ESCHERICHIA COLI TRANSPORT MODELING AT SOIL COLUMN SCALE AND WATERSHED SCALE

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

By

Chaozi Wang

August 2015

ESCHERICHIA COLI TRANSPORT MODELING AT SOIL COLUMN SCALE AND WATERSHED SCALE

Chaozi Wang, Ph.D.

Cornell University 2015

Understanding the transport of *Escherichia coli* at different scales is very important for water quality control and public health. Therefore, *E. coli* transport modeling was conducted at both soil column scale and watershed scale. At soil column scale, first, a small-scale rainfall experiment was conducted in which *E. coli* was mixed with a simple soil composed of sand (250-300 μm) and clay with the mass ratio of 9:1. By applying Hairsine-Rose model and Gao model to the microbial transport simulation, the solute-particle duality of microbes was discovered: a solute that does not diffuse is essentially a particle and a particle that does not settle out of suspension is essentially a solute. Then, the co-transport of clay and *E. coli* was investigated by simulated rainfall over two sets of saturated soil columns infused with *E. coli*: one set of columns consisted of sand and the other consisted of a 9:1 sand-clay mixture. Based on a combination of empirical and modeling results it was concluded that any role clay particles play in facilitating bacteria transport is offset by its role in decreasing the penetration depth and the effectiveness of raindrop impact. At last, the conclusions from the soil column scale studies were applied to the understanding and prediction of *E. coli* concentration and loading in a stream, and it was found that the level of *E. coli* in a stream is driven by four independent processes—runoff and erosion, microbial activity, shallow subsurface flow, and groundwater, and the importance of the processes at Fall Creek Watershed decreases by the order.

BIOGRAPHICAL SKETCH

Chaozi Wang was born in Beijing, the capital of China, in Sep. 9th 1989 and grown up there until she came to Ithaca. Chaozi Wang began to love Mathematics at 7 years old, love Physics at 13, love Chemistry at 14, love geography at 15, and love biology at 16. And she loves all these disciplines until now.

The time she graduated from her undergraduate university, Beijing Normal University, in Resources Science and Engineering, she found that the cross-disciplinary research in Soil and Water Lab in the Department of Biological and Environmental Engineering, Cornell University is just the subject crosses all the disciplines she loves.

At Soil and Water Lab, she studied preferential erosion of biochar for her Master's, and microbial transport for her PhD.

ACKNOWLEDGMENTS

I want to thank Professor M. Todd Walter for coming up with such interesting projects for me to work on, for patiently answering my bunch of questions from the day he became my advisor until now, for giving me a lot of suggestions in choosing courses, modifying experiment, analyzing data, finding a good model, and writing, especially writing the thesis, for letting me feel free to speak out my thoughts and discuss problems with him. I real enjoy working with him and being his student.

I want to thank Professor Jean-Yves Parlange for helping me find the right way to do the experiment, for always sitting at his office and willing to answer my questions, for teaching me the physical meaning of the transport equations, and for giving me very quick response whenever I send him an email. He makes me feel free to explore the unknown things and apply complicated methods, because he can always find a way to solve my problems.

I want to thank Professor Rebecca L. Schneider for giving me many good suggestions on the dissertation, especially on Chapter 4, and always encouraging me. I feel very happy working with her.

I want to thank Erik Rasmussen, Xing Wang, Professor Minghong Cheng, Sara Storrer, Pu Wang, Theresa Chu, and Grace Tan for helping with experiment, and thank Dr. Bin Gao for assistance with Gao solute model. I want to thank Allison Truhlar for proof reading Chapter 2 and Chapter 3 for me.

I want to thank China Scholarship Council for providing a four-year scholarship.

I want to thank my parents for giving me life and not preventing me to come to Cornell University to study, although I'm their precious only child.

I want to thank my husband, Pu Wang, for being such an excellent elder brother, who gives me a lot of guidance and is always willing to answer my all kinds of questions, and tolerate my all kinds of silly thoughts and behavior.

TABLE OF CONTENTS

CHAPTER 1 INTRODUCTION

The World Health Organization (WHO, 2011) considers water borne pathogens to be the most important water quality risk to address in the foreseeable future. Indeed, pathogens impair more kilometers of rivers and streams than any other pollutant for 303(d) listed water bodies, or water bodies failing to meet water quality standards for their designated uses (US EPA, 2010). *Escherichia coli* (*E. coli*) is a widely accepted indicator of water borne pathogen (WHO, 2011). Therefore, an improved understanding of the transport mechanisms of *E. coli* may help in developing strategies for mitigating pathogen loads between the landscape and surface waters.

At the watershed scale, climatic, hydrological, and water quality factors have all been shown to affect concentrations and/or loadings of *E. coli* in streams. Whitman et al. (2008) found that sunlight, season, snowmelt, and storms can influence *E. coli* in a stream. Vidon et al. (2008) stated that stream flow conditions (high flow or baseflow), discharge, precipitation (especially 7 day antecedent precipitation), turbidity, location (i.e. headwater or lower reaches) and temperature all showed significant correlation with either concentrations or loadings of *E. coli* in streams. Dwivedi et al. (2013) investigated the correlations of 13 water quality factors with respect to *E. coli* loading and concluded that temperature, dissolved oxygen, phosphate, ammonia, suspended solids and chlorophyll are the most important ones. Diurnal variability (Meays et al., 2006) and carbon dioxide (Gray, 1975) have been found to be relevant as well.

Existing watershed models (Benham et al., 2006; Walker and Stedinger, 1999) and other conceptual models (Wilkinson et al., 1995) that predict *E. coli* loads in streams generally require too much site specific information to easily employ. Artificial neural network models can be good

predictors of *E. coli* (Basant et al., 2010; Dwivedi et al., 2013), but often the underlying processes influencing *E. coli* transport are unclear. Statistical models offer some advantages to these other approaches in that they do not require full *a priori* understanding of the sources, sinks, transport processes, etc. Multiple linear regression (MLR) models (Nevers and Whitman, 2005) cannot avoid co-linearity, which is common in water quality data. Models like LOADEST estimate constituent loads utilizing only streamflow and time (Dwivedi et al., 2013; Runkel et al., 2004), however, as summarized above, *E. coli* levels depends not only on streamflow, but also on other, perhaps correlated, factors.

At plot scale, empirical study dominants, people apply microbes to plots and measure their vertical and/or horizontal movement. Roodsari et al. (2005) studied both vertical and horizontal transport of fecal coliform along 20% slopes. They found that vegetation largely increases infiltration, therefore, reduces horizontal transport but increases vertical transport. And it is more effective for soil with poorer drainage condition. Muirhead et al. (2006b) investigated the horizontal transport of *E. coli* and bromide across saturated soil, and found that *E. coli* behave as solute in this process. In addition, it was found that if *E. coli* was pre-attached to large particles (>45μm), it will settle out of the overland flow; but if *E. coli* was not pre-attached to large particles, it will remain in suspension. The experiments of Ferguson et al. (2007) included bare soil treatment, first as a control and then receiving *E. coli* enriched cowpats. Their objective was to determine how vegetation and microbe size influence microbial transport in runoff through statistical comparisons.

And at soil column scale, most studies are on microbial transport through vertical or horizontal soil columns or porous media chambers. Using vertical soil columns, people focus on the effect of

soil structure on microbial transport (Guber et al., 2005b; Safadoust et al., 2012a; Safadoust et al., 2012b). Using horizontal soil columns or porous media chambers, people focus on revealing the transport mechanisms and determining the important coefficients for the microbes (Barton and Ford, 1995, 1997; Olson et al., 2005; Reynolds et al., 1989; Sherwood et al., 2003).

In this dissertation, we investigate a rarely studied microbial transport process at soil column scale: splash erosion under raindrop impact, and try a new combination of methods to explain and predict the pathogen level in streams.

In order to elucidate the mechanisms controlling the transfer of microbes from soil into runoff under rainfall conditions, we designed a simple experiment to investigate the splash erosion process of *E. coli* from soil into overland flow and applied two models to explain this process. Our primary objective of Chapter 2 is to determine if *E. coli* transport can be simulated in the same way Gao et al. (2004) modeled solute transport into overland flow or if an erosion model, like the Hairsine-Rose model (1991), is needed to capture the particle-like characteristics of *E. coli*.

Both bacteria and clay particles fall into the colloid size range of 1 nm to 10 μm, and that *E. coli* preferentially attached to those soil particles within the size range 30-16 μm (Chrysikopoulos and Sim, 1996; Oliver et al., 2007; Vasiliadou and Chrysikopoulos, 2011a). Bacteria are commonly considered to be biocolloids (Kanti Sen and Khilar, 2006; Vasiliadou and Chrysikopoulos, 2011a).

Many researchers have studied colloid-facilitated contaminant transport (Abdel-Salam and Chrysikopoulos, 1995; Chen et al., 2005; Corapcioglu and Jiang, 1993; Grolimund and Borkovec, 2005; Ibaraki and Sudicky, 1995; Ouyang et al., 1996; Roy and Dzombak, 1997; Šimůnek et al., 2006; Smith and Degueldre, 1993) including bacteria-facilitated contaminant transport (Bekhit et al., 2009; Guiné et al., 2003; Kim and Corapcioglu, 2002a, b; Kim et al., 2003; Pang et al., 2005; Pang and Šimůnek, 2006). However there are few empirical studies of the transport of biocolloids, particularly bacteria, through association with mineral colloids.

Vasiliadou and Chrysikopoulos (2011a) and Chen (2012) investigated mineral colloids associated bacteria transport through soil columns. Vasiliadou and Chrysikopoulos (2011a) found that the kaolinite colloids inhibited the transport of *Pseudomonas putida* (*P. putida*), a rod-shaped bacteria of similar size as *E. coli*, because *P. putida* attached to kaolinite and kaolinite stayed attached to the solid matrix. Chen (2012) found that mineral colloids could either facilitate or retard the transport of *Salmonella typhimurium* and *E. coli* O157:H7 in soil depending on whether the mineral colloids were mobile or not, respectively.

Muirhead et al. (2006b) looked at the cotransport of *E. coli* and bromide in overland flow across saturated soil and concluded that they have similar transport mechanisms. It was also found that *E. coli* mainly attached to mineral particles smaller than 2 μm and, once mobilized, remained in suspension.

Based on the experimental design and conclusions in Chapter 2, in Chapter 3 we compare the transport of *E. coli* from soil into runoff when clay is and is not part of the soil matrix.

Partial least squares regression (PLS) is a powerful tool to select the most important variables among many highly correlated predictor variables (Esbensen et al., 2002). Thus, PLS is widely used in water quality modeling (Aguilera et al., 2000; Basant et al., 2010; Ortiz-Estarelles et al., 2001a; Ortiz-Estarelles et al., 2001b; Singh et al., 2007). Carroll et al. (2009) built PLS models using the data from possible sources to predict the data from the corresponding sinks to confirm or challenge possible source-sink relationships. Brooks et al. (2013) used PLS to predict fecal indicator bacteria on Great Lakes beaches using a large suite of possible predictors. Their predictors that might be relevant to streams included turbidity, air temperature, and antecedent rainfall (24h and 48h), and variables related to season (Julian date and month). Other variables were mostly relevant to beaches, like wave height, wind speed and wind direction. Like many statistical models, PLS is not necessarily reliable for revealing the underlying processes that control the *E. coli* levels (Brooks et al., 2013).

Principal component analysis (PCA) is a widely used method in classification. Many people used PCA for *E. coli* identification in food samples (Al-Holy et al., 2006; Al-Qadiri et al., 2006; de Sousa Marques et al., 2013; Siripatrawan et al., 2011), i.e. to differentiate *E. coli* from other microbes. A few researchers have used it to identify factors correlated to *E. coli* concentrations in field samples (Bech et al., 2014; Dwivedi et al., 2013). Dwivedi et al. (2013) used 6 variables in a PCA analysis, and found that PC1 corresponded mostly to physical factors, accounting for dissolved oxygen and temperature, and PC2 corresponded primarily to chemical and biological factors, accounting for phosphate and ammonia, suspended solids, and chlorophyll. Using PCA, Bech et al. (2014) found that leached fecal bacteria was negatively correlated with days after

slurry-manure application, which was consistent with Falbo et al. (2013). However, they also found that leached fecal bacteria is negatively correlated with soil water content, which, they admitted, is contrary to other researcher's findings. Moreover, they found that leached fecal bacteria was not correlated with precipitation, which is also contrary to previous findings.

However, to our knowledge, previous studies have not used PCA and PLS together to understand and model concentrations or loadings of *E. coli* in streams. We, thus, tested the applicability of applying PCA and PLS to stream *E. coli* concentrations and loadings to both predict and explain mechanisms controlling stream *E. coli* in Chapter 4.

