples (0-10 cm, 300 cm^3) were collected and composited from each plot using a hand auger that was cleaned between plots at a schedule of 0, 3, 5, 11, 13, 18, 32, 44, 58, and 72 days following manure incorporation. The composite samples were passed through a 2.38 mm

diameter sieve (No. 8; W.S. Tyler, St. Catherines, ON, Canada) to homogenize the samples and remove coarse particulate matter. Microbial analyses of the soils were then performed as described below.

2.2.2 Library-Independent Indicator Analysis

Nucleic acids were extracted from each soil sample using a PowerSoil DNA extraction kit (MoBio Laboratories Inc., Carlsbad, CA, USA) following manufacturer's instruction. Approximately 250 mg of each soil sample (wet weight) was processed, and the extracted DNA was stored at -20°C prior to analysis. Three genetic markers that have demonstrated high specificity to cattle feces were used in the analysis, with two being bovine-specific (*Bacteroidales* CowM2 and mitochondrial DNA, AcytB), and one being ruminant-specific (*Bacteroidales* BacR) (Table 2.2). The qPCR assays were performed on a CFX96 Touch system (Bio-Rad Laboratories Inc., Hercules, CA, USA). Each reaction was run in a total volume of 20 µL , containing 10 µL of 2x SsoFast Probes Supermix (Bio-Rad Laboratories Inc.), 250 µM forward and reverse primers, 100 µM dual-labeled fluorescent probe and 4 µL sample DNA. Eight tenfold serial dilutions of plasmid standards (10^0-10^7) were run in triplicates to generate the standard curves. All assays were initially decomposed by incubating at 95°C for 10 min, and subsequently run for 40 cycles of 95°C for 10 min and 60°C for 60 s. Blank DNA extraction controls, no template controls, and negative DNA controls (DNA from *E. coli* ATCC 25922 culture) were included in each run. Each sample was run in duplicate and the average values were used in further analysis.

Table 2.2. PCR primer and probe sequences used to detect host-associated *Bacteroidales* and mitochondrial DNA gene markers using quantitative PCR.

Primer Target	Primer Name	Primer Sequence	Length (bp)	Ref.
Bacteroidales (Ruminant)	BacR-F	5'-GCGTATCCAACCTTCCCG-3'	118	Reischer et al. 2006
	BacR-R	5'-CATCCCCATCCGTTACCG-3'		
	BacR-Probe	5'-FAM-CTTCCGAAAGGGAGATT-NFQ-MGB-3'		
Bacteroidales (Bovine)	CowM2-F	5'-CGGCCAAATACTCCTGATCGT-3'	92	Shanks et al. 2008
	CowM2-R	5'-GCTTGTTGCGTTCCTTGAGATAAT-3'		
	CowM2-Probe	5'FAM-AGGCACCTATGTCCTTTACCTCATCAACTACAGACA-BHQ1-3'		
Mitochondrial DNA (Bovine ACytB)	Consensus mtDNA-F	5'-GCAATACACTACACATCTGACACAACAA-3'	125	Baker-Austin et al. 2010
	Consensus mtDNA-R	5'CAGATAAAAAATGATGCTCCGTTTG-3'		
	Cow-specific probe	5'-FAM-CTCCTCTGTTACCCATATCTGCCGAGACG-BHQ1-3'		

2.2.3 *E. coli* Enumeration and Isolation

Twenty grams of the <2.38 mm soil fraction, as defined above, was resuspended in 180 mL of sterile peptone-saline (0.1% peptone, 0.85% NaCl; Fisher Bioreagents, Fair Lawn, NJ, USA) by vigorously hand-shaking for 60 seconds. The suspension was allowed to settle for 10 minutes prior to collecting the supernatant. Suitable dilutions were vacuum-filtered through 47 mm, 0.45 μm cellulose-nitrate membranes (Whatman Laboratory Division, Maidstone, England). The membranes and their retentate were placed on mFC basal media supplemented with 3-bromo-4-chloro-5-indolyl-β-glucopyranoside (BCIG; Inverness Medical, Ottawa, ON) and incubated for 2 h at 35˚C, followed by overnight at 44.5˚C. Distinctly separate, blue *E. coli* colonies were counted and converted to concentration (colony forming units) per gram of soil. Soil-borne *E. coli* colonies were collected for genotyping at 58 days following manure addition. The populations

occurring after 58 days following manure addition were assumed to represent the persistent subpopulation of *E. coli* since the concentrations at all treatment plots were reduced to below 10% of the initial concentrations determined on Day 0. The presumptive *E. coli* isolates were purified by transferring the colony onto Sorbitol-MacConkey agar (Oxoid, Ltd., Hampshire, England) and confirmed as *E. coli* through enzymatic (DMACA Indole; Difco Laboratories, Sparks, MD) and molecular procedures, described below. Prior to DNA extraction, indole-positive isolates were cultured in tryptic soy broth (TSB; Difco Laboratories, Sparks, MD) at 37˚C for 24 hours. Thirty *E. coli* isolates were taken from each plot, leading to the total analysis of 360 *E. coli* isolates.

2.2.4 DNA Extraction and Genetic Analysis

DNA was extracted from TSB cultures using prepGEM Bacteria DNA kits (ZyGEM Corporation, Ltd., Hamilton, New Zealand), following the manufacturer's protocols. Each isolate was identified as *E. coli* through the PCR-based phylogenetic grouping procedure developed by Clermont et al. (2000), which assigns isolates to one of four groups (A, B1, B2, and D) based on the presence or absence of three target gene sequences. Amplification conditions were used as described by Clermont et al. (2000), except the denaturation time was extended to 10 s and the annealing and extension time was extended to 15 s, to consistently amplify all three bands in the control strain *E. coli* ATCC 25922. Strain typing was performed using repetitive element palindromic (rep)-PCR using BOX A1R primers (BOX-PCR), following protocols and cycling conditions reported by Rademaker et al. (1998). To increase electrophoretic resolution, polyacrylamide gel electrophoresis (PAGE; Montreal Biotech Inc., Kirkland, PQ) was

used, with 3.5% gels run at 150 V for 195 minutes. *E. coli* strain ATCC 25922 was used as a positive control for intergel comparison and AmpliSize Molecular Ruler (50-2000 bp; Bio-Rad, Hercules, CA, USA) was used as a size standard. PAGE gels were stained with ethidium bromide and imaged using an ImageMaster VDS-CL documentation system (Amersham Pharmacia Biotech, Inc., UK).

2.2.5 Computer-Assisted Image Analysis and Cluster Assignment

BOX-PCR images were analyzed with GeneTools software (Syngene Ltd., Frederick, MD, USA). Band matching was performed using the rolling disk method for background subtraction. Band sizes between 200 and 2,000 bp were used in subsequent cluster analysis. The similarity of individual isolates was calculated using the curve-based Pearson's product-moment correlation coefficient. Dendrograms were created using GeneDirectory software (Syngene Ltd.) through unweighted pair group method with arithmetic mean (UPGMA) with a 0.8% threshold. Isolates exhibiting ≥87% similarity were classified as clonal strains, based on the minimum similarity value for intergel comparisons of gel patterns obtained for the *E. coli* ATCC 25922 control strain.

2.2.6 *E. coli* and FST Marker Decay Rates

The decay rate of the fecal indicators in soil was calculated using the equation:

$$C_t = C_0 e^{-kt} \qquad (2.1)$$

where C_t = the concentration of bacteria at time t, C_0 = the initial concentration of bacteria, k = the first-order decay coefficient (day^{-1}), and t = time (days) (Crane and Moore, 1986). The decay coefficient was calculated through linear regression, where the dependent value was $\ln(C_t/C_0)$, the independent variable was time (t), and the regression

coefficient was the decay coefficient (-k). A separate decay rate was calculated for each plot, which comprised the individual replicate values used for further statistical analysis. The significance of soil type and LDM application rate on the MST indicator decay rates was evaluated through two-way ANOVA using SigmaPlot software (v11.0; Systat Software Inc., Chicago, IL, USA). Pairwise multiple comparisons of significant (α = 0.05) factors was performed using Tukey's test.

2.2.7 *E. coli* Population Analysis

The diversity and estimated richness of the persistent subpopulation (defined as the strains remaining in the soil 58 days following manure incorporation) of *E. coli* in the soils was analyzed through rarefaction procedures described by Lu et al. (2005). Rarefaction curves were generated through the freeware program Analytical Rarefaction 1.3, available at http://www.uga.edu/strata/software/. The curves were plotted in SigmaPlot (v11.0, Systat Software Inc.), and the asymptotes were estimated using a one-site saturation ligand model. The asymptote (V_{max}) estimates the strain richness at sampling saturation, and the K_d value estimates the number of isolates required to capture half of the total richness. *E. coli* population richness was estimated for each plot, and the influence of slope position and manure application rate on *E. coli* richness was evaluated through two-way ANOVA in SigmaPlot (v11.0; Systat Software Inc.). The similarity of *E. coli* populations was visualized through ordinations of principle components analysis (PCA) using Hellinger transformed abundance data (Legendre and Gallagher, 2001). Ordinations were produced using CANOCO software (v4.5; Plant Research International, Wageningen, The Netherlands). The influence of soil type and LDM application rate on the proportional composition of *E. coli* phylogenetic groups and genotypic population

structure was evaluated using two-way non-parametric multivariate analysis of variance (NPMANOVA) in PAST software (Hammer et al., 2001). Euclidean distance was used directly for the phylogenetic group composition, and on Hellinger-transformed abundances for population structure.

2.3 Results

2.3.1 Fecal Source Tracking Indicator Decay

The FST indicator concentrations in the liquid dairy manure determined from samples taken during manure application were calculated as 3.14×10^5 CFU/g for *E. coli*, 6.85×10^5 gene copies (GC)/g for AcytB, 1.70×10^7 GC/g for BacR and 6.09×10^4 GC/g for CowM2. After application of the manure to the plots, the initial average soil concentrations according to the low and high application rates were 3.87×10^3 CFU/g and 1.23×10^4 CFU/g for *E. coli*, 2.18×10^4 GC/g and 6.97×10^4 GC/g for AcytB, 1.06×10^5 GC/g and 4.14×10^5 GC/g for BacR, and 3.14×10^3 GC/g and 8.50×10^3 GC/g for the CowM2 markers, respectively. The dilution factors for the indicators according to the low and high application treatments were 1:81 to 1:26 fold *E. coli*, 1:31 to 1:10 for AcytB, 1:160 to 1:41 for BacR, and 1:20 to 1:7 for CowM2. The control plots were negative for all host-specific markers over the course of the study, but *E. coli* was detected infrequently (30 to 40% of samples retrieved) at concentrations ranging between 2.5 and 32.5 CFU/g. No *E. coli* or host-specific markers were recovered in soil samples obtained prior to manure application. All concentrations presented are for wet weight soil. The detection limit for *E. coli* in the soil is 10 CFU/g and the detection limit for the MST markers is approximately 500 GC/g.

The *E. coli* decay rate constants ranged between 0.045 and 0.057 day^{-1} (Table 2.3). The decay rate constants for *E. coli* did not significantly vary according to soil type (p=0.200) or LDM application rate (p=0.511), and no significant interactive effects occurred (p = 0.197). The R^2 values for the curves ranged between 0.66 and 0.79, indicating moderate linear relationship (Table 2.3). The reduced linearity in the *E. coli* decay is largely the result of the regrowth or enhanced culturability observed between 11 to 13 days after manure incorporation in response to a precipitation event and increase in soil moisture content (Figure 2.1). Although the two blocks differed in soil moisture content, this did not lead to significant differences (p>0.05) in decay rates.

Ruminant-specific *Bacteroidales* BacR marker decay was fitted with log-linear models, and exhibited R^2 values ranging between 0.77 and 0.89 (Table 2.3). Although the BacR decay demonstrated slight non-linearity because of the high initial decay rate (Figure 2.2), advanced decay models (e.g. biphasic log-linear or Weibull) could not be adequately fitted to the concentration data for each plot because of the limited sample number obtained (n=5 per plot). However, the log-linear models provided adequate mathematical fits for downstream ANOVA to be performed on the obtained decay rate coefficients. The log-linear decay rate coefficients for the BacR marker ranged between 0.212 to 0.358 day^{-1} (Table 2.3). The BacR marker decay rate constants were significantly influenced by manure application rate (p=0.029), but not soil type (p=0.228) and no interaction effect occurred (p=0.181) between the explanatory factors. The high LDM application rate plots exhibited faster decay rates than the low application rate plots in both soil types (Figure 2.2). The *Bacteroidales* markers were neither observed to increase in copy numbers nor experience wash out due to precipitation (Figure 2.2).

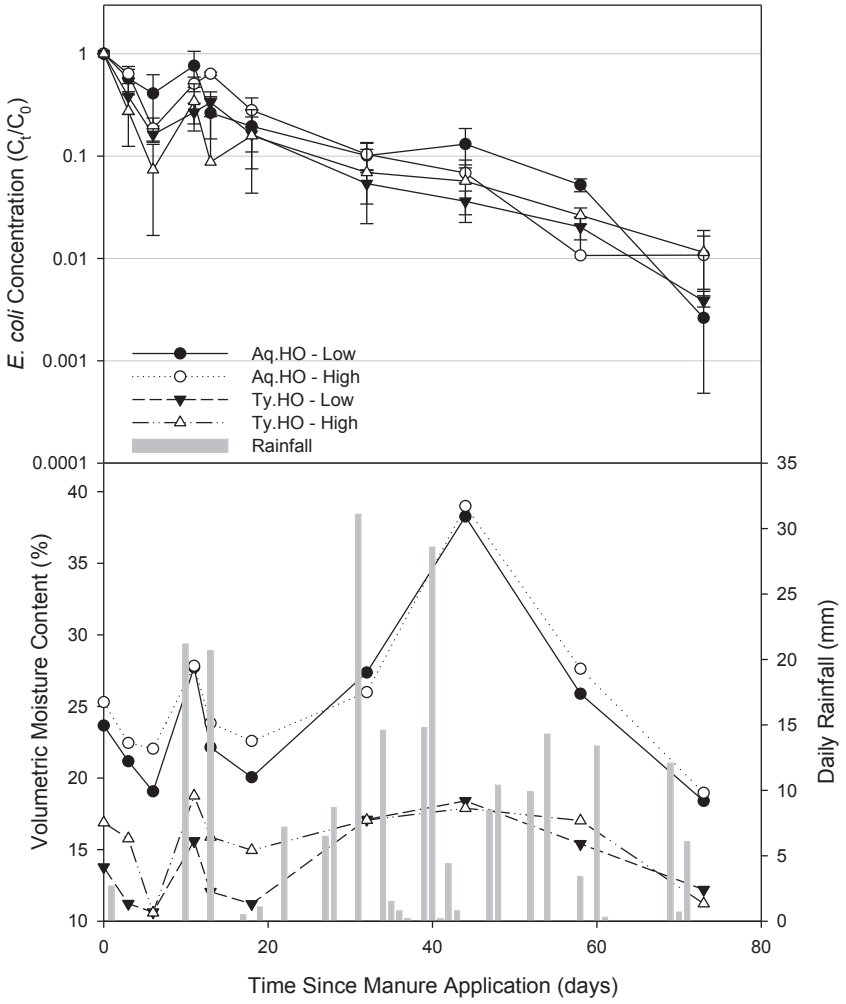


Figure 2.1. Average *E. coli* concentrations (normalized by the starting concentration (C_t/C_0)), volumetric moisture content (%), and daily rainfall (mm) for the experimental treatments measured over 76 days following manure incorporation. The experiment was conducted on Aquic Haplorthods (Aq.HO, toe-slope) and Typic Haplorthods (Ty.HO, up-slope) soil types under low and high liquid dairy manure application rates.

Decay rates for the bovine-specific *Bacteroidales* CowM2 marker could not be determined as the marker was only detected in the first (Day 0) or second (Day 3) sampling event for the majority of the plots and was last detected on Day 6 in one plot of each treatment (Table 2.3). The poor recovery of this marker in the topsoil reflects the low initial concentrations present in the manure ($6.1x10^4$ GC/g) and the amended topsoil (3.1×10^3 to 8.5×10^3 GC/g soil) relative to the other markers.

Table 2.3. Decay coefficients (k), coefficient of determination (R^2), and time to 1-log reduction (T_{90}), and last day of detection (LDD) of the marker for the replicate plots for culturable *E. coli*, and *Bacteroidales* and mitochondrial DNA gene markers in topsoil (0-10 cm) from the four experimental treatments. Arithmetic average values are presented with standard deviations (sd) in parentheses.

Soil Type – LDM Rate	*E. coli*			BacR			ACytB			Cow M2
	k (sd)	R^2 (sd)	T_{90} (sd)	k (sd)	R^2 (sd)	T_{90} (sd)	k (sd)	R^2 (sd)	T_{90} (sd)	LDD
Aq.HO – Low	0.045[a*] (0.030)	0.69 (0.17)	51 (10)	0.212 [a] (0.102)	0.89 (0.04)	11 (3)	0.088 [a] (0.005)	0.78 (0.14)	26 (2)	6,3,0
Aq.HO – High	0.047[a] (0.010)	0.66 (0.22)	49 (6)	0.358 [b] (0.092)	0.77 (0.11)	6 (3)	0.100 [a] (0.012)	0.82 (0.08)	23 (3)	6,3,3
Ty.HO – Low	0.051[a] (0.005)	0.77 (0.20)	45 (4)	0.236 [a] (0.033)	0.83 (0.08)	10 (1)	LDD = 3,0,0			6,0,0
Ty.HO – High	0.057[a] (0.009)	0.79 (0.07)	40 (3)	0.281 [b] (0.048)	0.82 (0.04)	8 (2)	0.135 [b] (0.016)	0.88 (0.04)	17 (1)	6,3,0

Aq.HO = Aquic Haplorthod (Toe-slope)
Ty.HO = Typic Haplorthod (Up-slope)
$k = - \ln(C_t/C_0)*t$ ($days^{-1}$)
$T_{90} = 2.303/k$ (days)
*Numbers in the same column followed by different letters are statistically different ($p<0.005$)

In the upslope soil (Typic Haplorthod) under low LDM application rates, the AcytB marker was last detected on the first sample event (day 0) for two of the plots and on the second sample event (day 3) for one of the plots. Hence, decay rate constants could not be calculated for this treatment. For the other treatments, the decay rate constants ranged between 0.088 and 0.135 day^{-1}, with strong linear relationships being observed with R^2 values ranging between 0.78 to 0.88 (Table 2.3). The decay rate constants for the high and low LDM treatments in the toe-slope soil (Aquic Haplorthod) were similar ($p= 0.517$). However, the decay rate in the upslope soil under high LDM application was significantly ($p<0.01$) higher than decay rates found in the toe-slope soils regardless of

the LDM application load. The AcytB marker degradation appeared unrelated to precipitation events during the experimental period (Figure 2.2).

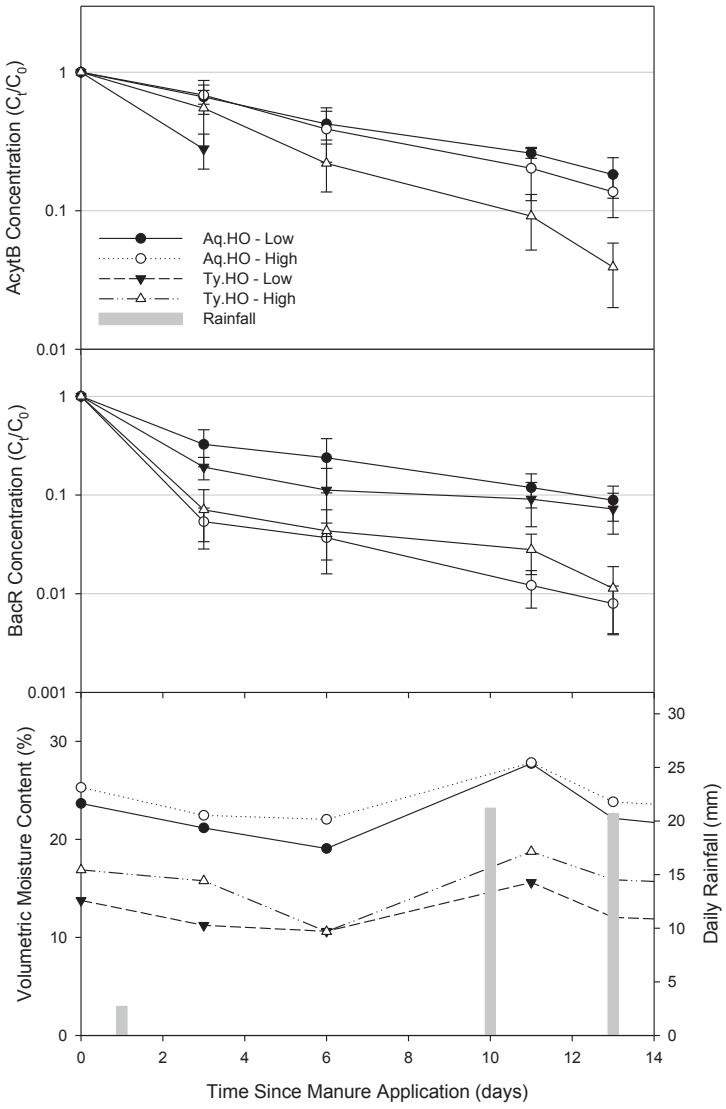


Figure 2.2. Average bovine-specific mitochondrial DNA (AcytB) and ruminant-specific *Bacteroidales* (BacR), volumetric moisture content (%), and daily rainfall (mm) for the experimental treatments measured over 13 days after manure incorporation (concentrations normalized by the starting concentrations (C_t/C_0)). The experiment was conducted on Aquic Haplorthods (Aq.HO, toe-slope) and Typic Haplorthods (Ty.HO, upslope) soil types under low and high liquid dairy manure application rates.

2.3.2 *E. coli* Population Structure

An ordination plot of the persistent *E. coli* genotypes according to the experimental treatments is presented in Figure 2.3. The axes of the ordination explains 51.9% of the variability in the population data, with the first component axis explaining 32.1% and the third component axis explaining 19.8%. The second component axis explained 21.1% of the variability, but offered poorer visualization of the data than the third axis. The treatment variables were included on the ordination as supplemental variables. The soil types follow the first component axis closely, suggesting that the first component corresponds to slope position. The *E. coli* populations appear to be separated according to soil type, as they are dispersed along the first component, but no apparent relationship exists between the loading rates. The results from NPMANOVA supported these observations, as the *E. coli* population structure was found to be significantly different ($p=0.012$) between soil types, while manure application rates ($p=0.121$) or the interaction between soil type and manure application rate ($p=0.341$) had no significant impact.

2.4 Discussion

2.4.1 Fecal Source Tracking Marker Decay

Effective FST markers must exhibit consistent or predictable environmental decay to be effective in water quality monitoring programs. The current study sought to identify whether FST markers in soils amended with LDM exhibited consistent rates of decay to determine their utility in, or confounding effect upon, non-point source water pollution studies. In the soils studied, the conventional fecal indicator *E. coli* was found to have relatively consistent decay rates between the different soil types and LDM application rates. The decay rate constants observed are consistent with those published in other

studies, which ranged between 0.02 and 0.238 day^{-1} in soils of various types (Habteselassie et al., 2008; Liang et al., 2011). The lack of significant differences in decay rates in the soils studied contrasts with other studies that found differences in decay among different soil types (Lau and Ingham, 2001; Natvig et al., 2002; Unc and Goss, 2006) or soil moisture contents (Chandler and Craven, 1980; Cools et al., 2001) where up to 6-fold faster decay was observed in drier, coarser textured soil types.

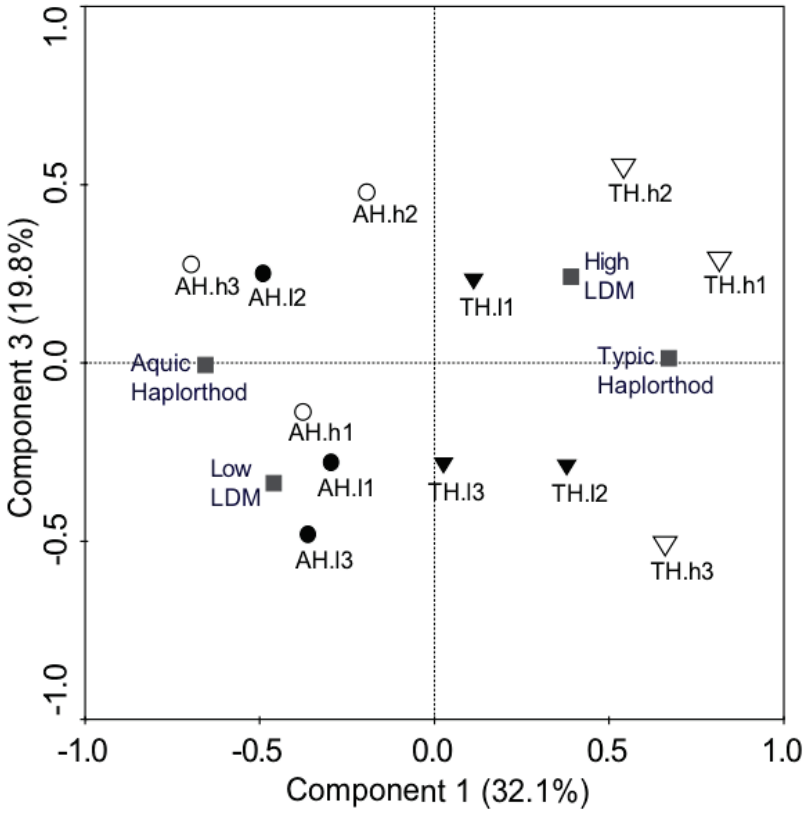


Figure 2.3. Principle components analysis of the persistent (>58 days following manure incorporation) *E. coli* genotypes according to the treatment conditions. The soil types studied are Aquic Haplorthods (AH) represented by circle symbols and Typic Haplorthods (TH) represented by inverted triangle symbols. Low (l) and high (h) liquid dairy manure (LDM) application rates are represented by black and white symbol colours, respectively. The numeric values following the LDM application designator (l, h) indicates the replicate plot number. Square symbols represent the treatment conditions that were included in the ordination as supplementary variables.

The soils used in previous studies were of different textural characters (silty clay loam vs. loamy sand), whereas this study used soils with moderate variations in fine

material content. Further, previous studies tended to compare the influence of fixed moisture contents under laboratory settings, whereas this study was field based with both soil types undergoing fluctuations in volumetric soil moisture throughout the study period. Vinten et al. (2002) also found no influence of temperature or moisture effects on *E. coli* decay rates in field soils. Previous studies also tended to use fresh manure, whereas in this study, manure had been retained in an under-barn collector cistern for several days prior to its application. Lu et al. (2005) reported that the population structure of stored manure changed in comparison to fresh manure. Manure storage lagoons represent a secondary, extra-intestinal habitat for *E. coli*, where the community population will undergo selective pressure favouring strains better adapted to the external environment. Moderate differences in soil texture and associated soil moisture content may not be sufficient to affect the rate of decay of these persisting *E. coli* strains, which is an important observation for predicting *E. coli* decay on a hillslope scale.

Numbers of recoverable *E. coli* increased in all field plots following a period of rainfall that occurred 14 to 16 days after manure application. Regrowth, or change in culturability, of FIB in manure amended soils has been observed previously (Zaleski et al., 2005; Sinton et al., 2007). *E. coli* regrowth typically occurs in the presence of metabolizable substrates and/or reduced competition from other microflora (Whitman et al., 2006). Zaleski et al. (2005) contend that *E. coli* responds to soil moisture increases only when a critical mass of organisms exists in the land-applied waste, but no growth response will occur if the population is below a threshold concentration. The observations in our study supports this assertion, since *E. coli* population densities increased in all treatments during a period of high rainfall within two weeks of manure

application, but no increases were observed during later rainfalls (32 - 44 days) when *E. coli* concentrations were lower ($2.45 \times 10^2 – 9.82 \times 10^2$ CFU/g) The exception, however, appeared to have been in the toe-slope soil type (Aquic Haplorthod) with higher moisture content, which demonstrated a slight increase in concentration 44 days after LDM application corresponding to a rainfall event.

The ruminant-specific *Bacteroidales* markers (BacR) were observed to occur in greater concentrations (1.7×10^7 vs. 6.1×10^4) than the bovine-specific *Bacteroidales* markers (CowM2). This observation is likely the result of the CowM2 marker targeting genes encoding secretive proteins (Shanks et al., 2008) whereas the BacR markers target the 16S rRNA gene (Reisher et al. 2006), which can occur up to 15 times in a bacterial genome (Vetrovsky and Baldrian, 2013). Raith et al. (2013) report that qPCR methods targeting 16S rRNA genes are consistently more sensitive and yield higher concentrations than those targeting non-ribosomal genes. Although highly specific, the use of CowM2 markers has been argued against in a muli-laboratory study due to issues of low sensitivity (Boehm et al. 2013), which may in part explain the lower concentrations observed in the cattle feces in comparison to BacR markers. In terms of estimating runoff from manure amended soils, the CowM2 marker was inferior to the BacR marker due to issues of low concentrations in the manure, initial soil concentrations, and rapid (<6 days) decay to non-detectable levels.

The ruminant-specific *Bacteroidales* (BacR) marker decay rate constants were consistent with those published in studies on water systems, which ranged between 0.22 to 1.61 day^{-1} (Sokolova et al., 2012). However, limited studies have been conducted to date on the prevalence of MST markers in soil systems, other than fecal indicator markers

by Rogers et al. (2011). The BacR marker decay rate was higher under elevated LDM application rates. Bell et al. (2009) reported that *Bacteroides* marker decay rates were unaffected by initial marker concentration, but were predominantly affected by temperature and biological activity within the environmental media. FST marker decay in water systems is consistently found to be greater in unsterile environmental media (Sokolova et al., 2012; Tambalo et al., 2012). High manure application rates could contribute to higher *Bacteroidales* decay rates by stimulating microbivorous protozoa and nematodes, leading to greater cell lysis (Forge et al., 2005); introducing or increasing bacterial biomass and the subsequent concentration of exogenous nucleases that decay free DNA (Blum et al., 1997); and/or adding organic acids that compete with free DNA for sorption sites on soil leading to greater bioavailability of the lysed DNA (Pietramallera et al. 2009). Additional studies are warranted to confirm the mechanisms of *Bacteroidales* marker decay in soils.

In contrast to the *Bacteroidales* markers, the bovine specific AcytB mitochondrial DNA marker was observed to have higher persistence in the toe-slope topsoil, and rapid decay in the upslope soil. These results suggest that the decay of mitochondrial DNA markers is influenced to a greater extent by soil texture and moisture status, rather than geochemical or biological activity affected by manure application rate. Decay rates in the upslope soil type may be greater because of desiccation-induced eukaryotic cell lysis and subsequent release of mtDNA (Huang and Tunnacliffe, 2004), and/or by greater DNA bioavailability resulting from reduced DNA sorption capacity in the coarser textured topsoil (Pietramallera et al., 2009). Of interest, the AcytB marker persisted longer in the soil environment than the *Bacteroidales* markers. Tambalo et al. (2012a) observed that a

canine-specific mitochondrial DNA marker persisted longer in water systems than a canine-associated *Bacteroidales* marker. Greater persistence of mtDNA in environmental media in comparison to *Bacteroidales* markers could be due to the extra protection afforded by the presence of organellar membranes against nuclease activity, or by differences in DNA adsorption to soil minerals and subsequent availability to exogenous nucleases (Pietramellara et al. 2009).

The fecal indicators studied differ in their decay rates as well as their responses to manure amendment application rates and soil type, as defined by hillslope position. Ultimately, the choice of which fecal indicator to use is reliant on the objectives of the water monitoring program. If runoff is likely to confound study objectives, then using indicators with low persistence in soils is recommended, which was observed to be the case with *Bacteroidales* markers in this study. If microbiological water quality deterioration from agricultural runoff is an expected outcome of the investigation, utilizing source tracking markers (e.g. mtDNA) or culturable organisms (e.g. *E. coli*) the demonstrate longer persistence in field soils is recommended. In any case, knowledge on the timing and rate of manure applications to cultivated areas within the study watershed is important for designing surface water quality monitoring programs.

2.4.2 Influence of Soil Type and Manure Application Rates on Persistent *E. coli* Subpopulations

To adequately define the contribution of non-point source pollution occurring from persistent strains of *E. coli* when using a library-dependent approach, the source libraries must capture the spatial heterogeneity of the target population of indicators. The intent of this study was to determine whether differences in soil type or LDM application rate

altered the structure of persistent (>58 day) subpopulations of *E. coli*, defined here as the viable *E. coli* strains isolated from soil plots 58 days following LDM application. Persistent *E. coli* genotypes were observed to vary significantly according to soil type, but not LDM application rate. The persistence of naturalized strains of *E. coli* in soils has been previously reported to be a function of soil nutrient status, organic matter content, and textural class (Ishii et al., 2006; Ishii and Sadowsky, 2008). The results reported here agree that hillslope position, with the concomitant differences in soil texture and moisture content, affects the population structure of persistent *E. coli* genotypes (Figure 2.3). However, nutrient status, as a function of LDM application rate, appears to have less effect at the population scale than on strain-specific responses. The *E. coli* populations present in the different soil types appear to be of the same relative complexity, as indicated by the estimated genotype richness. The differences in *E. coli* population structure between the soil types appear to be due to different genotypes being present between the upslope soil (30-32 strains) and the toe-slope soil (14-29 strains). Consequently, soil type, in terms of hillslope position and soil moisture status, is an important consideration when developing representative *E. coli* strain libraries from agricultural soils.

2.5 Conclusion

Differences in soil properties resulting from hillslope position did not affect the rate of decay of *E. coli* indicators, but influenced the population structure of the persistent (>58 days) *E. coli* genotypes present. Should studies attempt to reflect on the contribution of soilborne *E. coli* from manure amended soils in adjacent surface waters, the effect of hillslope position on *E. coli* population should be considered when developing

representative *E. coli* libraries specific to agricultural soils. Soil type also affected the persistence of the bovine mitochondrial DNA marker (AcytB), but had no effect on the ruminant-specific *Bacteroidales* marker (BacR). In contrast, the BacR marker decay rate was higher in soils that received a higher loading of liquid dairy manure, an effect not observed for the AcytB marker. Based on this study, soil type and agricultural practices appear to influence the decay rates of library-independent markers and consequently could be considered when predicting runoff from agricultural fields based on the type of marker used. However, further studies on this topic should be conducted to determine the extent to which different types of soils and agronomic practices influence MST marker persistence in field soils.

CHAPTER 3 ASSESSING THE INFLUENCE OF PERSISTENT SOIL-BORNE *E. COLI* ON TILE DRAINAGE EFFLUENT AND ADJACENT STREAM WATER USING MODELING AND GENOTYPING

3.1 Introduction

The application of livestock manure to agricultural fields represents a significant non-point pollution source, introducing a variety of microbiological and chemical contaminants to adjacent surface waters (Ritter, 2001). Land-applied manure seldom has undergone treatment, and the transport of manure-borne pathogens during precipitation events remains a critical issue for water quality impairment in agricultural watersheds (Goss and Richards, 2008). Although surface runoff is considered the primary contaminant transport mechanism from manure-amended soils, subsurface drainage effluent from tile-drained fields is another important pathway for transporting fecal microbes (Jamieson et al., 2002). Many agricultural fields in North America are artificially drained to provide suitable soil moisture conditions for agronomic production (Skaggs et al., 1994). Tile drainage systems can reduce surface runoff by increasing infiltration and soil water storage capacity, making subsurface drainage a more important pathway for contaminant transport (Skaggs et al., 1994; Watelet and Johnson, 1999). Subsurface drainage can increase health risks associated with manure-borne pathogens by providing a direct link between manure-amended soils and surface waters (Goss and Richards, 2008). In Atlantic Canada, frequent high precipitation events lead to elevated transport of manure-borne nutrients and fecal indicator bacteria (FIB) through tile drains (Kinley et al, 2007; Thiagarajam et al., 2007).

The microbiological quality of agricultural water is routinely monitored using FIB, with *Escherichia coli* most commonly used for freshwater systems. Major assumptions of FIB are that the indicator must demonstrate fecal specificity and must not be capable of proliferating in environmental media (Tallon, 2005). However, recent evidence suggests that certain strains of *E. coli* are capable of environmental persistence, surviving for long periods (even proliferating) in temperate soils and sediments; a transition referred to as "naturalization" (Ishii et al., 2006). The proliferation and potential naturalization of *E. coli* in temperate soils questions the continued use of *E. coli* for water quality monitoring (Brennan et al., 2010; Byappanahalli et al., 2012). Although typically present in higher concentration following manure application, *E. coli* has been observed in tile drainage systems well outside the range of decay expected in manure-amended soil (Brennan et al., 2010; Vanderzaag et al., 2010; Esseili et al., 2011). Previously attributed to wildlife frequenting tile drained fields (Geohring et al., 1998), evidence supporting the environmental adaptation of *E. coli* suggests that persistence in tile drainage effluent could represent naturalized strains rather than be indicative of recent fecal pollution. The survival and transport of naturalized strains appear to be influenced by a variety of factors, including soil moisture, clay content, soil nutrient status, temperature oscillations and precipitation events (Ishii et al., 2010; Brennan et al., 2010; Esseili et al., 2011). The degree to which potentially naturalized strains of *E. coli* in soils are contributing to tile drainage effluent following manure application to agricultural fields has not been definitively established.

The purpose of this study was to evaluate continuous *E. coli* loading from a tile drained, manure-amended agricultural field to an adjacent stream system over two

growing seasons, while concurrently determining the presence of manure-borne versus potentially naturalized *E. coli* strains in the tile drainage effluent. The specific objectives of this research were to: (i) test combined mechanistic hydrological (DRAINMOD) and statistical multivariate adaptive regression spline (MARS) models for predicting daily *E. coli* loads from the tile drain system; (ii) conduct a genetic population analysis to determine whether tile drain *E. coli* represent manure-amended soil or environmentally adapted strains; and (iii) to monitor the occurrence of *E. coli* strains identified in tile drainage effluents in an adjacent stream. This information is required to inform water quality monitoring programs that use *E. coli* as an indicator of fecal pollution because the detection of environmentally adapted *E. coli* can lead to overestimations of public health risk from agricultural water resources.

3.2 Materials and Methods

3.2.1 Study Area and Agronomic Practices

This study was conducted in a "working" 9.6-ha, field comprised of three subsurface drainage plots that is located within the Thomas Brook Watershed (TBW) of the Annapolis Valley of Nova Scotia. The TBW is part of the Agriculture and Agri-Food Canada's Watershed Evaluation of Beneficial Management Practices (WEBs) program. Only two of the three tile drainage plots were used (plots DT2, 4.2 ha, and DT3, 3.2 ha), because they provide consistent annual drainage volumes that could be collected at each sampling point. Each plot contains 100 mm diameter subsurface tiles located approximately 80 cm below the surface with lateral spacing ranging between 9 and 12 m. The soils are of the Pelton Series, which is classified as a Haplorthod (USSC). A toposequence from Typic to Aquic Haplorthods is represented in each plot. The soils

examined in this study consist of granular sandy loam topsoil (0-18 cm) over subangular blocky loam subsoil (18 – 72 cm) and firm, massive loam parent material (>72 cm). This field has been in a corn-corn-barley rotation, with conventional tillage and periodic chisel-plowing for 10 years (VanderZaag et al., 2010). The field was converted to pasture hay mix at the onset of the current study. On May 10, 2010, solid dairy cattle manure that was stockpiled at the upslope portion of the field over the winter was spread at a rate of 5 t/ha and incorporated with a chain harrow. The field was then seeded with a pasture mix that remained on the field over the course of the study. On June 12, 2011, liquid manure from the same herd was top-dressed on the field at a rate of 20 m^3/ha.

3.2.2 Field Sample Collection

Soil and tile drainage water sampling commenced three days following the solid manure application, and continued on a bi-weekly schedule until November 2011; thereby, capturing two full growing seasons and an overwinter period. During the 2011 growing season, tile drainage water was collected during an additional four storm events, with samples retrieved during peak tile drain flow, and the following day during flow recession. To allow for detection of *E. coli* population differences due to soil moisture content, soil was collected from both upslope and toe-slope zones. Triplicate samples of surface soil (0 – 15 cm) were collected with a cleaned hand auger and composited into a single, sterile container. Drainage water was collected in sterile 500 mL containers. When flow rates were high, the water was sampled as a composite of two 250 mL subsamples obtained approximately 60 s apart. Tile drainage water was analyzed for its temperature, pH, electrical conductivity, and dissolved oxygen with a multiparameter water quality sonde (600R, YSI Environmental Incorporated, San Diego, CA, USA). A 1

L sample was collected for laboratory-based measurement of turbidity (2100AN, Hach Company, Loveland, CO, USA), and total suspended solids (TSS). The latter by filtering a volume of water through glass microfiber filter (934-AH, Whatman, Maidstone, UK), and drying for a minimum of 4 h at 105°C. Water samples (500 mL) were collected from a stream system adjacent to the field. These samples were collected downstream from both tile drain outlets for the purposes of characterizing the influence of tile drainage effluent on the *E. coli* population in the stream system.

3.2.3 *E. coli* Enumeration

Soil samples were passed through a 2.38 mm diameter sieve (No. 8; W.S. Tyler, St. Catherines, ON, Canada) to homogenize the soil samples and remove coarse particulate matter. Twenty grams of the <2.38 mm fraction were resuspended in 180 mL of sterile peptone-saline by hand-shaking for 60 s. Two 10 mL subsamples of the supernatant were collected after a 10 minute settling time, and vacuum-filtered through 0.45 μm cellulose-nitrate membranes (Whatman Laboratory Division, Maidstone, England). The membranes were transferred to mFC basal media supplemented with 3-bromo-4-chloro-5-indolyl-β-glucopyranoside (BCIG; Inverness Medical, Ottawa, ON). The contents of tile drainage and stream water samples were collected on 0.45 μm cellulose-nitrate membranes prior to their transfer to mFC-BCIG agar plates. The samples were incubated for 2 h at 35˚C, then overnight at 44.5˚C. Distinctly separate, blue *E. coli* colonies were counted and converted to colony forming units per gram of soil (CFU/g soil) or CFU/100 mL water.