In Chapter 5 we will give a comprehensive summary of the conclusions we draw from the above three Chapters and look into the future of *E. coli* transport modeling.

CHAPTER 2 MODELING TRANSPORT OF *ESCHERICHIA COLI* FROM SOIL INTO OVERLAND FLOW UNDER RAINDROP IMPACT: SOLUTE OR PARTICLE?[1]

Abstract

Escherichia coli (*E. coli*) transport from soil under rain-splash erosion has not been well studied. One central question is whether *E. coli* will behave like a solute, which it is clearly not, or small particles, which are much more difficult to experimentally quantify. A small-scale rainfall experiment was conducted in which *E. coli* was mixed with a simple soil composed of sand (250-300 μm) and clay with the mass ratio of 9:1, respectively. The Gao solute model and the Hairsine-Rose erosion model were applied to the experiment. *E. coli* transport into overland flow was equally well modeled either as a small particle (Hairsine-Rose) or as a solute (Gao). This is perhaps expected given the similarity of the underpinning mechanisms and assumptions in the two models and we show that they are identical for very small or light particles. However, the solute model allows one to quantify *E. coli* concentration as colony-forming units (CFU) per volume of water rather than make assumptions about the mass of *E. coli*, which is necessary when modeling them as particles. We were not trying to determine whether *E. coli* attached to clay or not during this process, because the phase of *E. coli* does not affect its transport. We conclude that a solute that does not diffuse is essentially a "particle" and a particle that does not settle out of suspension is essentially a "solute."

[1] Wang, C., Parlange, J.-Y., Rasmussen, E.W., Wang, X., Chen, M., Walter, M.T., 2015. Modeling transport of *Escherichia coli* from soil into overland flow under raindrop impact: Solute or particle? J. Hydrol. (submitted for publication).

2.1 Introduction

The World Health Organization (WHO, 2011) considers water borne pathogens to be the most important water quality risk to address in the foreseeable future. Indeed, pathogens impair more kilometers of rivers and streams than any other pollutant for 303(d) listed water bodies, or water bodies failing to meet water quality standards for their designated uses (US EPA, 2010). Storm water is a primary transport pathway for many pathogens and increased concentrations are often correlated with large precipitation and snowmelt events (Falbo et al., 2013; Jamieson et al., 2003; Kistemann et al., 2002; Pettibone and Irvine, 1996; Simon and Makarewicz, 2009a, b; Traister and Anisfeld, 2006). An improved understanding of the transport mechanisms of microorganisms, like *Escherichia coli* (*E. coli*), may help in developing strategies for mitigating pathogen loads between the landscape and surface waters.

Previous studies of *E. coli* transport primarily focused on the attachment of *E. coli* to solid particles (Guber et al., 2005a, b; Muirhead et al., 2006a, b; Oliver et al., 2007), the transport of *E. coli* through soil columns (Barton and Ford, 1995, 1997; Jiang et al., 2007; Olson et al., 2005; Powelson and Mills, 2001; Schäfer et al., 1998; Sherwood et al., 2003; Smith et al., 1985), or the transport of *E.coli* in runoff over bare soil or through vegetated buffers (Ferguson et al., 2007; Muirhead et al., 2006a).

Recently, several researchers conducted rainfall experiments to investigate rain-driven erosion of manure slurry amended soil or cowpats and the associated transfer of *E. coli* into runoff (Ferguson et al., 2007; Kouznetsov et al., 2007; Muirhead et al., 2005, 2006c; Roodsari et al., 2005; Zyman and Sorber, 1988). The experiments of Ferguson et al. (2007) included bare soil treatment, first as a control and then receiving *E. coli* enriched cowpats. Their objective was to determine how vegetation and microbe size influence microbial transport in runoff through statistical comparisons. In contrast, our goal is to elucidate the mechanisms controlling the transfer of microbes from soil into runoff under rainfall conditions. Therefore, we designed a simple experiment to investigate the splash erosion process of *E. coli* from soil into overland flow and applied two mechanistic models to explain this process. Our primary objective is to determine if *E. coli* transport can be simulated in the same way Gao et al. (2004) modeled solute transport into overland flow or if an erosion model, like the Hairsine-Rose model (1991), is needed to capture the particle-like characteristics of *E. coli*. Of course, *E. coli* are colloidal particles, similar in size to clay, but it is difficult and expensive to quantify them as either a number-of-particles or a mass-of-particles. The simplest way to measure them is to culture samples on agar plates and enumerate the concentration as colony-forming units (CFU) per volume of sample. While this is a fairly standard and reasonably repeatable measurement technique, it is difficult to relate CFUs to bacteria numbers or masses, but they can be expressed as a concentration suitable for a solute model.

2.2 Experimental design

The experiment set-up (Figure 2.1) was similar to that utilized by Gao et al. (2004, 2005) and Wang et al. (2013), which was based on Heilig et al. (2001). This design allows us to make a

number of assumptions that simplify our modeling: the ponding water depth and rain intensity are constant throughout the experiment and the soil is a simple 9:1 mixture of kaolinite clay and dark sand, equivalent to loamy sand soil; we assume that sand settles out of the ponding water very rapidly and the clay effectively remains in suspension once ejected from the soil. We also assume *E. coli* remains in suspension, whether attached to clay particles or not, once ejected from the soil.

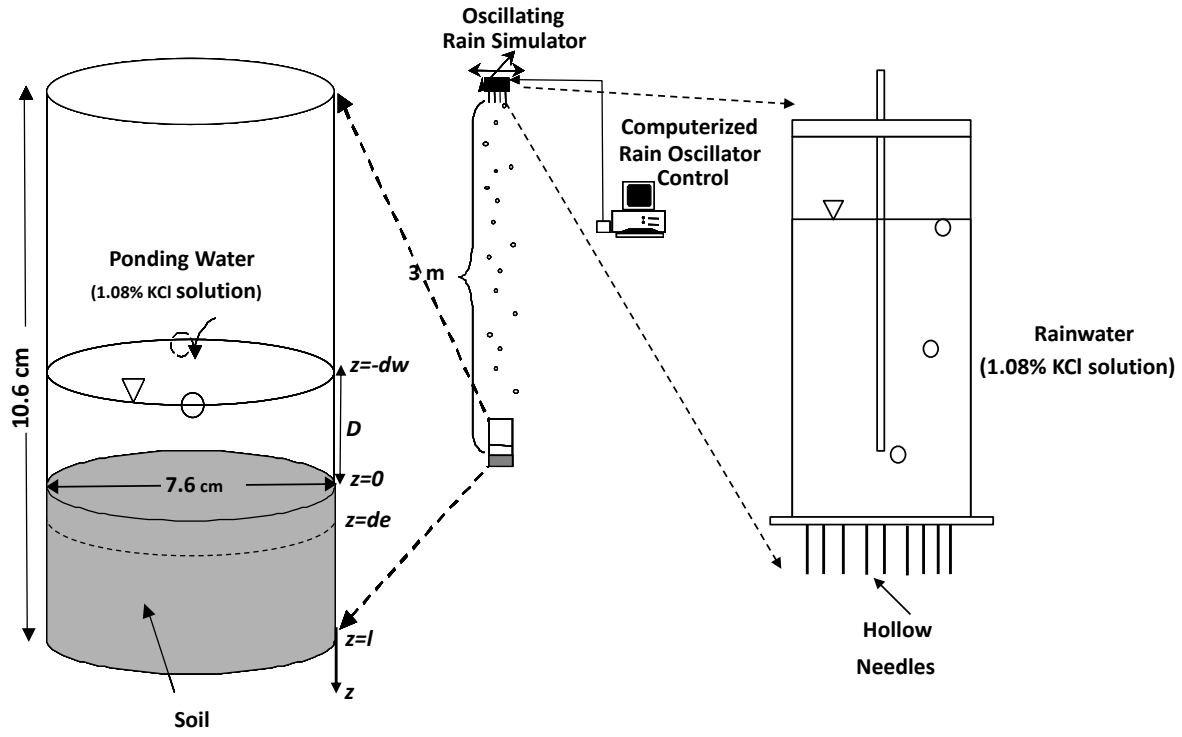


Figure 2.1 Experiment set-up, adapted from Gao et al. (2004, 2005). The ponded water (d_w) exchange layer (d_e) as defined in the Gao model are labeled next to the column.

The soil mixture consisted of 225 g sand (University Sand & Gravel, Brooktondale, NY) and 25 g clay (Kaolinite, Englehard Corp, NJ) in a 7.6-cm-diameter cylindrical plexiglass column (Figure

2.1). This soil mixture is similar to that used by Gao et al. (2004, 2005) with the exception of a slightly larger sand particle size used in this experiment (250-300 µm versus 198-212 µm).

Escherichia coli ATCC 25922, a nonpathogenic surrogate of pathogenic *E. coli* O157:H7 (Muirhead et al., 2006b; Salleh-Mack and Roberts, 2007; Sauer and Moraru, 2009), was grown in Tryptic Soy Broth (TSB; Becton, Dickinson and Company, Sparks, MD) for 18 hours at 37^{o}C on a shaking table. Two milliliters of this culture was then mixed with 60 ml 1.08% potassium chloride (KCl) solution. The resulting solution and the soil mixture were homogenized, and poured into a column.

The pre-saturated soil columns were packed on a shaking table. A 0.5 ml sample was immediately extracted from the ponded solution to determine the initial *E. coli* concentration in the soil water. The remaining ponded solution was poured off. The column was then placed under the rainmaker and *E. coli*-free 1.08 % KCl solution was gently added to pre-pond the experiment, so our steady-state runoff assumption would be valid. Another 0.5 ml sample was extracted from the clean ponded water to characterize the runoff at $t = 0$.

We used sterilized rainwater so that we did not introduce any unknown *E. coli.* A Marriott bottle was used (Figure 2.1) to seal the rainwater from the environment while maintaining a constant rainfall rate. The rainwater was 1.08% KCl solution, as deionized (DI) water would cause the cells to lyse.

The study area was protected with an umbrella until the rainfall was steady. A timer was started when the umbrella was removed. Samples of 0.5 ml were taken from the runoff at varying intervals depending on how rapidly we anticipated the changes in concentration based on trial experiments. The length of the sampling interval changed at 0, 2, 5, 10, and 20 min from the beginning of rainfall. The sampling intervals used at each time period were 15 s, 30 s, 1 min, 2 min, and 5 min, respectively. Rainfall lasted for a total of 30 minutes.

The concentration of bacteria in each runoff sample was determined by a dilution and inoculation procedure. Sterilized 1.08% KCl solution was used as the diluent. Three aliquots from each runoff sample were diluted to three different end ratios. Subsequently, each dilution was plated on *Escherichia coli* media with 4-methylumbelliferyl-β-D-glucuronide (EC-MUG; Neogen corporation, Lansing, MI) and incubated for 20 hours at 37^{o}C. CFUs were manually counted on each plate and converted to the bacteria concentration units (CFU/ml) in the original runoff sample. To verify that there were no unknown sources of *E. coli*, an additional experimental run was done without adding *E. coli* to the soil.

Despite the short duration of our experiments, we were concerned that the *E. coli* concentrations could be influenced by growth or die-off of the organisms. Growth or survival curves have been developed for *E. coli* in many different environments (Beversdorf et al., 2007; Gill and Delacy, 1991; Hwang et al., 2014; Muirhead and Littlejohn, 2009; Sagdic and Ozturk, 2013), but none

exist for conditions similar to our experiment. We diluted our culture of *E. coli* in TSB with different doses of 1.08% KCl solution. Twenty microliters from each diluted sample was inoculated on an EC-MUG plate every 2 hours over 10 hours. The plates were then incubated for 20 hours at 37^{o}C and enumerated to generate a population curve. There was a 40% increase in the population of *E. coli* over the 10 hours period, but only a 5% increase in the population of *E. coli* during the first two hours (data not shown). Thus, we assume growth or die-off of *E. coli* during our 30 min experiments is negligible, which is consistent with Oliver et al. (2007).

The concentration of clay in water samples was measured following Heilig et al. (2001) and Gao et al. (2003, 2004, 2005) . A spectrometer (Spectronic 1001, Bausch and Lomb) was used to measure the runoff samples at 546.1 nm. Samples from one experimental run without clay in the soil were also analyzed by spectroscopy in order to determine if the presence of *E. coli*, TSB, or KCl in the water samples interfered with the measurements of clay concentration. These samples were all below our detection limit, so we assume these substances did not substantially affect our clay concentration measurements.

The rainfall intensity and the ponding water depth were measured before and after each experimental run. As in previous studies using this set-up (Gao et al., 2004, 2005; Wang et al., 2013), a layer of almost pure sand develops on the surface of the soil, i.e., referred to as a shield layer in the context of erosion and as a mixing-layer (a.k.a exchange-layer) in the context of solute transfer. The shield layer depth and the dry weight of shield layer were measured after each experimental run. Although we attempted to keep conditions identical between experimental runs,

ponding depth, rainfall intensity, and initial *E. coli* concentrations varied a little from run to run due to uncontrollable variability; this was partially attributable to our need to set-up and run the experiments rapidly before the *E. coli* population could change.

2.3 Models

All parameters and variables for the Harisine-Rose and Gao models are described in Table 2.1 and Table 2.2.

2.3.1 Hairsine-Rose model

We used a version of the Hairsine-Rose model (Hairsine and Rose, 1991) developed by Heilig et al. (2001) that makes the simplifying assumption that light weight particles like clay do not settle out of the runoff.

With these assumptions, the first governing equation of the Harisine-Rose model, which describes the concentration of each particle class (c_i) in the ponded water, can be expressed as:

$$\frac{dc_i(t)}{dt} = \frac{1}{d_w}\left[\frac{ap}{I_t}(1 - H(t)) - pc_i(t)\right] \quad (2\text{-}1)$$

where $c_i(t)$ (g/ml) is the concentration of *E. coli* or each class of clay in the suspension at time t, d_w (cm) is ponding water depth, a (g/ml) is the soil detachability, p (cm/min) is rainfall intensity,

I_t is the total number of equal-mass particle classes, and $H(t)$ is the fraction of soil protected by shield layer at time t. The first term in the square brackets represents the rate at which i-class particles are ejected from the soil and the second term represents the rate at which they wash away in the overland flow. Because the rainfall-runoff is at steady state, the runoff rate is equal to the rainfall intensity, p.

Because we quantify *E. coli* as CFUs, we normalize the particle classes, I_t, as:

$$I_t = 1 + n + m \tag{2-2}$$

where n and m are number of mass classes of clay and sand, respectively, acquired by normalizing the initial total mass of clay and sand, respectively, to the initial total mass of *E. coli* in the soil column; note $m = 9n$ in our experiments. Following Hairsine and Rose (1991), it was assumed that every parcel of soil ejected into the runoff consisted of the same ratio of sand, clay, and *E. coli* as the initial mixture. Therefore:

$$\frac{1}{M_e(t)} = \frac{n}{M_c(t)} = \frac{m}{M_s(t)} \tag{2-3}$$

where $M_e(t)$ and $M_c(t)$ are the cumulative masses of ejected *E. coli* and clay per unit area at time t, respectively, and $M_s(t)$ the cumulative mass of ejected and re-deposited sand per unit area at time t (mass units are g/cm^2). To reconcile units for *E. coli*, we invoke an effective *E. coli* mass per CFU (M_{cfu}) to convert CFUs per unit area ($N_e(t)$) to $M_e(t)$:

$$M_e(t) = N_e(t)M_{cfu} \tag{2-4}$$

The $H(t)$ term in Eq. 2-1 is given by Sander et al. (1996) as:

$$H(t) = \frac{M_s(t)}{M_s^*} \tag{2-5}$$

where M_s^* (g/cm^2) is the mass of shield layer (sand) per unit area at complete shielding, i.e., when raindrop impact is prevented by the shield layer from ejecting underlying particles. Because we know the initial sand to clay ratio (9:1) and we run our experiments until no additional clay is eroding:

$$M_s^* = 9\int_0^T pc_c(t)dt \tag{2-6}$$

where subscripts *c* and *s* refer to clay and sand, respectively, and T is the time at which all the erodible clay has washed out of the experiment. Analogous to M_s^*, we define M_e^* (g/cm^2) and M_c^* (g/cm^2) to be the mass of total ejected *E. coli* and clay per unit area, respectively, when the shield layer is fully developed.

The second governing equation of the Hairsine-Rose model makes use of the observation that $\frac{dM_s(t)}{dt}$ is proportional to the rate at which clay and *E. coli* are ejected from the soil (Eq. 2-1). Therefore, $\frac{dM_s(t)}{dt}$ can be expressed as:

$$\frac{dM_s(t)}{dt} = \frac{m}{l_t} ap[1 - H(t)] \tag{2-7}$$

Inserting Eq. 2-5 into Eq. 2-7 and solving with the initial condition $M_s(0) = 0$ yields:

$$H(t) = 1 - \exp(-\frac{ma}{l_t M_s^*} pt) \tag{2-8}$$

Inserting Eq. 2-8 into Eq. 2-1 and solving it with the initial condition $c_i(0) = 0$ yields:

$$c_i(t) = \frac{\frac{a}{l_t d_w}}{\frac{1}{d_w} - \frac{ma}{l_t M_s^*}} \exp\left(-\frac{p}{d_w} t\right) \left\{\exp\left[\left(\frac{1}{d_w} - \frac{ma}{l_t M_s^*}\right) pt\right] - 1\right\} \tag{2-9}$$

Applying Eqs. 2-3 to Eq. 2-9:

$$c_c(t) = n \frac{\frac{a}{l_t d_w}}{\frac{1}{d_w} - \frac{na}{l_t M_c^*}} \exp\left(-\frac{p}{d_w} t\right) \left\{\exp\left[\left(\frac{1}{d_w} - \frac{na}{l_t M_c^*}\right) pt\right] - 1\right\} \tag{2-10}$$

$$c_e(t) = \frac{M_e^*}{M_c^*} n \frac{\frac{a}{l_t d_w}}{\frac{1}{d_w} - \frac{na}{l_t M_c^*}} \exp\left(-\frac{p}{d_w} t\right) \left\{\exp\left[\left(\frac{1}{d_w} - \frac{na}{l_t M_c^*}\right) pt\right] - 1\right\} \tag{2-11}$$

Then, using Eq. 2-4 to replace M_e^* in Eq. 2-11, we arrive at the final form of our simplified Hairsine-Rose model for *E. coli*:

$$c_e(t) = n \frac{N_e^* \times M_{cfu}}{M_c^*} \frac{\frac{a}{l_t d_w}}{\frac{1}{d_w} - \frac{na}{l_t M_c^*}} \exp\left(-\frac{p}{d_w} t\right) \left\{\exp\left[\left(\frac{1}{d_w} - \frac{na}{l_t M_c^*}\right) pt\right] - 1\right\} \tag{2-12}$$

Table 2.1 Summary of parameters and the ways they were determined.

Notation	Definition (Unit)	Value				
		run 1	run 2	run 3	run 4	run 5
a	Soil detachability [c] (g/ml)	4.500	0.350	0.800	0.450	0.450
C_o	Initial concentration of *E. coli* in soil [c] ($\times 10^6$ CFU/ml)	2.29	7.05	13.4	3.20	3.17
d_e	Exchange layer depth [c] (cm)	0.294	0.175	0.085	0.180	0.126
d_w	Ponding water depth [a] (cm)	0.825	0.800	0.900	0.950	0.950
I_t	Total number of normalized classes (when 1 colony=1 cell)	2450891	798451	420811	1757421	1764711
I_t	Total number of normalized classes (when 1 colony=1000 cell)	2451	801	421	1761	1761
K_p	Partition coefficient for dissolved and adsorbed *E. coli* (ml/g)	0	0	0	0	0
m	Number of classes representing sand (when 1 colony=1 cell)	2205801	718605	378729	1581678	1588239
m	Number of classes representing sand, m (when 1 colony=1000 cell)	2205	720	378	1584	1584
M_c^*	Eroded clay per unit area [b] (g/cm^2)	0.0453	0.0270	0.0131	0.0277	0.0194
n	Number of classes representing clay (when 1 colony=1 cell)	245089	79845	42081	175742	176471
n	Number of classes representing clay (when 1 colony=1000 cell)	245	80	42	176	176
N_e^*	Eroded *E. coli* per unit area [b] ($\times 10^5$ CFU/cm^2)	1.947	3.563	3.277	1.660	1.160
N_{eo}	Initial concentration of *E. coli* in soil water [a] ($\times 10^6$ CFU/ml)	1.58	3.40	3.85	1.03	3.15
p	Rainfall intensity [a] (cm/min)	0.28	0.276	0.26	0.26	0.24
θ	Soil water content by volume at saturation [a]	0.288	0.288	0.288	0.288	0.288
ρ_b	Bulk density of the soil [a] (g/cm^3)	1.543	1.543	1.543	1.543	1.543

[a] directly measured, see section 2.3 for details
[b] calculated from directly measured values, explained in section 2.2
[c] curve fitted, elaborated in section 2.4

Table 2.2 Summary of variables and complex parameters.

Notation	Definition (Unit)
$c_c(t)$	Concentration of clay in the suspension at time t (g/ml)

$c_e(t)$	Estimated concentration of *E. coli* in runoff at time *t* (g/ml)
C_e	Concentration of *E. coli* in exchange layer pore water (CFU/ml)
$c_i(t)$	Concentration of *E. coli* or each class of clay in the suspension at time *t* (g/ml)
C_o	Initial concentration of *E. coli* in soil water (both underlying soil and exchange layer) (CFU/ml)
C_w	Concentration of *E. coli* in runoff (CFU/ml)
e_r	Shown in Eq. 2-15
$H(t)$	Fraction of soil protected by shield layer at time *t*
$M_c(t)$	Cumulative masses of ejected clay per unit area at time *t*
M_{cfu}	Convert coefficient from CFU/ml to g/ml (g/CFU)
M_e^*	Maximum eroded *E. coli* per unit area (g/cm^2)
$M_e(t)$	Cumulative masses of ejected *E. coli* per unit area at time *t*
$M_s(t)$	Cumulative masses of ejected sand per unit area at time *t*, also equal to the mass of sand deposited on the surface per unit area at time *t*
M_s^*	Mass deposited sand (shield layer) at which no additional underlying soil can be eroded (g/cm^2)
$N_e(t)$	The counterpart of $M_e(t)$ with the unit of CFU/cm^2
t	Time (min)
T	The time at which all the erodible clay has washed out of the experiment (min)
z	Vertical axis, as shown in Figure 2.1
α	Shown in Eq. 2-15

2.3.2 Gao solute model

To model *E. coli* as a solute, we adopt the Gao solute model (Gao et al., 2004) and make the simplifying assumption of no diffusion from below the "exchange layer" (i.e., no change in concentration in the soil water below the exchange layer). Note that the exchange layer has the same depth as the final shield layer in the Heilig et al. (2001) version of the Hairsine-Rose (1991) model. The Gao et al. (2004) model is conceptually similar to the Hairsine-Rose (1991) model but instead of keeping track of the masses of particles ejected from or deposited on the soil surface, this model performs a mass balance of the solutes ejected from the mixing layer into the ponded water (the layers of the Gao model are labeled in Figure 2.1). All concentrations are in units of CFU/ml.

Exchange layer: $$\alpha d_e \frac{dC_e}{dt} = -e_r C_e \quad (2\text{-}13)$$

Ponded water: $$d_w \frac{\mathrm{d}C_w}{\mathrm{dt}} = e_r C_e - pC_w \quad (2\text{-}14)$$

where $$e_r = \frac{ap\theta}{\rho_b}, \alpha = \rho_b K_p + \theta \quad (2\text{-}15)$$

Initial conditions: $$C_e = C_o, C_w = 0 \quad (2\text{-}16)$$

where C_e and C_w are concentrations of *E. coli* in exchange layer pore water and runoff (a.k.a., ponded water), respectively; C_o is the initial concentration of *E. coli* in soil (both underlying soil and exchange layer); d_e (cm) is exchange layer depth; t (min) is time; p (cm/min) rainfall intensity; d_w (cm) is ponding water depth; a (g/cm^3) soil detachability; θ is volumetric soil water content at saturation; ρ_b (g/cm^3) is bulk density of the soil; and K_p (ml/g) is partition coefficient for adsorbed and "dissolved" (i.e., not adsorbed to sand) *E. coli*; here K_p is assumed to be 0, which means no *E. coli* are adsorbed to sand. We assume all pore water ejected from the exchange layer is replaced with clean rainwater (see Gao et al. 2004 for a full analysis and justification of this assumption).