3.2.4 DRAINMOD Hydrological Model Parameterization and Calibration

The DRAINMOD (v. 6.0, Skaggs, 1980) hydrological model initially was constructed and calibrated for the study field using weather data collected from Environment Canada's Greenwood, Nova Scotia weather station from January 2007 to December 2008 (Campbell, 2009). Data was collected at daily time steps and entered into the model on an hourly basis using the utility program available within the model, assuming that daily rainfall was uniformly spread. During calibration, model outputs were compared to tile drainage outflow rates measured continuously using the tipping bucket method between May 2007 and December 2008 (Campbell, 2009). Drainage volume, upward flux and infiltration parameters were calculated by an internal DRAINMOD subroutine, which uses soil water characteristics of each soil layer to produce the drainage volume for each layer ranging from the surface to the bottom of the soil profile. For predicting drainage in cold conditions, soil temperature and snow prediction parameters were included, and were obtained from a previously conducted study in Atlantic Canada (Luo et al., 2000). Sensitivity analysis was performed on twelve high ranking parameters identified for DRAINMOD by Wang et al. (2005) prior to model calibration. Sensitive parameters were adjusted and observed, and simulated tile drainage outflow was qualitatively and quantitatively compared with observed drainage values to calibrate the model. The field soils are relatively consistent between the two drainage plots and the same soils data were used in simulations for both plots. The primary difference between the plots is surface slope, where the slopes are approximately 1.5% for 300 m in DT2 and 2.4% for 214 m in DT3. These topographical differences were accounted for in the DRAINMOD simulations. Daily minimum and maximum temperatures were used to calculate potential evapotranspiration (ET) in DRAINMOD using the Thornthewaite method. Monthly ET

adjustment factors were used to improve the Thornthwaite ET predictions. Final adjustment values were 0.1 from October to March, 1.5 from April to June, and 1.0 from July to September. Final values for all calibrated input parameters are listed in Table 3.1.

3.2.5 DRAINMOD Hydrological Model Simulations

For the current study, a HOBO meteorological station (Model E-348-UA-002-08; Onset®, Cape Cod, MA, USA) was installed at the southwest perimeter of the field for collecting site-specific air temperature and precipitation data. Readings were taken every 15 s and averaged on 10 min intervals. Site-specific weather data was used to simulate hydrology during the study period. Hourly precipitation values were calculated from the collected data and entered into the DRAINMOD program. Crop data used in model simulations was a pasture mix, with growth commencing on May 10 of 2010, and May 1, 2011, for a total of 159 and 168 days used as the growing period. At each sampling event, tile drainage rates were measured using a stopwatch and a bucket of known volume to validate the outflow of the model drainage predictions.

3.2.6 Soil *E. coli* Decay Models

Decay models were developed from the observed soil *E. coli* concentrations to assist in the creation of a continuous time series of soil-borne *E. coli* concentrations as an input for the tile drain *E. coli* prediction model. From the concentration data obtained, the *E. coli* decay rate kinetics were best represented with a log-linear decay model (1) during the 2010 solid manure addition, and with a biphasic log-linear decay model (2) following the 2011 liquid manure application:

$$C_t = C_0 * e^{-kt} \quad (4.1)$$

$$C_t = C_0[(f)\exp^{-k_1 t}][(1-f)\exp^{-k_2 t}] \quad (4.2)$$

where C_t = *E. coli* concentration at time t (CFU/g soil), C_0 = initial *E. coli* concentration at the time of manure application (CFU/g soil), k = decay rate coefficient (d^{-1}), t = day following manure application (d). In the biphasic decay model, k_1 = decay rate coefficient for the fast decay phase (d^{-1}), k_2 = decay rate coefficient for the second decay phase (d^{-1}), and f = fraction of *E. coli* in the fast decay phase. The models were fitted using the log-linear and biphasic decay model options available in the GInaFit program produced and distributed by Geeraerd et al. (2005). The calculated C_t values were used as the daily soil concentration of *E. coli* in the multivariate adaptive regression spline (MARS) model for predicting tile drain concentrations.

3.2.7 Statistical Prediction of Tile Drainage *E. coli* Concentration and Daily Load

A MARS model was constructed for predicting waterborne *E. coli* discharging from each drainage plot. The candidate explanatory variables included in the modeling process were daily precipitation (m^3/d), average daily air temperature (°C), natural log of soil *E. coli* concentration (lnCFU/g) calculated as a daily time series from the decay models, and the DRAINMOD outputs of tile drainage rate (m^3/d), infiltration rate (m^3/d), evapotranspiration rate (m^3/d), and surface runoff rate (m^3/d). All of the hydrological variables were converted to m^3/d by multiplying the length/day of the hydrological variable by the area of the respective drainage field (4.2 ha for DT2, 3.2 ha for DT3). Drainage field number (DT2 vs. DT3) and manure type (solid vs. liquid) were input as categorical explanatory variables. The MARS model was produced in MATLAB (v. 12, The MathWorks Inc., Natick, MA, USA) using the ARESLab toolbox produced and distributed by Jekabsons (2011).The model developed was piecewise-linear, allowed 1st

degree interactions, had a maximum of 21 basis functions, allowed backward pruning, and had a generalized cross-validation penalty of 3.

Table 3.1. Drainage system design, soil temperature and hydrological calibration parameters used in DRAINMOD 6.0.

Parameter	Value
Drainage System Parameters	
Drain depth (cm)	80
Drain space (cm)	1050
Effective radius of drains (cm)	0.5
Actual distance to impermeable layer (cm)	100
Equivalent depth from drain to impermeable layer (cm)	17.99
Drainage coefficient (cm day^{-1})	2.5
Kirkham's coefficient, G	13.3
Initial depth to water table (cm)	75
Maximum surface storage (cm)	0.5
Kirkham's depth for flow to drains (cm)	0.5
Soil Temperature and Snow Prediction Parameters	
Thermal conductivity function coefficients (W/m °C)	a = 0.55; b = 1.96
Phase lag for daily air temperature (h)	9
Soil temperature at the bottom of the profile (°C)	7
Rain/snow dividing temperature (°C)	0
Snowmelt base temperature (°C)	1.5
Snowmelt coefficient	5
Critical ice content	0.2
Hydrological Calibration Parameters	
Saturated hydraulic conductivity (cm hr^{-1})	
1^{st} soil layer (0 – 26 cm)	5.70
2^{nd} soil layer (26 – 31 cm)	4.25
3^{rd} soil layer (31 – 100 cm)	1.17
4^{th} soil layer (100 – 200 cm)	0.10
Heat Index	75

3.2.8 Model Validation Statistics

The performance of the DRAINMOD hydrological model and the MARS *E. coli* concentration model were evaluated with the coefficient of determination (R^2), the Nash-Sutcliffe efficiency (NSE), and the percent bias (PBIAS). The coefficient of determination describes the proportion of variance in the observed data explained by the predictive model. The values range from 0 to 1, with higher values indicating less error variance. R^2 values greater than 0.5 are considered acceptable for hydrological modeling (van Liew et al., 2007). The NSE contrasts the magnitude of the residual variance with the measured data variance, indicating how well the plot of observed to predicted values fit the 1:1 line. NSE ranges between $-\infty$ and 1.0, with 1.0 indicating perfect model performance, values exceeding 0.5 indicating satisfactory model performance, and negative values indicating that the observed mean is a better predictor than the simulated value (Moriasi et al. 2007). The percentage bias measures the average tendency of the predictions to be greater or less than the observed values. The optimal PBIAS value is 0.0, where positive values indicate model underestimation bias and negative values indicate model overestimation bias (Moriasi et al., 2007).

3.2.9 *E. coli* Isolation and Genotyping

Up to 30 *E. coli* colonies were isolated for strain typing from each soil extract, tile drainage effluent and stream water membrane filtration mFC-BCIG plate. Presumptive *E. coli* were purified on Sorbitol-MacConkey agar and confirmed as *E. coli* through enzymatic (DMACA Indole; Difco Laboratories, Sparks, MD) and molecular procedures. Prior to DNA extraction, indole-positive isolates were cultured in tryptic soy broth (TSB) at 37˚C for 24 h. In total, 642 *E. coli* isolates were obtained from manure-amended soil,

489 isolates were retrieved from DT2 tile drainage effluent, 503 isolates were obtained from DT3 tile drainage effluent, and 696 isolates were collected from the stream.

Nucleic acid was extracted from the TSB cultures using prepGEM Bacteria DNA kits (ZyGEM Corporation, Ltd., Hamilton, New Zealand). Each isolate was identified as *E. coli* through the phylogenetic grouping procedure of Clermont et al. (2000), which assigns isolates to one of four phylogenetic groups based on the presence of three target genes. Amplification conditions described by Clermont et al. (2000) were used, except the denaturation time was extended to 10 s and the annealing and extension time was extended to 15 s to consistently amplify all three bands in the control strain *E. coli* ATCC 25922. Strain typing was performed using repetitive element palindromic (rep)-PCR using BOX A1R primers (BOX-PCR), following protocols reported by Rademaker et al. (1998). Polyacrylamide gel electrophoresis was used, with 3.5% gels run at 7.5 V/cm for 195 min. *E. coli* ATCC 25922 was used as a positive control for intergel comparison, and AmpliSize Molecular Ruler (50-2000 bp; Bio-Rad, Hercules, CA, USA) was used as a size standard. The PAGE gels were stained with ethidium bromide and recorded using an ImageMaster VDS-CL documentation system (Amersham Pharmacia Biotech, Inc., UK). For purposes of this study, *E. coli* that displayed different BOX A1R-PCR banding patterns were considered to be genetically distinct.

3.2.10 Computer-Assisted Image Analysis and Cluster Assignment

BOX-PCR images were analyzed with GeneTools software (Syngene Ltd., Frederick, MD, USA). Band matching was performed using the rolling disk method for background subtraction and band sizes between 50 and 2,000 bp were used in subsequent cluster analysis. Cluster analysis was performed with GeneDirectory software (Syngene Ltd.)

using unweighted pair group method with arithmetic mean (UPGMA), through profile comparisons with Pearson's product moment correlation coefficient and a 1% threshold. Isolates exhibiting ≥85% similarity were classified as clonal strains, as established with intergel comparisons of *E. coli* ATCC 25922.

3.2.11 Soil *E. coli* Population Similarity

The similarity of *E. coli* populating the manure amended soil was tested for significant differences ($p<0.05$) using non-parametric multivariate analysis of variance (NPMANOVA) in PAST software (Hammer et al., 2001). Soil samples were grouped according to drainage field (n=22 sample time points), and each field was further separated based on solid manure addition (n=5) and liquid manure addition (n=17). The liquid manure application was further grouped according to the two-stage decay curve, where group 1 represented the populations present in the rapid die-off phase (n=5) and group 2 represented the population in the slower die-off phase (n=12). *E. coli* strain abundance data, in terms of the number of isolates displaying a certain strain type, for the upslope and toe-slope positions samples were composited within each drainage plot, and the slope positions were not analyzed separately. NPMANOVA was conducted on Hellinger transformed data using Euclidean distance (Legendre and Gallagher, 2001), and 999 permutations were performed.

3.2.12 Tile Drain *E. coli* Population Similarity and Comparison to Manure-Amended Soil Strains

The *E. coli* strains obtained from the soil following both solid and liquid manure additions in 2010 and 2011, respectively, were compiled into a single library (n = 642). All *E. coli* isolates retrieved from the soil were assumed to be derived from manure

sources and not from wildlife or potentially naturalized populations, as the soil-borne *E. coli* experienced typical decay to non-detectable(<1 CFU/g) concentrations in both growing seasons. The isolates retrieved from the tile drains were compared against the soil strain library for each sampling event. Groupings of tile drainage samples were identified by the percentage of isolates that matched the manure-amended soil library. Three groups were created: >80% of the isolates matching the soil library, 50 – 80% of the isolates matching the soil library, and <50% of the isolates matching the soil library. Combined with separations based on manure type (solid vs. liquid) six manure-based groups were created. The groups were projected onto a principle coordinate analysis (PCoA) ordination to visualize *E. coli* population similarities in PAST software (Hammer et al., 2001) using Euclidean distance and Hellinger transformed strain abundances. One-way NPMANOVA was used to statistically compare the manure-influenced groups as identified above, using Euclidean distances on Hellinger transformed abundance data and 999 permutations.

3.2.13 Detection of Tile Drain *E. coli* Strains in Adjacent Stream Water

The percentage contribution of tile drainage *E. coli* strains to the waterborne *E. coli* population in the adjacent stream was calculated by dividing the number of *E. coli* isolates that matched the library of the tile effluent strains by the total number of isolates collected from each water sample. The tile effluent library contained all of the strains identified in DT2 and DT3, irrespective of their relationship to manure amendments.

3.3 Results and Discussion

3.3.1 Hydrological Modeling of Tile Drainage Systems

The DRAINMOD model satisfactorily predicted the daily drainage rates occurring throughout the study period. Although the predictive performance of the calibration data was moderate for site DT2 and poor for DT3, the predictive performance of the validation data, which encompassed the entire study period, was good for both DT2 and DT3, exhibiting R^2 and NSE values in the range of 0.60 to 0.66 (Table 3.2). The poor performance of the calibration data in contrast to the validation data is likely a result of the quality of the weather data used in the model. The data used for model calibration were retrieved from the nearest regional monitoring site (approximately 24 km from study site), and were systematically split into hourly increments using a utility program in DRAINMOD. In contrast, the validation data was composed of site-specific temperature and precipitation data, and simulations were run using precise hourly precipitation inputs. With the validation data, the model tended to under-predict the tile drainage outflow by approximately 25% according to the positive PBIAS values (Table 3.2), which is an acceptable error for hydrological simulations (Moriasi et al., 2007).

Table 3.2. Hydrological (DRAINMOD) and *E. coli* concentration (MARS) model performance statistics.

Evaluation Statistic	DRAINMOD – Tile Drainage				MARS – *E. coli* Concentration			
	Calibration		Validation		Calibration		Validation	
	DT2	DT3	DT2	DT3	DT2	DT3	DT2	DT3
R^2	0.42	0.22	0.66	0.63	0.90	0.85	0.69	0.86
NSE	0.40	0.21	0.64	0.60	0.90	0.85	0.67	0.73
PBIAS	-0.15	24.13	26.5	25.9	-5.15	5.71	-6.37	-9.99

3.3.2 Soil *E. coli* Concentration and Decay

Soil-borne *E. coli* concentrations were determined at different hillslope positions (upslope versus toe-slope) at each sampling event and were used to model the decay of *E. coli* in the soil of each drainage plot (soil *E. coli* decay curve fits are shown in the Supplementary Material). Following solid manure application, the initial concentration of *E. coli* was 456 and 658 CFU/g soil in drainage fields DT2 and DT3, respectively. Log-linear decay models for the solid manure application exhibited R^2 values of 0.916 and 0.886 for DT2 and DT3, respectively (Supplement 3.1 and 3.2). The decay rate coefficients were 0.170 day^{-1} for DT2 and 0.181 day^{-1} for DT3, which are consistent with published values for agricultural soils (Lau et al., 2001; VanderZaag et al., 2010). Following the liquid dairy manure application, initial concentrations of soil-borne *E. coli* were 8930 and 4550 CFU/g soil in drainage fields DT2 and DT3, respectively. The subsequent decrease in *E. coli* concentrations followed a biphasic decay model with R^2 of 0.905 and 0.914 for DT2 and DT3, respectively (Supplement 3.3 and 3.4). In drainage field DT2, the decay rate coefficients were 0.376 day^{-1} in the first stage and 0.014 day^{-1} in the second stage, with the fraction of *E. coli* in the first stage being 0.987. In drainage field DT3, the decay rate coefficients were 0.369 day^{-1} in the first stage and 0.013 day^{-1} in the second stage, with the fraction of *E. coli* in the first stage being 0.989. The proportion of *E. coli* observed in the fast decay phase and the calculated biphasic decay rate coefficients are within previously reported values for manure amended soils (Vinten et al., 2002). Soil *E. coli* concentrations were an order of magnitude greater after the liquid dairy manure amendment, likely a result of the higher application rate of liquid manure and stockpiling of the solid manure prior to application. *E. coli* was detectable in the soil

for 124 days following liquid manure application and 30 days following solid manure application. The longer persistence of liquid manure *E. coli* echoes previous reports of longer persistence of *E. coli* derived from liquid manure in comparison to solid manure in agricultural soils (Unc and Goss, 2006).

3.3.3 Tile Drainage *E. coli* Prediction Model

The MARS model developed for predicting the natural log (ln) of tile drainage *E. coli* concentrations (lnCFU-Drainage; lnCFU/100 mL) included the natural log of soil *E. coli* concentrations (lnCFU-Soil; lnCFU/g), air temperature (°C), tile drainage rate (m^3/d), evapotranspiration rate (m^3/d), and infiltration rate (m^3/d) as predictor variables according to the equation:

lnCFU-Drainage = 1.905 + 1.271×max(0, lnCFU-Soil – 2.303) – 0.781×max (0, AirTemp – 16.8) – 0.0548×max(0, 45.24-Drainage) + 0.558×max(0,AirTemp-10.8) + 0.0323×max(0, 37.8-ET) + 0.00283×max(0, Infiltration-390.4) – 0.00285×max(0,390.4-Infiltration)

The model intercept sets the initial concentration of effluent *E. coli* at 1.905 lnCFU/100 mL (~7 CFU/100 mL). The modeled drainage *E. coli* concentration is positively correlated to soil *E. coli* if concentrations exceed 2.303 lnCFU/g soil (~10 CFU/g soil), evapotranspiration is low (<37.8 m^3/d), infiltration is high (>390.4 m^3/d), and daily average air temperature is above 10.8°C, but below 16.8°C. The modeled drainage *E. coli* concentrations are negatively associated with high average daily temperatures (>16.8°C), low drainage values (<45.24 m^3/d) and low infiltration values (<390.4 m^3/d). The MARS model showed excellent predictive performance in the calibration data and good predictive performance for the validation data in both plots (Table 3.2). The model

tended to over-predict drainage *E. coli* within the range of 5-10% for both plots, according to the calculated PBIAS (Table 3.2). In general, the predicted *E. coli* concentrations fit the observed values well (Figure 3.1).

The application of manure or biosolids to tiled fields is known to increase drainage *E. coli* concentration because these amendments increase the reservoir of *E. coli* in soils (Jamieson et al., 2002; Vinten et al., 2004; Esseili et al., 2012; Frey et al., 2013). The MARS model suggests that the influence of soil-borne *E. coli* is concentration dependent, where concentrations below 10 CFU/g soil have limited effect on drainage *E. coli* concentrations. Although soil *E. coli* concentrations are important, microbial transport to tile drains is driven by preferential flow through soil macropores that are activated by infiltrating precipitation (Geohring et al., 1998; Thiagarajan et al., 2007; Samarajeewa et al., 2012). In this study, the MARS model set a threshold infiltration rate of 390.4 m^3/d for increasing drainage *E. coli* concentrations. This corresponds to a rainfall of approximately 9 to 12 mm when back-calculated for the area of plots DT3 and DT2, respectively. Rainfall-induced increases in tile drainage *E. coli* concentrations have been reported previously (Oliver et al., 2005; Frey et al., 2013). Here the MARS model suggests that a minimum precipitation is required to influence drainage *E. coli* concentrations. Average daily air temperature also affected drainage *E. coli* concentrations. Tile drainage *E. coli* concentrations increased with average daily temperatures between 10.8°C and 16.8°C, but decreased at higher temperatures (>16.8°C). Ishii et al. (2010) observed that soil moisture and soil temperature affected the growth of naturalized *E. coli* in soils, although soil moisture influence was negligible at

15°C. Average daily air temperatures between 10.8 – 16.8°C may be within the ideal range to limit soil moisture effects on soil-borne *E. coli* in the system studied.

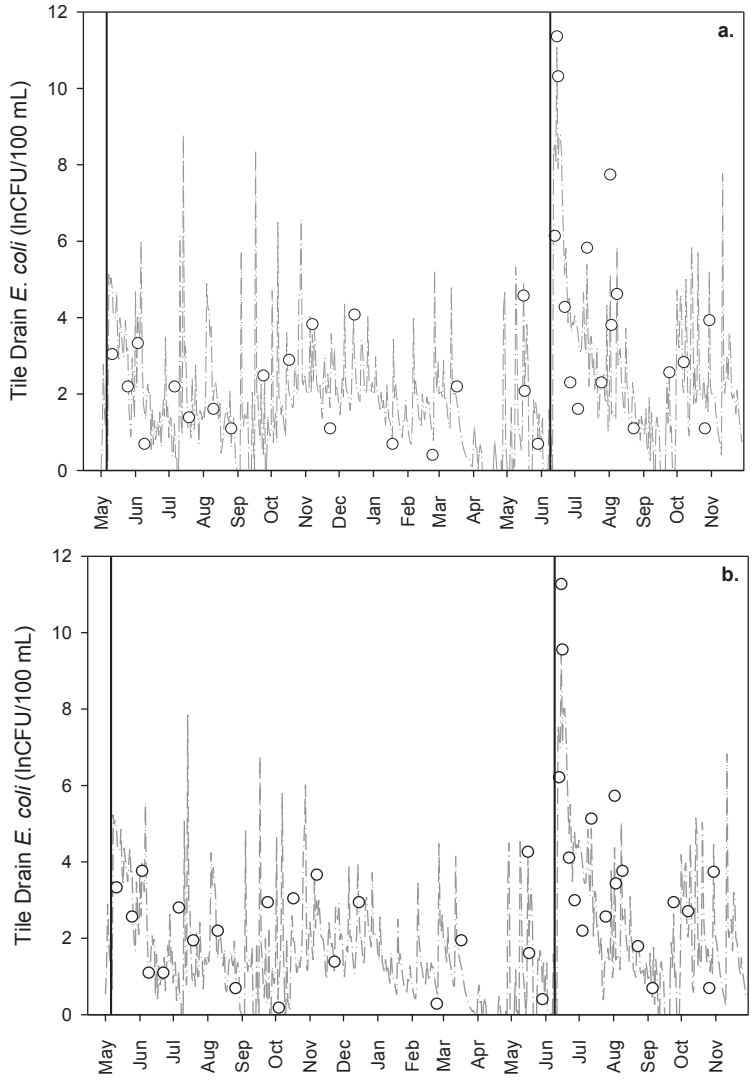


Figure 3.1. *E. coli* concentration estimates with MARS models for drainage plots DT2 (a) and DT3 (b). Grey lines represent the predicted *E. coli* concentration (lnCFU/100 mL) over the study period and open circles represent the observed concentrations. Vertical lines represent the manure application dates.

3.3.4 Daily *E. coli* Loading from the Tile Drainage Fields

The combined mean daily *E. coli* load discharging from both tile drain plots on a monthly basis was calculated using the MARS *E. coli* prediction model. The daily *E. coli* load

(logCFU/d) and tile drain discharge volumes (m^3/d) were separated according to precipitation events exceeding or below the threshold infiltration identified in the MARS model (≥10 mm). The daily *E. coli* loads varied by orders of magnitude over the study period, with the *E. coli* loads in the growing season (May – September) being generally lower than the non-growing season (October – April), except for a short time period following manure application (Figure 3.2). Daily *E. coli* loads from the drainage plots increased by one to three orders of magnitude following manure application (Figure 3.2), corresponding with previous studies that observed higher *E. coli* concentrations following manure application (Vinten et al., 2004; Pappas et al., 2008; Samarajeewa et al., 2012).

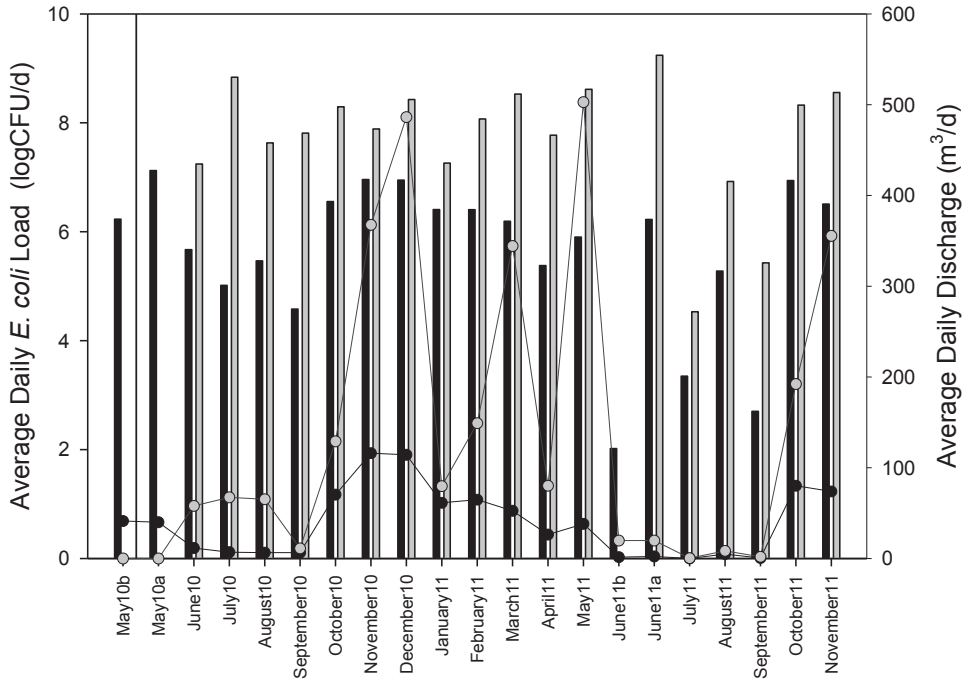


Figure 3.2. Geometric mean drainage discharge volumes and daily *E. coli* load (logCFU/d) in tile drain effluent composited from the two drainage plots separated by month over the study period. Black bars represent the daily *E. coli* load below the threshold precipitation events (<10 mm), and grey bars represent daily *E. coli* loads above the threshold precipitation (>10 mm) identified in the MARS model. Black and grey line points represent the geometric mean daily tile drainage occurring below or above the threshold precipitation, respectively. The months of May 2010 and June 2011 were split as time points before (b) and after (a) manure application.

Although the tile drainage *E. coli* concentrations were highest following manure amendments in late spring (May – June), the mean daily *E. coli* load was elevated in the non-growing season as a result of higher drainage rates during this period.

Previous studies have reported consistent low tile drainage *E. coli* concentrations that appear to be independent of manure or biosolids amendments (Brennan et al., 2010; VanderZaag et al., 2010; Aislabie et al., 2011; Esseili et al., 2012). Our results demonstrate that when considered on a flow-weighted basis, these low concentrations still result in high daily loads to adjacent surface water when tile drainage rates are elevated. Given that the *E. coli* present in soil and drainage water during the non-growing period may represent naturalized strains that are distinct from strains of fecal origin (Brennan et al., 2010; Byappanahalli et al., 2012), artificially high risk may be attributed to tile drains in agricultural water monitoring programs during these periods (Whitman et al., 2006; Goss and Richards, 2008).

3.3.5 *E. coli* Population Structure in Manure-Amended Soil

Soil *E. coli* populations grouped according to manure consistency (solid vs. liquid) and biphasic decay curves (phase 1 vs. phase 2) were compared to interpret differences occurring in tile drainage *E. coli* populations. The overall NPMANOVA probability that the groups are the same is $p = 0.0001$. In a pairwise comparison, the soil populations were significantly different between the solid and liquid manure for both tile drains, but the *E. coli* populations did not significantly vary between the decay phases (Table 3.3). These results indicate that differences in soil *E. coli* population structure, in terms of the presence and abundance of strain types, were greatest between liquid and solid manure amendments, despite their being sourced from the same dairy herd. This was expected

since *E. coli* populations have been found to change over time in manure holding tanks (Lu et al., 2005; Duriez and Topp, 2007). The population structure of the soil *E. coli* was not significantly different ($p > 0.05$) for any group between the tile drain plots (Table 3.3), suggesting that the plots transported similar *E. coli* strains to the drain outlets.

Table 3.3. Pairwise soil *E. coli* population similarity results determined by one-way NPMANOVA. The groupings were based on tile drainage plot (DT2 and DT3), manure type (solid and liquid), and soil *E. coli* decay phase (1 vs. 2). Bold values indicate significant population differences.

Groupings	DT2 Solid	DT2 Liquid 1	DT2 Liquid 2	DT3 Solid	DT3 Liquid 1	DT3 Liquid 2
DT2 – Solid		**0.0271**	**0.0042**	0.1	**0.0302**	**0.0044**
DT2 – Liquid 1	**0.0271**		0.1522	**0.0272**	0.4057	0.0791
DT2 – Liquid 2	**0.0042**	0.1522		**0.0033**	0.1648	0.1237
DT3 – Solid	0.1	**0.0272**	**0.0033**		**0.0294**	**0.0046**
DT3 – Liquid 1	**0.0302**	0.4057	0.1648	**0.0294**		0.3781
DT3 – Liquid 2	**0.0044**	0.0791	0.1237	**0.0046**	0.3781	

3.3.6 Percentage of Manure-Amended Soil *E. coli* among Tile Drain Isolates

The percentage of *E. coli* isolates retrieved from the tile drains that matched the manure-amended soil (MAS) library throughout the study period is presented in Figure 3.3. Over 80% of tile drain *E. coli* isolates examined in each tile drainage water sample were comprised of MAS strains during the first 21 days following solid manure application (12-25 isolates/sample), and for 12 days following liquid dairy manure application (19-30 isolates/sample). The tile drainage populations contained over 50% MAS strains (6-18 isolates/sample) for 50 days following solid manure application. Following liquid manure

application, the tile drainage *E. coli* populations contained over 50% MAS strains (22-29 isolates/sample) for 20 days during non-storm conditions, but up to 60 days during precipitation events, suggesting MAS strain transport to tile drains was influenced by infiltration. Esseili et al. (2012) also observed a greater influence of biosolids-amendment on drainage *E. coli* during storm events, and reduced influence during low flow conditions. The tile drainage *E. coli* populations were minimally influenced (10 – 40%) by MAS strains (6-26 isolates/sample) until early September in both years. Effectively no MAS strains were recovered in the autumn and winter months (0-30 isolates/sample) for both tile drainage systems, except for the storm event captured in late October 2011. The longer impact of MAS strains on drainage water content following liquid manure application reflects the longer persistence of soil-borne *E. coli* observed in this growing season. Although no MAS strains were observed in the tile effluents during the non-growing season, the daily *E. coli* load was equivalent to, or exceeded, the daily load during the growing season when tile drains were influenced by manure amendments (Figure 3.2). This suggests that *E. coli* strains other than those included in the MAS library were contributing to *E. coli* loads. Naturalized *E. coli* have been found previously in temperate soils that are genetically distinct from those attributed to livestock and wildlife fecal sources (Byappanahalli et al., 2006; Ishii et al., 2006). Since the strains observed were distinct from the MAS library, they potentially represent a reservoir of naturalized strains in the tile drainage plots. However, it is possible that the strains identified are uncharacterized manure strains or were introduced through wildlife sources or through manure applications in previous years.

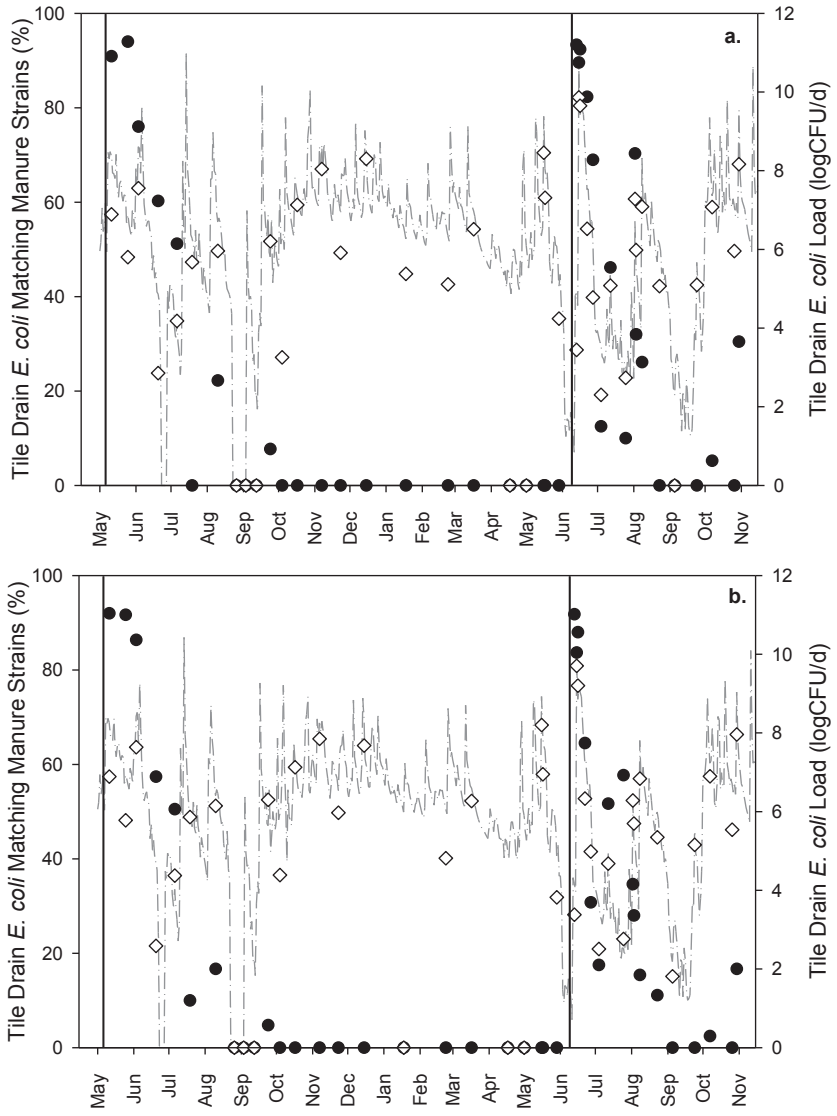


Figure 3.3. Percentage of the *E. coli* isolates in each tile drain sample that match the manure-amended soil library. Black dots indicate the percentage of tile drain *E. coli* isolates that match the soil library, grey diamonds represent the observed *E. coli* daily load, and the grey line indicates the predicted daily *E. coli* load.

3.3.7 Similarity of Tile Drainage *E. coli* Populations

The six groups of tile drainage water samples (Table 3.4) whose *E. coli* composition was influenced by manure application were projected onto a PCoA ordination, which served to explain 32% of the population variability (Figure 3.4).

Table 3.4. Pairwise one-way NPMANOVA results for the six groups of tile drain *E. coli* populations that were distinguished by the percentage of strains identified as belonging to soil amended with solid cattle manure or liquid dairy manure. Bold values indicate statistically significant differences.

Population Grouping	Solid >80%	Solid 50-80%	Solid <50%	Liquid >80%	Liquid 50-80%	Liquid <50%
Solid >80%		0.5955	**0.0015**	**0.0225**	**0.0405**	**0.003**
Solid 50-80%	0.5955		**0.047**	**0.0225**	**0.03**	**0.003**
Solid <50%	**0.0015**	**0.047**		**0.0015**	**0.0075**	0.381
Liquid >80%	**0.0225**	**0.0225**	**0.0015**		**0.0075**	**0.0015**
Liquid 50-80%	**0.0405**	**0.03**	**0.0075**	**0.0075**		0.128
Liquid <50%	**0.003**	**0.003**	0.381	**0.0015**	0.128	

Tile drainage samples that contained high percentages of MAS strains (>80%) tended to separate according to the consistency of the manure applied (solid vs. liquid), and were distant from the other groupings. The >80% samples were found to be statistically different between the manure types and from the other groups in NPMANOVA analysis (Table 3.4), with the exception that the solid manure >80% samples were similar to the solid manure 50-80% samples. Thus, the *E. coli* strains reaching the tile drains differed based on the condition of the manure applied, reflecting differences in soil *E. coli* population structure (Table 3.3). Following both manure applications, tile effluent samples containing low percentages of MAS strains (<50%) tended to converge (i.e. showed similar population composition) and exhibited dissimilarity to the samples containing higher percentages of MAS strains (Figure 3.4). Further, tile drain samples containing low percentages (<50%) of MAS strains were found to be statistically similar through NPMANOVA analysis (Table 3.4). This suggests that tile drain *E. coli* populations that are not influenced by manure amendments tend to contain similar strains

regardless of the season or type of manure type applied, providing further evidence for a reservoir of naturalized *E. coli* in this system. Similarly, Esseili et al. (2012) observed high similarity within background *E. coli* populations that were disconnected from biosolids amendments, and low similarity between biosolids-influenced and background *E. coli* populations in tile drain effluent. The results of the current study and others suggest that *E. coli* strains that populate tile drainage effluent outside of the influence of manure or biosolids application tend to be similar over time, and are distinct from strains introduced through soil amendments.

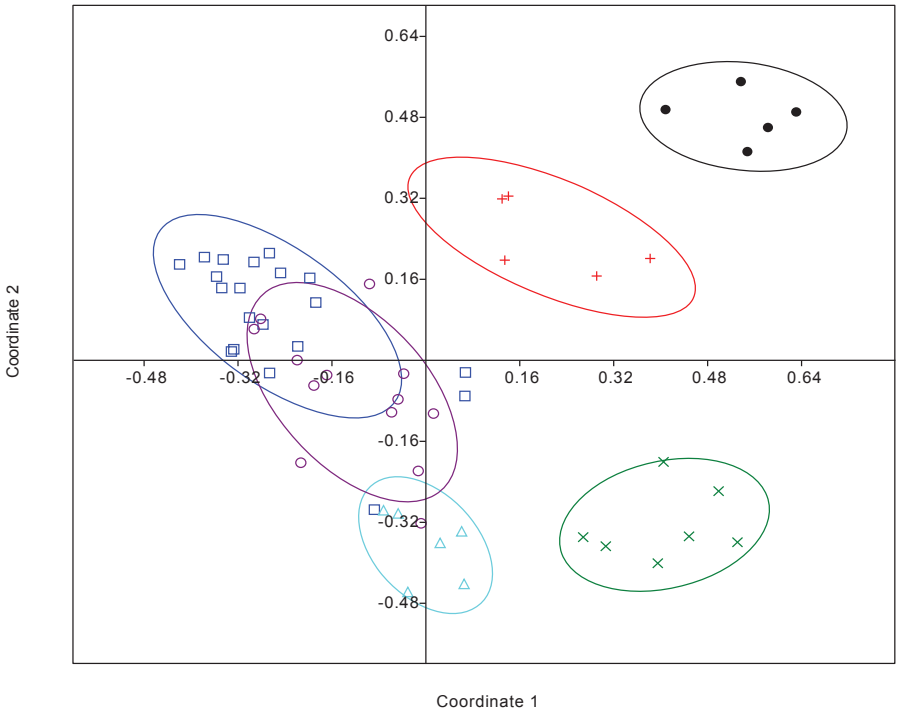


Figure 3.4. Principle coordinates analysis of the tile drain *E. coli* populations grouped according to manure type and percentage of strains matching the soil library. Groups present include: >80% solid manure strains (•), 50-80% solid manure strains (+), <50% solid manure strains (†), >80% liquid manure strains (**x**), 50-80% liquid manure strains (Δ), and <50% liquid manure strains (**o**).

3.3.8 Detection of Tile Drainage Strains in Adjacent Stream Water

Tile drainage *E. coli* strains were regularly observed in the adjacent stream system throughout the study period (Figure 3.5). On average, 13% of the waterborne *E. coli*

isolates collected from the stream matched the library of *E. coli* strains collected from the tile drainage outlets over the study period. The occurrence of tile drainage *E. coli* in stream *E. coli* populations appeared to be independent of manure applications (Figure 3.5). When MAS *E. coli* strains were dominant (>50% *E. coli* isolates) in the tile drainage effluent, the contribution of tile drainage *E. coli* strains to the stream water *E. coli* population averaged 14% (range: 0%- 35%) following solid manure addition and 5% (range: 0%- 13%) following the liquid manure addition. Comparatively, when the tile drainage effluents were dominated by non-MAS strains (<50% MAS strains), the tile drainage *E. coli* strains averaged 19% (range: 0%-50%) of the waterborne *E. coli* following solid manure addition and 12% (range: 0%-27%) following liquid manure addition. These results coincide with Kon et al. (2009) who observed environmentally adapted *E. coli* strains contributing 15.5% – 22.5% of the waterborne *E. coli* population on a temporally stable basis.

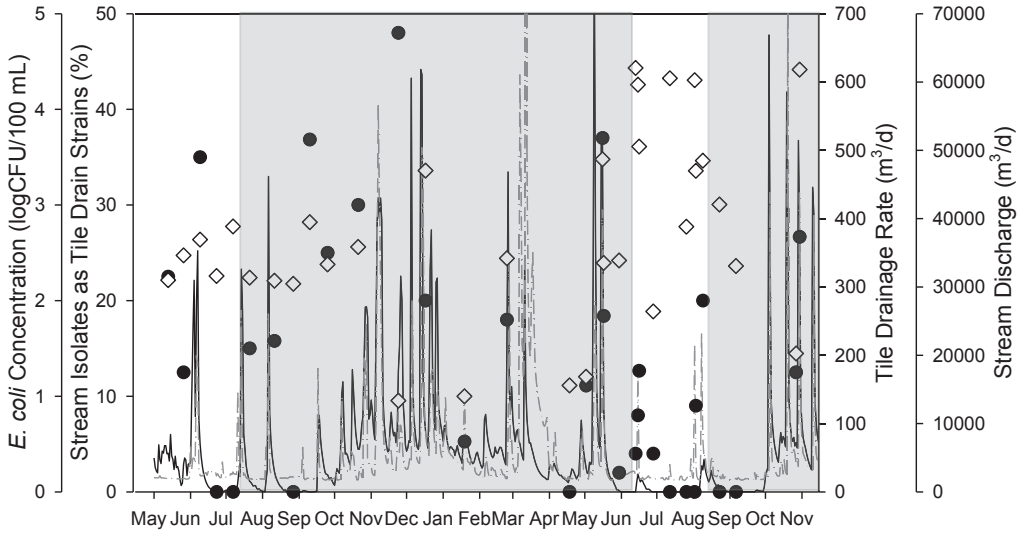


Figure 3.5. Concentration of *E. coli* in the stream adjacent to the tile drainage field (◊) and the percentage of stream *E. coli* isolates that match those obtained from the tile drainage effluent (•). The black lines represent daily tile drainage rates (m^3/d) and the grey lines represent the daily stream discharge (m^3/d). The grey sections represent the time points in which the tile drainage *E. coli* populations were represented by <50% manure-amended soil strains.