The above equations can be solved analytically for C_e and C_w (details are shown in Appendix A) to reach the final form of the simplified Gao solute model:

$$C_e = C_o \exp\left(-\frac{ap}{\rho_b d_e} t\right) \quad (2\text{-}17)$$

$$C_w = C_o \frac{ap\theta}{\rho_b d_w\left(\frac{p}{d_w} - \frac{ap}{\rho_b d_e}\right)}\left\{\exp\left[\left(\frac{p}{d_w} - \frac{ap}{\rho_b d_e}\right)t\right] - 1\right\}\exp\left(-\frac{p}{d_w}t\right) \quad (2\text{-}18)$$

2.4 Model application

We applied Eq. 2-10 to the observed clay concentrations by fitting *a*. All other parameters were either directly measured or calculated from measured parameters (Table 2.1). We found that the d_e calculated by $d_e = 10M_c^*/\rho_b$ worked better than trying to directly measure the exchange layer depth; indeed this is analogous to how we determine M_s^*. Also fitting C_o so that the cumulative modeled *E. coli* lost equaled the experimental data worked better than trying to use the directly measured C_o (results not shown). We believe these are inherently sensitive parameters to measure, especially C_o which is based on a plating method that is known to be somewhat variable (Hedges, 2002; Oliver et al., 2005).

It is not clear how to determine M_{cfu} but if we express our results normalized to C_o which is expressed as the product of the initial CFUs per unit area (N_{eo}) and conversion factor, M_{cfu}, then this factor cancels out. However, *n* and *m* are related to M_{cfu} by Eqs. 2-3 and 2-4. As long as n is much larger than one, and, therefore, *m* = 9n is also much larger than one, $n{:}\, m{:}\, I_t = 1{:}\, 9{:}\, 10$ holds and the actual value of M_{cfu} does not really matter. Consider two cases, one in which each cell in the sample grows into one colony and another where every 1000 cells in the sample grows into one colony. If one colony equals one cell, then according to Neidhardt et al. (1990), $M_{cfu} = 9.5\times10^{-}$

^{13} g/cell × 1 cell/CFU = 9.5×10^{-13} g/CFU. If one colony equals 1000 cells, then M_{cfu} = 9.5×10^{-13} g/cell × 1000 cell/CFU = 9.5×10^{-10} g/CFU. Then, by Eqs. 2-3 and 2-4 we can get *n*, *m*, and I_t (Table 2.1). We can see that in either of the extreme cases, $n{:}\,m{:}\,I_t = 1{:}\,9{:}\,10$ is true. And so will it be true in any case within the range of one colony forming unit equals 1 to 1000 cells.

2.5 Results

The clay data were well behaved (Figure 2.2), with the possible exception of the first run (Figure 2.2a). As has been shown in similar experiments, the Hairsine-Rose model captures the dynamics of the erosion process very well (e.g., Heilig et al. 2001, Gao et al. 2005).

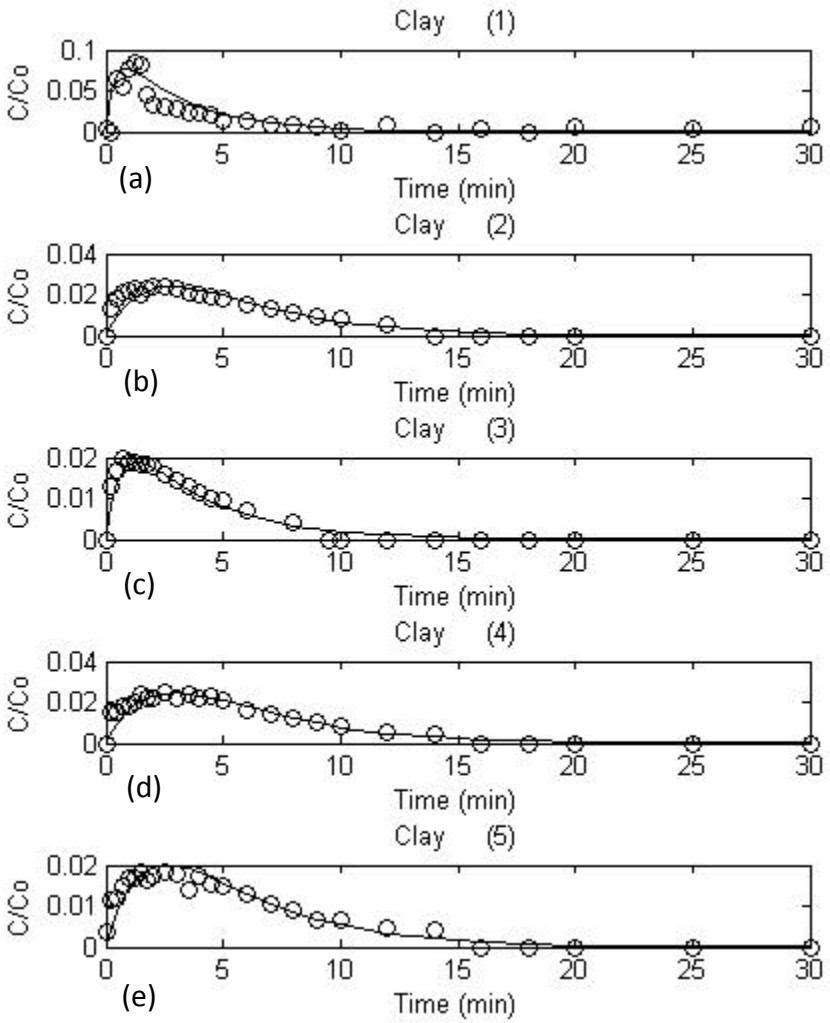


Figure 2.2 Measured relative clay concentrations (circles) fitted by Hairsine-Rose model (lines). Each graph is for a separate experiment.

As expected, the *E. coli* data were noisier than the clay data, in part because of the inherent variability in plate counts (Hedges, 2002; Oliver et al., 2005). However, despite the relatively noisy data, both the Hairsine-Rose erosion model and the Gao solute model work well for simulating *E. coli* concentrations in these experiments (Figure 2.3).

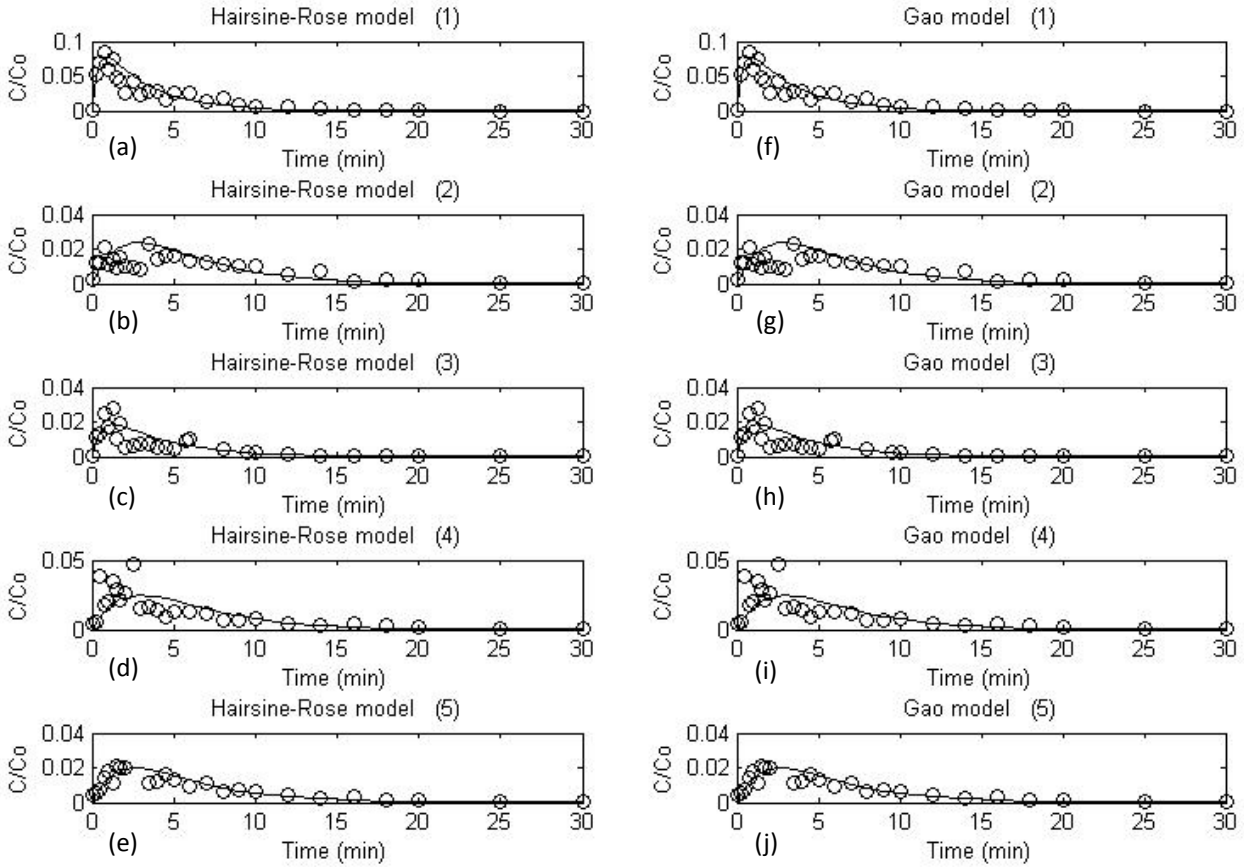


Figure 2.3 Measured relative *E. coli* concentrations (circles) fitted by Hairsine-Rose model (left column) and Gao model (right column). The experiments are in the same order as in Figure 2.2.

2.6 Discussion

Comparing clay (Figure 2.2) and *E. coli* (Figure 2.3), we can see that they showed similar overall trends and concentrations peaked at about the same times. Thus, it makes sense that the erosion model (Hairsine-Rose) worked well because we are essentially estimating the concentration of *E.*

coli by scaling the corresponding estimated clay concentration with total eroded mass or number-of-cells per unit area (M_c^* and N_e^*). Indeed, one wonders if the *E. coli* are simply attached to the clay, which explains why the patterns are so similar (the authors are currently addressing this in a separate study). One nice thing about this model was that only one parameter needed to be calibrated; the main problematic aspect was how to interpret CFUs in the context of a particle mass.

The Gao solute model also described our observations well. The advantage of this model is that units of concentration are preserved. This might also suggest that *E. coli* can be considered as an effective "solute" in this transport process. Indeed, we assumed there was no attachment of *E. coli* to sand by assuming $K_p = 0$. Note, our plating method did not distinguish freely suspended *E. coli* from those attached to clay. The unfortunate issue with this model was that we were not able to make a good direct measure of C_o, which introduced a parameter determined *a posteriori*.

The Gao solute model and the Hairsine-Rose erosion model generated essentially the same results for *E. coli*, although from different points of view. Conceptually, it may not be too surprising that these models gave such similar results because both invoke an exponentially decaying mass of *E. coli* in the top layer of the soil and assume they are being ejected proportionally to the rain intensity. Comparing Eq. 2-8 with Eq. 2-17 we see:

$$1 - H(t) \approx \frac{C_e}{C_o} \tag{2-19}$$

$$\exp\left(-\frac{ma}{I_t M_s^*} pt\right) \approx \exp\left(-\frac{ap}{\rho_b d_e} t\right) \tag{2-20}$$

Note that $\rho_b d_e$ is the mass of dry soil (1 part clay and 9 parts sand) in exchange layer per unit area and M_s^* is the mass of sand in exchange layer per unit area, so $\frac{I_t M_S^*}{m}$ is also the mass of dry soil (1 part clay and 9 parts sand) in exchange layer per unit area, when $n{:}\, m{:}\, I_t = 1{:}\, 9{:}\, 10$. Accounting for this, it can be shown that Eqs. 2-1 and 2-7, the governing equations for Heilig's version of the Hairsine-Rose model, are the same as Eqs. 2-13 and 2-14, the governing equations for the Gao model. Briefly, from Eq. 2-5 (of the Hairsine-Rose model) and Eq. 2-19 we can see that $\frac{M_S(t)}{M_S^*} = 1 - \frac{C_e}{C_o}$, so Eq. 2-7 can be expressed as:

$$\frac{dM_S^*(1-\frac{C_e}{C_o})}{dt} = \frac{m}{I_t} ap \cdot \frac{C_e}{C_o} \tag{2-21}$$

which can be rearranged:

$$\frac{dC_e}{dt} = -\frac{m}{I_t M_S^*} ap \cdot C_e \tag{2-22}$$

Since $\rho_b d_e = \frac{I_t M_S^*}{m}$, Eq. 2-22 is the same as Eq. 2-13 from the Gao model. Similarly, recognizing that $c_e(t) = M_{cfu} C_w$ we can rewrite the other governing equation from the Hairsine-Rose model, Eq. 2-1, as:

$$\frac{dM_{cfu}C_w}{dt} = \frac{1}{d_w}\left[\frac{ap}{I_t} \cdot \frac{C_e}{C_o} - pM_{cfu}C_w\right] \tag{2-23}$$

By dimensional analysis $\frac{1}{I_t} = \frac{\theta}{\rho_b} \cdot C_o M_{cfu}$, so Eq. 2-23 can be rewritten as Eq. 2-14, the governing equation in the Gao model for the mass of *E. coli* in the ponded water. So, the two models used in this study are identical (the corresponding items are shown in Table 2.3) when applied to our

experiments. To confirm this, we applied the Gao model to clay data and it also worked well (see Appendix A Figure A1).

Table 2.3 Corresponding items in Hairsine-Rose model and Gao model.

Hairsine-Rose model	Gao model
$c_e(t)/M_{cfu}$	C_w
$H(t) = \frac{M_s(t)}{M_s^*}$	$1 - \frac{C_e}{C_o}$
$\frac{I_t M_s^*}{m}$	$\rho_b d_e$
$\frac{1}{I_t M_{cfu}}$	$\frac{\theta}{\rho_b} \cdot C_o$

So, is *E. coli* a solute or a particle? For the sake to the transport processes we considered here, *E. coli* shows properties of a unique solute, most notably it does not appear to exhibit the diffusion properties that Gao et al. (2004, 2005) and Walter et al. (2007) observed for true solutes in similar experiments. Indeed, *E. coli* are much larger than ions, and a number of previous studies have shown that *E. coli* do not diffuse like true solutes (Oliver et al., 2005). So, *E. coli* cannot diffuse freely in the soil like typical solutes. Perhaps we should coin the concept of microbial solute-particle duality. In other words, a solute that does not diffuse is essentially a "particle" and a particle that does not settle out of suspension is essentially a "solute."

One difference between our experiments and similar previously published experiments (Gao et al., 2004, 2005; Heilig et al., 2001) is that our soil detachability parameter, a, varied for the same type of soil (Table 2.1). We believe this was due to the fact that we had to prepare the soil columns very gently and quickly so that the *E. coli* concentrations did not change due to growth or die-off; although in the end we calculated C_o rather than using the directly measured value because the

population of *E. coli* in that initial sample appeared to have changed, i.e., our directly measured initial concentrations (data not shown) were uncorrelated with what we calculated, in some cases higher and in others lower. So, we believe we had variable density soils from experiment to experiment.

In the Gao model, we used calculated d_e, which was smaller than the measured d_e at the side of the column. Upon closer inspection, we found that the depth of deposited sand at the side of the column was systematically deeper than in the middle. We speculate that a sand grain deposited near the wall may be less likely to be re-ejected by a raindrop, thus there is a systematic accumulation in this part of the column.

The concentrations of clay and *E. coli* from clay-sand mixture all followed the Heilig et al. (2001) version of the Hairsine-Rose model, which is based on the assumption that small particles, once ejected from the soil, remain in suspension for the duration of the rain event. The agreement between the modeled and measured *E. coli* suggests that the bacteria, specifically *E. coli* ATCC 25922, were generally not attached to sand. If they were, we would expect the model to over-predict *E. coli* concentrations as some *E. coli* would be sequestered in the shield layer. Similarly, the fact that the Gao solute model worked with $K_p = 0$ indicates that almost none of the *E. coli* were attached to sand. This is consistent with Oliver et al. (2007) who found that only small fractions (2%) of *E. coli* attached to particles ≥31 μm diameter; our sand is 250-300μm.

We did not differentiate the freely suspended *E. coli* from those potentially attached to clay particles in runoff, because mathematically it makes no difference. *E. coli*, roughly a 2 μm by 1

μm rod (Neidhardt et al., 1990), was of similar size as clay particles, and both *E. coli* and clay are in the size range of colloidal particles, from 1nm to 1μm (Buddemeier and Hunt, 1988; McGechan and Lewis, 2002a). This means that each *E. coli* cell transports as a colloidal particle whether it adsorbed to one or more clay particles or not. Oliver et al. (2005) found that the presence of clay resulted in "colloid facilitated transport" of *E. coli* but we did not see any solid evidence in our experiments to either support or contradict this. Our results are more similar to those of Muirhead et al. (2005), who found that less than 10% of *E. coli* was attached to solids in experiments involving rainfall on cowpats. We compared our results to Ferguson et al. (2007) and, qualitatively, our data look similar to theirs for *E. coli* from bare soil. A figure of our data in a format similar to Figure 2 (control part) from Ferguson et al. (2007) is included in Appendix A (Figure A2).

It is obvious that the relative concentration of *E. coli* showed more scatter than that of clay (Figure 2.2 and Figure 2.3). This is because the method used to measure concentration of *E. coli* was less precise than that used to measure concentration of clay. For clay, each runoff sample was usually diluted by a 1:5 or 1:10 ratio; whereas, for *E. coli*, the runoff samples were usually diluted by 1:20 and then a 20 μl subsample from the 1 ml diluted sample was inoculated on a plate. This means that the error of concentration of clay measured by the spectrometer was amplified by 5 or 10 times in the final results, while the error of the concentration of *E. coli* measured by counting the colony-forming units in the plates was amplified by $20 \times 50 = 1000$ times. And the dilution ratio had to be adjusted frequently from experiment to experiment as well as sample to sample in order to keep the CFU counts within method-approved ranges (Oliver et al., 2007). Indeed, the plate counting is especially imprecise when trying to relate it to a mass of *E. coli* (Hedges, 2002; Oliver et al., 2005).

Our experimental run-times were very short (30 min) in order to neglect considering biological processes, such as growth and microaggregates forming. Future research should consider incorporating these. We are also currently working on a project to assess the role of clay in facilitating or inhibiting *E. coli* transport from soil into runoff.

2.7 Conclusions

We found that the transport of *E. coli* from soil into overland flow is equally well described by Gao solute transport model and Hairsine-Rose soil erosion model. Thus we proposed the microbial solute-particle duality, in which microbes can be modeled as non-diffusing "solutes." We were unable, with these experiments, to determine if *E. coli* transport was facilitated by eroding clay. But our results suggest that most of the *E. coli* were not attached to sand particles. Our results suggest that management practices that reduce raindrop impact are most likely to minimize microbial transfer from soil into storm runoff, e.g., maintaining a vegetative or residue cover.

Future studies on microbial transport from soil into overland flow are recommended to (1) do the same experiment with no or little clay in the soil, (2) use timely and non-destructive methods like magnetic resonance imaging, (3) find better ways to separate freely suspended microbes and those attached to very fine particles like clay, (4) investigate the influence of the interactions between microbes and between microbes and soil constituents, and (5) investigate other microbes and compare to our findings of *E. coli* ATCC 25922.

CHAPTER 3 TRANSPORT OF *ESCHERICHIA COLI* UNDER RAINDROP IMPACT: THE ROLE OF CLAY[2]

Abstract

Many researchers have investigated colloid-facilitated contaminant transport. However, the co-transport of mineral colloids and bacteria has not received much attention. We used simulated rainfall over saturated soil columns infused with *Escherichia coli* (*E. coli*) to examine the transport of *E. coli* into overland flow under raindrop impact. One set of columns consisted of sand and the other consisted of a 9:1 sand-clay mixture. Previous research showed that the non-diffusive form of the Gao solute model simulated colloid transfer between soil and overland flow, and it worked well for these experiments as well. We conducted ANOVA analysis for the two types of soil, and we also applied the Gao model using average experimental soil parameters for the sand and sand-clay columns, respectively, and average overall initial *E. coli* concentration from all experiments. Both of the ANOVA analysis and the average modeled curve showed that, there was substantially more *E. coli* transferred from the sand column than from the sand-clay, because the raindrops penetrated deeper into the soil-media and the proportion of soil water in the total ejected mass is larger. The primary role of clay appears to be changing the soil texture, making the soil harder to penetrate and containing less soil water, which reduces *E. coli* transport into overland flow. Thus, soil texture plays an important role in the co-transport of mineral colloid and bacteria from soil into overland flow under raindrop impact.

[2] Wang, C., Parlange, J.-Y., Schneider, R.L., Rasmussen, E.W., Wang, X., Chen, M., Walter, M.T., 2015. Transport of *Escherichia coli* under raindrop impact: The role of clay. J. Hydrol. (submitted for publication).

3.1 Introduction

Both bacteria and clay particles fall into the colloid size range of 1 nm to 10 □m (Chrysikopoulos and Sim, 1996; Vasiliadou and Chrysikopoulos, 2011b). Bacteria are commonly considered to be biocolloids (Kanti Sen and Khilar, 2006; Vasiliadou and Chrysikopoulos, 2011a).

Many researchers have studied colloid-facilitated contaminant transport (Abdel-Salam and Chrysikopoulos, 1995; Chen et al., 2005; Corapcioglu and Jiang, 1993; Grolimund and Borkovec, 2005; Ibaraki and Sudicky, 1995; Ouyang et al., 1996; Roy and Dzombak, 1997; Šimůnek et al., 2006; Smith and Degueldre, 1993) including bacteria-facilitated contaminant transport (Bekhit et al., 2009; Guiné et al., 2003; Kim and Corapcioglu, 2002a, b; Kim et al., 2003; Pang et al., 2005; Pang and Šimůnek, 2006). However there are few empirical studies of the transport of biocolloids, particularly bacteria, through association with mineral colloids.

Two separate studies have investigated mineral-colloid-associated bacteria transport using soil columns. Vasiliadou and Chrysikopoulos (2011a) found that the kaolinite colloids inhibited the transport of *Pseudomonas putida* (*P. putida*), a rod-shaped bacteria of similar size as *E. coli*, because *P. putida* attached to kaolinite and kaolinite stayed attached to the solid matrix. Working with both *Salmonella typhimurium* and *E. coli* O157:H7, Chen (2012) found that mineral colloids either facilitated or retarded the transport of these bacteria in soil, depending on whether the mineral colloids were mobile or immobile, respectively.

An additional two experiments have investigated overland flow related mineral-colloid-associated transport of *E. coli*. Muirhead et al. (2006b) looked at the cotransport of *E. coli* and bromide in overland flow across saturated soil and concluded that they have similar transport mechanisms. It was also found that *E. coli* mainly attached to mineral particles smaller than 2 µm and, once mobilized, remained in suspension. In Chapter 2, we used lab experiments and the Gao et al. (2004) solute model to study mineral colloid and bacteria co-transport from soil into overland flow under raindrop impact (splash erosion). They drew similar conclusion to those of Muirhead et al. (2006b), noting that a particle that does not settle out of the water can be treated as a solute and a solute that does not diffuse is essentially a particle. But in Chapter 2 we were unable to specifically identify the role of clay in the transport of *E. coli*. It was speculated that whether the *E. coli* were attached to the clay or not was not relevant to the transport processes, i.e., whether attached or not, both were colloidal. The goal of this study was to compare the transport of *E. coli* from soil into runoff when clay is and is not part of the soil matrix.

3.2 Experimental design

We used the same experimental set-up as in Chapter 2 (Figure 3.1) and the same analytical procedures. These methods are described in brief in the section below. The only difference here is that we use a second soil media composed of pure sand (250-300 µm sand) in addition to the 9:1 sand-clay mixture (250-300 µm sand, kaolinite clay).

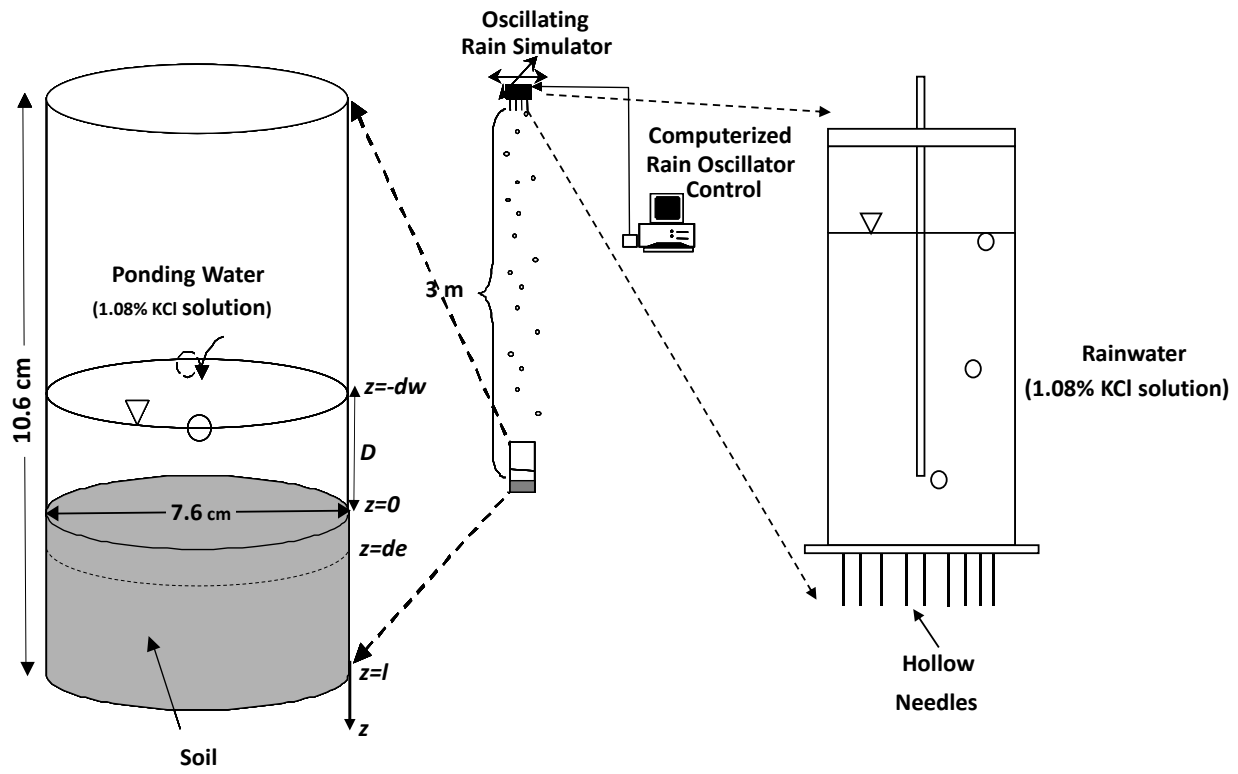


Figure 3.1 Experimental set-up, same as Chapter 2, adapted from Gao et al. (2004, 2005).