The detection of tile drainage *E. coli* strains in the adjacent stream was found to be significantly ($p < 0.05$) linearly correlated with daily *E. coli* load from the tile drains ($r = 0.55$) and exhibited an exponential relationship with the logarithm of daily discharge (Figure 3.6). The influence of tile drainage effluent on the adjacent stream appears to be hydrologically controlled and not dependent on *E. coli* introduced through manure applications. Our observations indicate that the exponential relationship between drainage rate and tile drain *E. coli* contribution to the adjacent stream occurs when either manure-borne or putatively naturalized *E. coli* dominate the tile drain effluent. Tile drainage rates tend to be higher in the non-growing season and when putatively naturalized *E. coli* dominates the effluent. Thus, there may be an overestimation of health risks in agricultural water monitoring programs if *E. coli* distinct from manure-borne strains are being detected during periods outside of the active influence of manure amendments.

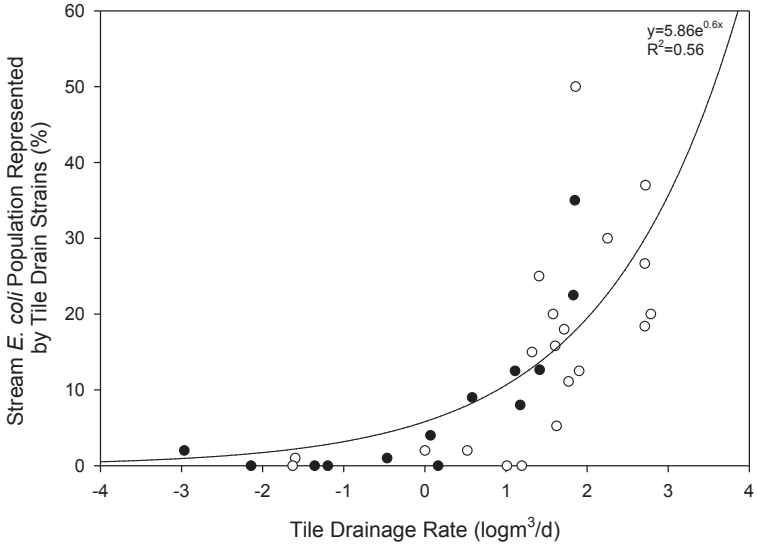


Figure 3.6. Relationship between tile drainage rates and the percentage of stream *E. coli* isolates represented by tile drain *E. coli* strains. Black circles denote time points in which the tile drainage effluent was dominated (>50% composition) by manure-amended soil strains, and open circles indicate time points where the tile drainage effluent was dominated by *E. coli* strains not found in manure-amended soils.

3.4 Conclusion

Increases in tile drainage *E. coli* concentrations were found to be dependent on soil *E. coli* concentrations resulting from manure amendments, air temperature and hydrological fluctuations. *E. coli* concentrations in the tile effluent tended to increase following manure amendments, and strains populating the manure amended soils dominated the tile drainage effluent for up to 55 days following manure application. However, persistent *E. coli* was observed throughout the study period and their populations contained strains that were genetically distinct from manure amended soil strains. On a flow-weighted basis, the daily *E. coli* loads were highest during the non-growing season when discharge rates were high and tile drain effluents were dominated by *E. coli* strains that were distinct from manure amendments. The contribution of tile effluent *E. coli* to the waterborne *E. coli* populations in the adjacent stream was hydrologically controlled and independent of whether tile drainage effluents were dominated by MAS *E. coli* strains or non-MAS strains. These findings raise concern for the continued use of generic *E. coli* concentrations as the sole indicator for monitoring agricultural water quality from manure amended fields. The detection of *E. coli* in tile drainage effluent and adjacent stream waters should be treated cautiously during periods when the *E. coli* population is not influenced by manure amendments.

CHAPTER 4 CHARACTERIZING SPATIAL STRUCTURE OF SEDIMENT *E. COLI* POPULATIONS TO INFORM SAMPLING DESIGN

Materials in this chapter have been published in **Environmental Monitoring and Assessment** and are being reproduced in this thesis with permission from the publisher, **Springer**.

Piorkowski, G.S., R.C. Jamieson, L. Truelstrup Hansen, G.S. Bezanson, and C.K. Yost, 2013. Characterizing spatial structure of sediment *E. coli* populations to inform sampling design. Environmental Monitoring and Assessment, doi: 10.1007/s10661-013-3373-2.

4.1 Introduction

Fecal contamination of water resources introduces enteric pathogens, which are subsequently transmitted to human populations through recreational exposure and consumption of inadequately processed drinking water or irrigated food crops (Solomon et al. 2002; Reynolds et al. 2008; Yoder et al. 2008). *Escherichia coli* is the most commonly used fecal indicator bacterium (FIB) for monitoring freshwater resources, where it is assumed to have high fecal specificity and exhibit an environmental fate and transport similar to fecal pathogens (Tallon et al. 2005). *E. coli* remains the primary FIB for water quality monitoring programs and for simulating microbial dynamics in deterministic watershed models, such as SWAT and WATFLOOD (Dorner et al 2005, Kim et al. 2010). However, the appropriateness of *E. coli* as a FIB has been questioned given recent evidence that *E. coli* is capable of long-term persistence in forest soils, beach sands and freshwater sediments (Ishii et al. 2006; Whitman et al. 2006; Ishii et al. 2007; Halliday and Gast 2011). Recognition of persistent sediment-borne *E. coli* in watershed models through process-based representations of particle settling and resuspension has

improved the efficacy of model simulations (Rehmann and Soupir 2009; Pandy et al. 2012).

Uncertainty in predicted stream water *E. coli* concentrations using these models has been attributed in part to uncertainty in *E. coli* concentrations within stream sediments, which can vary by orders of magnitude (Cho et al. 2010; Kim et al. 2010). Although sediment *E. coli* concentration is an important parameter when modeling waterborne *E. coli*, and attempts have been made to model temporal variability in sediment *E. coli* concentrations (Kim et al. 2010), sediment sampling programs for model parameterization often do not reflect upon the spatial variability possible within a stream reach. Single or composite grab samples are typically taken at the point of water collection, which is to represent the concentrations and population structure of sediment-borne *E. coli* along the investigated stream reach (Rehmann and Soupir 2009; Ouattra et al. 2011; Yakirvech et al. 2013). Fluvial sediments are characterized by considerable spatial heterogeneity in sediment properties brought about by differential patterns in water velocity and shear stress (Gordon et al. 2004). Since elevated concentrations of organic matter and percentages of silt and clay affect the environmental persistence of *E. coli* (Burton et al. 1987; Davies et al. 1995; Craig et al. 2004; Haller et al. 2009; Garzio-Hadzick et al. 2010), spatial variation in sediment *E. coli* concentrations may exist along a stream reach and lead to uncertainty in monitoring and modeling programs designed to link sediment- and waterborne *E. coli*. Greater understanding of the spatial variability existing in sediment *E. coli* concentrations could improve model performance by incorporating representative estimates of *E. coli* concentration within a stream reach. Model performance has been improved by incorporating spatial heterogeneity in particle

properties (Pandey et al. 2012), and better representation of sediment *E. coli* concentrations may further improve model performance.

Another important consideration in modeling *E. coli* distribution is the potential for persistent sediment *E. coli* to dominate the microbial budget of waterborne *E. coli* in streams (Kim et al. 2010). Previous studies have attempted to associate sediment-borne *E. coli* populations to waterborne *E. coli* strains in an effort to link persistent *E. coli* strains to water quality impairment, often concluding that high variability in sediment and waterborne *E. coli* populations limited study outcomes (Kinzelman et al. 2004; Lu et al. 2004; Wu et al. 2009). In these studies, sediment sampling occurred at the point of water collection without consideration of possible spatial variability in *E. coli* population structure within the stream reach. Strain-dependent variability in the response of *E. coli* to environmental media has been demonstrated previously (Topp et al. 2003; Anderson et al. 2005; Pachepsky et al. 2008). Failure to represent the spatial variability inherent in a source population can affect confidence in source assignment of waterborne *E. coli* (Lu et al. 2004; Johnson et al. 2004). Insight into the spatial patterns of *E. coli* populations within streambed sediments would aid in guiding sampling design for studies aiming to relate clonal *E. coli* populations in sediments to waterborne *E. coli* strains.

The present study evaluated the influence of sediment properties and streambed geomorphology on the concentration and population structure of sediment-borne *E. coli* in a stream draining a rural, mixed-use watershed. The objectives of the study were to: (i) assess differences in *E. coli* populations among watershed monitoring sites and streambed morphological features; (ii) examine the degree to which spatial and environmental variables explain streambed *E. coli* population variability, (iii) assess the stability of the

observed spatial patterns during baseflow and following stormflow. This information was required to help design future studies examining temporal alterations in streambed strain composition and the relationship between sediment and waterborne *E. coli*.

4.2 Materials and Methods

4.2.1 Site Description and Sampling Design

Sampling was conducted in the Thomas Brook Watershed, which is part of Agriculture and Agri-Food Canada's Watershed Evaluation of Beneficial Management Practices (WEBs) program, located in the Annapolis Valley of Nova Scotia. Previous studies have concluded that high surface water *E. coli* concentrations are a critical water quality issue in this watershed (Jamieson et al. 2003; Sinclair et al. 2008). Three stream reaches located downstream from permanent monitoring locations (Sites 2, 3 and 4) were selected for study: Site 2 is downstream from a large dairy operation; Site 3 is influenced by low-density residential development; and Site 4 is in a mixed land-use area downstream from both Sites 2 and 3 (Figure 4.1). At each stream reach, five morphological features were identified at stream meanders (point bars and bank scours), pool-riffle sequences and straight segments (i.e., runs). Each stream reach was sampled in the following order: riffle, pool, point bar, bank scour and run (Figure 4.2).

At each morphological feature, triplicate sediment samples (200 to 300 g) were retrieved from the sediment-water interface (0 to 5 cm depth) using a lever-action grab sampler with a 950 cm^3 bucket that was rinsed with stream water between each sample. The retrieved samples were stored at 4˚C and cultured for *E. coli* within 24 hours of collection. Geographic coordinates of each sample location were obtained with a

HiPerGa GPS system (Topcon Positioning Systems Inc., Livermore, CA, USA) for use in spatial statistics. Flow velocity was measured at each feature using a FlowTracker Acoustic Doppler Velocimeter (SonTek/YSI, San Diego, CA, USA). Channel geometry and water velocity are summarized for each sampling site and event in Table 4.1.

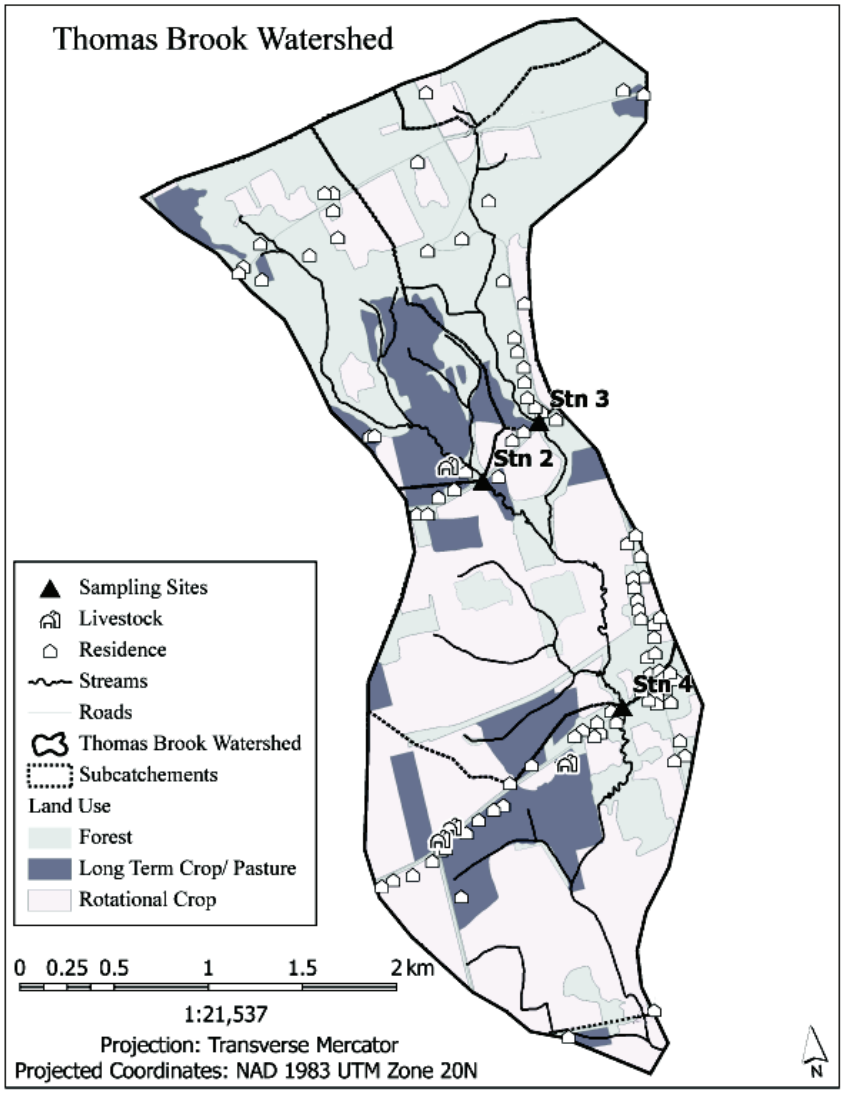


Figure 4.1. Plan view of the Thomas Brook Watershed (Somerset, Nova Scotia, Canada) illustrating the location of permanent monitoring sites and associated land uses.

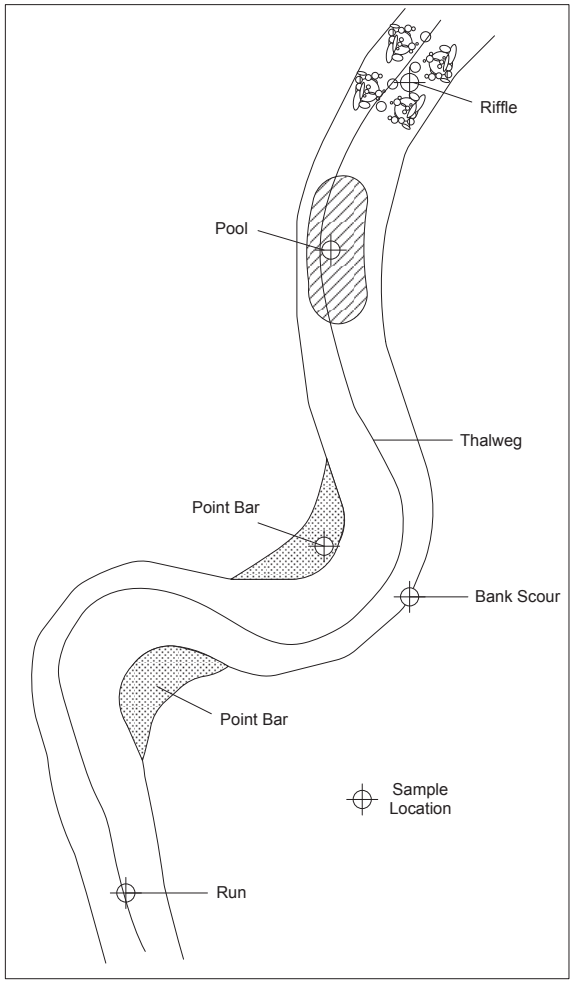


Figure 4.2. Generalized schematic representing the study stream reaches, denoting morphological features sampled and sampling sequence (Riffle to Run).

Duplicate sampling events were conducted to assess the stability of the observed spatial structure. The first sampling event occurred in August 2010 during a period of prolonged baseflow (~2 weeks), and the second event was conducted in September 2010, three days following a storm event (40 mm over 24 h) that lead to a 5 to 8 fold increase in discharge in the investigated reaches. Previous storm hydrographs generated for Thomas Brook indicate stormflow recession within 24 to 48 hours following a precipitation event. Sampling three days following the storm event ensured that no significant sediment redistribution was occurring at the time of sampling.

Table 4.1. Land use, channel properties and discharge for the monitoring sites studied in the Thomas Brook Watershed, NS, Canada. Values represent the discharge and channel properties at the run morphological features. The geometric mean, minimum and maximum discharges are only year the study occurred (2010).

Stream Reach	Primary Land Use	Geometric Mean Discharge - 2010 (m^3/s)	Minimum Discharge - 2010 (m^3/s)	Maximum Discharge - 2010 (m^3/s)	Cross-Sectional Dimensions [width(m) x depth(m)]		Median Particle Diameter (μm)	
					Base	Post-Storm	Base	Post-Storm
Site 2	Agricultural	0.022	0.005	0.504	1.9 x 0.46	2.2 x 0.51	168	116
Site 3	Residential	0.036	0.009	0.503	1.7 x 0.12	1.7 x 0.14	2120	829
Site 4	Mixed	0.094	0.016	2.498	2.1 x 0.09	2.5 x 0.11	1866	631

4.2.2 *E. coli* Enumeration and Isolation

Sediment *E. coli* concentration was measured using membrane filtration. Prior to analysis, sediment samples were sieved (2.38 mm, No. 8; W.S. Tyler, St. Catherines, ON, Canada) to remove coarse particles and homogenize the samples. Twenty grams of the <2.38 mm fraction were resuspended in 180 mL of sterile peptone-saline (0.1% peptone, 0.85% NaCl) by hand-shaking for 60 seconds. The supernatant was collected after a settling time of 10 minutes, filtered through 0.45 μm cellulose-nitrate membranes (Whatman Laboratory Division, Maidstone, England), and incubated on mFC basal media supplemented with 3-bromo-4-chloro-5-indolyl-β-glucopyranoside (BCIG; Inverness Medical, Ottawa, ON) for 2 h at 35˚C, then overnight at 44.5˚C. Distinctly separate, blue *E. coli* colonies were counted and converted to concentration per dry weight of sediment. Presumptive *E. coli* were purified on Sorbitol-MacConkey agar (Oxoid, Ltd., Hampshire, England) and confirmed as *E. coli* through enzymatic

(DMACA Indole; Difco Laboratories, Sparks, MD) and molecular procedures described below. Prior to DNA extraction, indole-positive isolates were cultured in tryptic soy broth (TSB; Difco Laboratories) at 37˚C for 24 hours. Twelve *E. coli* isolates were taken from each replicate sample of the morphological features, yielding 36 isolates per morphological feature per site. This generated a total of 540 isolates for each sampling event and a total of 1080 isolates.

4.2.3 DNA Extraction and Genetic Analysis

DNA was extracted from the broth cultures using prepGEM Bacteria DNA kits (ZyGEM Corporation, Ltd., Hamilton, New Zealand) following manufacturer's protocols. Each isolate was genetically identified as *E. coli* through the phylogenetic grouping procedure developed by Clermont et al. (2000), which assigns isolates to one of four phylogenetic groups based on the presence of three target genes. Amplification conditions described by Clermont et al. (2000) were used, except the denaturation time was extended to 10 s and the annealing and extension time was extended to 15 s to consistently amplify all three bands in the control strain *E. coli* ATCC 25922. Strain typing was performed using repetitive element palindromic (rep)-PCR using BOX A1R primers (BOX-PCR), following protocols reported by Rademaker et al. (204). Polyacrylamide gel electrophoresis was used, with 3.5% gels run at 7.5 V/cm for 195 minutes. *E. coli* strain ATCC 25922 was used as a positive control for intergel comparison, and AmpliSize Molecular Ruler (50-2000 bp; Bio-Rad, Hercules, CA, USA) was used as a size standard. The PAGE gels were stained with ethidium bromide and imaged using an ImageMaster VDS-CL documentation system (Amersham Pharmacia Biotech, Inc., UK).

4.2.4 Computer-Assisted Image Analysis and Cluster Assignment

BOX-PCR images were analyzed with GeneTools software (Syngene Ltd., Frederick, MD, USA). Band matching was performed using the rolling disk method for background subtraction and band sizes between 50 and 2,000 bp were used in subsequent cluster analysis. Cluster analysis was performed with GeneDirectory software (Syngene Ltd.) through unweighted pair group method with arithmetic mean (UPGMA), using the Dice coefficient of similarity and a 1% threshold. Isolates exhibiting ≥90% similarity were classified as clonal strains.

4.2.5 Sediment Analyses

The >2.38 mm sieved fraction was combined with the sieved material (<2.38 mm) for determining particle size distribution (PSD). The reconstituted sample was dried at 105˚C for 24 h and mechanically sieved to determine mass distributions for particle classes between 6 mm and 0.25 mm. For samples where the <0.25 mm fraction was > 10% of the total dry mass of the sediment, a laser *in-situ* scanning and transmissometer (LISST-100X, Sequoia Scientific Inc., Bellevue, WA, USA) was used to determine the PSD between 0.0025 and 0.25 mm. Here, approximately 100 mg of the <0.25 mm sediment was dispersed in 10 mL of 5% hexametaphosphate solution, diluted with 110 mL of deionized water in the LISST-100X mixing chamber attachment, and analyzed for 30 repetitions. Particle size concentrations were converted to proportions by dividing the average for each class by the sum of all classes. The mass of each class was calculated by multiplying the class proportion with the final mass of the <0.25 mm sieve fraction. All mass data were combined and particle size properties were calculated with GRADISTAT

software (version 4.0; Blott and Pye 2001). The calculated values included geometric mean diameter, sorting, percent sand (>0.063 mm), percent silt/clay (<0.063 mm), $D_{75/25}$ (ratio of interquartile particle diameters), median particle diameter (D_{50}) and effective particle size (D_{10}), where 10% of the particles in that sample (by weight) are of a smaller diameter. Organic carbon concentration was determined by the dichromate redox titration method outlined by Skjemstad and Baldock (2008). Approximately one gram of dry sediment from the <0.25 mm sample was used in the analysis.

4.2.6 Statistical Analysis of *E. coli* Concentrations

The significance of spatial location on sediment *E. coli* concentration was assessed using two-way analysis of variance (ANOVA) in SigmaPlot (version 11.0, SYSTAT Software Inc., Chicago, IL, USA), where monitoring site and morphological feature served as the main factors. Tukey's test was used for post-hoc determination of significant differences ($p \leq 0.05$) among factors. Prior to analysis, *E. coli* concentrations were computed per gram of wet weight sediment and log transformed. Least absolute shrinkage and selection operator (LASSO) regression was performed in MATLAB (Version 2012a; The MathWorks Inc., Natick, MA, USA) to determine the environmental variables that explained sediment *E. coli* concentration. LASSO regression was performed on normalized variables to examine the relative effect of the variables by removing measurement scale, and the model chosen had the lowest mean squared error calculated through 10-fold cross validation

4.2.7 Richness and Similarity of *E. coli* Populations in Stream Morphological Features

The richness of *E. coli* strains observed in the stream morphological features was estimated through rarefaction procedures described by Lu et al. (2005). Rarefaction curves were generated through the freeware program Analytical Rarefaction 1.3, available at http://www.uga.edu/strata/software/. The rarefaction curves were plotted in SigmaPlot (v11.0, Systat Software Inc.), and the asymptotes were estimated using a one-site saturation ligand model. The asymptote (V_{max}) estimates the strain richness at sampling saturation, and the K_d value estimates the number of isolates required to capture 50% of the estimated richness. All isolates obtained from each morphological features (n=36) were used to build the rarefaction curves.

Principle Coordinates Analysis (PCoA) was conducted in PAST (version 2.11; Hammer et al. 2001) software to visualize similarities in *E. coli* population structure existing among monitoring sites and fluvial features. The intent of this analysis was to visualize patterns of *E. coli* population similarity based on the monitoring sites and streambed morphological features. Ordinations were generated using Euclidean distances on Hellinger transformed abundance data (Legendre and Gallagher 2001). All isolates (n = 36) obtained from each fluvial feature were composited to aid in visualization, and a transformation exponent of 4 was chosen.

4.2.8 Variation Partitioning Between Spatial and Environmental Variables

Variation partitioning was used to determine whether clonal *E. coli* populations exhibit spatial clustering or random distribution in the streambed, and identify whether these

distributions result from responses to environmental and/or spatial gradients. Spatial explanatory variables [S] were produced using topological-based Moran's eigenvector maps, generated from the MATLAB code produced and distributed by Griffith and Peres-Neto (2006). All of the listed PSD properties, organic carbon and water velocity were used as environmental explanatory variables [E]. Partial canonical correspondence analysis (pCCA) was used to evaluate the influence of environmental variables in the absence of spatial autocorrelation [E|S], and the influence of spatial location without the influence of environmental gradients [S|E]. These analyses were performed using CANOCO Software (version 4.5; Plant Research International, Wageningen, The Netherlands). For all analyses, biplot scaling was used focusing on inter-sample differences, and rare species were downweighted. Automatic selection of variables was conducted using 999 unrestricted Monte Carlo permutations. For pCCA analysis of environmental variables, only significant ($p<0.05$) spatial variables were used as covariables to retain high analytical power (Peres-Neto and Legendre 2010).

4.3 Results

4.3.1 Influence of Sampling Site and Fluvial Morphology on *E. coli* Concentration

For the baseflow sampling event, sediment *E. coli* concentrations among the fluvial morphological features were not significantly different ($p=0.0697$), but varied significantly among monitoring sites ($p=0.002$) (Fig. 3a). Following the storm event, *E. coli* concentration exhibited significant interactions ($p=0.002$) indicating that differences among morphological features varied by site. High sediment *E. coli* concentrations existed in the pools following stormflow, where it was highest at Sites 3 and 4, and

second highest at Site 2 (Fig. 3b). Overall, sediment *E. coli* concentrations were greater at all sites after the September storm event compared to the August baseflow sampling.

LASSO regression with normalized variables was used to determine the relative influence of environmental factors on sediment *E. coli* concentrations in the absence of measurement scale. The resultant beta coefficients demonstrate the magnitude of change in sediment *E. coli* given one standard deviation change in the predictor variable within the system studied, and should not be interpreted as broadly applicable regression coefficients. During baseflow, sediment *E. coli* concentrations were observed to be influenced by water velocity and effective particle size (D_{10}) according to the equation:

$$\ln(\mathrm{CFU/g}) = 4.573 - 0.0915(D_{10}) - 0.2745(\mathrm{Velocity})$$

Both variables have negative beta coefficients, suggesting that lower effective particle size (D_{10}) and velocity are associated with higher sediment *E. coli* concentrations. Velocity had a greater influence on sediment *E. coli* concentrations than texture during baseflow, according to the magnitude of the beta coeffcients.

Following stormflow, organic carbon, water velocity and median particle diameter (D_{50}) were included in the regression equation:

$$\ln(\mathrm{CFU/g}) = 6.6797 + 0.4837(\mathrm{OrgC}) - 0.1839(D_{50}) - 0.3224(\mathrm{Velocity})$$

Similar to baseflow, sediment texture and velocity are negatively associated with sediment *E. coli* concentrations, with velocity exhibiting greater influence than texture. Organic carbon is positively associated with *E. coli* concentrations, suggesting that higher organic carbon is associated with higher *E. coli* concentrations. Further, organic carbon appears to exhibit a greater relative effect on *E. coli* concentration than velocity and sediment texture. Average values of the sediment variables and velocity are included as

supplementary files in Appendix A for baseflow (Supplement 4.1) and stormflow (Supplement 4.2).

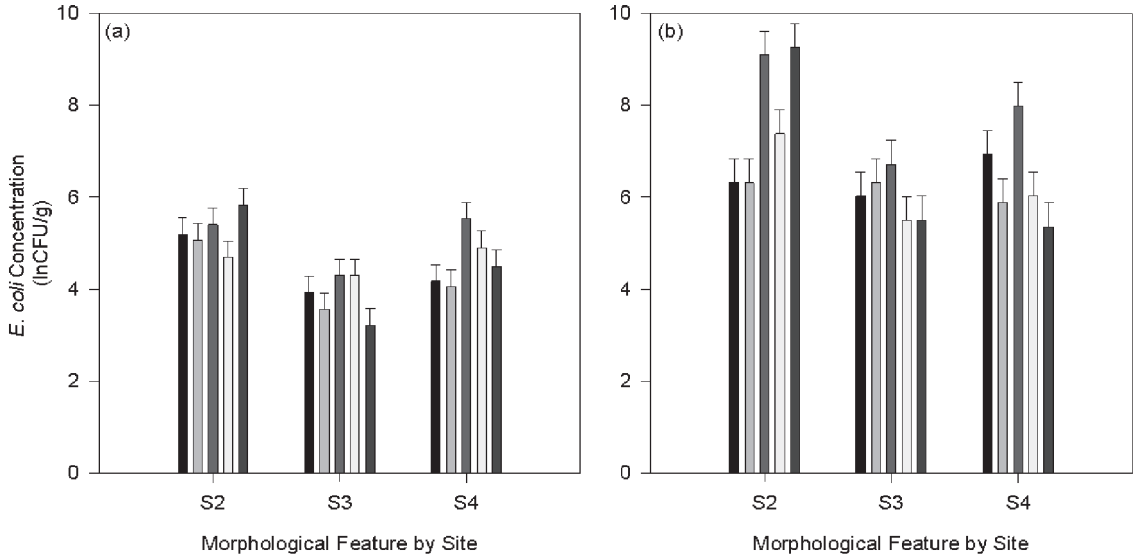


Fig. 4.3. Average sediment *E. coli* concentration (lnCFU/g) collected within fluvial morphological features sampled at each stream reach for the: (a) baseflow sampling event; and (b) post-stormflow sampling event. From left to right, the grayscale bars indicate point bar, bank scour, pool, riffle and run for all groupings. Error bars represent standard deviation (n=3).

4.3.2 *E. coli* Strain Similarity among Sites and Morphological Features

Repetitive element analysis (BOX-PCR) separated the 1080 *E. coli* isolates into 274 genotypes, with 82 genotypes uniquely identified during baseflow, 121 genotypes uniquely identified following stormflow, and 71 genotypes present on both sampling occasions. High diversity in *E. coli* genotypes was observed in the sediments studied. During baseflow, runs and bank scours exhibited the highest estimated *E. coli* genotype richness, followed by point bars, pools, and then riffles (Table 4.2). The sampling effort required to characterize 50% of the total genotypes present ranged from 41 to 100 isolates (Table 4.2). Following stormflow, all morphological features demonstrated an increase in the number of estimated genotypes present, ranging from 88 to 107 genotypes (Table

4.2). The sampling effort required to characterize 50% of the total strains also increased, ranging from 93 to 105 isolates. Within a stream reach, assuming a composite sample of morphological features is obtained, the number of genotypes present was 55 to 77 genotypes during baseflow, increasing to 79 to 117 genotypes following stormflow (Table 4.2). Sampling effort required to characterize 50% of the genotypes in a stream reach was 58 to 83 isolates during baseflow, and 77 to 122 isolates following stormflow.

Table 4.2. Estimated genotype richness and the number of isolates required to detect 50% of the estimated genotypes separated by baseflow and post-stormflow sampling periods. Values presented are for each morphological feature, as the average across stream reaches, and for each stream reach, as the average across the morphological features. Standard deviations are in parentheses.

Location	Baseflow		Post-Stormflow	
	Estimated Genotype Richness	No. Isolates to Detect 50% of Genotypes	Estimated Genotype Richness	No. Isolates to Detect 50% of Genotypes
Morphological Feature				
Point Bar	68 (±10)	76 (±14)	97 (±30)	95 (±31)
Bank Scour	72 (±5)	77 (±8)	99 (±49)	102 (±47)
Pool	57 (±10)	61 (±9)	88 (±14)	93 (±7)
Riffle	37 (±7)	41 (±5)	107 (±37)	105 (±37)
Run	96 (±48)	100 (±48)	91 (±15)	95 (±24)
Stream Reach				
Site 2	55 (±15)	58 (±14)	93 (±20)	94 (±19)
Site 3	77 (±43)	83 (±41)	79 (±17)	78 (±18)
Site 4	66 (±19)	72 (±22)	117 (±34)	122 (±27)

E. coli population similarity among the sites and stream morphological features was visualized with principle coordinates analysis (PCoA). For the baseflow event,

coordinates 1 and 2 explained 32.9% and 14.7% of the variance, respectively (Fig. 4.4). Samples from Sites 2 and 4 demonstrated clustering indicating higher population similarity, whereas samples from Site 3 were dissimilar on the basis of their distance from Sites 2 and 4 in the PCoA ordination. A common spatial pattern was evident among all sites, where populations associated with point bars, bank scour and pools exhibited similarity, whereas riffles exhibited a lower degree of similarity (i.e., greater multivariate distance) to other morphological features. The likeness of *E. coli* populations found in runs varied, as these populations were similar to the riffles at Site 2, depositional environments (i.e., pools and point bars) at Site 4, and dissimilar to all features at Site 3.

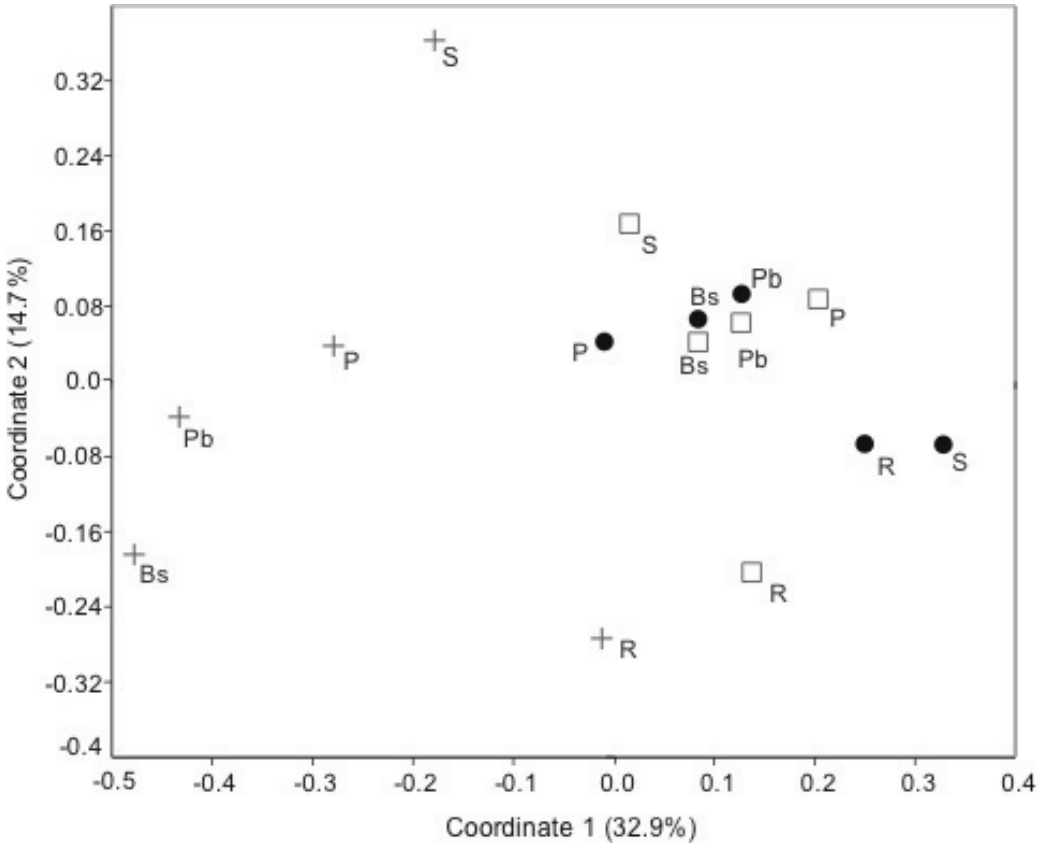


Fig. 4.4. Principle Coordinate Analysis ordination diagram to explain variations in the *E. coli* community composition for the baseflow sampling event. Black circles represent Site 2, crosses denote Site 3 and open boxes indicate Site 4. The morphological features are abbreviated as Pb for point bars, Bs for bank scours, P for pools, R for riffles and S for runs.

Following stormflow, coordinates 1 and 2 explained 33.0% and 22.4% of the *E. coli* population variance, respectively (Fig. 4.5). A similar spatial structure to the baseflow event was observed, where Sites 2 and 4 showed greater population similarity compared with Site 3, and the depositional areas exhibited a high degree of similarity. However, streambed features characterized by higher velocities tended to cluster together, particularly at Site 3, where populations at the bank scour, riffle and runs exhibited higher similarity. Bank scours at Sites 2 and 4 showed less similarity to the depositional areas than was observed during baseflow. The riffles of Sites 2 and 4 still clustered together, along with the run of Site 2. The run at Site 4 clustered with the depositional areas, similar to the baseflow event. Overall, other than differences in the bank scour feature, the spatial structure among the sites appeared relatively consistent between the sampling events.

4.3.3 Variation Partitioning to Determine the Influence of Spatial and Environmental Variables on *E. coli* Populations

For both sample events, spatial variables explained a greater proportion of variance in *E. coli* population structure than environmental variables, with spatial eigenvectors [S] explaining 26.9% of the population variance during baseflow and 31.7% following stormflow (Table 4.3). Comparatively, environmental variables [E] explained 9.2% of the variance during baseflow and 13.1% of the variance following stormflow. Variance in *E. coli* population structure explained jointly by environmental and spatial variables [$E \cap S$] was relatively low both for baseflow (1.8%) and following stormflow (2.9%), suggesting low spatial structuring of the environmental variables according to the spatial eigenvectors included in the analysis. The residual, or unexplained, variance [R] is 63.9%

during baseflow and 55.2% following stormflow, indicating that variables other than those modeled greatly influenced *E. coli* population structure.

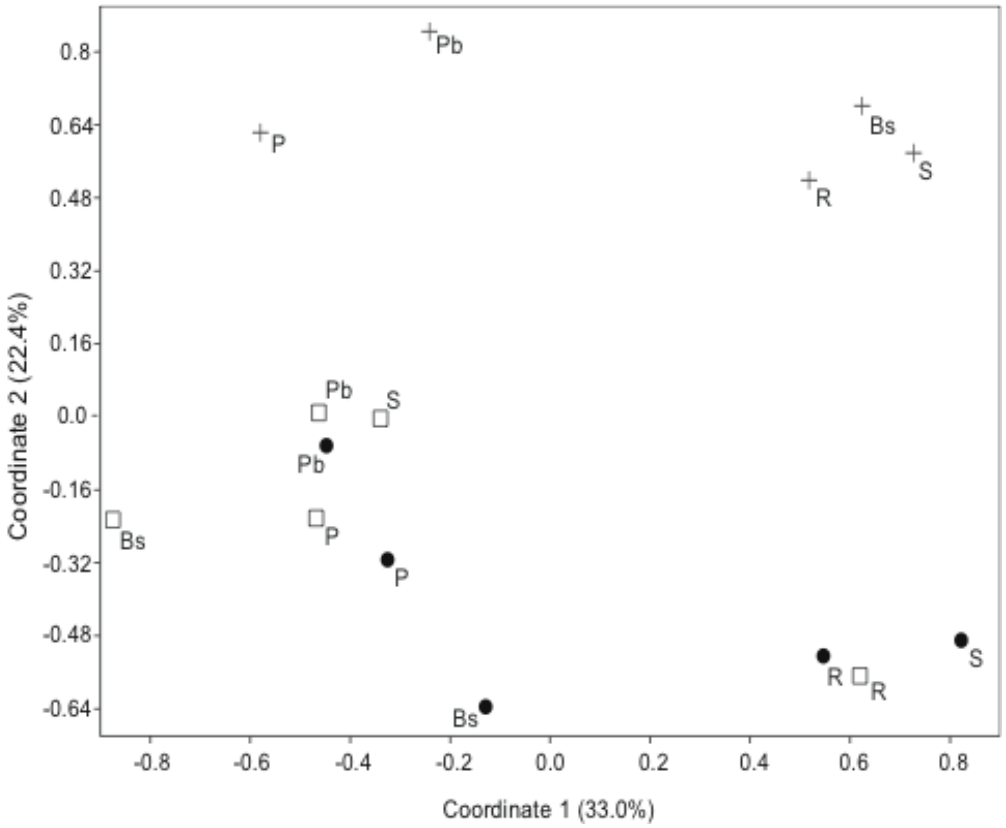


Fig. 4.5. Principle Coordinate Analysis ordination diagram to explain variations in the *E. coli* community composition for the post-stormflow sampling event. Black circles represent Site 2, crosses denote Site 3 and open boxes indicate Site 4. The morphological features are abbreviated as Pb for point bars, Bs for bank scours, P for pools, R for riffles and S for runs.