E. coli ATCC 25922, a popular nonpathogenic surrogate of pathogenic *E. coli* O157:H7 (Muirhead et al., 2006b; Salleh-Mack and Roberts, 2007; Sauer and Moraru, 2009), was grown in Tryptic Soy Broth (TSB) for 18 h at 37 C. Two milliliter of this culture was then mixed with 80 mL of 1.08% potassium chloride (KCl) solution and added to 225 g of soil.

The pre-saturated soil was packed into 7.6-cm-diameter plexiglass columns using a shaking table. A 0.5 ml sample was taken from the solution that ponded on the surface of the column during this procedure, before it was poured off. The columns were then placed under the rainmaker and *E. coli*-free KCl solution was gently added to pre-pond the experiment, so our steady-state runoff

assumption would be valid at $t = 0$. A 0.5 ml sample was extracted from this pre-ponded water to determine the initial *E. coli* concentration.

A Marriott bottle (Figure 3.1) was used to generate the rainfall. This container has the advantage of mostly seal the rainwater from the surrounding environment while maintaining a constant rain rate. The rainwater consisted of 1.08% KCl solution to avoid lysing of the cells. The study area was protected with an umbrella until the rainfall was steady. A timer was started when the umbrella was removed. Samples of 0.5 ml were taken from the runoff at varying intervals that transitioned at times 0, 2, 5, 10, and 20 min from the beginning of rainfall. The intervals used within each time period were 15 s, 30 s, 1 min, 2 min and 5 min, respectively. Rainfall lasted for a total of 30 minutes. In Chapter 2 we showed that the growth or die-off of *E. coli* during the 30 min rainfall experiment was negligible.

The concentration of bacteria in runoff samples was determined by a dilution and inoculation procedure. Sterilized 1.08% KCl solution was used to dilute each runoff sample. Samples were then plated on *Escherichia coli* media with 4-methylumbelliferyl-β-D-glucuronide (EC-MUG; Neogen Corporation, Lansing, MI) and incubated for 20 hours at 37^{o}C. CFUs were manually counted on each plate and converted to the bacteria concentration units (CFU/ml) in the original runoff sample. To verify that there were no unknown sources of *E. coli*, a control experimental run was done without adding *E. coli* to the soil.

A total of four experimental runs were completed. The rainfall intensity (p) and the ponding water depth (d_w) were measured before and after each experimental run and averaged to represent

constant values. The shield layer depth (d_e) was measured after each experimental run. The exchange layer for the clay-sand mixture was very distinct, because the underlying soil was lighter due to the white clay. Surprisingly, the exchange layer for the pure sand was visually distinguishable because a white precipitate formed in the soil below the exchange layer (see Chapter 2 for the figure), which allowed us to obtain the depth of the exchange layer. A subsequence analysis suggests that the white precipitate is mainly composed of *E. coli* cells. Although we attempted to keep conditions identical between experimental runs, ponding depth, rainfall intensity, and initial *E. coli* concentrations varied a little from run to run due to uncontrollable variability, especially initial *E. coli* concentrations.

3.3 Gao model

Chapter 2 showed that Gao solute model (Gao et al., 2004) simulates colloid transport between soil and runoff under rainfall impact if the diffusion is set to zero. The Gao et al. (2004) model performs a mass balance of the solutes ejected from the exchange layer (a.k.a. mixing layer) into the ponded water (Gao et al., 2004). The layers of the Gao model are shown in Figure 3.1. Assuming no diffusion of material from the underlying soil into the exchange layer, and, therefore, no alteration of the concentration of *E. coli* in the underlying soil, the Gao model for our experimental conditions can be written as:

Exchange layer: $$\theta d_e \frac{dC_e}{dt} = -e_r C_e \quad (3\text{-}1)$$

Ponded water: $$d_w \frac{\mathrm{d}C_w}{\mathrm{dt}} = e_r C_e - pC_w \quad (3\text{-}2)$$

where $$e_r = \frac{ap\theta}{\rho_b} \quad (3\text{-}3)$$

Initial conditions: $$C_e = C_o,\ C_w = 0 \quad (3\text{-}4)$$

where C_e and C_w are concentrations of *E. coli* (CFU/ml) in exchange layer pore water and runoff (i.e., ponded water), respectively; C_o is the initial concentration of *E. coli* in soil (CFU/mL); d_e (cm) is the exchange layer depth; t (min) is time; p (cm/min) is the rainfall intensity; d_w (cm) is the ponding water depth; a (g/cm^3) is the soil detachability; θ is the volumetric soil water content (saturated water content in our case); ρ_b (g/cm^3) is the bulk density of the soil. We assumed no *E. coli* is adsorbed to sand. Similar to the way Heilig et al. (2001) were able to generate an analytical solution to the Hairsine-Rose (1991) model for a simple experiment, we (Chapter 2) were able to analytically solve the Gao model for a similarly simple experiment:

$$C_e = C_o \exp\left(-\frac{ap}{\rho_b d_e} t\right) \quad (3\text{-}5)$$

$$C_w = C_o \frac{ap\theta}{\rho_b d_w \left(\frac{p}{d_w} - \frac{ap}{\rho_b d_e}\right)} \left\{\exp\left[\left(\frac{p}{d_w} - \frac{ap}{\rho_b d_e}\right) t\right] - 1\right\} \exp\left(-\frac{p}{d_w} t\right) \quad (3\text{-}6)$$

Similar to Chapter 2, the soil detachability, a, was used as a fitting parameter.

3.4 Results and discussion

Table 3.1 Summary of parameters and the ways they were determined (parameter values for clay-sand mixture were copied from Chapter 2)

Notation	Definition (Unit)	Value							
		Clay-Sand Mixture				Pure Sand			
		run 1	run 2	run 3	run 4	run 1	run 2	run 3	run 4
a	Soil detachability [c] (g/ml)	0.350	0.800	0.450	0.450	1.500	5.000	0.900	1.400
C_o	Initial concentration of *E. coli* in soil [c] ($\times 10^6$ CFU/ml)	7.05	13.4	3.20	3.17	1.42	2.73	1.41	9.56
d_e	Exchange layer (shield layer) depth [b] (cm)	0.175	0.085	0.180	0.126	0.656	0.501	0.532	0.413
d_w	Ponding water depth [a] (cm)	0.800	0.900	0.950	0.950	0.700	0.800	0.950	0.950
M_s^*	Weight of shield layer per unit area [a] (g/cm^2)	-	-	-	-	0.912	0.696	0.739	0.574
p	Rainfall intensity [a] (cm/min)	0.276	0.260	0.260	0.240	0.276	0.260	0.260	0.280
θ_s	Soil water content by volume at saturation [a]	0.288	0.288	0.288	0.288	0.416	0.416	0.416	0.416
ρ_b	Bulk density of the soil [a] (g/cm^3)	1.543	1.543	1.543	1.543	1.475	1.475	1.475	1.475

[a] directly measured, see section 3.2 for details
[b] calculated from directly measured values, explained in section 3.3
[c] curve fitted, elaborated in section 3.3

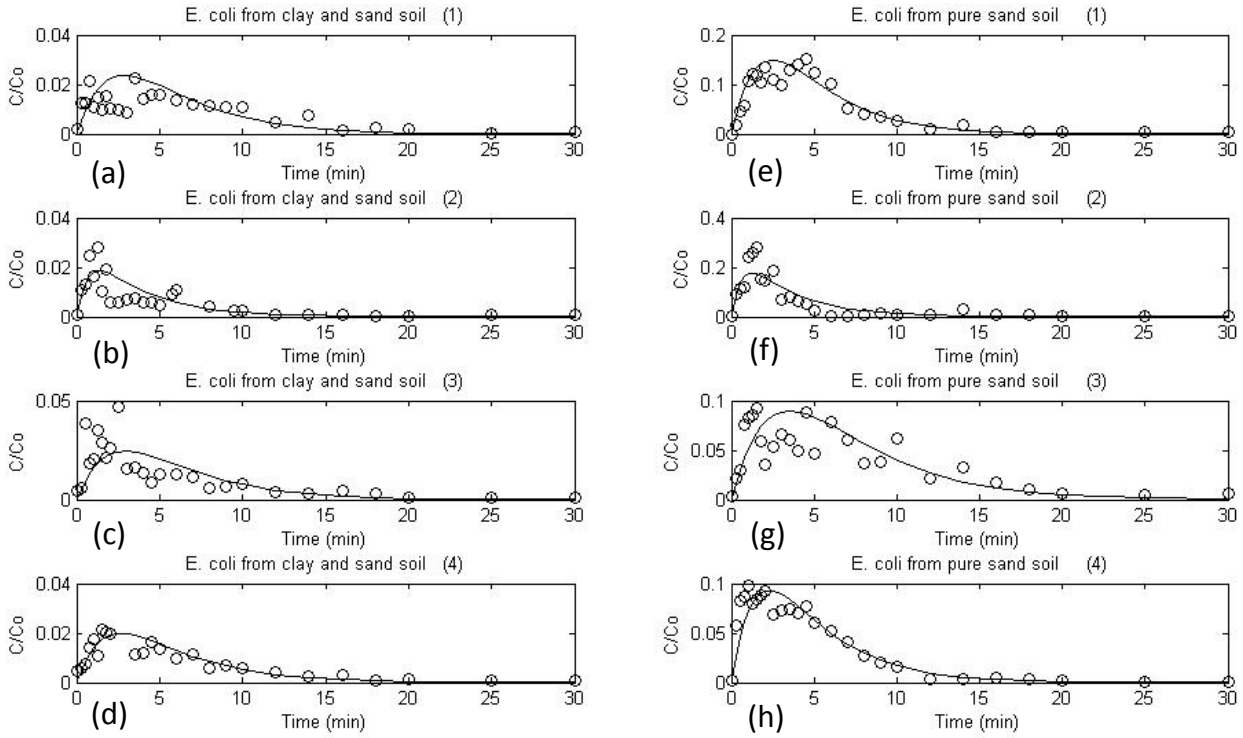


Figure 3.2 Gao solute model for relative concentration of *E. coli* from clay and sand soil (left column) and from pure sand soil (right column).

We initially applied the directly measured d_e and C_o values to the Gao model and got seemingly random and physically unrealistic results (see Appendix B Figure B2). We disassembled some of our soil columns and found that d_e for the sand-clay mixture was much thinner throughout the column than it was on the edge. Because we had measured the clay erosion as well as the *E. coli* concentrations, we were able to calculate d_e (see Chapter 2):

$$d_e = \frac{10 \int_0^\infty C_c dt}{\rho_b} \tag{3-7}$$

where C_c is the clay concentration in the runoff. The exchange layer depth for the sand only media appeared to be well represented by at the edge of the column measurements. The directly measured C_o were highly variable so we fitted C_o by forcing the integral under the model curve equal to the integral under the measured data for each experiment.

All the parameter definitions and values are listed in Table 3.1. The Gao solute model was able to capture the *E. coli* export from both the sand-clay and sand alone experiments (Figure 3.2). This supports our findings from Chapter 2 that the fit of Gao solute transport model (Gao et al., 2004) suggests that particles that don't settle or diffuse essentially behave as solutes.