A total of 26 Moran's eigenvectors with positive eigenvalues were produced through spatial analysis. For the baseflow event, spatial partial canonical coordinates analysis (pCCA) [S|E] revealed that Moran's eigenvectors 2 (ME-2) and 10 (ME-10) were significant in explaining strain variation. Comparatively, Moran's eigenvectors 2 (ME-2) and 7 (ME-7) explained *E. coli* strain variation following stormflow. Plots of these eigenvectors suggest that ME-2 represents spatial autocorrelation, or clustering, based on site location, and ME-7 and ME-10 represent spatial autocorrelation according

to morphological features (Supplement 4.3). Environmental pCCA [E|S] revealed that mean particle diameter ($p = 0.037$) and water velocity ($p = 0.025$) explained the variance in *E. coli* population structure during baseflow and following stormflow, respectively.

Table 4.3. Partitioned variance among environmental and spatial variables on *E. coli* strain composition during baseflow and following a stormflow event in an agricultural watershed.

Fraction*	Explained Variance (%)	Significant Variables** (p-value)
Baseflow		
[S]	26.9	
[S|E]	25.1	ME-2 (0.002) ME-10 (0.026)
[E]	9.2	
[E|S]	7.6	Mean diameter (0.037)
[E∩S]	1.8	
[R]	63.9	
Post-Stormflow		
[S]	31.7	
[S|E]	28.8	ME-2 (0.026) ME-7 (0.043)
[E]	13.1	
[E|S]	10.2	Velocity (0.025)
[E∩S]	2.9	
[R]	55.2	

*[S] represents the explained spatial variance, [S|E] is spatial variance after environmental variance is removed, [E] is the environmental variance, [E|S] is the environmental variance after significant spatial autocorrelation is removed, [E∩S] is the combined environmental and spatial variance and [R] is the residual variance.
**ME = Moran's eigenvector

4.4 Discussion

4.4.1 Influence of Sampling Site, Morphological Feature and Environmental Variables on *E. coli* Concentration

The observed differences in sediment *E. coli* concentrations among sampling sites, and the higher concentration of *E. coli* following stormflow, are likely associated with variable fecal loading into the stream system affected by upstream land use. Sinclair et al. (2008), in examining the same watershed, found that waterborne bacterial loading was highest in subcatchments containing livestock operations (i.e., Site 2), and increased throughout the growing season and during stormflow events. Likewise, waterborne *E. coli* concentrations have been reported to be greater during stormflow events and adjacent to livestock operations in other studies (McKergow and Davies-Colley 2010; Pachepsky and Shelton 2011).

During prolonged baseflow, *E. coli* concentrations were different among sites, but not among morphological features within a stream reach. This result is surprising considering that water velocity and effective particle size (D_{10}) were identified as important predictors in the LASSO regression equation, and that morphological features are defined by sediment textural differences brought about by differential velocity and shear stress distributions (Charlton 2008). Heterogeneity in sediment texture within the morphological features could explain the non-significant difference among morphological features. For example, Powell (1998) reported sediment fining from the head to tail of depositional bars, indicating textural difference within the same morphological feature. Although sediment properties were different among

morphological features (Supplement 4.1), considerable variability was observed within the samples retrieved.

Following stormflow, sediment *E. coli* concentrations were variable among the morphological features depending on the stream reach, although pools exhibited high *E. coli* concentrations at each site. Higher deposition in pools is expected since these features are characterized by lower velocity and settling of finer particles, which are in turn associated with greater *E. coli* attachment (Oliver et al. 2007). Similar to baseflow, water velocity and sediment texture (D_{50}) explained variance in sediment *E. coli* concentrations, but organic matter had greater relative influence on *E. coli* concentrations than both velocity and texture. The importance of organic matter following stormflow, but not during baseflow, supports the postulation of Pachepsky and Shelton (2011) that sediment *E. coli* concentrations are positively associated with organic matter following runoff events as bacteria enter the stream together with an influx of fecal organic matter, whereas organic production during baseflow is disassociated with aboveground inputs. In this study, organic carbon concentrations were higher following stormflow than during baseflow (Supplement 4.2), reflecting the possibility of runoff inputs.

Sediment textural differences reflect velocity and shear stress distributions within a stream reach. In our study, finer sediment (lower D_{10} and D_{50}) were associated with higher *E. coli* concentrations. The significance of median particle diameter and effective particle size on *E. coli* numbers as opposed to explicit percentages of silt and clay is in contrast to other studies that relate silt-clay percentage to sediment *E. coli* concentration (Haller et al. 2009; Garzio-Hadzick et al., 2010). High *E. coli* concentrations have been previously observed in fine sand sediments (Cinotto 2005), which may reflect greater

nutrient exchange via hyphoreic flow in the pore space of sandy sediments (Grant et al. 2011).

Although sediment texture and organic matter influence *E. coli* persistence in sediments, there appears to be no significant relationship between *E. coli* concentrations and fluvial morphological features during baseflow. The lack of statistical relationship could result from sediment heterogeneity and, consequently, statistical homogenization of *E. coli* persistence within a stream reach. Spatial structuring of *E. coli* concentration among morphological features within a stream reach only appears to be relevant following stormflow events, where recent inputs and limited die-off have occurred. Adequate characterization of sediment *E. coli* concentrations for the purposes of monitoring or modeling should capture the variability within a stream reach. There exists a clear relationship of *E. coli* concentrations to sediment properties and water velocity, but the relationship to particular streambed morphological features is not temporally stable. Sampling programs should focus on the collection of multiple samples of sediments consisting of a variety of textures for representative characterization of a reach, although targeting particular morphological features does not appear to be necessary.

4.4.2 *E. coli* Diversity and Strain Similarity among Sites and Morphological Features

Site-level differences in *E. coli* population structure were observed between Sites 2 and 3, presumably due to differences in fecal inputs: large dairy farm versus low-density residential. The *E. coli* population at Site 4 exhibited high similarity to Site 2, illustrating that the higher loading of fecal inputs at Site 2 has a greater influence on the downstream Site 4. Wu et al. (2009) also reported a similar association of downstream waterborne *E.*

coli isolates with isolates from upstream sediments. Walk et al. (2007) observed homogenous sediment *E. coli* populations across sites within beach sand, suggesting association with a well mixed waterborne population. The association between Sites 2 and 4 could be due to high similarity in waterborne strains owing to high fecal loading at the upstream site, but comparison with the waterborne *E. coli* population is required to support this assertion.

Within each stream segment, greater similarity in *E. coli* populations occurred among low water velocity depositional features compared to high velocity features. The population dissimilarity does not appear to be a function of strain richness, as all features demonstrated relatively consistent strain richness, other than the riffle features during baseflow. Thus, it appears that some characteristic of high water velocity features are selecting for different strains within the streambed, possibly due to the capacity for these strains to resist migration either through the production of, or association with, biofilms (Droppo et al. 2009; Hirotani and Yoshino 2010). Conversely, the selection for *E. coli* strains in depositional environments could be attributed to strain-specific differences in attachment to, and deposition of, suspended particles of various sizes (Oliver et al. 2007; Pachepsky et al. 2008). Strain-dependent survival of *E. coli* strains in sediments has been reported previously (Anderson et al. 2005). This study shows that sediment heterogeneity along a stream reach affects which strains persist in fluvial morphological features.

The observed spatial pattern appeared to be fairly consistent between the two sampling events, suggesting stability in the observed spatial structuring in terms of certain *E. coli* genotypes preferentially existing in depositional features while others preferentially exist at high velocity features. However, *E. coli* populations in bank scours

exhibited variability, where the populations were similar to depositional environments during baseflow, but not following stormflow. The similarity in bank scour populations to depositional environments during baseflow could result from comparable texture, as bank scour sediments in this system are fine textured and poorly sorted (Supplement 4.1 and 4..2), or as a consequence of dispersal from the point bar or pools, which are located in close proximity to the bank scours. Determining the relative importance of spatial and environmental factors in explaining *E. coli* population structure could provide evidence to the dominant force affecting *E. coli* populations in these environments.

4.4.3 Variation Partitioning to Determine the Influence of Spatial and Environmental Variables on *E. coli* Populations

Spatial autocorrelation, or clustering, was observed within monitoring sites on both sampling occasions indicating that *E. coli* populations within each site are more similar than among sites. In this context, spatial autocorrelation is the degree to which *E. coli* genotypes exhibit higher similarity as a function of sample distance rather than as a response to ecological gradients. However, population similarity was also observed between Sites 2 and 4, suggesting that the site-level autocorrelation results from the disconnected Sites 2 and 3. Indeed, Moran's eigenvector 2 (ME-2), which was significant at both sampling events, emphasizes the separation of Site 3 from Sites 2 and 4, further supporting the conjecture that high *E. coli* loading at Site 2 is contributing to the *E. coli* found at the downstream Site 4. In the absence of variance explained by environmental gradients, spatial autocorrelation was observed within morphological features, suggesting that *E. coli* populations within fluvial features are more similar than populations among features within a stream reach, perhaps due to dispersal limitations.

Although spatial variables dominated the explained variance, environmental gradients were also important for structuring *E. coli* populations. During baseflow, mean particle diameter was found to explain *E. coli* population variance suggesting that texture selects for different *E. coli* strains, possibly a result of differences in nutrient acquisition, predation, UV damage or biofilm association (Craig et al., 2004; Haller et al., 2009). Following stormflow, water velocity explained *E. coli* population variance, likely a result of differences in velocity-dependent particle settling behavior among the morphological features. Since strains vary in their attachment to particles of various sizes (Pachepsky et al. 2008), differential sedimentation could result in a strain sorting effect.

The combination of spatial and environmental influences on *E. coli* strain composition can be explained through ecological metacommunity theory. According to Cottenie (2005), the combined significance of spatial dependence in the absence of environmental gradients [S|E] and response to environmental gradients after removal of spatial autocorrelation [E|S] denotes mass-effect, or source-sink, ecological dynamics. For this structuring effect to occur within *E. coli* populations, strains must exhibit preference for ecological gradients and be subject to sufficient dispersal, such that they emigrate from environments where they are good competitors (source) to environments where they are bad competitors (sink; Leibold et al. 2004). Adaptation, or preferential deposition, of strains to particular sediment environments yields a source-effect, whereas dispersal occurring from frequent patterns of sediment resuspension or hyporheic exchange (Grant et al., 2011) disperses the strains to sink environments. The statistical methods used in this study demonstrated the importance of both spatial and ecological gradients in clonal *E. coli* population structure. However, future research should consider

explicit representation of spatial boundaries and downstream dispersal conditions inherent in stream systems by using dendritic ecological networks (Peterson et al. 2013).

The observed source-sink effect is an important consideration when characterizing sediment *E. coli* populations. Previous studies linking sediment *E. coli* to waterborne load have taken sediment samples at the point of water collection (Lu et al. 2004; Kinzelman et al., 2004; Wu et al. 2009). Considering the streambed environment affects strain-sorting among low velocity deposition and high velocity features, a targeted sampling campaign is required to capture the diversity of strains found within a reach in order to develop unbiased, representative sediment libraries required to confidently assign waterborne *E. coli* strains to sources (Johnson et al. 2004). Particular focus should be paid to depositional environments (point bars and pools) and riffles, since these environments demonstrated the greatest dissimilarity of *E. coli* populations.

4.5 Conclusion

Spatial heterogeneity of sediments and water velocity exerts a selective effect on *E. coli* concentration and population structure, although the differences are not necessarily reflected among fluvial morphological features within a stream reach. Sampling programs attempting to characterize sediment *E. coli* concentrations and population structures should consider heterogeneity in streambed properties, rather than collecting sediments at the point of water collection. At a minimum, representative characterization of *E. coli* concentrations should be performed at each stream reach of concern, particularly where different land use inputs occur, and should capture differences in sediment properties in depositional areas (pools and point bars) and higher velocity features (riffles and runs). High diversity in stream sediments requires substantial isolate

characterization, where up to 120 isolates are required to identify 50% of the isolates within a stream reach. Strain-sorting based on spatial and environmental variables should be accommodated in sampling programs, by collecting samples from depositional and high velocity features. Better representation of streambed *E. coli* concentrations and population structure can increase confidence in waterborne *E. coli* modeling and source assignment.

CHAPTER 5 REACH SPECIFIC EVIDENCE FOR TEMPORAL SHIFTS IN SEDIMENT *E. COLI* POPULATIONS AND HYDROLOGICAL FACTORS AFFECTING THEIR CONTRIBUTIONS TO THE WATER COLUMN

5.1 Introduction

Microbiological water quality is monitored through the use of fecal indicator bacteria (FIB) which are assumed to be derived from animal excreta and, as such, indicate the possible presence of gastrointestinal (enteric) pathogens in water resources (Field and Samadpour, 2007). *Escherichia coli* remains the dominant FIB used for monitoring and modeling the quality of freshwater resources. Waterborne *E. coli* can be derived from two major sources: (i) catchment sources transported via surface and subsurface runoff fluxes; and (ii) channel stores, or the reservoir of bacteria associated with the sediment bed (Rodgers et al., 2003; Jamieson et al., 2005; Wilkinson et al., 2011). Process-based water quality models have most effectively modeled waterborne *E. coli* concentrations by including both hillslope and in-stream hydrological processes that govern the hydrological connectivity between the water column and catchment or sediment stores (Rehmann and Soupir, 2009; Kim et al., 2010).

Many modeling studies assume that the exchange of *E. coli* at the sediment-water interface results from the suspension and transport of sediment particles, where FIB are irreversibly attached (Cho et al., 2010; Russo et al., 2011; Gao et al., 2011; Pandey et al., 2012). Pachepsky and Shelton (2011) recognized that sole reliance on sediment transport relationships to waterborne *E. coli* is a deficiency in such watershed models. Some studies have reported that elevated FIB concentrations can occur after sediment

redistribution events, postulating the importance of transient storage or hyporheic exchange (Grant et al., 2011; Ghimire and Deng, 2013; Yakirevich et al., 2013). Hyporheic bacterial transport is defined as the release of bacteria trapped in pore spaces by advective flux of water across the sediment-water interface (Grant et al. 2011). In effect, hyporheic exchange is manifested by the downwelling of stream water into the sediment bed and the subsequent upwelling of water at a downstream location, and is driven by variations in streambed topography, fluvial geomorphology, and sediment permeability (Boulton et al, 2010). The distinction between *E. coli* release occurring through resuspension versus hyporheic exchange is important as the transport of sediment-bound *E. coli* occurs during high-discharge events, whereas the release of *E. coli* through hyporheic exchange could elevate waterborne *E. coli* in the absence of sediment redistribution.

Although storm events can lead to *E. coli* loads being orders of magnitude higher than baseflow loads (Stumpf et al., 2010), some stream systems in agricultural watersheds demonstrate chronically elevated *E. coli* concentrations (Jamieson et al., 2003; Muirhead et al., 2004; Sinclair et al., 2008), resulting in restricted water usage for the purposes of irrigation during baseflow periods. Waterborne *E. coli* is generally assumed to be derived from fecal sources, but recent investigations indicate that *E. coli* is capable of proliferating in soil and freshwater sediments under certain conditions (Ishii et al., 2006; Byappanahalli et al., 2006). Surbeck et al. (2010) argue that FIB are not static contaminants, like sediments and nutrients, but are biological entities that experience differences in growth and transport based on ambient nutrient availability, competition with heterotrophic bacteria, and predation. If persistent *E. coli* in streambed sediments is

an important contributor to the overall waterborne *E. coli* population, greater understanding of the population structure, relatedness to fecal sources and potential environmental adaptation of sediment-borne strains would improve risk assessments done for agricultural waters.

The objectives of this study were to: (i) correlate sediment transport and hydrological variables with sediment *E. coli* populations, particularly their concentrations, strain richness (i.e. total number of genotypes present in a sample) and population similarities; (ii) determine the contribution of sediment populations relative to adjacent upland sources (dairy manure lagoon, residential septic systems, tile-drained cultivated field) to waterborne *E. coli* using a library-dependent microbial source tracking approach, and correlate their relative contribution to hydrological and sediment transport variables.

5.2 Materials and Methods

5.2.1 Site description and Sampling Design

This study was conducted within the Thomas Brook Watershed (TBW) of the Annapolis Valley of Nova Scotia. The TBW is part of the Agriculture and AgriFood Canada's (AAFC) Watershed Evaluation of Beneficial Management Practices (WEBs) program. Previous studies have concluded that high surface water *E. coli* levels are a critical water quality issue in this watershed (Jamieson et al. 2003; Sinclair et al. 2009). Three stream reaches were chosen for investigation based on their proximity to sources of fecal material and differences in their hydrological characteristics. Site TB2 is located downstream from a dairy farm (cattle fecal source), site TB3 is located downstream from low-density residential developments (human fecal source), and site TB6 is located

downstream from TB2 and TB3, and adjacent to a tile-drained agricultural field (Figure 5.1). All monitoring sites have sand textured sediments, but differ based on the coarseness of the sediments and water surface slope (S_f), combined with relative energy of the flowing waters. Sites TB2 and TB3 are located in higher slope areas of the watershed, with TB3 having the greatest S_f and coarsest sediments, whereas TB6 has a relatively low slope and finer textured sediments.

At each location, sediment and water samples were collected on a bi-weekly basis over the course of two growing seasons and an overwinter period (May 2010 – November 2011). Four storm events were also sampled during the 2011 growing season to determine hydrological variability in waterborne *E. coli* source contributions. Sediment samples were not obtained during the storm events. Triplicate sediment samples (200 to 300 g) were retrieved from the sediment-water interface (0 to 5 cm depth) using a lever-action grab sampler with a 950 cm^3 bucket that was rinsed with stream water between each sample. Samples were collected from depositional areas (pools or point bars), riffles and runs of each stream reach and composited in a single sterile container to accommodate heterogeneity in *E. coli* populations existing within stream reaches (Piorkowski et al., 2013a). Sediment particle size distribution and organic matter content were determined as reported by Piorkowski et al. (2013a).

Prior to sediment sampling, 500 mL water samples were collected approximately 10 m downstream of the sediment collection areas at the mid-point of the water column. Fecal material was obtained from the manure lagoon at the dairy farm adjacent to TB2 and from each of three residential septic tanks in closest proximity to the TB3 monitoring location in May, July, August and October 2011. Three dairy manure samples were

obtained from separate locations within the lagoon and composited. The retrieved samples were stored at 4˚C and cultured for *E. coli* within 24 hours of collection.

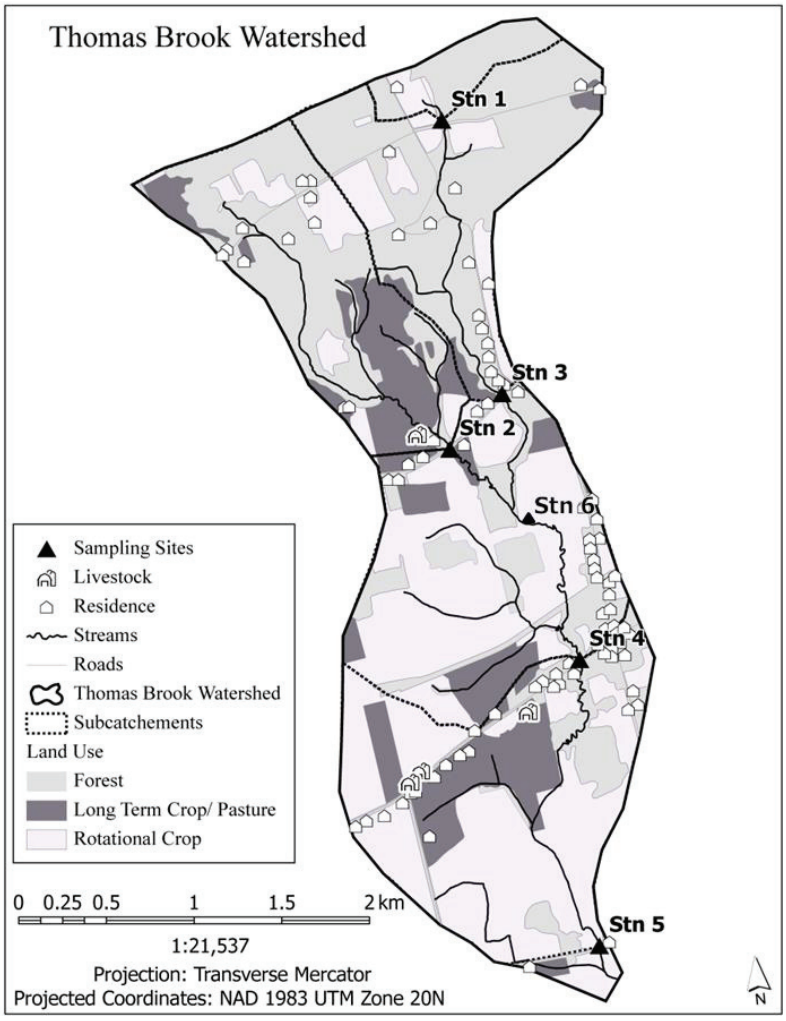


Figure 5.1 Map of the Thomas Brook Watershed identifying permanent monitoring locations and adjacent land uses. Sites 2, 3 and 6 were used investigated in this study.

At the time of sampling, flow velocity was measured using a FlowTracker Acoustic Doppler Velocimeter (SonTek/YSI, San Diego, CA, USA) and used to calculate volumetric discharge via the midsection method (Buchanan and Somers, 1969). The calculated discharge was used for validating continuous discharge records obtained from AAFC. At each stream reach, a continuous series of stage measurements were obtained

using either KPSI Series 169 pressure transducers (Pressure Systems, Inc., Hampton, VA) in conjunction with Campbell Scientific CR10X dataloggers (CSI, Logan, UT) or Global Water WL15 Water Level Loggers (Global Water Instrumentation, Inc., Gold River, CA); each recording water depths on an hourly basis. Stage-discharge relationships (Linsley et al., 1982) were developed for each site using pressure transducer data and discharge measurements collected at each site.

Table 5.1. Explanation and codes of the environmental variables used for correlating sediment *E. coli* population structure and contributions to the water column.

Variable Name	Units	Description
Water Quality		
EC	mS/cm	Electrical conductivity
pH	-	Logarithm of hydrogen ion activity
DO	mg/L	Dissolved oxygen
WatTemp	°C	Instantaneous water temperature (sonde)
TSS	mg/L	Total suspended solids
OM	%	Percentage organic matter: (VSS/TSS)*100
Turbidity	NTU	
TOC	mg/L	Total organic carbon
NO_3-N	mg/L	Nitrate-nitrogen
TP	mg/L	Total phosphorus
Sediment Transport and Hydrology		
Q_w	m^3/s	Volumetric stream discharge rate
τ_b	N/m^2	Bed shear stress
Avg. τ_b	N/m^2	Average bed shear stress between sampling events
Max. τ_b	N/m^2	Maximum bed shear stress between sampling events
Q_s	kg/s	Sediment discharge rate
ΣQ_s	kg	Cumulative sediment discharge between sampling events
ΣPpt	mm	Cumulative precipitation between sampling events
API-7	mm	7-day antecedent precipitation index
DailyPpt	mm	Precipitation on the day of sampling
TDRate	m^3/d	Daily volumetric tile drainage effluent rate
Sediment Variables		
OM	%	Organic matter percentage
Si-Cl	%	Silt-clay percentage
D_{50}	μm	Median particle diameter

At the time of sampling, the water column was analyzed for its temperature, pH, electrical conductivity, and dissolved oxygen with a multi-parameter water quality sonde (600R, YSI Environmental Incorporated, San Diego, CA, USA). A separate 1 L sample was collected for laboratory-based measurement of turbidity (2100AN, Hach Company, Loveland, CO, USA), and total suspended solids (TSS). The latter was achieved by filtering a volume of water through glass microfiber filter (934-AH, Whatman, Maidstone, UK), and drying for a minimum of 4 h at 105°C. Topographic surveys were conducted on each stream reach to determine cross-sectional geometry and water surface slope for each of the investigated stream reaches. The codes and variables used as correlating variables in further analysis are presented in Table 5.1. Channel geometry, hydrological measurements and water quality parameters are summarized for each sampling site in Table 5.2.

5.2.2 *E. coli* Enumeration and Isolation

Water and sediment *E. coli* concentrations were measured using membrane filtration. Prior to analysis, sediment samples were sieved (2.38 mm, No. 8; W.S. Tyler, St. Catherines, ON, Canada) to remove coarse particles and to further homogenize the samples. Twenty grams of the <2.38 mm fraction were resuspended in 180 mL of sterile peptone-saline (0.1% peptone, 0.85% NaCl) by hand-shaking for 60 seconds. The supernatant was collected after a settling time of 10 minutes, filtered through 0.45 μm cellulose-nitrate membranes (Whatman Laboratory Division, Maidstone, England), and the latter incubated on mFC basal media supplemented with 3-bromo-4-chloro-5-indolyl-β-glucopyranoside (BCIG; Inverness Medical, Ottawa, ON) for 2 h at 35˚C, then overnight at 44.5˚C. Distinctly separate, blue *E. coli* colonies were counted and

converted to concentration per dry weight of sediment. Presumptive *E. coli* were purified on Sorbitol-MacConkey agar (Oxoid, Ltd., Hampshire, England) and confirmed as *E. coli* through enzymatic (DMACA Indole; Difco Laboratories, Sparks, MD) and molecular procedures described below. Prior to DNA extraction, indole-positive isolates were cultured in tryptic soy broth (TSB; Difco Laboratories) at 37˚C for 24 hours. The average number of validated *E. coli* isolates retrieved per sample and total number of isolates collected per matrix are listed in Table 5.3.

5.2.3 DNA Extraction and Genetic Analysis

DNA was extracted from the TSB cultures using prepGEM Bacteria DNA kits (ZyGEM Corporation, Ltd., Hamilton, New Zealand) following manufacturer's protocols. Each isolate was genetically identified as *E. coli* through the phylogenetic grouping procedure developed by Clermont et al. (2000), which assigns isolates to one of four phylogenetic groups based on the presence of three target genes. Amplification conditions described by Clermont et al. (2000) were used, except the denaturation time was extended to 10 s and the annealing and extension time was extended to 15 s to consistently amplify all three bands in the control strain *E. coli* ATCC 25922. Strain typing was performed using repetitive element palindromic (rep)-PCR using BOX A1R primers (BOX-PCR), following protocols reported by Rademaker et al. (2004). Polyacrylamide gel electrophoresis (PAGE) was used, with 3.5% gels run at 7.5 V/cm for 195 minutes. *E. coli* strain ATCC 25922 was used as a positive control for intergel comparison, and AmpliSize Molecular Ruler (50-2000 bp; Bio-Rad, Hercules, CA, USA) was used as a size standard. The PAGE gels were stained with ethidium bromide and imaged using an ImageMaster VDS-CL documentation system (Amersham Pharmacia Biotech, Inc., UK).

Table 5.2. Contributing land use, channel geometry, hydrological characteristics, water quality, and sediment properties of the three investigated stream reaches (sites TB2, TB3, TB6) in the Thomas Brook Watershed, Nova Scotia, Canada.

Parameter	TB2		TB3		TB6	
	Mean	Range	Mean	Range	Mean	Range
Contributing Land Use						
Agricultural (%)	38.9		18.2		35.1	
Residential	2.9		3.9		3.4	
Forested	51.7		67.6		54.2	
Other	6.5		10.3		7.3	
Channel Geometry						
Width	3.0		3.2		4.0	
Slope	0.0011		0.0019		0.0004	
Manning's n	0.024		0.027		0.021	
Hydrology and Sediment Transport						
Q_w	0.017	0.003-0.952	0.026	0.009-0.462	0.045	0.019-1.36
τ_b	2.85	1.25-22.9	6.97	2.13-26.1	1.66	0.95-11.4
Q_s	0.029	0.004-0.953	0.080	0.027-1.325	0.011	0.004-0.435
Water Quality						
WatTemp	9.0	0.3-19.8	9.1	0.6-20.9	8.9	0.1-20.7
pH	7.4	6.5-8.1	7.3	6.4-7.8	7.2	6.5-7.7
DO	10.6	8.6-13.1	10.8	8.7-13.2	10.8	8.8-14.1
EC	0.30	0.09-0.50	0.33	0.17-0.50	0.35	0.14-0.69
TSS	9.3	0.7-43	13.6	0.9-68	17.1	0.8-145
Turbidity	6.5	0.3-35	7.2	0.4-36	16.1	0.5-124
Sediment Variables						
OM	8.45	7.0-11.3	6.61	5.0-8.0	7.03	5.2-8.9
Si-Cl	0.10	0.04-0.17	0.08	0.03-0.13	0.13	0.04-0.20
D_{50}	526		754		414	

5.2.4 Computer-Assisted Image Analysis and Cluster Assignment

BOX-PCR images were analyzed with GeneTools software (Syngene Ltd., Frederick, MD, USA). Band matching was performed using the rolling disk method for background subtraction and band sizes between 50 and 2,000 bp were used in subsequent cluster analysis. Cluster analysis was performed with GeneDirectory software (Syngene Ltd.) through unweighted pair group method with arithmetic mean (UPGMA), using Pearson's product moment correlation coefficient and a 1% threshold. Isolates exhibiting ≥86%

similarity were classified as clonal strains; a threshold determined by the lowest similarity of intergel matches for the control strain *E. coli* ATCC 25922. Those with ≤ 85% similarity were considered to be unrelated strain types. All isolates from the sediment communities at each respective monitoring location were compared in a single cluster analysis, and the abundance of each strain observed during each monitoring event was tabulated.

Waterborne *E. coli* isolates were compared against the source libraries (dairy manure, septic tanks, sediments, and tile drainage effluent) using the library matching function in GeneDirectory. For matching, each isolate was compared via UPGMA analysis against 200 of the closest matching isolates, with 3% error in band matching allowed. The rate of correct classification (RCC) was calculated for the dairy manure, septic tanks and tile drainage effluent sources. For each source, 100 isolates from each category were compared against the three catchment libraries. The RCC was calculated as the number of isolates which were correctly assigned to a given group by the total number of isolates in that group tested (n=100) and multiplying by 100. The RCC of the sediment sources could not be adequately calculated because of the hydrological continuity existing between the catchment sources and the sediment populations. All sediment isolates that matched the catchment sources were assumed to have originated from those sources.

5.2.5 Sediment Transport Calculations

Bed shear and sediment rating curves were developed for each stream reach using WinXSPRO software (v3.0, USDA Forest Service, Fort Collins, CO, USA). Cross-sectional geometry and surface water slopes were collected for each site through rod-and-

level surveying. User supplied Manning's *n* were used for the resistance equations. Typical *n* values for sand bed rivers were used and adjusted for stream meandering as discussed by Arcement and Schneider (1989). Sediment discharge was calculated using the Ackers-White model and the average D_{50} values obtained from the grain size distributions measured for each monitoring site. Power law functions were constructed using SigmaPlot software (v11.0, Systat Software Inc., Chicago, IL, USA) with stream discharge as the independent variable and bed shear stress and sediment discharge from the WinXSPRO output as the dependent variables. Continuous records of bed shear and sediment discharge were calculated using the resultant power law functions and the continuous record of discharge obtained from each of the three monitoring sites. Cross-sectional geometry, surface water slope, Manning's *n*, as well as geometric mean discharge, sediment discharge and bed shear are presented in Table 5.2.

5.2.6 Richness and Similarity of *E. coli* Populations

Total richness, or the total number of individual genotypes, of *E. coli* contained within each sediment sample were estimated through rarefaction procedures described by Lu et al. (2005). Rarefaction curves were generated through the freeware program Analytical Rarefaction 1.3, available at http://www.uga.edu/strata/software/. The rarefaction curves were plotted in SigmaPlot (v11.0, Systat Software Inc.), and the asymptotes were estimated using a one-site saturation ligand model. The asymptote (V_{max}) estimates the strain richness at sampling saturation. The proportion of the *E. coli* population characterized for each sample was calculated by dividing the number of strains detected in the sample by the estimated number of strains calculated using the rarefaction procedure outlined.

For the similarity analysis, sediment *E. coli* populations were temporally constrained to assess the degree of population change occurring between sampling events. *E. coli* population similarity was calculated using the Raup-Crick similarity index and 999 Monte Carlo permutations in PAST software (Hammer et al. 2001). The Raup-Crick similarity index is a probability-based procedure that compares the observed number of species occurring in two samples with the distribution of co-occurrences from *n* random replicates from the pool of samples (Raup and Crick, 1979). In this example, the index measures population similarity between two temporally constrained samples by evaluating the number of strains common to the two samples against the number of strains expected to be shared if the samples were randomly generated from pooling all of the strains identified from the collection site.

5.3 Results and Discussion

5.3.1 Coverage of *E. coli* Genotypic Diversity and Rates of Correct Classification

Holding tank manure displayed high *E. coli* genotype richness, but only 44% of the estimated numbers of genotypes were detected with the sampling done (Table 5.3). The human sources were better represented, with 85% of the estimated number of genotypes being characterized. Individual human or cow hosts typically harbor 1 to 10 *E. coli* genotypes at any time (Anderson et al., 2006; Gordon et al., 2010). Higher richness in *E. coli* populations of the dairy manure holding tank likely resulted from the higher number of cows (186 head) contributing to the holding tank as compared to the domestic septic tanks (1 – 4 residents per household).

Table 5.3. Number of samples, isolates collected per sample, estimated richness and percentage of strains detected for the *E. coli* populations inhabiting catchment sources, sediments, and the water column. Values presented are the averages with the range of values in parentheses.

		Number of Samples	Number of Isolates Collected per Sample	Total Number of Isolates Collected	Total Number of Genotypes Identified	Estimated Genotype Richness per Sample	Percentage of Genotypes Detected per Sample	Rate of Correct Classification*
Catchment Sources								
Dairy Manure Holding Tank		4	96	476	160	101 (68 – 139)	44 (37 – 50)	96%
Septic Tanks		4	32	388	59	15 (1 – 46)	85 (64 – 99)	96%
Tile Drainage System		33	18 (1 – 30)	991	123	24 (2 - 88)	62 (42 - 90)	100%
Stream Reaches								
TB2	Sediment	28	22 (11 – 29)	535	150	28 (6 – 68)	45 (25 – 78)	-
	Water	33	21 (11 – 26)	671	325	58 (11 – 196)	36 (9 – 61)	-
TB3	Sediment	28	18 (4 – 30)	307	124	26 (2 – 67)	44 (14 – 94)	-
	Water	33	21 (7 – 30)	665	285	55 (10 – 270)	36 (8 – 63)	-
TB6	Sediment	28	20 (3 – 29)	509	129	25 (6 – 59)	47 (13 – 82)	-
	Water	33	22 (15 – 29)	696	251	63 (12 – 254)	32 (7 – 67)	-

*Rates of correct classification were calculated only for the catchment *E. coli* sources. All sediment isolates matching catchment sources were assumed to be derived from the catchment source rather than being attributed to incorrect classification.

The stream reaches were found to exhibit similar trends in genotype richness between the sediment and water matrices at all three locations. The sediment samples contained 25 – 28 individual genotypes on average, with an average of 44% – 47% being detected (Table 5.3). The water samples exhibited higher *E. coli* strain richness, averaging 55 – 63 strains per sample and the average percentage of strains detected ranged between 32% – 36%. These results indicate that the genetic diversity of water samples is higher than sediment samples. Higher genetic diversity of waterborne *E. coli* versus sediment sources has been reported previously (McLellan, 2004; Brownell et al. 2007), presumably because sediment *E. coli* populations harbour persistent genotypes capable of survival and proliferation rather than the introduction of transient genotypes as observed in water samples (Ibekwe et al., 2011).

5.3.2 Temporal Fluctuations in Sediment *E. coli* Concentrations and Estimated Strain Richness

Sediment *E. coli* were detectable throughout the study period at all sites (Figure 5.2). Higher sediment *E. coli* concentrations were observed in stream reaches draining agricultural areas, with the highest concentrations detected adjacent to the dairy operation (site TB2, 2.1 logCFU/g), followed by the downstream cultivated field (TB6, 1.7 logCFU/g) and the low-density residential reach (TB3, 1.4 logCFU/g) (Table 5.4). Although differences in sediment *E. coli* concentrations among the stream reaches appears to be dependent on land-use, the average percentage of sediment *E. coli* isolates matching the catchment sources was relatively small for the dairy manure source at TB2 (3%) and septic system sources at TB3 (4%). No fecal (manure or septic) strains were

identified in the sediments at TB6. In contrast, the average percentage of sediment *E. coli* isolates that matched tile drainage effluent strains at TB6 was 33% (Table 5.4).

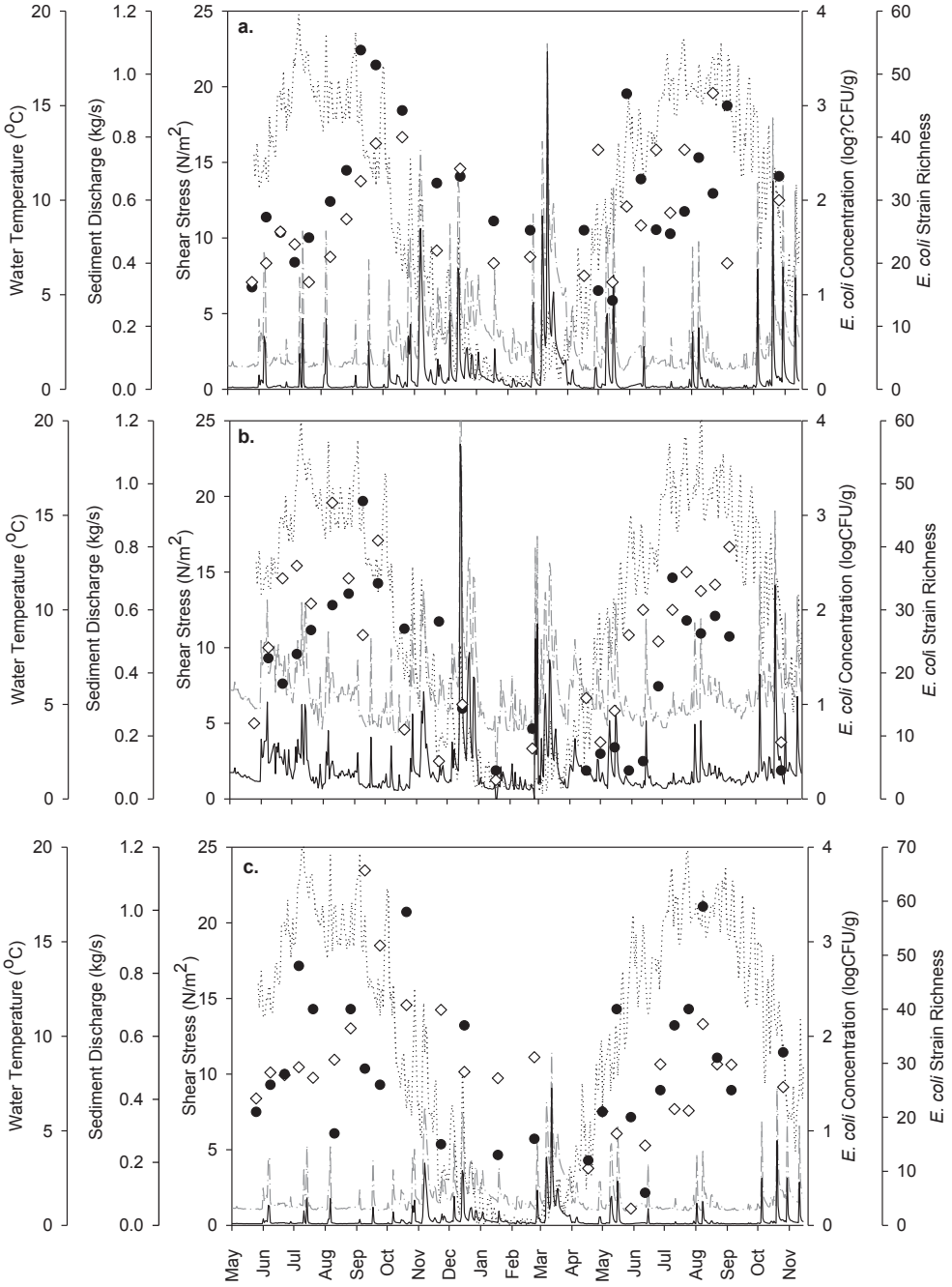


Figure 5.2. Sediment *E. coli* concentrations (•) and estimated strain richness (◊) over the course of the study, compared against average daily water temperature (dotted line),daily average bed shear stress (grey dash-dot line) and daily average sediment discharge (black line) at sites TB2 (a), TB3 (b) and TB6 (c).

Table 5.4. Pearson's correlation coefficients for sediment-borne *E. coli* populations (density, genotype richness, temporally-constrained percentage similarity, and contribution from adjacent fecal or tile drainage sources) relative to hydrological, sediment transport, water quality and sediment variables. Bold values are significant at $p<0.05$. (*E. coli* densities are in Log_{10} CFU/g).

	TB2				TB3				TB6			
	Density	Richness	Similarity	Fecal	Density	Richness	Similarity	Fecal	Density	Richness	Similarity	Tile
Mean	2.1	28	58	3	1.4	24	54	4	1.7	29	63	33
Range	0.9 – 3.6	17 - 47	8 - 98	0 - 26	0.3 - 3.2	3 - 47	15 - 96	0 - 20	0.2 – 3.8	6 – 59	2 - 99	0 - 90
Population Parameters												
Concentration	1				1				1			
Richness	**0.40**	1			**0.55**	1			**0.35**	1		
Similarity	0.07	-0.06	1		0.02	-0.16	1		0.25	0.02	1	
Fecal/Tile source	0.07	0.14	0.04	1	-0.06	0.19	0.01	1	0.01	0.08	**0.54**	1
Hydrological and Sediment Transport Variables												
Q_w	-0.28	-0.07	-0.30	-0.02	-0.16	-0.04	-0.37	0.25	-0.13	0.28	-0.05	0.11
τ_b	-0.31	-0.08	-0.37	-0.04	-0.16	-0.06	-0.37	0.25	-0.13	0.26	-0.05	0.11
Avg. τ_b	-0.21	-0.26	-0.27	-0.19	-0.07	0.02	-0.30	-0.03	-0.11	-0.19	-0.03	0.01
Max. τ_b	-0.09	-0.26	**-0.72**	-0.16	-0.28	-0.19	**-0.40**	-0.25	-0.09	-0.09	-0.03	0.08
Q_s	-0.30	-0.06	-0.32	-0.01	-0.10	-0.04	-0.27	0.10	-0.14	**0.38**	-0.05	-0.03
ΣQ_s	-0.14	-0.28	**-0.52**	-0.15	**-0.50**	**-0.41**	**-0.41**	-0.09	-0.08	-0.12	-0.01	0.08
ΣPpt	-0.09	-0.19	**-0.54**	-0.11	-0.17	-0.21	**-0.51**	0.02	0.09	-0.12	-0.11	0.08
API-7	0.06	0.12	**-0.66**	0.22	0.14	-0.20	**-0.39**	-0.27	0.29	0.15	**0.39**	0.12
TDRate	-	-	**-**	-	-	-	**-**	-	-0.05	-0.20	0.18	-0.25
Water Quality Variables												
WatTemp	0.21	0.16	0.11	0.25	**0.53**	**0.85**	-0.29	0.14	-0.08	0.32	0.15	-0.15
EC	**0.40**	0.24	**0.42**	0.13	**0.47**	0.18	0.15	0.08	**0.54**	-0.08	0.11	-0.04
DO	-0.32	-0.22	0.09	-0.23	**-0.55**	**-0.88**	0.25	-0.14	-0.13	**-0.41**	-0.12	-0.05
pH	-0.07	-0.17	0.30	0.05	0.10	**0.48**	-0.24	0.27	-0.11	-0.08	0.12	-0.02
TOC	0.06	-0.10	**0.62**	0.05	-0.19	0.20	-0.20	0.12	**-0.67**	-0.28	0.16	-0.20
NO_3-N	0.04	**0.38**	0.21	0.04	-0.33	-0.11	-0.03	-0.02	**-0.51**	-0.32	0.30	0.12
TP	0.32	0.24	-0.14	-0.29	-0.28	0.16	-0.03	-0.20	-0.30	-0.06	0.10	0.02
TSS	0.08	-0.01	-0.19	-0.22	-0.18	-0.33	-0.06	-0.06	-0.22	-0.05	0.01	-0.01
Turbidity	-0.04	-0.01	-0.16	-0.18	-0.22	-0.25	-0.07St	-0.29	-0.05	-0.06	0.03	-0.12
Sediment Variables												
OM	-0.30	**-0.36**	0.15	-0.11	**0.53**	0.10	**0.40**	-0.17	-0.05	0.11	**-0.38**	-0.05
Si-Cl	-0.24	-0.34	**0.40**	-0.15	0.29	-0.14	-0.31	-0.02	0.01	0.11	0.25	-0.02
D_{50}	0.25	0.05	-0.24	0.25	**-0.42**	0.01	-0.24	0.29	0.15	0.14	**-0.40**	-0.09

E. coli population shifts are known to occur between primary and secondary habitats, where the dominant strains in the host population are different from the dominant strains that occur in the secondary habitat (Whittam et al., 1989; Gordon et al., 2002). In a mesocosm experiment, Anderson et al. (2005) observed that fecal strains have lower capacity to survive in sediment and water systems than those isolated from environmental samples. This perhaps explains the much greater prevalence of tile drainage effluent strains within the sediment *E. coli* population in comparison to those attributed to fecal sources, since the tile drainage effluent is presumably enriched for genotypes more adapted to secondary environmental habitats (Piorkowski, unpublished).

General seasonal fluctuations in sediment *E. coli* concentrations and estimated strain richness were observed at all sites (Figure 5.2). Water temperature was significantly ($p<0.05$) correlated with sediment *E. coli* concentration and strain richness only at TB3 (Table 5.4). Kim et al. (2010) reported a strong seasonal relationship between sediment *E. coli* concentrations as a function of water temperature. *E. coli* decay rates are generally reduced at lower temperatures (Garzio-Hadzick et al., 2010), so lower concentrations of sediment *E. coli* populations may reflect the failure to detect *E. coli* in the viable-but-non-culturable state or reduced migration of *E. coli* to the sediment environment during the overwinter period. Correlations of sediment *E. coli* concentrations with strain richness support this hypothesis (Table 5.4). Since single hosts generally yield only a few strains (Anderson et al., 2006; Gordon et al., 2010), higher diversity in sediment *E. coli* is likely a function of higher wildlife densities during the summer period. However, higher discharge and sediment mobilization during the overwinter periods may explain the lower densities, since frequent redistribution of

sediment often leads to lower sediment *E. coli* densities (Muirhead et al. 2004; Jamieson et al., 2005; Droppo et al., 2011). The correlation of *E. coli* concentrations with electrical conductivity (EC) at all sites (Table 5.4) supports this assertion, as EC is a proxy variable for baseflow conditions.

Sediment organic matter had site-specific correlations with sediment *E. coli* concentrations and strain richness (Table 5.4). Percentages of organic matter were negatively correlated with *E. coli* richness at TB2, positively correlated with *E. coli* concentrations at TB3, and exhibited no correlation at TB6. Site TB2 has high organic matter content and finer texture than TB3 (Table 5.2). Increases in silt and organic matter at TB3 may improve the habitability of coarser sediments by increasing attachment substrates (Haller et al., 2009; Garzio-Hadzick et al., 2010). Conversely, increases in organic matter and silt in finer textured sediments with high initial organic matter may lower concentrations by stimulating antagonistic behaviour of the sediment microbial community, such as resource competition and predation (Surbeck et al., 2010). Byappanahalli et al. (2003) also observed that sediments with high organic detritus did not manifest greater *E. coli* concentrations compared with adjacent inorganic sediments.

5.3.3 Temporal Changes in Sediment *E. coli* Population Structure

Some *E. coli* strain types occurred at multiple time points in the sediment at all sites. However the same pattern was noted for strains associated with tile drainage effluent and fecal sources in the sediment environment, although to a lesser extent. Consequently, to conclude that particular *E. coli* strains were part of the autochthonous sediment environment at any site could not be confidently justified, as has been asserted for studies in soil and beach sands (Byappanahalli et al., 2006; Ishii et al., 2006; Ishii et al., 2007).

Frequent temporal changes in *E. coli* population structure were observed over the course of the study at all sites (Figure 5.3). The calculated percentage similarities fluctuated between 8% - 98% at TB2, 15% - 96% at TB3, and 7% - 99% at TB6. Whittam et al. (1989) reported that changes in *E. coli* population structure occurred through the immigration and local extinction of clones. Population turnover, as defined by marked declines in population similarity between two sampling events, at the upstream sites (TB2 and TB3) appeared to be dependent on sediment transport events since *E. coli* population similarity was negatively correlated with the 7-day antecedent precipitation index (API-7), maximum bed shear stress, cumulative sediment discharge, and cumulative precipitation between sampling events (Table 5.4). These observations demonstrate the importance of sediment redistribution as a driver for *E. coli* population shifts in high-energy stream reaches that have higher sediment redistribution. Thus, *E. coli* populations were more similar during baseflow conditions at these study sites. At fine scales of sampling (days), *E. coli* populations have been found to oscillate during a period of equilibration in soils and sediments, ultimately resulting in a stable population once equilibrated (Topp et al., 2003; Garzio-Hadzick et al., 2010). These oscillations are the result of strain-dependent responses to the secondary environmental matrix, with only a subset of the introduced strains being able to survive (Guber et al., 2007; Brownell et al., 2007). Due to the coarser temporal sampling in this study, these oscillations were not detected, but stability of the *E. coli* population inhabiting the sediments between redistribution events was observed.

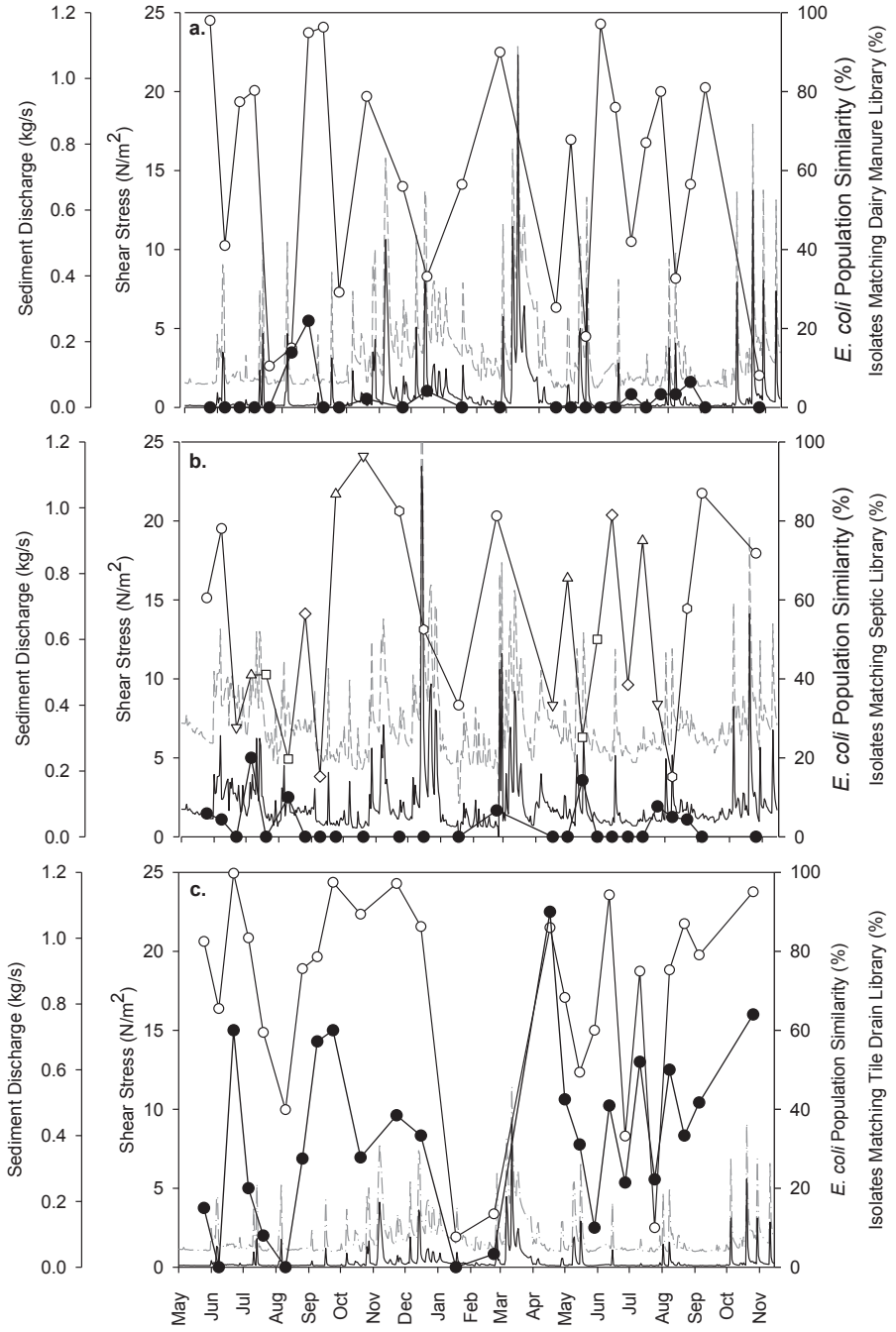


Figure 5.3. Percentage similarity of sediment *E. coli* populations between successive sampling events (o) and the percentage of sediment *E. coli* populations matching adjacent catchment sources (•) compared against daily average bed shear stress (grey dash-dot line) and daily average sediment discharge (black line) at sites TB2 (a), TB3 (b) and TB6 (c).

In the lower energy downstream reach (TB6), the similarity of sediment *E. coli* populations was less influenced by hydrological and sediment transport variables. The

strongest correlation (r = 0.54) existed between population similarity and the percentage of tile drainage isolates identified in the sediment environment, with positive correlations to API-7 and negative correlations to sediment organic matter and median grain size (Table 5.4). These correlations suggest that the sediment *E. coli* population was strongly influenced by the adjacent tile drained agricultural field, the connectivity of which is driven by antecedent precipitation. Changes in the *E. coli* population at TB6 were not correlated with hydrological or sediment transport variables, perhaps due to the relative stability of the sediment bed at this reach. In beach sands, *E. coli* populations have been reported to be stable (Ishii et al., 2007; Kon et al., 2007), with gradual population shifts resulting from strain migration into the sediments (Byappanahalli et al., 2006). The positive relationship between sediment and tile drainage effluent population implies that similarity is driven by the influence of tile drainage effluent strains, with immigration from upstream sources or wildlife possibly responsible for population turnover in this reach.

5.3.4 Waterborne *E. coli* Concentrations in the Thomas Brook Watershed

Waterborne *E. coli* concentrations frequently exceeded guidelines for irrigation water quality (100 CFU/100 mL; CCME, 1999) over the course of the study period (Figure 5.4). The percentage of samples exceeding guideline values were 82% at TB2, 51% at TB3 and 82% at TB6. Water collected adjacent to livestock operations and/or manure-amended fields tended to have higher *E. coli* concentrations than those sampled adjacent to low-density residential developments. This is in agreement with prior reports of greater concentrations of waterborne *E. coli* in subcatchments draining agricultural land

uses relative to forested or low-density residential developments (George et al., 2004; Servais et al., 2007; Tetzlaff et al., 2012). No significant correlation existed between sediment and waterborne *E. coli* concentrations at the stream reaches investigated, matching previous reports (Schilling et al., 2009; Pachepsky and Shelton, 2011). At all sites, *E. coli* concentrations followed predictable correlations with discharge, bed shear, sediment transport, daily precipitation, waterborne nutrients (particularly total phosphorus), total suspended solids and turbidity (Tables 5.5 and 5.6); relationships that are well supported in the literature (Rodgers et al., 2003; Cho et al, 2010; Pandey et al., 2012).

More interestingly, waterborne *E .coli* concentrations exhibited site-specific differences in their relationship with the identified sediment or catchment sources. At site TB2, *E. coli* concentrations were positively correlated with the dairy manure contribution, and were not related to sediment-identified strains (Table 5.5). Conversely, waterborne *E. coli* concentrations were positively correlated with the contribution of sediment *E. coli* strains at TB3, and not correlated with adjacent septic sources (Table 5.5). These observations indicate reach-specific differences in hydrologic connectivity between catchment and sediment sources of waterborne *E. coli*. At the downstream site,TB6, the waterborne *E. coli* concentration was positively correlated with the TB2 reach sources (both sediment and dairy manure) and negatively correlated with strains identified in the TB3 reach (Table 5.6), indicating that the contribution from TB2 dominates during high *E. coli* loading and TB3 reach dominates during periods of lower *E. coli* loading.

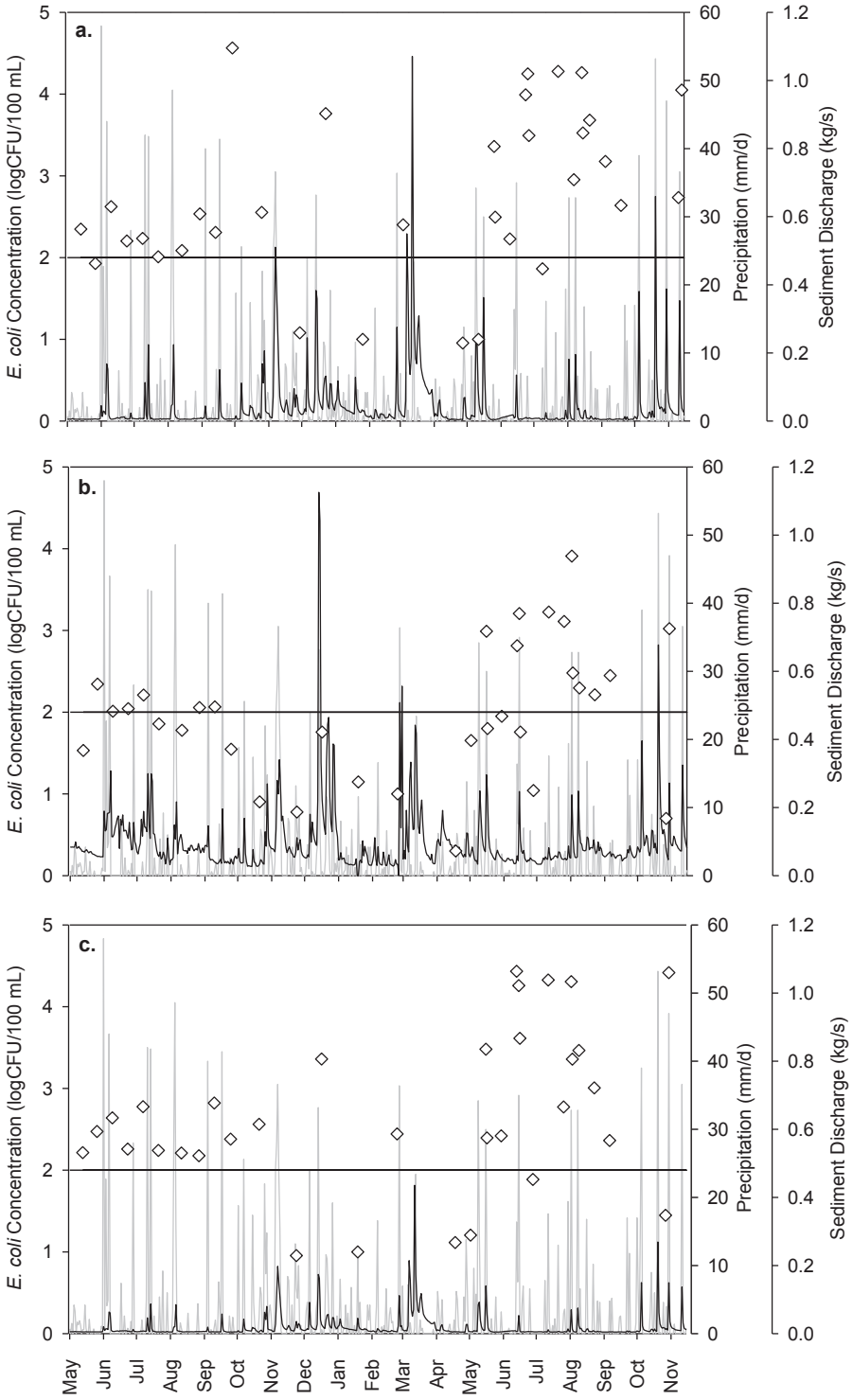


Figure 5.4. Waterborne *E. coli* concentrations (◊) over the course of the study compared against daily precipitation (grey bars) and daily average sediment discharge (black line) at sites TB2 (a), TB3 (b) and TB6 (c).

Table 5.5. Pearson's correlation coefficients for waterborne *E. coli* densities and source contributions from sediments and adjacent fecal sources in relation to sediment *E. coli* population parameters, hydrological and sediment transport variables, and water quality variables at sites TB2 and TB3. Bold values indicate significant correlations at $p<0.05$. (*E. coli* densities in $Log_{10}CFU/100$ mL).

	TB2				TB3			
	Density	Sediment	Cow Manure	Unknown	Density	Sediment	Septic Systems	Unknown
Mean	2.7	13	17	70	2.0	18	21	62
Range	1.0 – 4.6	0 - 50	0 - 68	6 - 100	0.3 – 3.9	0 - 90	0 - 80	0 - 100
Waterborne Concentration and Source Contribution								
Density	1				1			
Sediment	0.02	1			**0.38**	1		
Fecal	**0.36**	-0.07	1		0.14	-0.08	1	
Unknown	-0.30	**-0.59**	**-0.83**	1	-0.28	**-0.54**	**-0.42**	1
Sediment E. coli Population Variables								
Density	0.27	0.16	0.21	-0.29	0.30	0.26	-0.01	0.07
Richness	0.32	0.14	-0.29	0.17	**0.52**	0.12	0.16	-0.01
Hydrological and Sediment Transport Variables								
Q_w	**0.48**	0.17	**0.35**	**-0.38**	**0.40**	**0.47**	0.25	**-0.52**
τ_b	**0.48**	0.14	**0.34**	**-0.38**	**0.40**	**0.47**	0.25	**-0.52**
Avg. τ_b	-0.02	-0.06	0.13	-0.08	-0.02	0.35	**0.49**	**-0.39**
Max. τ_b	-0.12	0.14	**0.35**	**-0.37**	-0.33	0.25	0.24	-0.28
Q_s	**0.49**	0.15	**0.35**	**-0.38**	**0.40**	**0.47**	0.33	**-0.52**
ΣQ_s	**-0.35**	-0.11	-0.07	-0.08	**-0.57**	0.01	0.02	-0.01
ΣPpt	-0.28	0.26	0.06	-0.19	-0.25	-0.10	0.14	-0.04
API-7	0.20	0.05	-0.08	0.05	-0.09	0.03	**0.45**	-0.10
DailyPpt	**0.57**	0.09	**0.61**	**-0.43**	**0.67**	0.10	0.02	**-0.37**
Water Quality Variables								
WatTemp	0.24	-0.02	0.08	-0.02	**0.50**	0.19	0.04	-0.04
EC	-0.26	0.01	-0.32	0.26	-0.19	-0.29	-0.28	0.34
DO	**-0.49**	0.03	-0.14	0.08	**-0.52**	-0.17	-0.08	0.12
pH	-0.17	-0.20	-0.12	0.22	0.02	-0.26	0.02	0.12
TOC	**0.41**	-0.11	-0.22	0.23	0.07	0.17	-0.28	0.29
NO_3-N	0.05	0.06	0.11	-0.04	-0.06	0.15	0.02	0.20
TP	**0.78**	0.16	**0.54**	**-0.54**	**0.42**	-0.07	**0.36**	-0.14
TSS	**0.64**	0.16	**0.36**	**-0.39**	**0.59**	0.16	0.01	-0.23
Turbidity	**0.67**	0.16	**0.38**	**-0.37**	**0.48**	0.15	0.12	-0.22

Table 5.6.Pearson's correlation coefficients for waterborne *E. coli* density and source contributions at site TB6 in relation to *E. coli* population parameters, hydrological and sediment transport variables, and water quality variables. Bold values indicate significant correlations at $p<0.05$. (*E. coli* densities in $Log_{10}CFU/100$ mL).

	Density	Tile Drain Effluent	TB6 Sediment	TB2 Sediment	TB3 Sediment	Septic	Cow Manure	Unknown
Mean	2.9	14	1	27	12	1	1	42
Range	1.0 – 4.4	0 - 48	0 - 8	5 - 65	0 - 37	0 - 14	0 - 17	10 - 80
Population Parameters								
Density	1							
Tile drain effluent	-0.06	1						
TB6 sediment	0.28	0.13	1					
TB2 sediment	**0.61**	0.03	0.12	1				
TB3 sediment	**-0.39**	-0.07	-0.18	-0.05	1			
Septic	0.12	0.23	-0.14	-0.04	0.14	1		
Cow manure	**0.36**	0.13	0.29	**0.38**	-0.18	-0.05	1	
Unknown	0.01	**-0.36**	-0.03	**-0.72**	-0.11	0.02	-0.33	1
Sediment E. coli Population Variables								
Density	-0.05	**0.41**	0.09	-0.31	-0.02	-0.03	**-0.42**	-0.05
Richness	**0.39**	0.01	-0.17	-0.01	-0.20	0.33	0.11	0.13
Hydrological and Sediment Transport Variables								
Q_w	**0.52**	0.25	**0.45**	**0.42**	-0.01	0.16	**0.56**	-0.30
τ_b	**0.52**	0.24	**0.45**	**0.43**	-0.03	0.16	**0.55**	-0.33
Avg. τ_b	-0.02	0.29	0.11	0.32	0.32	-0.08	0.32	**-0.41**
Max. τ_b	-0.24	0.32	0.00	-0.06	0.04	-0.08	0.21	-0.14
Q_s	**0.52**	0.29	0.31	**0.42**	0.03	0.16	**0.56**	-0.27
ΣQ_s	-0.33	0.01	0.05	-0.14	0.18	-0.07	-0.05	0.02
Σ Ppt	-0.27	**0.47**	0.12	-0.20	**0.36**	-0.08	-0.06	-0.01
API-7	0.11	**0.56**	-0.06	0.06	0.05	**0.40**	0.06	-0.05
DailyPpt	**0.69**	-0.15	**0.67**	**0.38**	-0.12	-0.11	**0.59**	-0.21
TDRate	-0.16	**0.64**	-0.15	0.09	0.01	0.12	0.21	-0.33
TB2- τ_b	**0.52**	0.25	**0.45**	**0.46**	-0.01	0.16	**0.55**	**-0.35**
TB2- Q_s	**0.52**	0.22	**0.45**	**0.46**	0.00	0.14	**0.55**	-0.28
TB3- τ_b	**0.46**	0.10	0.15	0.32	0.03	0.21	0.29	-0.25
TB3- Q_s	**0.46**	0.05	0.26	0.32	-0.05	0.21	0.33	-0.16
Water Quality Variables								
WatTemp	0.33	**-0.51**	0.03	-0.10	**-0.38**	0.09	-0.20	0.30
EC	-0.32	-0.01	-0.27	**-0.52**	0.14	-0.06	**-0.36**	0.30
DO	**-0.55**	**0.42**	-0.17	0.00	**0.43**	0.04	0.07	-0.28
pH	-0.21	**-0.36**	-0.27	-0.13	0.11	-0.23	-0.31	0.17
TOC	0.24	-0.2	0.07	0.16	-0.04	0.29	**0.36**	-0.11
NO_3-N	**0.47**	-0.07	**0.39**	0.28	**-0.47**	0.13	**0.42**	-0.16
TP	**0.74**	-0.04	**0.44**	**0.55**	-0.29	-0.20	**0.88**	-0.34
TSS	**0.64**	0.04	**0.66**	**0.50**	-0.26	-0.16	**0.50**	-0.22
Turbidity	**0.70**	0.04	**0.69**	**0.41**	-0.33	-0.13	**0.55**	-0.20

5.3.5 Comparative Contribution of Sediment versus Catchment Sources of Waterborne *E. coli*

The contribution of dairy manure sources to the waterborne *E. coli* population ranged between 0-68% at TB2, the contribution of septic inputs ranged between 0-80% at TB3, and the combined input of dairy manure and septic inputs ranged between 0-31% at TB6 (Tables 5.5 and 5.6). Variable levels of fecal input into water loads have been observed in other studies. Human inputs have been measured at 1 to 20%, livestock at 10 to 65%, and wildlife inputs at 5 to 75% of the waterborne *E. coli* population in other studies (Whitlock et al., 2002; Ahmed et al., 2007; Somarelli et al., 2007; Kon et al., 2010; Liwimbi et al., 2010), reflecting the importance of site-specific hydrological connectivity between fecal sources and water monitoring locations. Environmentally adapted *E. coli* from secondary sources (e.g. sediments) have been reported to contribute up to 23% of the waterborne population (Kon et al., 2010), making a strong case for including environmentally adapted strains in microbial source tracking studies.

At the upstream sites, waterborne *E. coli* genotypes matching those isolated from the sediment beds ranged between 0 to 50% at TB2 and 0 to 90% at TB3 (Table 5.5). At the downstream site (TB6), the *E. coli* population inhabiting the reach sediments demonstrated low interaction with the water column (0 – 8%). Sediment at the upstream sites contributed substantially to the waterborne *E. coli* population, with 5 – 65% contribution from TB2 and 0 – 37% contribution from TB3. The total contribution from sediments at TB6 averaged 39% and ranged between 5 to 79%. The tile drainage effluent from the adjacent cultivated field contributed 0 – 48% of the *E. coli* load. At all sites, the percentage of unknown isolates were high, with an average of 70% at TB2, 62% at TB3,

and 42% at TB6. The unidentified components of the upstream reaches likely reflect wildlife inputs, but could also represent uncharacterized *E. coli* isolates from the fecal sources or environmentally adapted strains from upstream locations. The latter argument is supported by the waterborne *E. coli* population at TB6 being populated by upstream sediment *E. coli* genotypes.

5.3.6 Correlations of Reach-Specific Hydrology on Sediment Contributions to Waterborne *E. coli*

Sediment contributions to waterborne *E. coli* at the upstream sites demonstrated differences in hydrological connectivity between the sediment and the water column. The sediment contribution at site TB2 exhibited no significant correlation to any of the hydrological, sediment transport, or water quality variables included in the analysis (Table 5.5). However, the sediment contributions at TB2 did appear to be relatively consistent (Figure 5.5a). Conversely, the sediment contributions at TB3 were more influenced by sediment transport, being positively correlated with discharge, bed shear stress, and sediment discharge (Table 5.5). The importance of sediment transport on elevated waterborne *E. coli* concentrations is well established (Rehmann and Soupir, 2009; Wilkinson et al., 2011). However, sediment-borne *E. coli* have also been reported to populate the water column outside of the influence of sediment redistribution (Jamieson et al., 2004; Yakirevich et al., 2013), and models incorporating hyporheic exchange at the sediment-water interface have yielded improved *E. coli* prediction accuracy (Grant et al., 2011; Ghimire and Deng, 2013). Since sediment contributions at TB2 are consistent, yet unrelated to sediment transport, it appears that hyporheic

exchange between sediments and the water column may be important than sediment resuspension at this site.

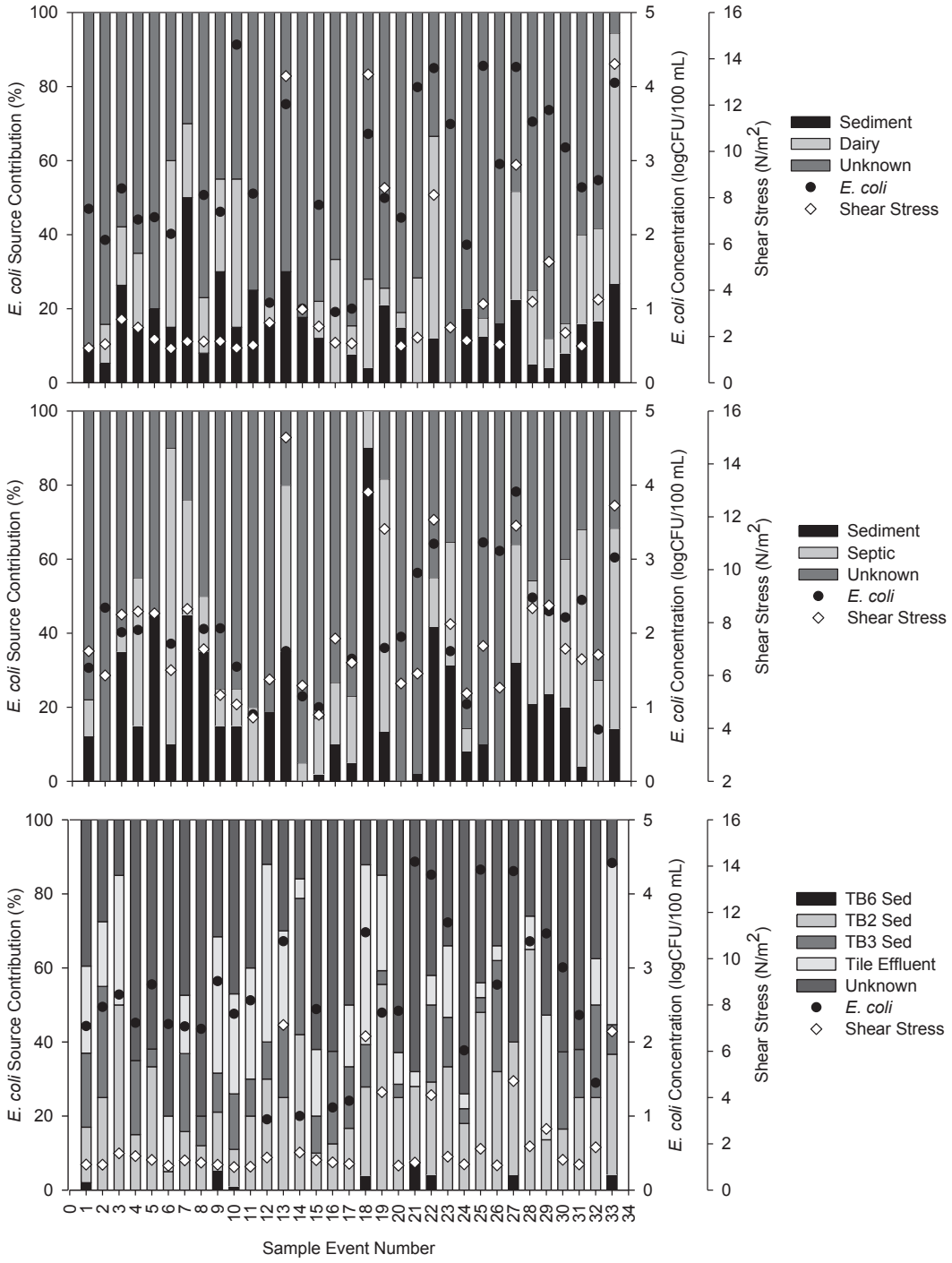


Figure 5.5. Graphical representation of the identified *E. coli* sources at each sampling event for sites TB2 (a), TB3 (b) and TB6 (c). Percentage contributions from reach sediments, catchment sources, upstream sediments (TB6 only) and unknown sources are included. Waterborne *E. coli* concentrations and daily average bed shear stress for each sampling event are also presented.

It is possible that the differences in *E. coli* transport behaviour between reaches TB2 and TB3 may in part result from strain-specific differences in particle attachment behaviour. Sediment-driven transport of *E. coli* is more important when a higher proportion of waterborne *E. coli* is attached to particles (Rehmann and Soupir, 2009). *E. coli* particle attachment has been found to be strain-specific (Pachepsky et al. 2008; Cook et al., 2011), and *E.coli* populations arising from human hosts have been found to have greater particle attachment capacity than those arising from livestock manure (Boutilier et al. 2009). Piorkowski et al. (2013b) reported higher particle attachment percentages in waterborne *E. coli* in stream reaches dominated by human versus livestock inputs. The stronger relationship between sediment *E. coli* contributions and sediment transport at TB3 may be in part due to the greater particle attachment efficiency of human-derived *E. coli*. Conversely, lower attachment efficiencies of livestock-associated *E. coli* at site TB2 may result in greater bacterial penetration into the sediment environment as a result of lower filtration efficiency of the sediment matrix (Page et al., 2012). The conjoint importance of particle attachment efficiencies and hyporheic exchange rates into sediment compartments warrants further investigation.

The *E. coli* populations inhabiting sediments at the downstream site (TB6) were infrequently detected in the water column, averaging 1% of the waterborne *E. coli* population. This observation may result from the low rates of sediment resuspension and hyporheic exchange occurring in this reach. Russo et al. (2011) also observed relative insignificance of resuspension on waterborne coliforms in a slow-moving stream and speculated that waterborne *E. coli* was dominated by catchment sources. At site TB6, fecal catchment sources constituted an average 2% of the waterborne population while

the tile drainage effluent averaged 14% of the waterborne *E. coli*. However, the waterborne *E. coli* population at TB6 comprised an average of 27% from the TB2 sediment *E. coli* population and 12% from the TB3 sediment population. Wu et al. (2009) also linked waterborne *E. coli* strains inhabiting upstream sediments to downstream water samples. These observations demonstrate that upstream sediments can play an important role in the waterborne *E. coli* population of slow-moving stream segments, even in the absence of sediment redistribution occurring within the investigated reach.

At TB6, the contributions of *E. coli* from upstream sediments demonstrated differences in hydrological connectivity that somewhat contradicted the results of the reach-specific investigations at TB2 and TB3. The TB2 sediment contributions to TB6 were correlated with discharge and sediment transport variables, whereas the contributions from TB3 correlated with API-7 and cumulative precipitation. Cho et al. (2010) reported that high *E. coli* concentrations within upstream sediment compartments could have a strong influence on downstream waterborne *E. coli* during sediment resuspension events. Although sediment-water exchange was not entirely influenced by sediment transport within the TB2 reach, the *E. coli* load disseminating from TB2 during storm events likely dominated the waterborne *E. coli* load at the downstream TB6 location. Conversely, TB3 sediments were associated with indicators of high precipitation but not immediate transport, suggesting a lag period. Boulton et al. (2010) discuss that sediment texture and stream morphology are important drivers of hyporheic exchange. The coarser textured sediments and particular morphological ffeatures of TB3 may have resulted in hyporheic processes occurring within or downstream from this reach, resulting in the observed lag in sediment contributions to the waterborne population at TB6.

Reach and catchment-scale hyporheic processes are known to create different ecological manifestations (Boulton et al., 1998). The relative effect of scale on hyporheic processes on sediment *E. coli* transport requires further investigation.

5.4 Conclusion

The Thomas Brook Watershed is a headwater agricultural watershed that demonstrates chronic *E. coli* contamination above irrigation water quality guidelines. This study sought to identify the relative contribution of sediment-borne *E. coli* to waterborne *E. coli* loads under variable stream flow regimes in comparison to fecal sources and agricultural fields. Reach-specific differences in sediment texture, flow conditions and sediment transport were seen to affect sediment *E. coli* populations. Greater sediment *E. coli* population turnover occurred in upstream reaches with more dynamic sediment transport behaviour. In contrast, immigration processes from adjacent catchment sources were important factors for population stability in low-energy stream reaches. Sediment *E. coli* strains were found in the water column at most sampling events, but the relative importance of sediment resuspension and speculated hyporheic exchange was reach-specific. Hyporheic processes were postulated to be operating at reach and catchment scales in this watershed. Further investigation on the contribution of hyporheic exchange to small- and large-scale transport is suggested.

CHAPTER 6 EVALUATION OF STATISTICAL MODELS FOR PREDICTING *E. COLI* PARTICLE ATTACHMENT IN FLUVIAL SYSTEMS

Materials in this chapter have been published in **Water Research** and are being reproduced in this thesis with permission from the publisher, **Elsevier**.

Piorkowski, G.S., R.C. Jamieson, G.S. Bezanson, L. Truelstrup Hansen and C.K. Yost, 2013b. Evaluation of statistical models for predicting *E. coli* particle attachment in fluvial systems. Water Research, doi: 10.1016/j.watres.2013.09.003.

6.1 Introduction

Fecal contamination of water resources, and possible co-occurrence of pathogens, is monitored through the use of fecal indicator bacteria (FIB). *Escherichia coli* are the recommended FIB for freshwater sources, and have been extensively monitored and modeled in these systems. Recent advances in simulating *E. coli* dynamics within complex watersheds have been made with deterministic models such as SWAT and WATFLOOD (Dorner et al., 2006; Kim et al., 2010). Model advancements have included explicit, process-based representations of *E. coli* particle interactions, and settling and resuspension of fecal microorganisms within surface water systems, in general leading to improved efficacy of these models for predicting surface water *E. coli* concentrations (Wu et al., 2009). To predict the rates of suspension and deposition, the percentage of waterborne *E. coli* that are attached to suspended particles needs to be estimated. This is known to be a sensitive parameter that contributes to model uncertainty (Pandey et al., 2012). In previous studies, the particle attached *E. coli* fraction has been estimated as a static parameter (Dorner et al., 2006), or dynamically predicted based on a linear correlation with total suspended solids (Bai and Lung, 2005), or the proportion of

suspended clay (Kim et al., 2010). Considering that the attachment parameter contributes to model uncertainty, these approaches for parameterizing attachment may not be adequate as suspended solid concentration, particle size distribution and particle composition is highly variable in fluvial systems. Also, other environmental factors may affect particle attachment.

E. coli particle attachment has been demonstrated to be dependent upon particle size (Soupir et al., 2010), suspended sediment load (Garcia-Armisen and Servais, 2009), water chemistry (Park et al., 2008), organic content of the suspended particles (Guber et al., 2007), and stormflow conditions (Characklis et al., 2005), with differences existing between the rising and falling limbs of the storm hydrograph (Krometis et al., 2007). Strain-dependent variability in *E. coli* particle attachment resulting from differential expression of genes encoding cellular surface properties has been reported (Pachepsky et al., 2008; Foppen et al., 2010). Therefore, *E. coli* derived from different fecal sources may demonstrate differences in particle attachment. Indeed, Boutilier et al. (2010) reported differences in *E. coli* attachment percentage as a function of wastewater type. Pachepsky et al. (2008) further suggested that variable attachment behaviour may result in environmental transport differences between pathogenic and non-pathogenic *E. coli*. Pathogenic *E. coli* have been reported to be prevalent in recreational and drinking water sources and exhibit a lack of correlation to culturable FIB (Duris et al., 2009; Huan et al., 2012). Understanding the occurrence of virulent *E. coli* and the nature of particle association of these strains are important steps toward evaluating risks associated with these organisms.

Since *E. coli* particle attachment is variable, assuming a constant attachment percentage on a watershed scale may increase uncertainty in water quality models. Use of statistical models for predicting *E. coli* particle attachment may decrease model uncertainty; however, linear regression models developed to predict *E. coli* attachment often demonstrate poor results (Bai and Lung, 2005; Characklis et al., 2005). A combination of water quality, flow, and land use factors are likely required to build more accurate predictive models. Further, linear models may not be appropriate as non-linear and non-monotonic relationships often exist with strain-dependent data. Environmental data obtained to build complex regression models often demonstrates multicollinearity and high dimensionality resulting from a high independent variable to sample ratio (Loh, 2011). Alternative types of statistical models are required to develop suitable prediction models for particle attachment.

Least angle shrinkage and selection operator (LASSO) offers independent variable selection, as in subset regression, while not suffering from overfitting as a result of multicollinearity and high dimensionality by introducing a bias term to the regression coefficients (i.e. regularizing), as in ridge regression (Tibshirani, 1996). Classification and regression trees (CART) are statistical prediction models based on recursive partitioning that accommodate non-monotonic relationships, account for interactions between explanatory variables, and do not assume linear relationships among variables (Parkhurst et al., 2005). Because of these benefits, CART analysis is increasingly being applied recreational water quality assessment (Stidson et al., 2012), and FIB concentration and pathogen occurrence in watersheds (Wilkes et al., 2011; Jones et al., 2012). Multivariate adaptive regression splines (MARS) are nonparametric models that

combine recursive partitioning and spline fitting, allowing for independent variable subset selection, while maintaining model continuity that is lacking in CART algorithms (Friedman, 1991).