Visual inspection of the plots suggests that *E. coli* in pure sand columns eroded more readily. We performed a series of ANOVA analysis to determine whether the characteristics of the clay/sand mix and the pure sand were significantly different. Compared to the clay/sand mix, the pure sand had a significantly larger penetration depth (d_e; $p = 0.0004$), peak amount of *E. coli* transport (via erosion; $p = 0.025$), total amount of *E. coli* transport between 0 and 10 min ($p = 0.014$) and total amount of *E. coli* transport between 0 and 30 min ($p = 0.0002$). However, the two soil types have similar soil detachability (a; $p = 0.125$), rainfall intensity (p; $p = 0.312$), ponding water depth (d_w; $p = 0.506$), peak time ($p = 0.907$), and integral from 0 to average peak time ($p = 0.663$). These results suggest that the soil without clay is easier to penetrate, which resulted in higher peak and total eroded amount of *E. coli*. In addition, pure sand soil has higher soil water content by volume at saturation (θ_s) and lower bulk density of the soil (ρ_b) than clay-sand soil (Table 3.1). Thus, with similar detachability and exposed to rainfall of similar intensity, more soil water and associated *E. coli* would be ejected from pure sand than the clay/sand mix.

Finally, we reran the Gao model using average soil properties for each of the soil media (sand-clay mixture vs. sand only) and applying the same initial *E. coli* concentration (Figure 3.3). From these results, it is obvious that clay does not facilitate *E. coli* transport. This is most likely due to the difference in the model parameters between the two soils, specifically, the reduced penetration depth, d_e, and reduced soil water content at saturation, θ_s, for the clay/sand mix compared to the pure sand.

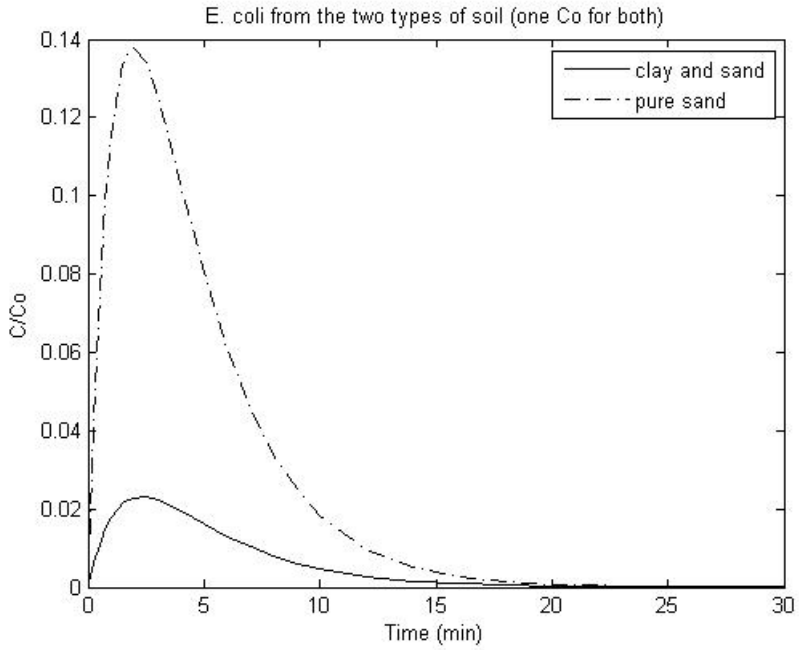


Figure 3.3 Compare the two soil conditions by model simulation: Gao model using average soil properties for each of the soil media (sand-clay mixture vs. sand only) and applying the same initial *E. coli* concentration.

3.5 Conclusion

We conclude based on a combination of empirical and modeling results that soils with increased clay content will transfer less bacteria into storm runoff compared to sandy soils, due to the role of clay in decreasing the effectiveness of raindrop impact at causing erosion.

CHAPTER 4 EXPLAINING AND MODELING THE CONCENTRATION AND LOADING OF *ESCHERICHIA COLI* IN A STREAM—A CASE STUDY

Abstract

Escherichia coli (*E. coli*) level in streams is important for public health. Therefore, being able to explain why *E. coli* level is sometimes high and sometimes low, and being able to predict the *E. coli* level in streams is of importance. We took Fall Creek as an example and found that complementarily using principle component analysis (PCA) and partial least squares regression (PLS) is a convenient and reliable way to find the drivers of *E.coli* and predict *E. coli* level. We found that runoff and erosion, microbial activity, shallow subsurface flow and groundwater seepage are the four processes driving the fate and transport of *E. coli* in a watershed. PLS model is very good for stormwater conditions (R^2=0.85 for log of *E. coli* concentration and R^2=0.90 for log of *E. coli* loading), while not very good for baseflow conditions. But in our case, both *E. coli* concentration and *E. coli* loading are higher in stormwater conditions, which means that we are good at predicting the hazardous events.

4.1 Introduction

Pathogen in surface waters is a serious international concern and *Escherichia coli* (*E. coli*) is a commonly used indicator organism for freshwater (Brooks et al., 2013; Dwivedi et al., 2013; Vidon et al., 2008). Climatic, hydrological, and water quality factors have all been shown to affect concentrations and/or loadings of *E. coli* in streams. Whitman et al. (2008) found that sunlight, season, snowmelt, and storms can influence *E. coli* in a stream. Vidon et al. (2008) stated that stream flow conditions (high flow or baseflow), discharge, precipitation (especially 7 day antecedent precipitation), turbidity, location (i.e. headwater or lower reaches) and temperature all

showed significant correlation with either concentrations or loadings of *E. coli* in streams. Dwivedi et al. (2013) investigated the correlations of 13 water quality factors with respect to *E. coli* loading and concluded that temperature, dissolved oxygen, phosphate, ammonia, suspended solids and chlorophyll are the most important ones. Diurnal variability (Meays et al., 2006) and carbon dioxide (Gray, 1975) have been found to be relevant as well.

Existing watershed models (Benham et al., 2006; Walker and Stedinger, 1999) and other conceptual models (Wilkinson et al., 1995) that predict *E. coli* loads in streams generally require too much site specific information to easily employ. Artificial neural network models can be good predictors of *E. coli* (Basant et al., 2010; Dwivedi et al., 2013), but often the underlying processes influencing *E. coli* transport are unclear. Statistical models offer some advantages to these other approaches in that they do not require full *a priori* understanding of the sources, sinks, transport processes, etc. Multiple linear regression (MLR) models (Nevers and Whitman, 2005) cannot avoid co-linearity, which is common in water quality data. Models like LOADEST estimate constituent loads utilizing only streamflow and time (Dwivedi et al., 2013; Runkel et al., 2004), however, as summarized above, *E. coli* levels depend not only on streamflow, but also on other, perhaps correlated, factors.

Partial least squares regression (PLS) is a powerful statistical tool for selecting the most important variables among many highly correlated predictor variables (Esbensen et al., 2002). Thus, PLS is widely used in water quality modeling (Aguilera et al., 2000; Basant et al., 2010; Ortiz-Estarelles et al., 2001a; Ortiz-Estarelles et al., 2001b; Singh et al., 2007). Carroll et al. (2009) built PLS

models using the data from possible sources to predict the data from the corresponding sinks to confirm or challenge possible source-sink relationships. Brooks et al. (2013) used PLS to predict fecal indicator bacteria on Great Lakes beaches using a large suite of possible predictors. Their predictors that might be relevant to streams included turbidity, air temperature, and antecedent rainfall (24h and 48h), and variables related to season (Julian date and month). Other variables were mostly relevant to beaches, like wave height, wind speed and wind direction. Like many statistical models, PLS is not necessarily reliable for revealing the underlying processes that control the *E. coli* levels (Brooks et al., 2013).

Principal component analysis (PCA) is a widely used method in classification. Many people used PCA for *E. coli* identification in food samples (Al-Holy et al., 2006; Al-Qadiri et al., 2006; de Sousa Marques et al., 2013; Siripatrawan et al., 2011), i.e. to differentiate *E. coli* from other microbes. A few researchers have used it to identify factors correlated to *E. coli* concentrations in field samples (Bech et al., 2014; Dwivedi et al., 2013). Dwivedi et al. (2013) used 6 variables in a PCA analysis, and found that PC1 corresponded mostly to physical factors, accounting for dissolved oxygen and temperature, and PC2 corresponded primarily to chemical and biological factors, accounting for phosphate and ammonia, suspended solids, and chlorophyll. Using PCA, Bech et al. (2014) found that concentrations of leached fecal bacteria were negatively correlated with the number of days after slurry-manure application, which was consistent with Falbo et al. (2013). However, they also found that leached fecal bacteria was negatively correlated with soil water content, which, they admitted, is contrary to other researcher's findings. Moreover, they

found that leached fecal bacteria was not correlated with precipitation, which is also contrary to previous findings.

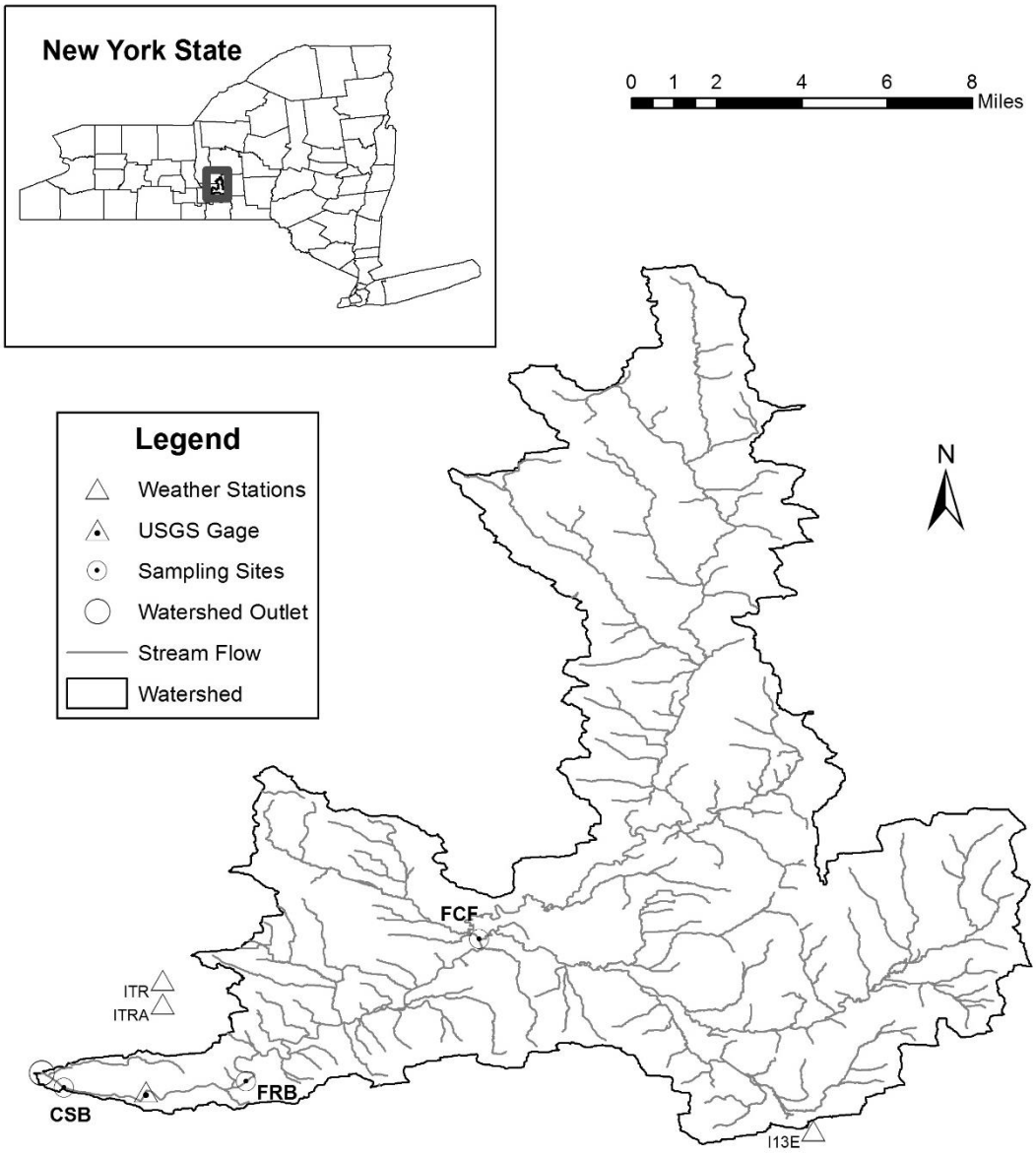
However, to our knowledge, previous studies have not used PCA and PLS together to understand and model concentrations or loadings of *E. coli* in streams. We, thus, tested the applicability of applying PCA and PLS to stream *E. coli* concentrations and loadings to both predict and explain mechanisms controlling stream *E. coli*.