Using statistical predictive models that are not confounded by overfitting, multicollinearity or restriction to linear relationships may allow for a more confident approach to parameterizing *E. coli* particle attachment in watershed models. The objectives of this study were to: (i) identify hydrological, water quality, particle-property, and land use variables important for defining the *E. coli* particle attachment and virulent *E. coli* presence and attachment behaviour; and (ii) construct and evaluate statistical models (LASSO, CART and MARS models) for predicting total *E. coli* particle attachment, and virulent *E. coli* presence in surface waters.

6.2 Materials and Methods

6.2.1 Sampling Location and Strategy

Sampling was conducted in the Thomas Brook Watershed, which is part of Agriculture and Agri-Food Canada's Watershed Evaluation of Beneficial Management Practices (WEBs) program, and located in the Annapolis Valley of Nova Scotia, Canada. Significant surface water *E. coli* loading has been previously documented in this watershed (Sinclair et al., 2008). Four monitoring locations (Sites 2, 3, 4 and 6) were selected for this study: Site 2 is downstream from a large dairy operation; Site 3 is influenced by low-density residential development; Site 4 is in a mixed land-use area; and Site 6 is located below a cultivated field that receives annual dairy manure applications (Figure 1). Watershed drainage area and contributing land-use information

for each site were used as explanatory variables in the regression models, and were calculated from a GIS database developed for the Thomas Brook Watershed (Table 6.1).

Table 6.1. Summary of response and explanatory variables used in the regression models, including variable abbreviations, units and descriptions.

Variable Name	Units	Description	Range of Values
Water Quality			
EC	mS/cm	Electrical conductivity	0.09 – 0.53
pH	-	Logarithm of hydrogen ion activity	6.40 – 7.90
DO	mg/L	Dissolved oxygen	5.2 –12.35
WatTemp	°C	Instantaneous water temperature (sonde)	5.9 – 18.3
DailyWatTemp	°C	Average daily water temperature (datalogger)	5.9 – 18.3
Particle Load			
TSS	mg/L	Total suspended solids	1.12 – 390
VSS	mg/L	Volatile suspended solids	0.25 - 80
%OM	%	Percentage organic matter: (VSS/TSS)*100	5 - 96
Turbidity	NTU		1.4 - 354
Particle Size Distribution			
PartConcn	µL/L	Total volumetric particle concentration	10.8 – 500
%Sand	%	Percentage of size classes >63 um by concentration	9.8 – 85.2
%Silt/Clay	%	Percentage of size classes <63 um by concentration	14.8 – 90.2
GeoMean	µm	Geometric mean particle diameter	5.3 – 46.4
D_{50}	µm	Median particle diameter	15.0 – 488.5
D_{10}	µm	Effective particle size	3.0 – 41.2
Sorting		Measure of dispersion in particle diameters	2.9 – 13.9
$D_{75/25}$		Ratio of interquartile particle diameters	3.6 – 21.6
D_{75}-D_{25}	µm	Interquartile range of particle diameters	25.4 – 465.3
Hydrology			
Discharge	m^3/s	Volumetric stream discharge	0.006 – 2.57
FlowStage		Baseflow, Rising Limb, Falling Limb	
DailyRain	mm	Daily accumulated rainfall	0 - 47
MaxIntensity	mm/15 min	Maximum rain intensity on the day of sampling	0 - 10
AvgPpt	mm	Average daily rainfall over 3 d, 5 d and 7 d prior to sampling	0 – 11.8
TotalPpt	mm	Accumulated rainfall over 3 d, 5 d and 7 d prior to sampling	0.6 – 59.4
Land Use			
DrainArea	km^2	Drainage area contributing to monitoring location	1.18 – 4.82
%Resid	%	Percentage of total drainage area that is residential	2.9 – 5.3
%Agric	%	Percentage of total drainage area that is agricultural	18.2 – 47.2
%Forest	%	Percentage of total drainage area that is forested	41.3 – 67.6
%Other	%	Percentage of drainage area occupied by other uses	6.2 – 10.3

Grab samples (500 mL) were collected from the cross-sectional midpoint of each stream segment for bacteriological and particulate analysis on a bi-weekly schedule from early June to late October 2012. Sampling also occurred during four storm events (>20 mm rainfall), where samples were retrieved during the rising and falling limbs of the stream hydrograph. Over the course of the study, fifteen sampling events were conducted, yielding 60 samples from the four monitoring sites. The retrieved samples were stored at 4˚C and cultured for *E. coli* within 24 hours of collection.

6.2.2 Hydrologic and Water Quality Monitoring

At each sampling event, water quality and flow velocity were measured. Water quality parameters (pH, temperature, dissolved oxygen, conductivity) were measured instantaneously with a multi-parameter water quality sampling sonde (600R, YSI Environmental Incorporated, San Diego, CA, USA). Water velocity was measured with a FlowTracker Acoustic Doppler Velocimeter (SonTek/YSI, San Diego, CA, USA), and the discharge was calculated using the mid-section method. Water temperature was collected continuously using HOBO temperature/light pendant dataloggers (Model E-348-UA-002-08; Onset®, Cape Cod, MA, USA) installed at each monitoring location. Meteorological variables (air temperature, precipitation, relative humidity, barometric pressure, wind speed and direction) were collected continuously throughout the study period with a HOBO weather station datalogger (Model E-348-H21-001; Onset®) installed near Site 6. Readings were collected every 30 seconds and averaged into 15 minute intervals. Precipitation and air temperature were used directly or as the source information for indices included in Table 6.1.

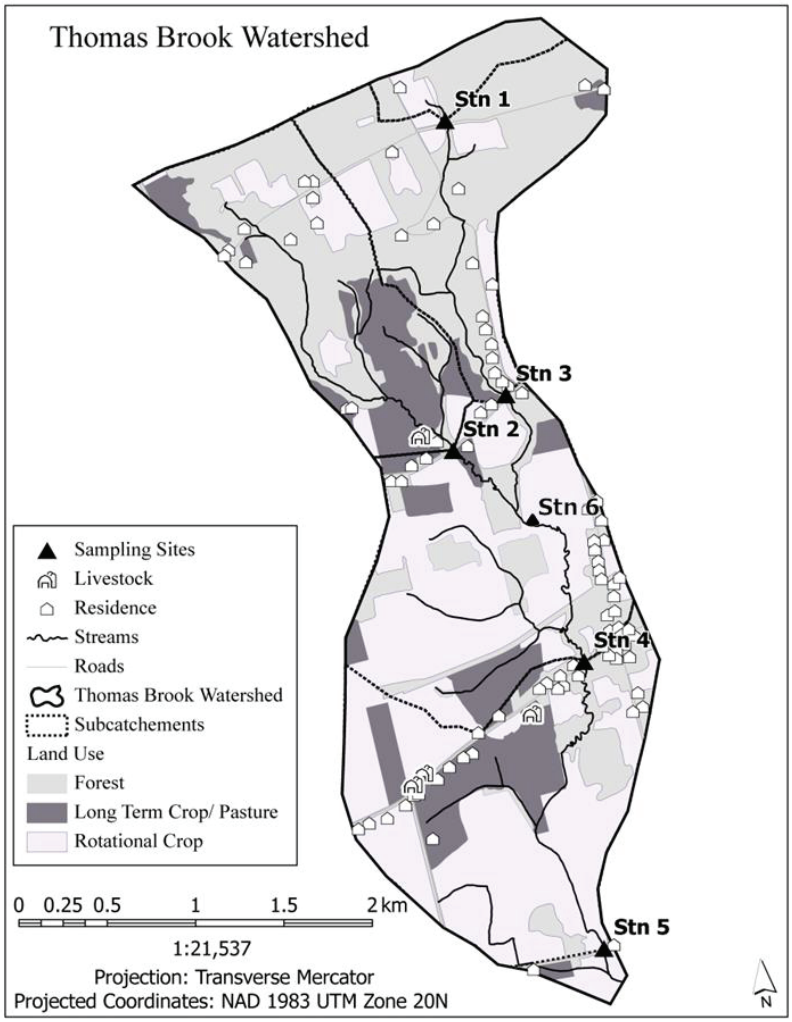


Figure 6.1. Map of the Thomas Brook Watershed identifying permanent monitoring locations and adjacent land uses. Sites 2, 3, 4 and 6 were investigated in this study.

6.2.3 Particle Analysis

Several particle characteristics were analyzed for each sample. Turbidity was analyzed with a laboratory turbidimeter (2100AN, Hach Company, Loveland, CO, USA). Total suspended solids (TSS) and volatile suspended solids (VSS) were analyzed by filtering a volume of water through glass microfiber filter (934-AH, Whatman, Maidstone, UK), and drying for a minimum of 4 h at 105°C (TSS), followed by combustion in a muffle furnace at 550˚C for 20 minutes (VSS). Particle size distribution between 2.5 µm and 500

μm was determined in the laboratory using a laser *in situ* scanning and transmissometer (LISST-100X, Sequoia Scientific Inc., Bellevue, WA, USA). Particle size properties were calculated with GRADISTAT software (version 4.0; Blott and Pye, 2001), and the derived variables are presented in Table 6.1.

6.2.4 *E. coli* Enumeration and Attachment Calculation

The percentages of particle-attached *E. coli* were measured using fractional membrane filtration. Water samples were first filtered through a 5 μm nitrocellulose membrane (EMD Millipore, Billerica, MA, USA), with the retentate containing the particle-attached fraction. The filtrate was then passed through a 0.45 μm cellulose-nitrate membrane (Whatman Laboratory Division, Maidstone, UK) to obtain the unattached fraction. Since the size of *E. coli* cells typically ranges between 1-3 μm, a 5 μm nominal pore opening was selected to allow free cells to pass through the filter. Moreover, the filter type chosen allowed for direct incubation on growth media without the need for an elution step. The membranes were incubated on mFC basal media supplemented with 3-bromo-4-chloro-5-indolyl-β-glucopyranoside (BCIG; Inverness Medical, Ottawa, ON) for 2 h at 35˚C, then incubated overnight at 44.5˚C. Distinctly separate, blue *E. coli* colonies were counted and converted to colony forming units per 100 mL. The percentage of *E. coli* attached to particles was calculated as:

$$\% \text{ Attachment} = CFU_{5\mu m}/100 \text{ mL} \div (CFU_{5\mu m}/100 \text{ mL} + CFU_{0.45\mu m}/100 \text{ mL}) \times 100 \quad (6.1)$$

6.2.5 DNA Extraction and Pathogen Marker Detection

Following enumeration, the membranes were transferred to 15 mL centrifuge tubes filled with 10 mL sterile phosphate-buffered saline (EMD Millipore) supplemented with 0.05%

Tween 80 (Fisher Bioreagents, Fair Lawn, NJ, USA), and vortexed for 10 minutes. The membranes were removed and the samples were centrifuged for 10 minutes at 4000 rpm. The pellets were collected with sterile pipette tips, and transferred to UltraClean Soil DNA Extraction Kit tubes (MoBio Laboratories, Inc., Carlsbad, CA, USA), and DNA was extracted following manufacturer's instruction. Pathogenic *E. coli* were detected through a multiplex PCR procedure developed by Toma et al. (2003), which analyzes for the presence of genes associated with enteropathogenic (EPEC; *eae*), enteroinvasive (EIEC; *ipaH*), enteroaggregative (EAEC; *aggR*), enterohemorrhagic (EHEC; *eae* + *stx*), and enterotoxigenic (ETEC; *elt*, *est*) *E. coli*. Reaction conditions described by Toma et al. (2003) were followed, and PCR products were separated on 1.5% agarose gels. *E. coli* virulence gene presence was analyzed in both the particle-attached (>5 μm) and unattached (<5 μm) sample fractions (n=60 for each fraction).

6.2.6 Data Splitting for Statistical Analysis

For model development, the data was split into training (n=48) and validation (n=12) data sets. The validation data sets were randomly extracted from the bulk data matrix. For each regression model type, fifteen different combinations of training and validation data sets were used to test hypotheses surrounding the equivalence of model performance statistics among the model types. The same combinations of training and validation data sets were used for each model type to limit potential confounding factors. The occurrence of any type of virulence gene in a sample was considered a presence in the data matrix, and no attempt was made to model each virulence marker separately.

6.2.7 Regression Model Approaches

Regularized linear regression was used to develop a generalized linear model through penalized shrinkage of the coefficients, which reduces the effects of covariance among explanatory variables by adding a bias term to the coefficient. L1 penalization, developed by Tibshirani (1996), and termed least angle shrinkage and selection operator (LASSO), offers simultaneous coefficient shrinkage and parameter selection by reducing unimportant variables to absolute zero. LASSO regression models were constructed in MATLAB (v. 12; The MathWorks Inc., Natick, MA, USA), using the `lasso` function in the Statistics Toolbox. All variables listed in Table 6.1 were included as candidate explanatory variables, and 5-fold cross-validation was used to select the optimum LASSO model.

Classification and regression trees are suited to environmental data as they do not assume distributions between explanatory and response variables, can be adapted to different types of response variables, are invariant to transformations of explanatory variables, and are easy to interpret (De'Ath and Fabricius, 2000). Several forms of classification and regression trees exist based on their recursive partitioning algorithms. The generalized, unbiased, interaction detection and estimation (GUIDE) approach offers independent variable selection that is less biased than the standard CART algorithm developed by Brieman et al. (1984), and is well suited for high variable-to-sample number problems (Loh, 2011). The classification and regression trees used in this study were produced with GUIDE Classification and Regression Trees and Forests software (version 13.4; http://www.stat.wisc.edu/~loh/guide.html). The percentage of particle attachment was estimated using stepwise least square linear regression. Classification

models for predicting the presence-absence of *E. coli* virulence markers were estimated by simple classification trees, with linear and interaction splits selected. Both types of trees were pruned by 5-fold cross validation, and the smallest node size was set to n=4.

The MARS approach to fitting nonlinear functions divides the range of explanatory variables into subsets using 'knots', which are defined by recursive partitioning, then fits a linear segment through each knot to create a continuous function (Friedman, 1991). Model fitting and variable selection is performed by forward selection steps that identify several knots, followed by backward pruning to remove unnecessary knots, typically through cross-validation. The MARS models were produced in MATLAB (v. 12, The MathWorks Inc.) using the ARESLab toolbox produced and distributed by Jekabsons (2011).The models developed were piecewise-linear, allowed 2nd degree interactions, had a maximum of 21 basis functions, allowed backward pruning, and had a generalized cross-validation penalty of 3.

6.2.8 Regression Model Performance

Models created for predicting percentages of *E. coli* particle attachment were tested using statistical metrics described by Moriasi et al. (2007) and Burnham and Anderson (2004). The coefficient of determination (R^2) estimates the collinearity between observed and predicted values, and ranges between 0 (no linear relationship) and 1 (perfect linear relationship). The Nash-Sutcliffe efficiency (NSE) indicates the degree to which the observed data fits the simulated data. NSE values range between 1 (perfect fit) and -∞, where negative values mean that the average value is a better predictor than the model. The root mean square error-observations standard deviation ratio (RSR) represents a normalized error index statistic, where RSR values of 0 indicate perfect model

performance. Akaike's information criterion with second-order bias correction (AIC_c) was used to evaluate the models based on their relative goodness of fit adjusted by the number of parameters incorporated into the models. AIC_c is an index for model selection and does not indicate how well the predicted data fits the observations. The preferred model is the one with the highest R^2 and NSE values, and the lowest AIC_c and RSR values.

The models created for predicting *E. coli* virulence marker presence were assessed using binomial statistical metrics described by Manel et al. (2001). Calculated statistics included the prediction success (proportion of cases correctly predicted), sensitivity (proportion of true positives predicted), specificity (proportion of true negatives predicted) and Cohen's kappa – the extent to which models predict occurrence at rates that are better than chance expectation. Sensitivity declines when negative predictions are actually positive (false negative), and specificity declines when positive predictions are actually negative (false positive). Cohen's kappa values of 0.0 to 0.4 indicate slight to fair model performance, 0.4 to 0.6 indicates moderate performance, 0.6 to 0.8 is good performance, and 0.8 to 1.0 indicates excellent model performance.

Fifteen models were generated of each regression type (LASSO, GUIDE, and MARS) for each response variable (percentage attachment, presence of particle-attached *E. coli* containing virulence genes, and the presence of *E. coli* containing virulence genes in the unattached state). These replicates were used to test hypotheses about the equality of the performance indicators. Performance metrics for each model type were statistically compared within the training and validation data sets using one-way analysis of variance

(ANOVA) in SigmaPlot software (version 11.0, SYSTAT Software Inc., Chicago, IL, USA). Post-hoc comparisons were performed using Tukey's test.

6.3 Results and Discussion

6.3.1 *E. coli* Particle Attachment in the Study Watershed

E. coli particle attachment percentages ranged between 48.2% and 91.4%, with an average of 67.3%. The observed values were generally higher than comparable studies that reported attachment ranges between 20 to 40% during baseflow and 50 to 70% during stormflow (Characklis et al., 2005; Krometis et al., 2007). This is due to analytical differences, since these studies used centrifugation techniques to interpret *E. coli* attachment to settleable particles. *E. coli* can be associated with organic flocs (Droppo et al., 2009), which are less dense than mineral particles and are not as settleable in fluvial systems. Boutilier et al. (2010) noted low deposition of particle-associated *E. coli* in wastewater treatment wetlands when analyzing attachment on a filtration basis, reflecting the non-settleable character of the particles. However, attachment of *E. coli* to particles of any sort enhances their survival (Burton et al., 1987), which is an important consideration in fate and transport modeling. While numerous methods have been used to assess *E. coli* particle attachment, there is no universally agreed upon analytical method nor definition of attachment. Regardless of analytical technique, this study demonstrates the utility of the regression models for predicting *E. coli* attachment.

6.3.2 Environmental Factors Correlated with Total *E. coli* Particle Attachment

All of the regression methods used concluded that *E. coli* particle attachment is related to a combination of water quality, hydrological, meteorological, land use, particle concentration and particle size parameters (Table 6.2). Hydrological and meteorological variables tended to have relatively minor influence on *E. coli* particle attachment as defined by the number of models that included these variables, except in the LASSO models. Land use and particle properties (concentration, organic content, and size distribution) were included in the majority of models for all regression approaches. In the LASSO models, stream discharge, percentage of residential land, organic matter content, and effective particle size (D_{10}) exhibited frequent correlations to the percentage of particle attachment. Previous studies have created linear models for predicting *E. coli* particle attachment based on discharge and storm hydrographs with limited to moderate success (Characklis et al., 2007; Krometis et al., 2007). There appears to be some positive linear relationship among discharge and percentage attachment, but the non-linear models infrequently selected discharge or other hydrological variables as predictors.

Instead, non-linear models tended to select particle properties that are in turn influenced by discharge conditions and consequent shear stresses. The GUIDE models tended to select land use, particle properties and water quality as predictor variables. Electrical conductivity (EC), 5-day average water temperature, percentage residential land use, particle organic content (percentage organic matter, VSS) and particle size (D_{50}, geometric mean diameter) were frequently included in the GUIDE models. The MARS models frequently contained some measure of land use, particle size distribution, and particle concentration and organic content. Water quality was selected less often, and hydrological variables were infrequently selected. The particular parameters selected

Table 6.2. Frequency of explanatory variables included in the LASSO, GUIDE, and MARS models used to predict *E. coli* particle attachment and virulence marker presence. Fifteen models of each type were created for each response variable, and the cell values reflect the number of times the explanatory variable was included in a prediction model. Bold-underline values indicate the number of models that included any variable from respective environmental categories.

Explanatory Variable	*E. coli* Particle Attachment Models			*E. coli* Virulence Marker Models					
				Particle Attached			Unattached		
	LASSO	GUIDE	MARS	LASSO	GUIDE	MARS	LASSO	GUIDE	MARS
Water Quality									
Any Variable (n)	**3**	**12**	**9**	**15**	**10**	**11**	**15**	**15**	**15**
pH		3	1	6	2	2		1	2
EC	2	4	4	15	3	8	15	11	15
DO	1	2	3	5	1	1		2	4
WatTemp		1	6		1	8			9
5DWatTemp		5		4	8	2	4	5	3
Hydrology and Meteorology									
Any Variable (n)	**11**	**6**	**2**	**3**	**4**	**5**	**4**	**3**	**2**
Discharge	8	2	1	1		2			1
FlowStage			1			2			
MaxIntensity		2			1		2	1	1
DailyRain		1		3			1		
5DTotalPpt	5					4			
AirTemp5d		1			4		2	1	
Land-Use									
Any Variable (n)	**9**	**10**	**13**	**7**	**0**	**7**	**7**	**0**	**5**
DrainArea		1	3	1		1	1		1
%Forest		1	4	2		2			2
%Agric	1	2	2				1		2
%Resid	8	7	4	5		5	6		
Particle Concentration and Organic Content									
Any Variable (n)	**12**	**10**	**13**	**11**	**8**	**10**	**1**	**2**	**6**
Turbidity	6	2	8	5				1	2
PartConcn	1	3	1	1	1	4			
TSS	2	3	5		2	3			2
VSS		3	3		3	1	1	1	2
%OM	8	5	11	11	2	7			3
Particle Size Distribution									
Any Variable (n)	**10**	**12**	**15**	**5**	**8**	**12**	**5**	**1**	**2**
D_{50}		4	2						
D_{10}	9	3	6		3	5		1	1
D_{75}-D_{25}		1	2			2			
$D_{75/25}$	1	3	5	4		1	5		
GeoMean	1	2	6		2	2			1
Sorting		3	2	2	2	3			
%Silt			1		2	8	1		1

within each category of environmental variables were relatively evenly distributed, although percentage organic matter, turbidity, water temperature, D_{10} and geometric mean diameter figured prominently.

In all models, contributing land use was important for predicting particle attachment. To our knowledge, land use variables have not been included in predictive models in previous studies. The models tended to indicate that rising percentages of residential land use positively influenced attachment percentage. Not necessarily a causative factor, the land use designations served to divide the monitoring sites into subcatchments with different *E. coli* sources and ratios of attached vs. free bacteria. In the study watershed, subcatchments with higher residential land use are associated with lower agricultural land use. *E. coli* associated with runoff from agricultural fields has been reported to have a low percentage of attachment (Muirhead et al., 2006; Soupir and Mostaghimi, 2010). Higher percentages of attachment have been observed in domestic septic tank effluent in comparison to dairy wastewater effluent (Boutilier et al., 2010). Contributing fecal source appears to have an effect on *E. coli* attachment and should be included in prediction models.

Previous attempts at relating *E. coli* attachment to particle concentration have resulted in weak predictive models (Bai and Lung, 2005; Characklis et al., 2005). Garcia-Armisen and Servais (2009) also observed complex *E. coli* – particle relationships where *E. coli* was positively correlated with TSS at concentrations below 50 mg/L, but no such correlation was observed at higher sediment loads. Suspended sediments (indicated as turbidity, TSS and particle concentration) were found to correspond to *E. coli* particle attachment in all models. However, the relationship between suspended sediments and *E.*

coli particle attachment appears to be complex, particularly regarding particle size. In the LASSO models, linear relationships between particle attachment and effective size (D_{10}) were positive, but the relationship to geometric mean was negative. In the MARS models, measures of particle size dispersion (ratio of interquartile diameters, particle sorting) were often included, further suggesting that complexity within the particle sizes can affect the percentages of particle attachment.

The organic matter content of suspended particles also appears to influence *E. coli* particle attachment percentages, and was negatively associated with particle attachment in all models. Thus, inorganic particles were associated with higher *E. coli* attachment in this study. In contrast to our observations, Guber et al. (2007) reported preferential *E. coli* attachment to organic particles. It is possible that attachment to small colloidal organic matter (< 5 μm), which is abundant in fluvial systems (Ran et al., 2000), may have occurred in our study but would not have been detected by the method used. More importantly, interactive effects observed between organic matter and land use, flow stage and particle diameter in the non-linear models suggests that *E. coli* attachment to organic particles is not a simple linear process and may depend on organic matter source, fluvial hydraulics and particle size.

6.3.3 Performance Statistics for *E. coli* Particle Attachment Models

Performance statistics for the LASSO, GUIDE, and MARS models are summarized in Table 6.3. The non-linear GUIDE and MARS models outperformed the linear LASSO model as they had higher R^2 and NSE, and lower RSR values. The MARS models demonstrated better overall prediction than the GUIDE and LASSO models. However, the mid-range values of R^2 and NSE indicate the MARS models had moderate overall

prediction accuracy. The AIC_c values were equivalent for all models, suggesting that the number of parameters included were not sufficiently different to affect the relative goodness of fit for any regression model type.

Table 6.3. Model performance statistics for the regression models constructed to predict *E. coli* particle attachment percentage. Performance statistics were calculated for fifteen models of each type created using different combinations of training and data sets, and summarized as ranges, mean values and standard deviations (SD).

Performance Statistic		Training			Validation		
		LASSO	GUIDE	MARS	LASSO	GUIDE	MARS
R^2	Mean (SD)	0.16 [a] (0.08)	0.28 [b] (0.11)	0.52 [c] (0.17)	0.12 [a] (0.14)	0.18 [a] (0.11)	0.26 [b] (0.08)
	Range	0.05 - 0.32	0.11 - 0.46	0.30 - 0.82	0.00 - 0.53	0.04 - 0.41	0.11 - 0.40
NSE	Mean (SD)	0.15 [a] (0.09)	0.34 [b] (0.09)	0.48 [c] (0.18)	-0.15 [a] (0.17)	0.24 [b] (0.12)	0.28 [b] (0.15)
	Range	0.03 - 0.32	0.17 - 0.47	0.26 - 0.83	-0.56 - 0.05	0.01 - 0.41	0.10 - 0.46
RSR	Mean (SD)	0.96 [a] (0.05)	0.87 [b] (0.09)	0.76 [b] (0.26)	1.07 [a] (0.08)	1.17 [b] (0.14)	1.08 [ab] (0.32)
	Range	0.42 - 1.54	0.74 - 1.00	0.85 - 1.05	0.97 - 1.26	0.99 - 1.55	0.51 - 1.44
AIC_c	Mean (SD)	95.1 [a] (3.2)	95.6 [a] (5.3)	89.8 [a] (10.5)	57.1 [a] (22.1)	56.9 [a] (22.1)	53.4 [a] (18.0)
	Range	87.4 - 100.2	85.1 - 102.9	79.3 - 124.6	23.2 - 89.4	24.8 - 92.1	27.6 - 82.7

R^2 = collinearity between observed and predicted values. Values range between 0 (no linear relationship) and ±1 (perfect positive or negative relationship).

NSE = magnitude of residual variance to the measured data variance, indicating the degree to which the observed data fits the simulated data. Values range between 1 (perfect fit) and -∞, where negative values infer the average value is a better predictor than the model.

RSR = normalized error index statistic. Lower values indicate better model performance.

AIC_c = relative goodness of fit adjusted by the number of parameters incorporated into the. Lower values represent the preferred model.

[a,b,c] indicates significantly ($p<0.05$) different mean values . Separate ANOVAs were run on each performance statistic for each training or validation data set.

Through this study, it is apparent that linear regression techniques are inadequate for predicting *E. coli* particle attachment, presumably because of the complex non-linear relationships that exist between *E. coli* particle attachment and environmental variables.

The LASSO models demonstrated moderately-weak correlation to the observed values, which is similar to other studies that observed weak linear correlation to discharge (Krometis et al., 2007) and moderate linear correlation to particle concentration (Characklis et al., 2005). In the present study, the MARS models were most effective in predicting *E. coli* particle attachment, possibly as a result of continuous distributions rather than data partitioning as in the case of the GUIDE models. Although the MARS models demonstrated moderate performance, the performance statistics used compared the predicted values to average particle attachment. Thus, using these models to parameterize watershed models may offer improvements over using single static parameter values that are based on average particle attachment. Due to the relative ease of development and interpretation, it is possible that MARS models could be incorporated into watershed models to help parameterize particle attachment based on monitoring location (e.g., land use), water quality conditions, and particle properties.

6.3.4 Occurrence of *E. coli* Containing Virulence Markers

The presence of *E. coli* containing virulence markers in the particle-attached and unattached fractions were found in 44% of all samples. Virulence markers were observed in the attached state in 24% of the samples, in the unattached state in 13% of the samples, and in both the attached and unattached states in 7% of the samples. The EPEC marker (*eae*) was observed in 40% of the total samples, and the ETEC marker (*elt*) was observed in 7% of all samples, with 3% of the samples having both EPEC and ETEC markers. The ETEC marker was only observed during one of the storm events (35 mm precipitation). In other studies, Huan et al. (2012) observed EPEC genes in 10% of source waters and reported significant correlation of this group with water quality (pH and turbidity).

Masters et al. (2011) observed EPEC genes in 65% of freshwater samples during the rainy season in Australia. Although EPEC genes appear to be abundant in surface waters, Hunter (2003) reported low waterborne etiology of EPEC infections, presumably because of the high dose required for infection. ETEC genes have been observed in 30 to 35% of surface water samples (Begum et al., 2007; Masters et al. 2011), and in up to 67% in samples taken from endemic areas (Lothigius et al., 2007). These studies, however, focused on water systems exhibiting greater fecal contamination than the watershed evaluated in this study, as indicated by higher fecal indicator levels and the prevalence of virulence genes.

6.3.5 Environmental Factors Correlated with the Presence of *E. coli* Virulence Markers in the Attached and Unattached Fractions

Several environmental factors correlated with the presence of *E. coli* virulence markers (EPEC/ETEC) including water quality, hydrology, meteorology, land use, and particle properties (Table 6.2). For the attached fraction, the logistic LASSO models revealed that the presence of EPEC/ETEC markers were negatively associated with the organic content of suspended particles and positively associated with land use, predominantly the percentage residential land, and water quality, particularly electrical conductivity (EC). In the GUIDE models, 5-day average water and air temperatures, as well as particle properties predicted the presence of EPEC/ETEC markers attached to suspended particles. Marker presence tended to be negatively correlated with percentage organic matter. In the MARS models, water quality, land use and particle properties were positively associated with EPEC/ETEC marker occurrence. Frequently occurring variables were EC, water temperature, and the percentage of residential land use. The

LASSO models revealed that unattached EPEC/ETEC marker presence was predominantly explained by EC, 5-day average water temperature, maximum precipitation intensity, percentage residential land, and the ratio of interquartile diameters (Table 6.2). Important variables selected for predicting unattached EPEC/ETEC using the GUIDE approach were EC and 5-day average water and air temperatures. A combination of EC, percentage agricultural land, maximum rain intensity, and D_{10} values were included in the MARS models.

In all models, EC was a prominent explanatory variable and positively associated with EPEC/ETEC marker presence. Although EC affects bacterial particle attachment by suppressing the electron double layer (Jamieson et al., 2005), the significance of this variable is likely related to discharge conditions. In fluvial systems, EC is lower during stormflow when stream flow is dominated by precipitation and runoff, which has a lower proportion of solubilized minerals than groundwater or interflow (Schleppi et al., 2006). The prevalence of EPEC in this system appears to be associated with low flow conditions, where groundwater, tile drains and interflow dominate the stream flow. Consequently, EPEC could be entering the system through domestic effluent percolating through septic fields, manure-borne EPEC in agricultural soils transported through tile drains or groundwater, or cattle or wildlife entering the stream during low flow conditions. Duris et al. (2009) reported a lack of correlation between virulent *E. coli* presence and FIB in surface water. Our finding that EPEC occurs primarily during low flow conditions supports this study, as elevated *E. coli* concentrations occur during high discharge events in this watershed (Sinclair et al, 2008). Point biserial correlation coefficients were weakly negative for *E. coli* concentrations and particle attached

(r_{pb} = -0.004) or unattached (r_{pb} = -0.082) EPEC/ETEC, suggesting no correlation exists between EPEC/ETEC virulence markers and *E. coli* concentrations.

Water temperature figured prominently in the GUIDE and MARS models, and exhibited non-monotonic relationships as EPEC/ETEC was associated with both low and high temperature conditions. Francy et al. (2013) also reported that water temperature is an important variable when modeling pathogen occurrence at freshwater beaches, but found that temperature was not included in all models. Our results show that water temperature is an important predictor for EPEC/ETEC occurrence, but its relationship is nonlinear. Overall, particle properties tended to have little predictive influence for the unattached EPEC/ETEC marker presence compared with either the particle attached ETEC/EPEC markers or the percentage of particle attachment.

6.3.6 Regression Model Performance for Predicting Pathogen Occurrence

For predicting the presence or absence of particle-attached EPEC/ETEC markers, the MARS models outperformed the GUIDE and logistic LASSO models with the training data sets, as evidenced by the prediction success and Cohen's kappa (Table 6.4). With the validation data sets, the GUIDE and MARS models demonstrated fair model performance and predictive ability, whereas the LASSO models demonstrated poor predictive performance. The three regression model types performed equivalently in terms of model sensitivity (i.e. true positive prediction), but the MARS model had higher specificity (i.e. less false positives) than the LASSO and GUIDE models in both the training and validation data sets.

Table 6.4. Model performance statistics for the regression models created for predicting the presence of particle-attached and unattached *E. coli* virulence markers. Model performance statistics were calculated from fifteen models of each type and summarized as ranges, mean values and standard deviations (SD).

Test Statistic		Training			Validation		
		LASSO	GUIDE	MARS	LASSO	GUIDE	MARS
Particle-Attached markers							
Prediction Success	Mean (SD)	0.64 [a] (0.12)	0.74 [a] (0.14)	0.87 [b] (0.11)	0.58 [a] (0.10)	0.65 [a] (0.07)	0.65 [a] (0.11)
	Range	0.36-0.82	0.47-0.91	0.53-0.98	0.36-0.75	0.45-1.0	0.5-0.9
Sensitivity	Mean (SD)	0.74 [a] (0.11)	0.75 [a] (0.27)	0.71 [a] (0.28)	0.38 [a] (0.34)	0.58 [a] (0.23)	0.47 [a] (0.12)
	Range	0.0-1.0	0.0-1.0	0.0-1.0	0. -1.0	0.33-1.0	0.33-0.66
Specificity	Mean (SD)	0.59 [a] (0.18)	0.71 [a] (0.20)	0.92 [b] (0.12)	0.64 [a] (0.11)	0.70 [ab] (0.15)	0.79 [b] (0.13)
	Range	0.18-0.86	0.32 0.94	0.58-1.0	0.44-0.86	0.42-1.0	0.60-1.0
Cohen's Kappa	Mean (SD)	0.31 [a] (0.15)	0.47 [b] (0.18)	0.69 [b] (0.23)	0.01 [a] (0.25)	0.28 [b] (0.20)	0.38 [b] (0.15)
	Range	0.09-0.55	0.18-0.79	0.1-0.94	0.1-0.47	0.15-1.0	0.14-0.68
Unattached markers							
Prediction Success	Mean (SD)	0.67 [a] (0.10)	0.65 [a] (0.07)	0.81 [b] (0.08)	0.65 [a] (0.15)	0.60 [a] (0.13)	0.64 [a] (0.11)
	Range	0.36-0.78	0.54-0.84	0.67-0.95	0.25-0.52	0.36-0.82	0.50-0.83
Sensitivity	Mean (SD)	0.60 [a] (0.15)	0.65 [a] (0.16)	0.61 [a] (0.17)	0.40 [a] (0.35)	0.51 [a] (0.19)	0.44 [a] (0.18)
	Range	0.38-1.0	0.38-0.82	0.29-0.85	0.0-1.0	0.20-1.0	0.20-1.0
Specificity	Mean (SD)	0.69 [a] (0.18)	0.67 [a] (0.15)	0.93 [b] (0.07)	0.71 [ab] (0.21)	0.65 [a] (0.17)	0.82 [b] (0.16)
	Range	0.38-1.0	0.38-0.82	0.29-0.85	0.0-1.0	0.38-1.0	0.50-1.0
Cohen's Kappa	Mean (SD)	0.25 [a] (0.12)[a]	0.28 [a] (0.12)	0.52 [b] (0.23)	0.08 [a] (0.24)	0.17 [a] (0.14)	0.23 [a] (0.18)
	Range	0.01-0.46	0.07-0.38	0.0-0.89	0.0-0.42	0.0-0.42	0.0-0.57

Prediction Success = proportion of cases correctly predicted
Sensitivity = proportion of true positives
Specificity = proportion of true negatives
Cohen's kappa = extent to which models predict occurrence at rates that are better than chance expectation. Values of 0.0 to 0.4 indicate slight to fair performance, 0.4 to 0.6 indicates moderate performance, 0.6 to 0.8 is good performance, and 0.8 to 1.0 indicates excellent performance.
[a,b,c] indicates significantly ($p<0.05$) different mean values . Separate ANOVAs were run on each performance statistic for each training or validation data set.

In general, the models created to predict the presence of unattached EPEC/ETEC markers exhibited comparable performance to the particle-attached EPEC/ETEC marker models (Table 6.4). The model types demonstrated equivalent sensitivity in the training and validation data sets. The MARS model again demonstrated greater specificity with the training data sets, and remained greater than the GUIDE model, but equivalent to the LASSO model with the validation data sets. The MARS models had good model performance and outperformed the GUIDE and LASSO models with the training data sets. However, all three models performed equivalently with the validation data sets. In virulent bacteria prediction, models with a tendency to predict false negatives (i.e. lower sensitivity) could underestimate the risk associated with a water supply, and models with a tendency to predict false positives (i.e. low specificity) could overestimate risks leading to unnecessary restrictions on water usage. The model types demonstrated equivalent sensitivity, but the MARS models tended to have higher specificity, which could lead to more efficient water management practices.

6.4 Conclusion

This study evaluated the use of alternative statistical models for dynamically predicting *E. coli* particle attachment and the occurrence of *E. coli* virulence markers in the particle-attached and unattached fractions. In all model types, *E. coli* particle attachment and virulence marker presence in the attached and unattached subpopulations correlated with a combination of water quality, land use, particle concentration, and organic matter properties, suggesting complex dynamics of waterborne particle association. The non-linear GUIDE and MARS models offered moderate to good prediction performance for *E. coli* particle attachment and outperformed the linear LASSO model indicating that

models which accommodate complex ecological data are better suited for predicting *E. coli* particle attachment. The MARS model offered the highest prediction performance of the regression model types. MARS models also outperformed GUIDE and logistic LASSO models for predicting the occurrence of waterborne EPEC and ETEC markers (thus, their host pathogens) in the attached and unattached fractions. Non-linear GUIDE and MARS models may be useful for dynamic parameterization of *E. coli* particle attachment in watershed models. A validation study on multiple watersheds may help with the development of a universal model for predicting *E. coli* particle attachment.

CHAPTER 7 CONCLUSION

Despite evidence demonstrating that *E. coli* is capable of long term persistence and "naturalization" in soils and sediments of temperature climates, this organism remains the standard FIB for water quality monitoring and modeling programs. The agricultural watershed studied here demonstrates chronic *E. coli* contamination. Persistent and putatively naturalized *E. coli* were observed in the sediments and tile drainage effluents of a cultivated field within the Thomas Brook Watershed. The *E. coli* strains found in secondary sources (tile drainage effluent, streambed sediments) were observed to contribute to waterborne *E. coli* to a similar extent as primary fecal sources (livestock manure holding tanks, residential septic systems), suggesting that environmental *E. coli* has a profound impact on waterborne *E. coli* in this watershed. The research presented here contributs valuable information regarding the population ecology of *E. coli* within the soil and sediment matrices of an agricultural watershed.

7.1 Major Contributions

- Most studies concerning the differential survival and persistence of *E. coli* in agricultural soils report strain-specific responses to select conditions in laboratory experiments. The field study conducted as part of this work demonstrated population level *E. coli* responses to soil types, as reflected by hillslope position and associated variation in soil texture and moisture content. It is speculated that strain-specific responses to soil systems lead to significant differences in persistent population structure at the different hillslope positions.

- In previous studies, *E. coli* has been found to persistently populate tile drainage effluent in the absence of recent applications of manure amendments. Here, culturable *E. coli* was isolated from tile effluents throughout two growing seasons and an overwinter period. It was found that *E. coli* associated with manure-amended soils dominated tile drain effluent for up to 55 days following manure addition. Strain types of *E. coli* distinct from those detected in manure-amended soils populated the tile drainage system during the remainder of the study period. Tile drainage *E. coli* contributed up to 48% of the waterborne *E. coli* population in the adjacent stream, and the contribution was exponentially related to tile drainage rates, regardless of whether drainage effluent was populated by manure-associated or putatively naturalized *E. coli*. The detection of naturalized *E. coli* in tile effluents and adjacent streams may affect the results of water quality monitoring programs.
- Spatial patterns of *E. coli* populations in the sediment beds of heterogenic stream morphological features have not been previously investigated. It was found that *E. coli* concentrations were not significantly different among the features during baseflow, but differed markedly following stormflow events. Variation partitioning revealed that streambed *E. coli* population structure was explained by both spatial and environmental variables, indicating source-sink dynamics operate among sediment *E. coli* populations. Thus, *E. coli* demonstrates adaptation to particular sediment environments and are subsequently transported to "sink" environments where they are ill-adapated.
- Temporal investigations into *E. coli* population composition yielded observations that indicate sediment *E. coli* population turnover is driven by reach-specific factors.

E. coli population turnover in higher energy reaches correlated with sediment transport, supporting the argument of source-sink dynamics in controlling sediment *E. coli* populations. In a lower energy reach with less active bed sediments, *E. coli* population similarity correlated with the influence of adjacent catchment sources (tile drainage effluent) suggesting the importance of immigration on *E. coli* population structure at this location.

- The hydrological connectivity between sediment and waterborne *E. coli* populations also demonstrated reach-specificity. Sediment *E. coli* isolates correlated with sediment transport rates only at the reach with the highest bed shear and sediment transport rates. The waterborne *E. coli* populations of the lowest energy reach had minor contributions from reach sediments, but high contributions from upstream sediment populations. Reach-specific sediment resuspension and hyporheic exchange rates appear to affect the connectivity between sediment and waterborne *E. coli* populations.
- Recursive-based nonparametric regression modeling approaches were found to adequately predict *E. coli* concentrations in tile drainage effluent and particle attachment behaviour in fluvial waters. In particular, MARS models were most effective, presumably because of their integration of recursive partitioning with linear functionality to create continuous distributions. *E. coli* behaviour may be difficult to predict through simpler models because of strain-specific differences in survival and transport behaviour.

7.2 Future Directions

- The persistence of *E. coli* in surface water sediments remains an issue of concern for the continued use of *E. coli* as an indicator bacterium in water quality monitoring and modeling programs. This study identified that *E. coli* exhibits source-sink dynamics and speculated that reach- and catchment-scale hyporheic processes may be important for governing *E. coli* fate and transport in fluvial systems. Further research into hyporheic processes using conservative and microbial tracers may elucidate the transport patterns occurring in dendritic stream networks, particularly if connected stream reaches of different hydrological character are investigated.
- Disjoint stream reaches with differing catchment sources of *E. coli* were found to exhibit differences in waterborne *E. coli* particle attachment and relative influences of sediment transport on waterborne populations. It was speculated that strain-specific differences in particle attachment, as a function of divergent host populations, may be contributing to the observed phenomena. Further investigation into the particle attachment behaviour and sediment penetration among *E. coli* isolated from different host species would support the assertions made in this work.
- Tile drainage effluent was observed to contribute to waterborne *E. coli* within an adjacent stream irrespective of whether the effluent was populated with bacteria associated with manure amendments or with putatively naturalized genotypes. It was speculated that the detection of naturalized strains would lead to misrepresentation of health risks associated with the water resource. Further investigation into the occurrence of an array of human pathogens and the development of risk models

would allow for a better appreciation of the potential risks from tile effluents on agricultural waters.

BIBLIOGRAPHY

Agriculture and Agri-Food Canada (AAFC). 2006. Manure management on cropland: farm environmental management survey (FEMS). http://www4.agr.gc.ca/AAFC-AAC/display-afficher.do?id=1314803628665&lang=eng.

Ahmed, W., J. Stewart, T. Gardner, D. Powell, P. Brooks, D. Sullivan, and N. Tindale, 2007. Sourcing faecal pollution: a combination of library-dependent and library-independent methods to identify human faecal pollution in non-sewered catchments. Water Research, 41: 3771-3779.

Aislabie, J., M. McLeod, J. Ryburn, A. McGill, and D. Thornburrow, 2011. Soil type influences the leaching of microbial indicators under natural rainfall following application of dairy shed effluent. Soil Research, 49: 270-279.

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

Anderson, M.A., J.E. Whitlock, and V.J. Harwood, 2006. Diversity and distribution of *Escherichia coli* genotypes and antibiotic resistance phenotypes in feces of humans, cattle and horses. Applied and Environmental Microbiology, 72(11): 6914-6922.

Arcement, G.J., and V.R. Schneider, 1989. Guide for selecting Manning's roughness coefficients for natural channels and floodplains. United States Geological Survey Water-supply Paper 2339: 1-38.

Bai, S., and W-S. Lung, 2005. Modeling sediment impact on the transport of fecal bacteria. Water Research 39(20), 5232-5240.

Baker-Austin, C., R. Rangdale, J. Lowther, and D.N. Lees, 2010. Application of mitochondrial DNA analysis for microbial source tracking purposes in shellfish harvesting waters. Water Science and Technology, 61(1): 1-7.

Ball Coelho, B.R., R.C. Roy, E. Topp, and D.R. Lapen, 2007. Tile water quality following liquid swine manure application into standing corn. Journal of Environmental Quality, 36: 580-587.

Ballesté, E., X. Bonjoch, L.A. Belanche, and A.R. Blanch. 2010. Molecular indicators used in the development of predictive models for microbial source tracking. Applied and Environmental Microbiology, 76(6): 1789-1795.

Barnes, B. and D.M. Gordon, 2004. Coliform dynamics and the implications for source tracking. Environmental Microbiology, 6(5): 501-509.

Begum, Y.A., K.A. Talukder, G.B. Nair, S.I. Khan, A-M. Svennerhol, R.B. Sack, F. Qadri, 2007. Comparison of entertoxigenic *Escherichia coli* isolated from surface water and diarrhoeal stool samples in Bangladesh. Canadian Journal of Microbiology 53(1), 19-26.

Beloin, C., A. Roux and J-M. Ghigo, 2008. *Escherichia coli* biofilms. Current Topics in Microbiology and Immunology, 322: 249-289.

Bell, A., A.C. Layton, L. McKay, D. Williams, R. Gentry, and G.S. Sayler, 2009. Factors influencing the persistence of fecal *Bacteroides* in stream water. Journal of Environmental Quality, 38: 1224-1232.

Blott, S.J. and K. Pye, 2001. GRADISTAT: a grain size distribution and statistics package for the analysis of unconsolidated sediments. Earth Surface Processes and Landforms 26, 1237-1248.

Blum, S.A., M.G. Lorenz, and W. Wackernagel, 1997. Mechanism of retarded DNA degradation and prokaryotic origin of DNases in nonsterile soils. System. Applied Microbiology, 20:513-521.

Boehm, A.B., L.C. Van de Werfhorst, J.F. Griffith, P.A. Holden, J.A. Jay, O.C. Shanks, D. Wang, and S.B. Weisburg, 2013. Performance of forty-one microbial source tracking methods: a twenty-seven lab evaluation study. Water Research, http://dx.doi.org/10.1016/j.watres.2012.12.046.

Boulton, A.J., S. Findlay, P. Marmonier, E.H. Stanley, and H.M.Valett, 1998. The functional significance of the hyporheic zone in streams and rivers. Annual Review of Ecology and Systematics, 29: 59-81.

Boulton, A.J., T. Datry, T. Kasahara, M. Mutz, and J.A. Stanford, 2010. Ecology and management of the hyporheic zone: stream-groundwater interactions of running

waters and their floodplains. Journal of the North American Benthological Society, 29(1): 26-40.

Boutilier, L., R. Jamieson, R. Gordon, C. Lake, and W. Hart, 2009. Adsorption, sedimentation and inactivation of *E. coli* within wastewater treatment wetlands. Water Research 43(17), 4370-4380.

Brennan, F.P., V. O'Flaherty, G. Kramers, J. Grant, and K.G. Richards, 2010. Long-term persistence and leaching of *Escherichia coli* in temperate maritime soils. Applied and Environmental Microbiology, 76(5): 1449-1455.

Brieman, L., J.H. Friedman, R.A. Olshen, and C.J. Stone, 1984. Classification and Regression Trees. CRC Press.

Brownell, M.J., V.J. Harwood, R.C. Kurz, S.M. McQuaig, J. Lukasik, and T.M. Scott, 2007. Confirmation of putative stormwater impact on water qulaiyt at a Florida beach by microbial source tracking methods and structure of indicator organism populations. Water Research, 41: 3747-3757.

Burton, G.A., Jr., D. Gunnison, and G.R. Lanza, 1987. Survival of pathogenic bacteria in various freshwater sediments. Applied and Environmental Microbiology 53, 633-638.

Byappanahalli, M., M. Fowler, D. Shively, and R. Whitman, 2003. Ubiquity and persistence of *Escherichia coli* in a Midwestern coastal stream. Applied and Environmental Microbiology, 69: 4549-4555.

Byappanahalli, M.N., R.L. Whitman, D.A. Shively, W.T.E. Ting, C.C. Tseng, and M.B. Nevers, 2006. Seasonal persistence and population characteristics of *Escherichia coli* and enterococci in deep backshore sand of two freshwater beaches. Journal of Water and Health, 4(3): 313-320.

Byappanahalli, M.N., T. Yan, M.J. Hamilton, S. Ishii, R.S. Fujioka, R.L. Whitman, and M.J. Sadowsky. 2012. The population structure of *Escherichia coli* isolated from subtropical and temperate soils. Science of the Total Environment, 417-418: 273-279.

Campbell, K. 2009. Fate and transport of subsurface nitrate through an agricultural field and forested riparian zone. M.Sc. thesis. Dalhousie University, Halifax, NS.

Canadian Council of Ministers of the Environment (CCME), 1999. Canadian water quality guidelines for the protection of agricultural water uses. In: Canadian environmental quality guidelines, 1999, Canadian Council of Ministers of the Environment, Winnipeg.

Carozzi, M., R.M. Ferrara, G. Rana, and M. Acutis. 2013. Evaluation of mitigation strategies to reduce ammonia losses from slurry fertilization on arable lands. Science of the Total Environment, 449: 126-133.

Chandler, D.S. and J.A. Craven. 1980. Relationship of soil moisture to survival of *Escherichia coli* and *Salmonella typhimurium* in soils. Australian Journal of Agricultural Research, 31: 547-555.

Characklis, G.W., M.J. Dilts, O.D. Simmons III, C.A. Likirdopulos, L-A. Krometis, and M.D. Sobsey, 2005. Microbial partitioning to setllable particles in stormwater. Water Research 39(9), 1773-1782.

Charlton, R., 2008. *Fundamentals of fluvial geomorphology*. Routledge, New York, NY. 234 pp.

Cho, K.H., Y.A. Pachepsky, J.H. Kim, A.K. Guber, D.R. Shelton, and R. Rowland, 2010. Release of *Escherichia coli* form the bottom sediment in a first-order creek: Experiment and reach-specific modeling. Journal of Hydrology, 391: 322-332.

Cinotto, P.J., 2005. Occurrence of fecal-indicator bacteria and protocols for identification of fecal-contaminated sources in selected reaches of the West Branch Brandywine Creek, Chester County, Pennsylvania. U.S. Geological Survey Scientific Investigations Report 2005-5039. 91 p.

Clermont, O., S. Bonacorsi, and E. Bingen, 2000. Rapid and simple determination of the *Escherichia coli* phylogenetic group. Applied and Environmental Microbiology, 66(10): 4555-4558.

Cook, K.L., C.H. Bolster, K.A. Ayers, and D.N. Reynolds, 2011. *Escherichia coli* diversity in livestock manures and agriculturally impacted stream waters. Current Microbiology, 63: 439-449.

Cools, D., R. Merckx, K. Vlassak, and J. Verhaegen, 2001. Survival of *E. coli* and *Enterococcus* spp. derived from pig slurry in soils of different texture. Applied Soil Ecology, 17: 53-62.

Cottenie, K., 2005. Integrating environmental and spatial processes in ecological community dynamics. Ecology Letters, 8: 1175-1182.

Craig, D.L., H.J. Fallowfield, and N.J. Cromar, 2004. Use of microcosms to determine persistence of *Escherichia coli* in recreational coastal water and sediment and validation with in situ measurements. Journal of Applied Microbiology, 96: 922-930.

Davies, C.M., J.A.H. Long, M. Donald, M. and N.J. Ashbolt, 199). Survival of fecal microorganisms in marine and freshwater sediments. Applied and Environmental Microbiology, 61(5): 1888-1896.

De'Ath, G. and K.E. Fabricius, 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis. Ecology 81(11), 3178-3192.

Dorner, S.M., W.B. Anderson, R.M. Slawson, N. Kouwen, and P.M. Huck, 2006. Hydrologic modeling of pathogen fate and transport. Environmental Science and Technology 40(15), 4746-4753.

Droppo, I.G., S.N. Liss, D. Williams, T. Nelson, C. Jaskot and B. Trapp, 2009. Dynamic existence of waterborne pathogens in river sediment compartments: implications for water quality regulatory affairs. Environmental Science and Technology, 43(6): 1737-1743.

Droppo, I.G., B.G. Krishnappan, S.N. Liss, C. Marvin and J. Biberhofer, 2011. Modelling sediment-microbial dynamics in the South Nation River, Ontario, Canada: towards the prediction of aquatic and human health risk. Water Research, 45: 3797-3809.

Duriez, P. and E. Topp, 2007. Temporal dynamics and impact of manure storage on antibiotic resistance patterns and population structure of *Escherichia coli* isolates

from a commercial swine farm. Applied and Environmental Microbiology, 73(17): 5486-5493.

Duris, J.W., S.K. Haack, and L.R. Fogarty, 2009. Gene and antigen markers of shiga-toxin producing *E. coli* from Michigan and Indiana river water: occurrence and relation to recreational water quality criteria. Journal of Environmental Quality 38, 1878-1886.

Dwivedi, D., B.P. Mohanty, and B.J. Lesikar, 2013. Estimating *Escherichia coli* loads in streams based on various physical, chemical and biological factors. Water Resources Research, 49: 2896-2906.

Edge, T.A. and S. Hill, 2007. Multiple lines of evidence to identify the sources of fecal pollution at a freshwater beach in Hamilton Harbour, Lake Ontario. Water Research, 41: 3585-3594.

Esseili, M.A., I.I. Kassem, V. Sigler, K. Czajkowski and A. Ames, 2012. Genetic evidence for the offsite transport of *E. coli* associated with land application of Class B biosolids on agricultural fields. Science of the Total Environment, 433: 273-280.

Field, K.G. and M. Samadpour, 2007. Fecal source tracking, the indicator paradigm, and managing water quality. Water Research, 41: 3517-3538.

Foppen, J.W., G. Lutterodt, W.F.M. Roling, and S. Uhlenbrook, 2010. Towards understanding inter-strain attachment variations of *Escherichia coli* during transport in saturated quartz sand. Water Research 44, 1202-1212.

Forge, T.A., S. Bittman, and C.G. Kowalenko, 2005. Responses of grassland soil nematodes and protozoa to multi-year and single-year applications of dairy manure slurry and fertilizer. Soil Biology and Biochemistry, 37(10): 1751-1762.

Francy, D.S., E.A. Stelzer, J.W. Duris, A.M.G. Brady, J.H. Harrison, H.E. Johnson, and M.W. Ware, 2013. Predictive models for *Escherichia coli* concentrations at inland lake beaches and relationship of model variables to pathogen detection. Applied and Environmental Microbiology 79(5), 1676-1688.

Frey, S.K., E. Topp, B.R. Ball, M. Edwards, N. Gottschall, M. Sunohara, E. Zoski, and D.R. Lapen, 2013. Tile drainage management influences on surface-water and

groundwater quality following liquid manure application. Journal of Environmental Quality, 42(3): 881-892.

Friedman, J.H., 1991. Multivariate adaptive regression splines. The Annals of Statistics, 19(1): 1-67.

Gao, G., R.A. Falconer, R.A. and B. Lin, 2011. Numerical modeling sediment-bacteria interaction processes in the Severn Estuary. Journal of Water Resource and Protection, 3, 22-31.

Garcia-Armisen, T., and P. Servais, 2009. Partitioning and fate of particle-associated *E. coli* in river waters. Water Environment Research, 81(1): 21-28.

Garzio-Hadzick, A., D.R. Shelton, R.L. Hill, Y.A. Pachepsky, A.K. Guber, and R. Rowland, 2010. Survival of manure-borne *E. coli* in streambed sediment: effects of temperature and sediment properties. Water Research, 44: 2753-2762.

Geeraerd, A.H., V.P. Valdramidis, and J.F. Van Impe, 2005. GinaFiT, a freeware tool to assess non-log-linear microbial survivor curves. International Journal of Food Microbiology, 102: 95-106.

Geohring, L.D., P.E. Wright, and T.S. Steenhuis, 1998. Preferential Flow of Liquid Manure to subsurface drains. In: *Drainage in the 21st Century: Food Production and the Environment*. Proceedings of the 7th Annual Drainage Symposium, Orlando FL,Mar 8-10. Pp 1-8. ASAE, St. Joseph, MI.

George, I., A. Anzil, and P. Servais, 2004. Quantification of fecal coliform inputs to aquatic systems through soil leaching. Water Research, 38: 611-618.

Ghimire, B. and Z. Deng, 2013. Hydrograph-based approach to modeling bacterial fate and transport in rivers. Water Research, 47: 1329-1343.

Gordon, D.M., S. Bauer, and J.R. Johnson, 2002. The genetic structure of *Escherichia coli* populations in primary and secondary habitats. Microbiology, 148: 1513-1522.

Gordon, N.D., T.A. McMahon, B.L. Finlayson, C.J. Gippel, and R.J. Nathan, R.J. 2004. *Stream Hydrology: an Introduction for Ecologists*, 2nd Edition. John Wiley & Sons, Ltd.

Goss, M. and C. Richards, 2008. Development of a risk-based index for source water protection planning, which supports the reduction of pathogens from agricultural activity entering water resources. Journal of Environmental Management, 87: 623-632.

Grant, S.B., R.M. Litton-Mueller and J.H. Ahn, 2011. Measuring and modeling the flux of fecal bacteria across the sediment-water interface in a turbulent stream. Water Resources Research, 47(5*):* doi:10.1029/2010WR009460.

Griffith, D.A and P.R. Peres-Neto 2006. Spatial modeling in ecology: the flexibility of eigenfunction spatial analyses. Ecology*,* 87(10): 2603-2613.

Guber A.K., Y.A. Pachepsky, D.R. Shelton and O.Yu, 2007. Effect of bovine manure on fecal coliform attachment to soil and soil particles of different sizes. Applied and Environmental Microbiology, 73: 3363-3370.

Habteselassie, M., M. Bischoff, E. Blume, B. Applegate, B. Reuhs, S. Brouder, and R.F. Turco, 2008. Environmental controls on the fate of *Escherichia coli* in soil. Water, Air, and Soil Pollution, 190(1-4): 143-155.

Haller, L., E. Amedegnato, J. Pote and W. Wildi, 2009. Influence of freshwater sediment characteristics on persistence of fecal indicator bacteria. Water, Air and Soil Pollution*,* 203:217-227.

Halliday, E. and R.J. Gast, 2011. Bacteria in beach sands: an emerging challenge in protecting coastal water quality and bather health. Environmental Science and Technology, 45: 370-379.

Hammer, O., D.A.T. Harper and P.D. Ryan, 2001. PAST: Paleontological Statistics Software Package for Education and Data Analysis. Palaeontologica Electronica 4(1): 9 pp.

Harwood, V.J., 2007. Assumptions and limitations associated with microbial source tracking methods. In: J.W. Santo Domingo and M.J. Sadowsky (eds.): Microbial Source Tracking. ASM Press, Washington, D.C., USA.

He, L-M., J. Lu and W. Shi, 2007. Variability of fecal indicator bacteria in flowing and ponded waters in southern California: implications for bacterial TMDL development and implementation. Water Research, 41(14): 3132-3140.

Hirotani, H. and M. Yoshino, 2010. Microbial indicators in natural biofilms developed in the riverbed. Water Science and Technology, 62(5): 1149-1153.

Huan, S-W., B.M. Hsu, Y-J. Su, D-D. Ji, W-C. Lin, J-L. Chen, F-C. Shih, P-M. Kao, and Y-C. Chiu, 2012. Occurrence of diarrheagenic *Escherichia coli* genes in raw water of water treatment plants. Environmental Science and Pollution Research, DOI 10.1007/s11356-012-0777-4.

Huang, Z. and A. Tunnacliffe, 2004. Response of human cells to dessication: comparison with hyperosmotic stress response. Journal of Physiology, 558: 181-191.

Hunter, P.R., 2003. Drinking water and diarrhoeal disease due to *Escherichia coli*. Journal of Water and Health 1(2), 65-72.

Ibekwe, A.M., S.E. Murinda, and A.K. Graves, 2011. Genetic diversity and antimicrobial resistance of *Escherichia coli* from human and animal sources uncovers multiple resistances from human sources. PLoS ONE, 6(6): e20819.

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. Applied and Environmental Microbiology, 72(1): 612-621.

Ishii, S., D.L. Hansen, R.E. Hicks, and M.J. Sadowsky, 2007. Beach sand and sediments are temporal sinks and sources of *Escherichia coli* in Lake Superior. Environmental Science and Technology, 41: 2203-2209.

Ishii, S. and M.J. Sadowsky. 2008. *Escherichia coli* in the environment: implications for water quality and human health. Microbes and Environments, 23(2): 101-108.

Ishii, S., T. Yan, H. Vu, D.L. Hansen, R.E. Hicks, and M.J. Sadowsky, 2010. Factors controlling long-term survival and growth of naturalized *Escherichia coli* populations in temperate field soils. Microbes and Environments, 25(1):8-14.

Jamieson, R.C., R.J. Gordon, K.E. Sharpes, G.W. Stratton, and A. Madani, 2002. Movement and persistence of fecal bacteria in agricultural soils and subsurface drainage water: a review. Canadian Biosystems Engineering, 44: 1.1-1.9.

Jamieson, R., R. Gordon, S. Tattrie, and G. Stratton, 2003. Sources and persistence of fecal coliform bacteria in a rural watershed. Water Quality Research Journal of Canada, 38: 33-47.

Jamieson, R.C., D.M. Joy, H. Lee, R. Kostaschuk, R., and R.J. Gordon, 2004. Persistence of enteric bacteria in alluvial streams. Journal of Environmental Engineering and Science, 3: 203-212.

Jamieson, R. D.M. Joy, H. Lee, H., R. Kostaschuk, and R. Gordon, 2005. Transport and deposition of sediment-associated *Escherichia coli* in natural streams. Water Research, 39: 2665-2675.

Jekabsons, G., 2011. ARESLab: Adaptive Regression Splines toolbox for Matlab/Octave. Available at http://www.cs.rtu.lv/jekabsons/

Johnson, L.A.K., M.B. Brown, E.A. Carruthers, J.A. Ferguson, J.A., P.E. Dombek, and M.J. Sadowsky, 2004. Sample size, library composition and genotypic diversity among natural populations of *Escherichia coli* from different animals influence accuracy of determining sources of fecal pollution. Applied and Environmental Microbiology, 70(8): 4478-4485.

Jones, R.M., L. Liu and S. Dorevitch, 2012. Hydrometeorological variables predict fecal indicator bacteria densities in freshwater: data-driven methods for variable selection. Environmental Monitoring and Assessment, doi 10.1007/s10661-012-2716-8.

Kelsey, H., D.E. Porter, G. Scott, M. Neet, and D. White, 2004. Using geographic information systems and regression analysis to evaluate relationships between land use and fecal coliform bacterial pollution. Journal of Experimental Marine Biology and Ecology, 298: 197-209.

Kim, J-W., Y.A. Pachepsky, D.R. Shelton, and C. Coppock, 2010. Effect of streambed bacteria release on E. coli concentrations: monitoring and modeling with the modified SWAT. Ecological Modelling, 221: 1592-1604.

Kinley, R.D., R.J. Gordon, G.W. Stratton, G.T. Patterson, and H.J. Hoyle, 2007. Phosphorus losses through agricultural tile drainage in Nova Scotia, Canada. Journal of Environmental Quality, 36(2): 469-477.

Kinzelman, J., S.L. McLellan, A.D. Daniels, S. Cashin, A. Singh, S. Gradus, S. and R. Bagley, 2004. Non-point source pollution: determination of replication versus persistence of *Escherichia coli* in surface water and sediments with correlation of levels to readily measurable environmental parameters. Journal of Water and Health: 2(2): 103-114.

Kon, T., S.C. Weir, E.T. Howell, H. Lee, and J.T. Trevors, 2007. Genetic relatedness of *Escherichia coli* isolates in interstitial water from a Lake Huron (Canada) Beach. Applied and Environmental Microbiology, 73(6): 1961-1967.

Kon, T., S.C. Weir, E.T. Howell, H. Lee, and J.T. Trevors, 2009. Repetitive element (REP)-polymerase chain reaction (PCR) analysis of *Escherichia coli* isolates from recreational waters of southeastern Lake Huron. Canadian Journal of Microbiology, 55(3): 269-276.

Krometis, L-A.H., G.W. Characklis, O.D. Simmons III, M.J. Dilts, C.A. Lidirdopulos, and M.D. Sobsey, 2007. Intra-storm variability in microbial partitioning and microbial loading rates. Water Research 41(2), 506-516.

Laliberte, P., and D.J. Grimes, 1982. Survival of *Escherichia coli* in lake bottom sediments. Applied and Environmental Microbiology, 43(3): 623-628.

Lau, M.M. and S.C. Ingham, 2001. Survival of faecal indicator bacteria in bovine manure incorporated into soil. Letters of Applied Microbiology, 33: 131-136.

Legendre, P. and E.D. Gallagher, 2001. Ecologically meaningful transformations for ordination of species data. Oecologia, 129(2): 271-280.

Leibold, M.A., M. Holyoak, N. Mouquet, P. Amarasekare, J.M. Chase, and M.F. Hoopes, 2004. The metacommunity concept: a framework for multi-scale community ecology. Ecology Letters, 7: 601-613.

Liang, Z., Z. He, C.A. Powell, and P.J. Stoffella. 2011. Survival of *Escherichia coli* in soil with modified community composition. Soil Biology and Biochemistry, 43(7): 1591-1599.

Liwimbi, L., A.K. Graves, D.W. Israel, E. van Heugten, B. Robinson, C.W. Cahoon, and J.F. Lubbers, 2010. Microbial source tracking in a watershed dominated by swine. Water, 2: 587-604.

Loh, W-Y., 2011. Classification and regression trees. WIREs Data Mining and Knowledge Discovery, 1: 14-23.

Lothigius, A., A. Janzon, Y. Begum, A. Sjoling, F. Qadri, A-M. Svennerholm, and I. Bolin, 2007. Entertoxigenic *Escherichia coli* is detectable in water samples from an endemic area by real-time PCR. Journal of Applied Microbiology 104, 1128-1136.

Lu, L., M.E. Hume, K.L. Sternes, and S.D. Pillai, 2004. Genetic diversity of *Escherichia coli* isolates in irrigation water and associated sediments: implications for source tracking. Water Research, 38(18): 3899-3908.

Lu, Z., D. Lapen, A. Scott, A. Dang, and E. Topp, 2005. Identifying host sources of fecal pollution: diversity of *Escherichia coli* in confined dairy and swine production systems. Applied and Environmental Microbiology, 71(10): 5992-5998.

Luo, W., R.W. Skaggs, and G.M. Chescheir, 2000. DRAINMOD modifications for cold conditions. Transactions of the ASAE, 43(6), 1569–1582.

Manel, S., H.C. Williams, and S.J. Ormerod, 2001. Evaluating presence-absence models in ecology: the need to account for prevalence. Journal of Applied Ecology 38, 921-931.

Masters, N., A. Wiegand, W. Ahmed, and M. Katouli, 2011. *Escherichia coli* virulence genes profile of surface waters as an indicator of water quality. Water Research 45, 6321-6333.

Mawdsley, J.L., R.D. Bardgett, R.J. Merry, B.F. Pain, and M.K. Theodorou, 1995. Pathogens in livestock waste, their potential for movement through soil and environmental pollution. Applied Soil Ecology, 2: 1-15.

McKergow, L.A. and R.J. Davies-Colley, 2010. Stormflow dynamics and loads of *Escherichia coli* in a large mixed land use catchment. Hydrologic Processes, 24: 276-289.

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

Meals, D.W. and D.C. Braun, 2006. Demonstration of methods to reduce *E. coli* runoff from dairy manure application sites. Journal of Environmental Quality, 35: 1088-1100.

Meals, D.W., S.A. Dressing, and T.E. Davenport, 2009. Lag time in water quality response to best management practices: a review. Journal of Environmental Quality, 39(1): 85-96.

Moriasi, D.N., J.G. Arnold, M.W. Van Liew, R.L. Bingner, R.D. Harmel, and T.L. Veith, 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Transactions of the ASABE, 50(3): 885-900.

Muirhead, R.W., R.J. Davies-Colley, A.M. Donnison, and J.W. Nagels, 2004. Fecal bacteria yields in artificial flood events: quantifying in-stream stores. Water Research, 38: 1215-1224.

Natvig, E.E., S.C. Ingham, B.H. Ingham, L.R. Cooperband, and T.R. Roper, 2002. *Salmonella enterica* serovar *typhimurium* and *Escherichia coli* contamination of root and leaf vegetables grown in soils with incorporated bovine manure. Applied and Environmental Microbiology, 68(6): 2737-2744.

Oliver, D.M., L. Heathwaite, P.M. Haygarth, and C.D. Clegg, 2005. Transfer of *Escherichia coli* to water from drained and undrained grassland after grazing. Journal of Environmental Quality, 34: 918-925.

Oliver, D.M., C.D. Clegg, A.L. Heathwaite, and P.M. Haygarth, 2007. Preferential attachment of *Escherichia coli* to different particle size fractions of an agricultural grassland soil. Water, Air and Soil Pollution, 185: 369-375.

Ouattara, N.K., J. Passerat, and P. Servais, 2011. Faecal contamination of water and sediment in the rivers of the Scheldt drainage network. Environmental Monitoring and Assessment, 183: 243-257.

Pachepsky, Y.A., A.M. Sadeghi, S.A. Bradford, D.R. Shelton, A.K. Guber and T. Dao, 2006. Transport and fate of manure-borne pathogens: modeling perspective. Agricultural Water Management, 86: 81-92.

Pachepsky, Y.A., Yu, O., J.S. Karns, D.R. Shelton, A.K. Guber, and J.S. van Kessel, 2008. Strain-dependent variations in attachment of *E. coli* to particles of different sizes. International Agrophysics, 22: 61-66.

Pachepsky, Y.A. and D.R. Shelton, 2011. *Escherichia coli* and fecal coliforms in freshwater and estuarine sediments. Critical Reviews in Environmental Science and Technology, 41: 1067-1110.

Page, R.M., S. Scheidler, E. Polat, P. Svoboda, and P. Huggenberger, 2012. Faecal indicator bacteria: groundwater dynamics and transport following precipitation and river water infiltration. Water, Air and Soil Pollution, 223: 2771-2782.

Pandey, P.K., M.L. Soupir, and C.R. Rehmann, 2012. A model for predicting resuspension of *Escherichia coli* from streambed sediments. Water Research, 46: 115-116.

Pappas, E.A., R.S. Kanwar, J.L. Baker, J.C. Lorimor, and S.K. Mickelson, 2008. Fecal indicator bacteria in subsurface drain water following swine manure application. Transactions of the ASABE, 51(5): 1567-1573.

Park, S-J., C-G. Lee, and S-B. Kim, 2008. Quantification of bacterial attachment-related parameters in porous media. Environmental Engineering Research 13(3), 141-146.

Parkhurst, D.F., K.P. Brenner, A.P. Dufour, and L.J. Wymer, 2005. Indicator bacteria at five swimming beaches – analysis using random forests. Water Research 39(7), 1354-1360.

Peres-Neto, P.R. and P. Legendre, 2010. Estimating and controlling for spatial structure in the study of ecological communities. Global Ecology and Biogeography, 19: 174-184.

Peterson, E.E., J.M. Ver Hoef, D.J. Isaak, J.A. Falke, M-J. Fortin, C.E. Jordan, K. McNyset, P. Monestiez, A.S. Ruesch, A. Sengupta, N. Som, E.A. Steel, D.M. Theobald, C.E. Torgersen, and S.J. Wenger, 2013. Modelling dendritic ecological networks in space: an integrated network perspective. Ecology Letters, 16: 707-719.

Pietramellara, G., J. Ascher, F. Borgogni, M.T. Ceccherini, G. Guerri and P. Nannipieri, 2009. Extracellular DNA in soil and sediment: fate and ecological relevance. Biology and Fertility of Soils, 45: 219-235.

Piorkowski, G.S., R.C. Jamieson, L. Truelstrup Hansen, G.S. Bezanson, and C.K. Yost, 2013a. Characterizing spatial structure of sediment *E. coli* populations to inform sampling design. Environmental Monitoring and Assessment, doi: 10.1007/s10661-013-3373-2.

Piorkowski, G.S., R.C. Jamieson, G.S. Bezanson, L. Truelstrup Hansen and C.K. Yost, 2013b. Evaluation of statistical models for predicting *E. coli* particle attachment in fluvial systems. Water Research, doi: 10.1016/j.watres.2013.09.003.

Powell, D.M., 1998. Patterns and processes of sediment sorting in gravel-bed rivers. Progress in Physical Geography, 22(1): 1-32.

Rademaker, J.L.W., F.J. Louws, J. Versalovic and F.J. de Brujin, 2004. Characterization of the diversity of ecologically important microbes by rep-PCR genomic fingerprinting. In: G.A. Kowalchuk, F.J. de Brujin, J.M. Head, A.D. Akkermans, and J.D. van Elsas (eds). Molecular microbial ecology manual (Vol. 1 and 2). Dordrecht, The Netherlands. Pp 611-643.

Ran, Y., J.M. Fu, G.Y. Sheng, R. Beckett, and B.T. Hart, 2000. Fractionation and composition of colloidal and suspended particulate materials in rivers. Chemosphere 41(1-2), 33-43.

Raith, M.R., C.A. Kelty, J.F. Griffith, A. Schriewer, S. Wuertz, S. Mieszkin, M. Gourmelon, G.H. Reischer, A.H. Farnleitner, J.S. Ervin, P. Holden, D.L. Ebentier, J.A. Jay, D. Wang, A.B. Boehm, T.G. Aw, J.B. Rose, E. Balleste, W.G. Meijeri, M. Sivaganesan, and O.C. Shanks, 2013. Comparison of PCR and quantitative real-time PCR methods for the characterization of ruminant and cattle fecal pollution sources. Water Research, http:/dx.doi.org/10.1016/j.watres.2013.03.061

Raup, D.M. and R.E. Crick, 1979. Measurement of faunal similarity in paleontology. Journal of Paleontology, 53(5): 1213-1227.

Rehmann, C.R. and M.L. Soupir, 2009. Importance of interactions between the water column and the sediment for microbial concentrations in streams. Water Research, 43: 4579-4589.

Reischer, G.H., D.C. Kasper, R. Steinborn, R.L. Mach, and A.H. Farnleitner, 2006. Quantitative PCR method for sensitive detection of ruminant fecal pollution in freshwater and evaluation of this method in alpine karstic regions. Applied and Environmental Microbiology, 72: 5610-5614.

Reischer, G.H., D. Kollanur, J. Vierheilig, C. Wehrspaun, R.L. Mach, R. Sommer, H. Stadler, and A.H. Farnleitner, 2011. Hypothesis-driven approach for the identification of fecal pollution sources in water resources. Environmental Science and Technology, 45: 4038-4045.

Reynolds, K.A., K.D. Mena, and C.P. Gerba, 2008. Risk of waterborne illness via drinking water in the United States. Reviews of Environmental Contamination and Toxicology, 192: 117-158.

Ritter, W.F., 2001. Nonpoint source pollution and livestock manure management. In: W.F. Ritter and A. Shirmohammadi (eds.). Agricultural nonpoint source pollution: watershed management and hydrology. CRC Press Inc., Boca Raton, FL, USA. Pp. 136-164.

Rodgers, P., C. Soulsby, C. Hunter, and J. Petry, 2003 Spatial and temporal bacterial quality of a lowland agricultural stream in northeast Scotland. Science of the Total Environment, 314-316: 289-302.

Rogers, S.W., M. Donnelly, L. Peed, C.A. Kelty, S. Mondal, Z. Zhong and O.C. Shanks, 2011. Decay of bacterial pathogens, fecal indicators, and real-time quantitative genetic markers in manure-amended soils. Applied and Environmental Microbiology, 77(14): 4839-4848.

Roslev, P. and A.S. Bukh, 2011. State of the art molecular markers for fecal source tracking in water. Applied Microbiology and Biotechnology, 89: 1341-1355.

Russo, S.A., J. Hunn and G.W. Characklis, 2011. Considering bacteria-sediment associations in microbial fate and transport modeling. Journal of Environmental Engineering, 137: 697-706.

Samarajeewa, A.D., S.M. Glasauer, J.D. Lauzon, I.P. O'Halloran, G.W. Parkin, and K.E. Dunfield, 2012. Bacterial contamination of tile drainage water and shallow groundwater under different application methods of liquid swine manure. Canadian Journal of Microbiology, 58: 668-677.

Santo Domingo, J.W., D.G. Bambic, T.A. Edge, and S. Weurtz, 2007. Quo vadis source tracking? Towards a strategic framework for environmental monitoring of fecal pollution. Water Research, 41(16): 3539-3552.

Schilling, K.E., Y-K. Zhang, D.R. Hill, C.S. Jones and C.F. Wolter, 2009. Temporal variations of *Escherichia coli* concentrations in a large Midwestern river. Journal of Hydrology, 365: 79-85.

Schleppi, P., P.A. Waldner, and B. Fritschi, 2006. Accuracy and precision of different sampling strategies and flux integration methods for runoff water: comparisons based on measurements of the electrical conductivity. Hydrological Processes 20(2), 395-410.

Servais, P., T. Garcia-Armisen, I. George, and G. Billen, 2007. Fecal bacteria in the rivers of the Seine drainage network (France): Sources, fate and modeling. Science of the Total Environment, 375: 152-167.

Sinclair, A., D. Hebb, R. Jamieson, R. Gordon, K. Benedict, K. Fuller, G.W. Stratton, and A. Madani, 2008. Growing season surface water loading of fecal indicator organisms within a rural watershed. Water Research 43, 1199-1206.

Sinton, L.W., R.R. Braithwaite, C.H. Hall, and M.L. Mackenzie, 2007. Survival of indicator and pathogenic bacteria in bovine feces on pasture. Applied and Environmental Microbiology, 73(24): 7917-7925.

Skaggs, R.W., 1980. DRAINMOD reference report. USDA-SCS South Nation Technical Center, Fort Worth, TX, USA. Available at: www.bae.ncsu.edu/soil_water/documents/drainmod/chapter1.pdf

Skaggs, R.W., M.A. Breve, and J.W. Gilliam, 1994. Hydrologic and water quality impacts of agricultural drainage. Critical Reviews in Environmental Science and Technology. 24(1): 1-32.

Skjemstad, J.O. and J.A. Baldock, 2008. Total and organic carbon. In: Carter MR, Gregorich, ED (eds). *Soil Sampling and Methods of Analysis*, 2nd Edition. Canadian Society of Soil Science, Taylor & Francis Group, Boca Raton, FL. Pp. 225-237.

Smith, J.E., and J.M. Perdek, 2004. Assessment and management of watershed microbial contaminants. Critical Reviews in Environmental Science and Technology, 34: 109-139.

Sokolova, E., J. Astrom, T.J.R. Pettersson, O. Bergstedt, and M. Hermansson, 2012. Decay of Bacteroidales genetic markers in relation to traditional fecal indicators for water quality modeling of drinking water sources. Environmental Science and Technology, 46: 892-900.

Solomon, E.B., S. Yaron, and K.R. Matthews, 2002. Transmission of *Escherichia coli* O157:H7 from contaminated manure and irrigation water to lettuce plant tissue and its subsequent internalization. Applied and Environmental Microbiology, 68(1): 397-400.

Somarelli, J.A., J.C., Makarewicz, R. Sia, and R. Simon, 2007. Wildlife identified as major source of *Escherichia coli* in agriculturally dominated watersheds by BOX A1R-derived genetic fingerprints. Journal of Environmental Management, 82(1): 60-65.

Soupir, M.L. and S. Mostaghimi, 2010. *Escherichia coli* and enterococci attachment to particles in runoff from highly and sparsely vegetated grassland. Water, Air and Soil Pollution 39, 1019-1027.

Stidson, R.T., C.A. Gray, and C.D. McPhail, C.D., 2012. Development and use of modeling techniques for real-time bathing water quality predictions. Water and Environment Journal 26, 7-18.

Stoeckel, D.M. and V.J. Harwood. 2007. Performance, design, and analysis in microbial source tracking studies. Applied and Environmental Microbiology, 73(8): 2405-2415.

Stumpf, C.H., M.F. Piehler, S. Thompson, and R.T. Noble, 2010. Loading of fecal indicator bacteria in North Carolina tidal creek headwaters: hydrographic patterns and terrestrial runoff relationships. Water Research, 44: 4704-4715

Surbeck, C.Q., S.C. Jiang, and S.B. Grant, 2010. Ecological control of fecal indicator bacteria in an urban stream. Environmental Science and Technology, 44: 631-637.

Tallon, P., B. Magajna, C. Lofranco, and K.T. Leung, 2005. Microbial indicators of faecal contamination in water: a current perspective. Water, Air and Soil Pollution, 166: 139-166.

Tambalo, D.D., T. Boa, K. Liljebjelke, and C.K. Yost. 2012a. Evaluation of two quantitative PCR assays using *Bacteroidales* and mitochondrial DNA markers for tracking dog fecal contamination in waterbodies. Journal of Microbiological Methods, 91: 459-467.

Tambalo, D.D., B. Fremaux, T. Boa, and C.K. Yost. 2012b. Persistence of host-associated *Bacteroidales* gene markers and their quantitative detection in an urban and agricultural mixed prairie watershed. Water Research, 46: 2891-2904

Tetzlaff, D., R. Capell, and C. Soulsby, 2012. Land use and hydroclimatic influences on faecal indicator organisms in tow large Scottish catchments: towards land use-based models as screening tools. Science of the Total Environment, 434: 110-122.

Thiagarajan, A., R. Gordon, A. Madani and G.W. Stratton, 2007. Discharge of *Escherichia coli* from agricultural surface and subsurface water: tillage effects. Water, Air and Soil Pollution, 182: 3-12.

Tibshirani, R. J., 1996. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society, Serial B 58(1), 267–288.

Tiedje, J.M., A.J. Sexstone, T.B. Parkin, N.P. Revsbech, and D.R. Shelton, 1984. Anaerobic processes in soil. Plants and Soil, 76: 197-212.

Toma, C., Y. Lu, N. Higa, N. Nakasone, I. Chinen, A. Baschkier, M. Rivas, M. Iwanaga, 2003. Multiplex PCR assay for identification of human diarrheagenic *Escherichia coli*. Journal of Clinical Microbiology 41(6), 2669-2671.

Topp, E., M. Welsh, Y-C. Tien, A. Dang, G. Lazarovits, K. Conn, and H. Zhu, 2003. Strain-dependent variability in growth and survival of *Escherichia coli* in agricultural soil. FEMS Microbiology Ecology, 44(3): 303-308.

Unc, A. and M.J. Goss, 2006. Culturable *Escherichia coli* in soil mixed with two types of manure. Soil Science Society of America Journal, 70(3): 763-769.

VanderZaag, A.C., K.J. Campbell, R.C. Jamieson, A.C. Sinclair and L.G. Hynes, 2010. Survival of *Escherichia coli* in agricultural soil and presence in tile drainage and shallow groundwater. Canadian Journal of Soil Science, 90(3): 495-505.

Van Elsas J.D., A.V. Semenov, R. Costa and J.T. Trevors J.T., 2011. Survival of *Escherichia coli* in the environment: fundamental and public health aspects. ISME Journal, 5(2): 173-183.

Van Liew, M., T.L. Veith, D.D. Bosch and J.G. Arnold, 2007. Suitability of SWAT for the conservation effects assessment project: a comparison on USDA-ARS experimental watersheds. Journal of Hydrologic Engineering, 12(2): 173-189.

Větrovskỳ, T. and P. Baldrian, 2013. The variability of the 16S rRNA gene in bacterial genomes and consequences for bacterial community analysis. PLoS ONE, 8(2): e57923. doi:10.1371/journal.pone.0057923.

Vinten, A.J.A., D.R. Lewis, D.R. Fenlon, K.A. Leach, R. Howard, I. Svoboda, and I. Ogden, 2002. Fate of *Escherichia coli* and *Escherichia coli* O157 in soils and drainage water following cattle slurry application at 3 sites in southern Scotland. Soil Use and Management, 18: 223-231.

Vinten, A.J.A., J.T. Douglas, D.R. Lewis, M.N. Aitken, and D.R. Fenlon, 2004. Relative risk of surface water pollution by *E. coli* derived from faeces of grazing animals compared to slurry application. Soil Use and Management, 20: 13-22.

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

Wang, X., M.A. Youssef, R.W. Skaggs, J.D. Atwood, and J.R. Frankenberger, 2005. Sensitivity analyses of the nitrogen simulation model DRAINMOD-N II. Transactions of the ASAE, 48(6), 2205–2212.

Watelet, A. and P.G. Johnson, 1999. Hydrology and water quality of the Raisin River: Overview of impacts of recent land and channel changes in Eastern Ontario. Canadian Journal of Water Quality Research. 34(3):361-390

Weurtz, S., D. Wang, G.H. Reischer, and A.H. Farnleitner, 2011. Library-independent bacterial source tracking methods. In: C. Hagedorn et al. (eds.) Microbial Source Tracking: Methods, Applications, and Case Studies. Springer Science. Pp 61-112.

Whitlock, J.E., D.T. Jones, and V.J. Harwood, 2002. Identification of sources of fecal coliforms in an urban watershed using antibiotic resistance analysis. Water Research, 36: 4273-4282.

Whitman, R.L., M.B. Nevers, and M.N. Byappanahalli, 2006. Examination of watershed-wide distribution of *Escherichia coli* along Southern Lake Michigan: an integrated approach. Applied and Environmental Microbiology, 72(11): 7301-7310.

Whittam, T.S., 1989. Clonal dynamics of *Escherichia coli* in its natural habitat. Antonie van Leeuwenhoek, 55: 23-32.

Wilkes, G., T.A. Edge, V.P.J. Gannon, C. Jokinen, E. Lyautey, N.F. Neumann, N. Ruecker, A. Scott, M. Sunohara, E. Topp, and D.R. Lapen, 2011. Associations

among pathogenic bacteria, parasites and environmental and land use factors in multiple mixed-use watersheds. Water Research 45, 5807-5825.

Wilkinson, R.J., L.A. McKergow, R.J. Davies-Colley, D.J. Ballantine, and R.G. Young, 2011. Modelling storm-event *E. coli* pulses from the Motueka and Sherry Rivers in the South Island, New Zealand. New Zealand Journal of Marine and Freshwater Research, 45(3): 369-393.

Wu, J., P. Rees, S. Storrer, K. Alderisio and S. Dorner, 2009. Fate and transport modeling of potential pathogens: the contribution from sediments. Journal of the American Water Resources Association, 45(1): 35-44.

Yakirevich, A., Y.A. Pachepsky, A.K. Guber, T.J. Gish, D.R. Shelton, and K.H. Cho, 2013. Modeling transport of *Escherichia coli* in a creek during and after artificial hig-flow events; Three-year study and analysis. Water Research, 47: 2676-2688.

Yoder, J.S., M.C. Hlavsa, G.F. Craun, V. Hill, V. Roberts, and P.A. Yu, P.A., 2008. Surveillance for waterborne disease and outbreaks associated with recreational water use and other aquatic facility-associated health events – United States. MMWR Surveillance Summaries, 57(9): 1-29.

Zaleski, K.J., K.L. Josephson, C.P. Gerba and I.L. Pepper, 2005. Survival, growth and regrowth of indicator and pathogenic bacteria in biosolids, compost, soil, and land applied biosolids. Journal of Residuals Science and Technology, 2(1): 49-63.

APPENDIX A SUPPLEMENTARY MATERIAL

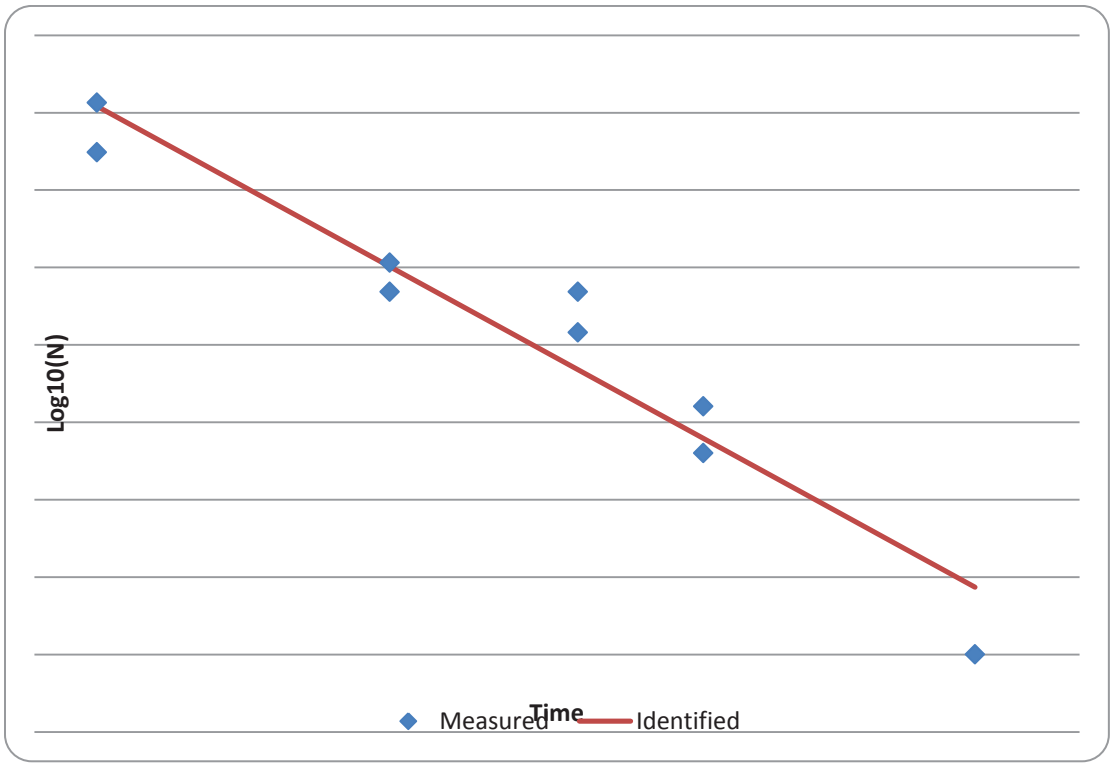
Supplement 3.1. Inactivation curve results for solid manure addition in drainage plot DT2:

Parameters	Parameter values	Standard Error
kmax	0.170	0.018
LOG10(N0)	2.76	0.20

Mean Sum of Squared Error	0.0952
Root Mean Sum of Squared Error	0.3086
R-Square	0.9263
R-Square adjusted	0.9158

Inactivation model identified

N= N0 * exp(-kmax * t)



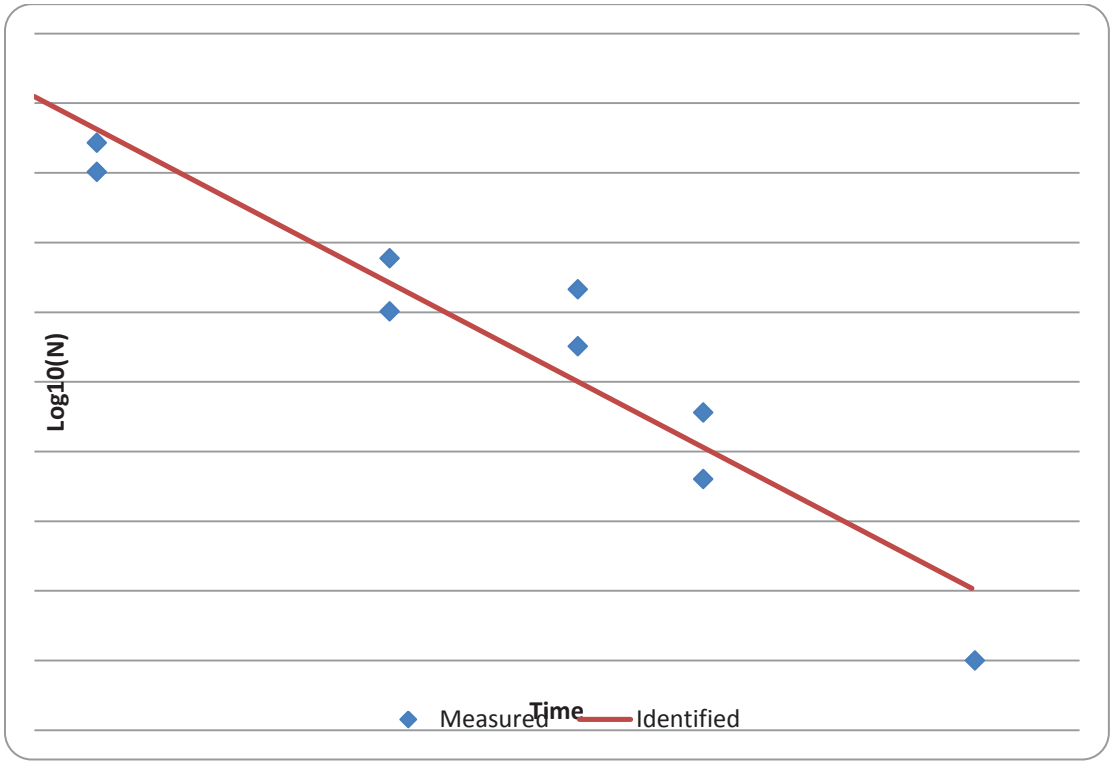
Supplement 3.2. Inactivation curve results for solid manure addition in drainage plot DT3:

Parameters	Parameter values	Standard Error
kmax	0.181	0.023
LOG10(N0)	3.05	0.26

Mean Sum of Squared Error	0.1497
Root Mean Sum of Squared Error	0.3869
R-Square	0.9005
R-Square adjusted	0.8863

Inactivation model identified

N= N0 * exp(-kmax * t)



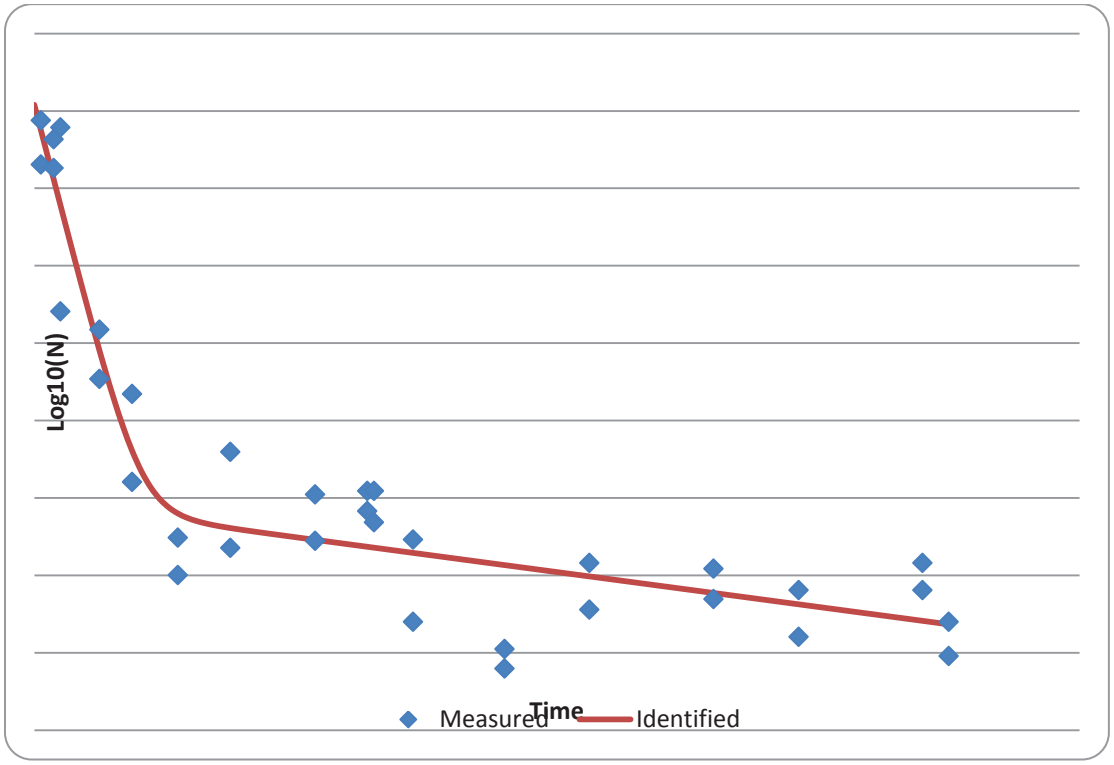
Supplement 3.3. Inactivation curve results for liquid dairy manure addition in drainage plot DT2:

Parameters	Parameter values	Standard Error
f	0.987	0.0014
kmax1	0.376	0.067
kmax2	0.014	0.004
LOG10(N0)	4.11	0.19

Mean Sum of Squared Error	0.1083
Root Mean Sum of Squared Error	0.3291
R-Square	0.9135
R-Square adjusted	0.9048

Inactivation model identified

log10(N)=log10(N0)+log10(f*exp(-kmax1*t)+(1-f)*exp(-kmax2*t))



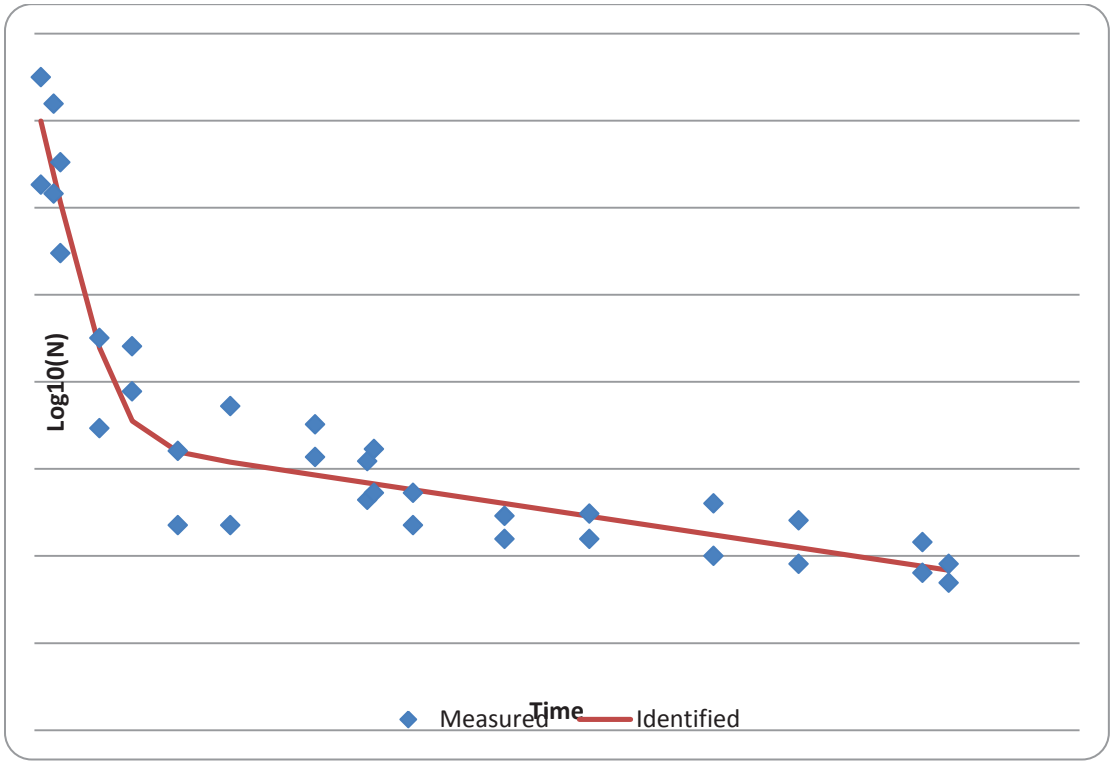
Supplement 3.4. Inactivation curve results for liquid dairy manure addition in drainage plot DT3:

Parameters	Parameter values	Standard Error
f	0.989	0.00427
kmax1	0.369	0.066
kmax2	0.013	0.003
LOG10(N0)	3.66	0.14

Mean Sum of Squared Error	0.0578
Root Mean Sum of Squared Error	0.2404
R-Square	0.9214
R-Square adjusted	0.9136

Inactivation model identified

log10(N)=log10(N0)+log10(f*exp(-kmax1*t)+(1-f)*exp(-kmax2*t))

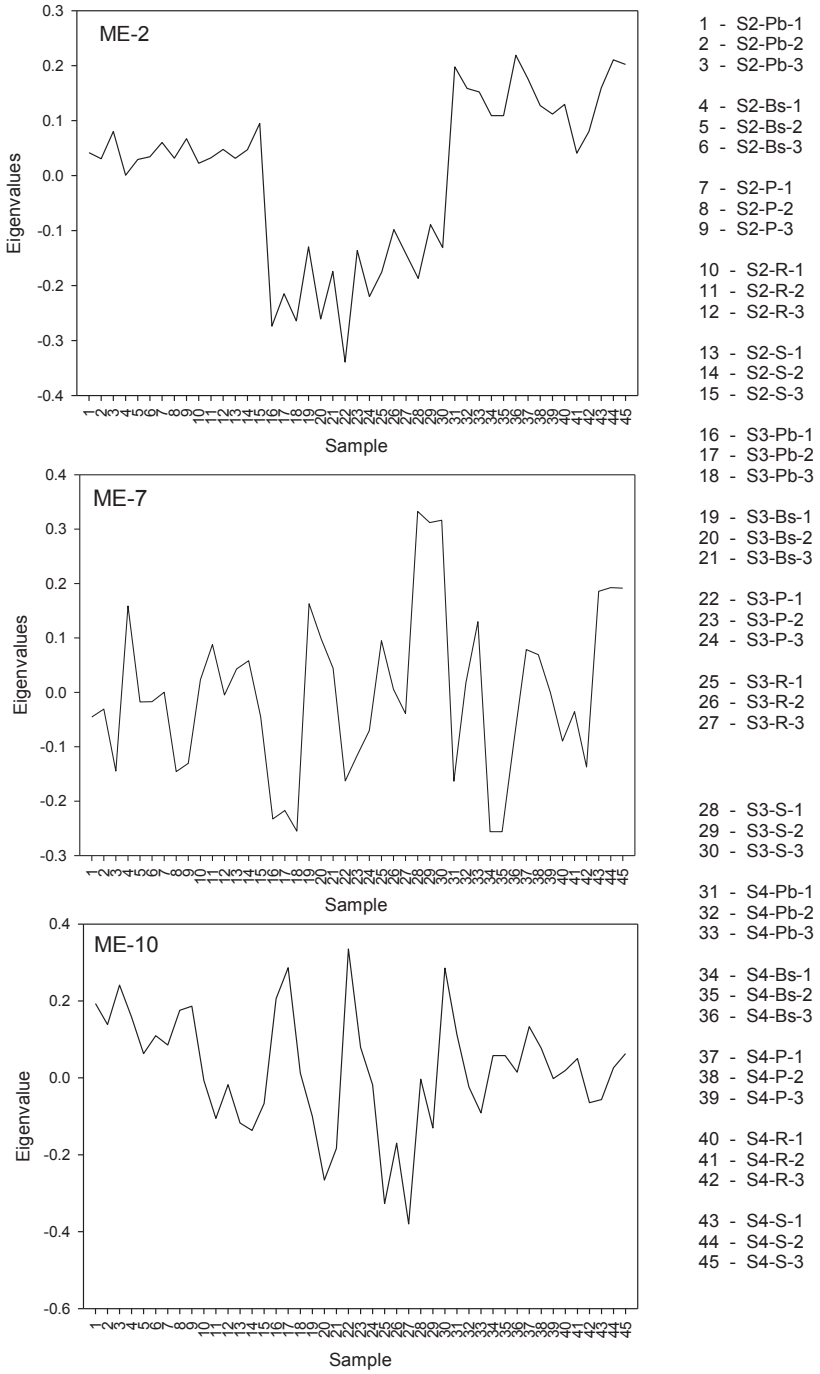


Supplement 4.1. Sediment properties and water velocity at the morphological features of each stream reach for the baseflow sampling event. The morphological features included in the study were point bars (Pb), bank scours (Bs), pools (P), riffles (R), and runs (S). Cell values represent arithmetic means (±standard deviation).

Site	Organic Carbon (mg/g)	D_{10} (µm)	D_{50} (µm)	D_{90} (µm)	Sorting	Sand (%)	Silt (%)	Velocity (m/s)
Site 2								
Pb	7.1 (±1.2)	69.2 (±24.6)	679.3 (±83.8)	1680.8 (±522.5)	3.73 (±1.0)	0.89 (±0.03)	0.10 (±0.02)	0.02
Bs	10.2 (±7.2)	3.6 (±0.5)	182.2 (±29.6)	591.8 (±353.2)	4.69 (±2.1)	0.54 (±0.07)	0.46 (±0.08)	0.03
P	7.2 (±2.1)	62.4 (±5.4)	420.6 (±70.8)	781.6 (±158.0)	2.85 (±0.1)	0.86 (±0.01)	0.12 (±0.02)	0.01
R	6.4 (±2.0)	106.5 (±96.8)	1461.5 (±1825.7)	1953.9 (±238.1)	3.05 (±1.6)	0.90 (±0.05)	0.09 (±0.06)	0.07
S	11.7 (±4.2)	66.7 (±3.7)	167.9 (±102.8)	614.4 (±235.2)	4.44 (±0.4)	0.92 (±0.02)	0.08 (±0.02)	0.02
Site 3								
Pb	4.8 (±1.4)	110.6 (±25.7)	643.8 (±485.5)	1428.8 (±1083.5)	2.43 (±0.4)	0.93 (±0.02)	0.07 (±0.01)	0.05
Bs	9.5 (±5.9)	21.7 (±26.7)	167.5 (±93.3)	1728.7 (±2213.0)	5.20 (±2.9)	0.70 (±0.16)	0.29 (±0.16)	0.09
P	7.3 (±1.2)	83.0 (±15.0)	549.3 (±140.5)	1282.1 (±467.6)	2.56 (±0.4)	0.91 (±0.01)	0.09 (±0.01)	0.03
R	5.0 (±1.4)	282.5 (±20.2)	1823.4 (±507.6)	5523.2 (±540.8)	2.47 (±0.5)	0.95 (±0.02)	0.04 (±0.02)	0.07
S	4.6 (±0.9)	310.8 (±81.8)	1819.5 (±548.6)	3382.0 (±769.4)	2.08 (±0.4)	0.96 (±0.01)	0.04 (±0.01)	0.05
Site 4								
Pb	12.1 (±7.9)	129.6 (±119.8)	413.5 (±11.3)	890.1 (±75.9)	2.33 (±0.7)	0.92 (±0.04)	0.08 (±0.04)	0.08
Bs	7.2 (±6.1)	3.0 (±0.1)	17.9 (±2.5)	447.5 (±43.2)	7.12 (±0.4)	0.30 (±0.01)	0.70 (±0.01)	0.13
P	13.8 (±7.0)	53.8 (±54.2)	453.6 (±139.4)	948.5 (±111.6)	3.48 (±1.5)	0.82 (±0.05)	0.17 (±0.04)	0.01
R	9.3 (±3.6)	127.4 (±142.0)	950.3 (±77.5)	1726.1 (±154.0)	4.31 (±0.6)	0.90 (±0.05)	0.10 (±0.04)	0.03
S	13.0 (±2.1)	248.1 (±150.3)	866.0 (±235.5)	1492.8 (±155.2)	2.20 (±2.3)	0.94 (±0.07)	0.05 (±0.07)	0.08

Supplement 4.2. Sediment properties and water velocity at the morphological features of each stream reach for the post-stormflow sampling event. The morphological features included in the study were point bars (Pb), bank scours (Bs), pools (P), riffles (R), and runs (S). Cell values represent arithmetic means (±standard deviation).

Site	Organic Carbon (mg/g)	D_{10} (μm)	D_{50} (μm)	D_{90} (μm)	Sorting	Sand (%)	Silt (%)	Velocity (m/s)
Site 2								
Pb	6.2 (±0.6)	31.8 (±13.6)	428.7 (±33.8)	1069.1 (±182.5)	4.01 (±0.7)	0.84 (±0.03)	0.13 (±0.02)	0.01
Bs	8.2 (±1.0)	3.8 (±0.5)	103.5 (±46.6)	416.7 (±53.2)	5.67 (±0.5)	0.57 (±0.09)	0.42 (±0.09)	0.02
P	20.8 (±4.9)	11.4 (±5.6)	266.9 (±115.1)	1033.1 (±127.4)	3.43 (±1.4)	0.78 (±0.03)	0.22 (±0.04)	0.01
R	11.7 (±1.1)	29.3 (±3.9)	578.8 (±157.7)	1364.0 (±295.1)	4.23 (±1.1)	0.87 (±0.04)	0.12 (±0.04)	0.03
S	12.6 (±1.4)	25.4 (±14.5)	116.2 (±86.4)	640.5 (±98.0)	3.85 (±1.0)	0.78 (±0.04)	0.18 (±0.02)	0.01
Site 3								
Pb	4.7 (±0.6)	67.7 (±24.8)	378.9 (±16.3)	829.0 (±30.6)	2.70 (±0.3)	0.90 (±0.02)	0.10 (±0.02)	0.09
Bs	14.3 (±4.9)	9.9 (±4.6)	225.0 (±186.9)	1139.1 (±775.1)	5.71 (±1.7)	0.64 (±0.06)	0.31 (±0.10)	0.16
P	10.8 (±3.9)	51.3 (±46.8)	359.6 (±111.6)	1033.4 (±295.3)	3.67 (±1.2)	0.82 (±0.06)	0.13 (±0.08)	0.02
R	5.4 (±0.2)	189.1 (±73.00)	1161.4 (±499.9)	5098.0 (±1139.4)	3.11 (±0.7)	0.89 (±0.02)	0.06 (±0.01)	0.08
S	8.65 (±2.48)	182.0 (±132.11)	829.1 (±35.0)	4655.0 (±1279.4)	3.74 (±1.6)	0.79 (±0.15)	0.08 (±0.06)	0.06
Site 4								
Pb	11.4 (±5.8)	202.1 (±50.24)	473.9 (±57.6)	889.6 (±43.3)	1.81 (±0.1)	0.96 (±0.01)	0.04 (±0.01)	0.04
Bs	3.7 (±0.1)	2.9 (±0.1)	17.5 (±2.1)	448.0 (±35.3)	7.36 (±0.3)	0.33 (±0.01)	0.67 (±0.01)	0.07
P	22.4 (±3.8)	16.9 (±6.1)	447.5 (±94.4)	994.9 (±177.7)	4.79 (±0.4)	0.81 (±0.03)	0.19 (±0.03)	0.01
R	16.4 (±1.6)	101.3 (±25.8)	414.0 (±148.1)	1826.4 (±89.8)	3.95 (±1.3)	0.86 (±0.07)	0.11 (±0.03)	0.05
S	8.4 (±1.1)	285.1 (±61.0)	631.4 (±15.4)	1320.5 (±139.7)	2.05 (±0.2)	0.94 (±0.01)	0.04 (±0.01)	0.19



Supplement 4.3. Plots of statistically significant Moran's eigenvectors determined through partial canonical correspondence analysis. Moran's eigenvector 2 (ME-2) was significant for both sampling events, ME-7 was significant following stormflow and ME-10 was significant during baseflow. Site number is indicated as S2, S3 and S4. Morphological features point bars (Pb), bank scours (Bs), pools (P), riffles (R) and runs (S). Replicate samples within morphological features are labeled as 1, 2 or 3.

APPENDIX B COPYRIGHT PERMISSION LETTERS

Copyright permission letters/Licence Agreements via RightsLink are provided for the following papers:

1. Piorkowski, G.S., R.C. Jamieson, L. Truelstrup Hansen, G.S. Bezanson, and C.K. Yost, 2013. Characterizing spatial structure of sediment *E. coli* populations to inform sampling design. Environmental Monitoring and Assessment, doi: 10.1007/s10661-013-3373-2.

2. Piorkowski, G.S., R.C. Jamieson, G.S. Bezanson, L. Truelstrup Hansen and C.K. Yost, 2013. Evaluation of statistical models for predicting *E. coli* particle attachment in fluvial systems. Water Research, 47(17): 6701-6711.

SPRINGER LICENSE
TERMS AND CONDITIONS

Dec 20, 2013

This is a License Agreement between Gregory S Piorkowski ("You") and Springer ("Springer") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Springer, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number	3272611354395
License date	Nov 19, 2013
Licensed content publisher	Springer
Licensed content publication	Environmental Monitoring and Assessment
Licensed content title	Characterizing spatial structure of sediment E. coli populations to inform sampling design
Licensed content author	Gregory S. Piorkowski
Licensed content date	Jan 1, 2013
Type of Use	Thesis/Dissertation
Portion	Full text
Number of copies	6
Author of this Springer article	Yes and you are a contributor of the new work
Order reference number	
Title of your thesis / dissertation	Genetic Diversity of Escherichia coli in Soils and Sediments of an Agricultural Watershed and their Spatiotemporal Influences on Water Quality
Expected completion date	Dec 2013
Estimated size(pages)	200
Total	0.00 USD
Terms and Conditions	

Introduction

The publisher for this copyrighted material is Springer Science + Business Media. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time

that you opened your Rightslink account and that are available at any time at http://myaccount.copyright.com).

Limited License
With reference to your request to reprint in your thesis material on which Springer Science and Business Media control the copyright, permission is granted, free of charge, for the use indicated in your enquiry.

Licenses are for one-time use only with a maximum distribution equal to the number that you identified in the licensing process.

This License includes use in an electronic form, provided its password protected or on the university's intranet or repository, including UMI (according to the definition at the Sherpa website: http://www.sherpa.ac.uk/romeo/). For any other electronic use, please contact Springer at (permissions.dordrecht@springer.com or permissions.heidelberg@springer.com).

The material can only be used for the purpose of defending your thesis, and with a maximum of 100 extra copies in paper.

Although Springer holds copyright to the material and is entitled to negotiate on rights, this license is only valid, subject to a courtesy information to the author (address is given with the article/chapter) and provided it concerns original material which does not carry references to other sources (if material in question appears with credit to another source, authorization from that source is required as well).

Permission free of charge on this occasion does not prejudice any rights we might have to charge for reproduction of our copyrighted material in the future.

Altering/Modifying Material: Not Permitted
You may not alter or modify the material in any manner. Abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of the author(s) and/or Springer Science + Business Media. (Please contact Springer at (permissions.dordrecht@springer.com or permissions.heidelberg@springer.com)

Reservation of Rights
Springer Science + Business Media reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

Copyright Notice:Disclaimer
You must include the following copyright and permission notice in connection with any reproduction of the licensed material: "Springer and the original publisher /journal title, volume, year of publication, page, chapter/article title, name(s) of author(s), figure number(s), original copyright notice) is given to the publication in which the material was originally published, by adding; with kind permission from Springer Science and Business Media"

Warranties: None

Example 1: Springer Science + Business Media makes no representations or warranties with respect to the licensed material.

Example 2: Springer Science + Business Media makes no representations or warranties with respect to the licensed material and adopts on its own behalf the limitations and disclaimers established by CCC on its behalf in its Billing and Payment terms and conditions for this licensing transaction.

Indemnity
You hereby indemnify and agree to hold harmless Springer Science + Business Media and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.

No Transfer of License
This license is personal to you and may not be sublicensed, assigned, or transferred by you to any other person without Springer Science + Business Media's written permission.

No Amendment Except in Writing
This license may not be amended except in a writing signed by both parties (or, in the case of Springer Science + Business Media, by CCC on Springer Science + Business Media's behalf).

Objection to Contrary Terms
Springer Science + Business Media hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and Springer Science + Business Media (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.

Jurisdiction
All disputes that may arise in connection with this present License, or the breach thereof, shall be settled exclusively by arbitration, to be held in The Netherlands, in accordance with Dutch law, and to be conducted under the Rules of the 'Netherlands Arbitrage Instituut' (Netherlands Institute of Arbitration).***OR:***

All disputes that may arise in connection with this present License, or the breach thereof, shall be settled exclusively by arbitration, to be held in the Federal Republic of Germany, in accordance with German law.

Other terms and conditions:

v1.3

If you would like to pay for this license now, please remit this license along with your payment made payable to "COPYRIGHT CLEARANCE CENTER" otherwise you will be invoiced within 48 hours of the license date. Payment should be in the form of a check or money order referencing your account number and this invoice number RLNK501163045.
Once you receive your invoice for this order, you may pay your invoice by credit card. Please follow instructions provided at that time.

Make Payment To:
Copyright Clearance Center
Dept 001
P.O. Box 843006
Boston, MA 02284-3006

For suggestions or comments regarding this order, contact RightsLink Customer Support: customercare@copyright.com or +1-877-622-5543 (toll free in the US) or +1-978-646-2777.

Gratis licenses (referencing $0 in the Total field) are free. Please retain this printable license for your reference. No payment is required.

ELSEVIER LICENSE
TERMS AND CONDITIONS

Dec 20, 2013

This is a License Agreement between Gregory S Piorkowski ("You") and Elsevier ("Elsevier") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

Supplier	Elsevier Limited The Boulevard,Langford Lane Kidlington,Oxford,OX5 1GB,UK
Registered Company Number	1982084
Customer name	Gregory S Piorkowski
Customer address	1360 Barrington Street Halifax, NS B2Y 2Y4
License number	3272610703722
License date	Nov 19, 2013
Licensed content publisher	Elsevier
Licensed content publication	Water Research
Licensed content title	Evaluation of statistical models for predicting*Escherichia coli*particle attachment in fluvial systems
Licensed content author	Gregory Piorkowski,Rob Jamieson,Greg Bezanson,Lisbeth Truelstrup Hansen,Chris Yost
Licensed content date	1 November 2013
Licensed content volume number	47
Licensed content issue number	17
Number of pages	11
Start Page	6701
End Page	6711
Type of Use	reuse in a thesis/dissertation
Intended publisher of new work	other

Portion	full article
Format	electronic
Are you the author of this Elsevier article?	Yes
Will you be translating?	No
Title of your thesis/dissertation	Genetic Diversity of Escherichia coli in Soils and Sediments of an Agricultural Watershed and their Spatiotemporal Influences on Water Quality
Expected completion date	Dec 2013
Estimated size (number of pages)	200
Elsevier VAT number	GB 494 6272 12
Permissions price	0.00 USD
VAT/Local Sales Tax	0.00 USD / 0.00 GBP
Total	0.00 USD
Terms and Conditions	

INTRODUCTION

1. The publisher for this copyrighted material is Elsevier. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at http://myaccount.copyright.com).

GENERAL TERMS

2. Elsevier hereby grants you permission to reproduce the aforementioned material subject to the terms and conditions indicated.

3. Acknowledgement: If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies. Suitable acknowledgement to the source must be made, either as a footnote or in a reference list at the end of your publication, as follows:

"Reprinted from Publication title, Vol /edition number, Author(s), Title of article / title of chapter, Pages No., Copyright (Year), with permission from Elsevier [OR APPLICABLE SOCIETY COPYRIGHT OWNER]." Also Lancet special credit - "Reprinted from The Lancet, Vol. number, Author(s), Title of article, Pages No., Copyright (Year), with permission from Elsevier."

4. Reproduction of this material is confined to the purpose and/or media for which permission is hereby given.

5. Altering/Modifying Material: Not Permitted. However figures and illustrations may be altered/adapted minimally to serve your work. Any other abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of Elsevier Ltd. (Please contact Elsevier at permissions@elsevier.com)

6. If the permission fee for the requested use of our material is waived in this instance, please be advised that your future requests for Elsevier materials may attract a fee.

7. Reservation of Rights: Publisher reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

8. License Contingent Upon Payment: While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by publisher or by CCC) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and publisher reserves the right to take any and all action to protect its copyright in the materials.

9. Warranties: Publisher makes no representations or warranties with respect to the licensed material.

10. Indemnity: You hereby indemnify and agree to hold harmless publisher and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.

11. No Transfer of License: This license is personal to you and may not be sublicensed, assigned, or transferred by you to any other person without publisher's written permission.

12. No Amendment Except in Writing: This license may not be amended except in a writing signed by both parties (or, in the case of publisher, by CCC on publisher's behalf).

13. Objection to Contrary Terms: Publisher hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you,

which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and publisher (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.

14. Revocation: Elsevier or Copyright Clearance Center may deny the permissions described in this License at their sole discretion, for any reason or no reason, with a full refund payable to you. Notice of such denial will be made using the contact information provided by you. Failure to receive such notice will not alter or invalidate the denial. In no event will Elsevier or Copyright Clearance Center be responsible or liable for any costs, expenses or damage incurred by you as a result of a denial of your permission request, other than a refund of the amount(s) paid by you to Elsevier and/or Copyright Clearance Center for denied permissions.

LIMITED LICENSE

The following terms and conditions apply only to specific license types:

15. **Translation**: This permission is granted for non-exclusive world **English** rights only unless your license was granted for translation rights. If you licensed translation rights you may only translate this content into the languages you requested. A professional translator must perform all translations and reproduce the content word for word preserving the integrity of the article. If this license is to re-use 1 or 2 figures then permission is granted for non-exclusive world rights in all languages.

16. **Website**: The following terms and conditions apply to electronic reserve and author websites:
Electronic reserve: If licensed material is to be posted to website, the web site is to be password-protected and made available only to bona fide students registered on a relevant course if:
This license was made in connection with a course,
This permission is granted for 1 year only. You may obtain a license for future website posting,
All content posted to the web site must maintain the copyright information line on the bottom of each image,
A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx or the Elsevier homepage for books at http://www.elsevier.com , and
Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.

17. **Author website** for journals with the following additional clauses:

All content posted to the web site must maintain the copyright information line on the bottom of each image, and the permission granted is limited to the personal version of your paper. You are not allowed to download and post the published electronic version of your article (whether PDF or HTML, proof or final version), nor may you scan the printed edition to create an electronic version. A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx . As part of our normal production process, you will receive an e-mail notice when your article appears on Elsevier's online service ScienceDirect (www.sciencedirect.com). That e-mail will include the article's Digital Object Identifier (DOI). This number provides the electronic link to the published article and should be included in the posting of your personal version. We ask that you wait until you receive this e-mail and have the DOI to do any posting.

Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.

18. **Author website** for books with the following additional clauses:
Authors are permitted to place a brief summary of their work online only.
A hyper-text must be included to the Elsevier homepage at http://www.elsevier.com . All content posted to the web site must maintain the copyright information line on the bottom of each image. You are not allowed to download and post the published electronic version of your chapter, nor may you scan the printed edition to create an electronic version.

Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.

19. **Website** (regular and for author): A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx. or for books to the Elsevier homepage at http://www.elsevier.com

20. **Thesis/Dissertation**: If your license is for use in a thesis/dissertation your thesis may be submitted to your institution in either print or electronic form. Should your thesis be published commercially, please reapply for permission. These requirements include permission for the Library and Archives of Canada to supply single copies, on demand, of the complete thesis and include permission for UMI to supply single copies, on demand, of the complete thesis. Should your thesis be published commercially, please reapply for permission.

21. **Other Conditions**:

v1.6

If you would like to pay for this license now, please remit this license along with your payment made payable to "COPYRIGHT CLEARANCE CENTER" otherwise you will be invoiced within 48 hours of the license date. Payment should be in the form of a check or money order referencing your account number and this invoice number RLNK501163033.
Once you receive your invoice for this order, you may pay your invoice by credit card. Please follow instructions provided at that time.

Make Payment To:
Copyright Clearance Center
Dept 001
P.O. Box 843006
Boston, MA 02284-3006

For suggestions or comments regarding this order, contact RightsLink Customer Support: customercare@copyright.com or +1-877-622-5543 (toll free in the US) or +1-978-646-2777.

Gratis licenses (referencing $0 in the Total field) are free. Please retain this printable license for your reference. No payment is required.