4.2 Site description and data availability

Fall Creek watershed (Figure 4.1) is a 333.5 km^2 watershed in central upstate New York that drains into Cayuga Lake. We used water quality data accrued by the Community Science Institute (CSI, 2014). We picked three CSI monitoring sites with longest time series of *E. coli* data (11 years): Fall Creek Freeville (FCF), Freese Road Bridge (FRB), and Cayuga Street Bridge (CSB) (Figure 4.1). Fall Creek Freeville is 11.9 miles from stream mouth and located immediately upstream of the confluence of Fall Creek and Virgil Creek (Community Science Institute, 2014). Freese Road Bridge is 3.7 miles from stream mouth and located upstream of Cornell water treatment plant intake and downstream of eroding stream banks at Monkey run (Community Science Institute, 2014). While Cayuga Street Bridge is 0.7 miles from stream mouth, where is close enough to the mouth of watershed and has no lake water interference, and located downstream of urban area (Community Science Institute, 2014). A total of 13 water quality analytes, were selected from the Community Science Institute data base (2014), including: *E. coli* concentration (colonies/100 ml)

and total coliform concentration (colonies/100 ml), together with alkalinity (mg $CaCO_3$/L), chloride (mg/L), total hardness (mg $CaCO_3$/L), pH, total Kjeldahl nitrogen (mg-N/L), nitrate-+nitrite-nitrogen (as mg-N/L), soluble reactive phosphorus (µg-P/L), total phosphorus (µg-P/L), specific conductance (µS/cm), total suspended solids (mg/L), and turbidity (NTU). The CSI (2014) water quality samples were labeled either stormwater or baseflow, which provided convenience for our later analysis.

Fall Creek Watershed



Created by Chaozi Wang on 11/12/2014
Data from USGS, the Community Science Institute, and NOAA
Projected Coordinate System: NAD 1983 UTM Zone 18N

Figure 4.1 Map of the Fall Creek Watershed. The circles with a dot are the Community Science Institute sampling sites used in this analysis: Fall Creek Freeville (FCF), Freese Road Bridge (FRB), and Cayuga Street Bridge (CSB).

Table 4.1 Notation

Variable	Meaning (Unit)
7dP	seven day antecedent precipitation (mm)
airT	instantaneous air temperature (℃)
alkalinity	alkalinity (mg $CaCO_3$/L)
B	1 for baseflow, and 0 for stormwater
Cl	chloride concentration (mg/L)
CSB	1 for data collected at Cayuga Street Bridge (CSB), and 0 for data collected at the other two sites
discharge	instantaneous discharge scaled by drainage area (cfs)
Ecoli	*E. coli* concentration (colonies/100 mL)
Ecoliload	*E. coli* load (colonies/s), product of Ecoli and discharge
F	fall, Sep 1st to Nov 14th, default season option
FCF	Fall Creek Freeville sampling site, default site option
FRB	1 for data collected at Freese Road Bridge (FRB), and 0 for data collected at the other two sites
hardness	total hardness (mg CaCO3/L)
KNT	total Kjeldahl nitrogen (as N, mg/L)
NNN	nitrate-nitrite-nitrogen (as N, mg/L)
pH	pH
SC	specific conductance (µS/cm)
Season	1 for winter (W), 2 for spring (Sq), 3 for summer (Su), and 4 for fall (F)
TSS	total suspended solids (mg/L)
Sp	1 for data collected during spring (March 1st to May 31st), and 0 for data collected during other seasons
SRP	soluble reactive phosphorus (as P, µg/L)

Su	1 for data collected during summer (June 1st to Aug 31st), and 0 for data collected during other seasons
Tcoliform	total coliform (colonies/100 ml)
TP	total phosphorus (as P, µg/L)
turbidity	turbidity (NTU)
W	1 for data collected during winter (Nov 15th to Feb 28th or 29th), and 0 for data collected during other seasons

The instantaneous discharge at each of the three water sampling locations was calculated by scaling the instantaneous discharge at USGS gage (#04234000) (USGS, 2007, 2014) on Fall Creek with the drainage area corresponding to the three water quality monitoring sites (Figure 4.1).

Since about half of the water quality data points do not have water temperature, we used instantaneous air temperature from weather stations closest to the sampling sites (Figure 4.1): mostly from Ithaca 13 E (I13E) and from one of two weather stations associated with the Ithaca Tompkins Regional Airport (ITRA, ITR) when I13E was missing data (NOAA, 2014). The seven day antecedent precipitations were acquired by summing up all the precipitations in the seven days preceding each water quality sampling date.

We eliminated all data for days in which discharge or climate data were unavailable. The number of data points used were: $n = 32$ at FRB, $n = 39$ at CSB, and $n = 34$ at FCF. We also excluded three outliers of extremely high reported *E. coli* concentration, because for these three data points other water quality analytes were not extremely high or low. The total number of sample site-days was 102.

4.3 Methodology

All the variables we used and the meaning of each variable is listed in Table 4.1. We have 8 categorical variables: (1) B, the variable represents the category of baseflow and stormwater , 1 for baseflow and 0 for stormwater; (2) W, 1 for data collected during winter (Nov 15th to Feb 28th or 29th), and 0 for data collected during other seasons; (3) Sp, 1 for data collected during spring (March 1st to May 31st), and 0 for data collected during other seasons; (4) Su, 1 for data collected during summer (June 1st to Aug 31st), and 0 for data collected during other seasons; (5) F, the default option for season, i.e. when all of W, Sp and Su are 0, it means F; (6) FRB, 1 for data collected at Freese Road Bridge (FRB), and 0 for data collected at the other two sites; and (7) CSB, 1 for data collected at Cayuga Street Bridge (CSB), and 0 for data collected at the other two sites; (8) FCF, the default option for site, i.e. when both of FRB and CSB are 0, it means FCF. They are either 1 or 0, so Table 4.2 shows the counts of 1 for each of them.

The other variables are continuous variables, including alkalinity, chloride (Cl), *E. coli* concentration (Ecoli), total hardness, pH, total kjeldahl nitrogen (KNT), nitrate-nitrite-nitrogen (NNN), total suspended solids (TSS), soluble reactive phosphorus (SRP), specific conductance (SC), total coliform (Tcoli), total phosphorus (TP), turbidity, discharge, air temperature (airT), seven day antecedent precipitation (7dP), and *E. coli* loading (Ecoliload) (Table 4.1).

We performed a one-way ANOVA to see if there were any significantly different groups of data that should be treated separately in subsequent analyses. We also calculated Pearson correlation coefficients for comparisons between *E. coli* (concentrations and loadings) and other the variables in order to provide basic information about correlations within the dataset.

Since the water quality analytes, discharge and weather variables have different units, before PCA and PLS analysis, each variable was normalized to its mean and scaled by its standard deviation, i.e. autoscaled, according to Esbensen et al. (2002).

PLS was implemented in MATLAB to take discharge, weather factors, and water quality analytes other than total coliform as predictors to estimate concentration and loading of *E. coli*. PLS uses Y matrix (dependent variables) as a starting point to iteratively find the optimal principle components (PCs) to project X matrix (independent variables) and Y matrix interdependently (Esbensen et al., 2002).

PCA was implemented in MATLAB to identify independent driving processes of the patterns of *E. coli* observed in the stream, and, further, to provide better understanding of the variation of *E. coli* level in the stream together with PLS analysis.

4.4 Results

4.4.1 Basic information

We did one-way ANOVA for *E. coli* levels to see if it is necessary to first divide the data into subgroups and then did analysis. Stormwater flows (n=34) had significantly higher discharge (p=5.83×10^{-7}), *E. coli* concentrations (p=1.79×10^{-5}), and *E. coli* loading (p=9.48×10^{-5}) than baseflow (n=68) (Figure 4.2a-c). The *E. coli* concentrations at the three sites were not significantly different (n=31, 38 and 33; p=0.71) (Figure 4.2d). So, henceforth we differentiate stormwater data from baseflow data, but group the data from the three sites together. Note that for one-way ANOVA, the means used are arithmetic means, but for easier comparison with Vidon et al. (2008), we also calculated geometric mean of the variables (Table 4.2).

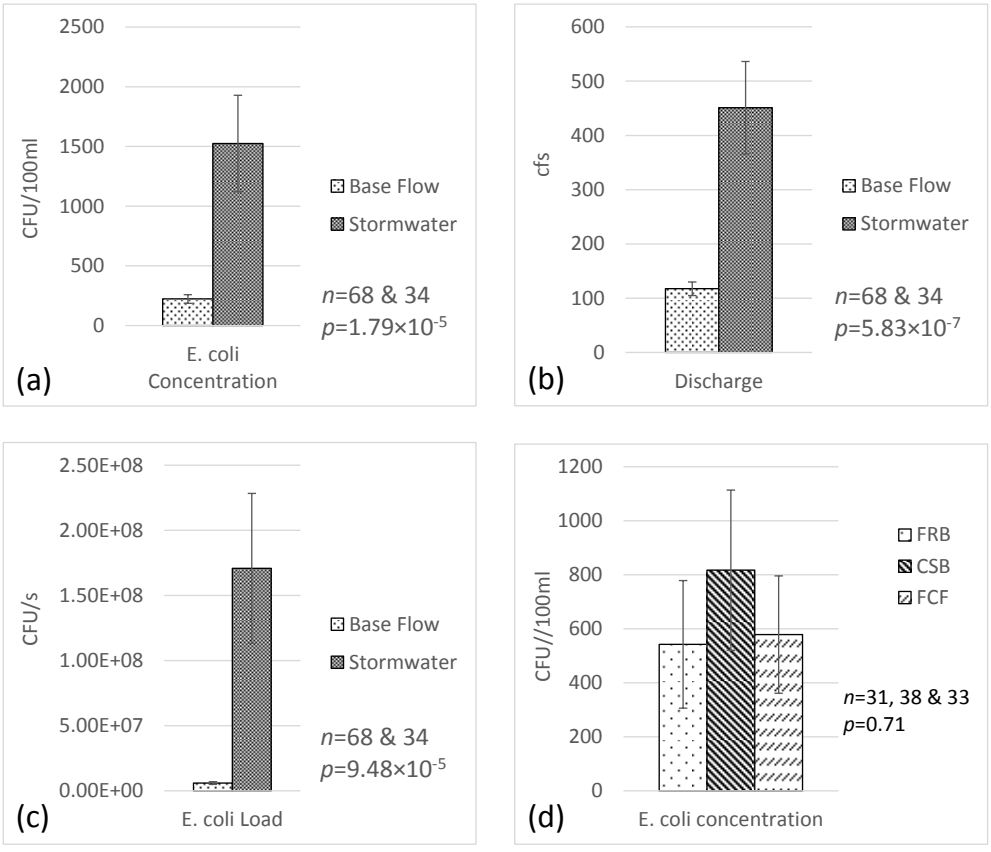


Figure 4.2 Comparing baseflow and stormwater conditions in terms of (a) *E. coli* concentration, (b) discharge, and (c) *E. coli* loading. Compare three sites in terms of *E. coli* concentration (d). The bars are arithmetic means and the error bars are standard errors.

The summary of all the variables are shown in Table 4.2. We did one-way ANOVA for all the continuous variables, Table 4.2 shows arithmetic mean, *p* value, and geometric mean of each of them. Except for chloride, nitrate-nitrite-nitrogen, air temperature, and seven day antecedent precipitation, all the other variables are significantly different between the two flow conditions. This further confirms our decision of dividing the dataset into two subgroups to do further analysis. Alkalinity, total hardness, pH and specific conductance are significantly higher in baseflow conditions than in stormwater conditions, which indicates that they are diluted and/or neutralized by rainwater. Whereas, total Kjeldahl nitrogen, total suspended solids, soluble reactive phosphorus, total coliform, total phosphorus and turbidity, besides previously mentioned *E. coli* concentration, discharge and *E. coli* loading, are significantly higher in stormwater conditions than in baseflow conditions, which indicates that their transport ways share some common with *E. coli*.

Table 4.2 Summary of all the variables.

	Counts for the categorical variables	
	Baseflow	Stormwater
W	13	15
Sp	14	6
S	20	9
F	21	4
FRB	21	10
CSB	24	14
FCF	23	10
	One-way ANOVA	Geometric mean

	Mean for baseflow	Mean for stormwater	*p*	Baseflow	Stormwater
alkalinity (mg $CaCO_3$/L)	133	107	**9×10^{-6}**	132	101
Cl (mg/L)	25.8	23.6	0.1243	25.0	22.6
Ecoli (colonies/100 mL)	223	1524	**2×10^{-5}**	98.6	548
hardness (mg $CaCO_3$/L)	159	128	**2×10^{-6}**	157	122
pH	7.95	7.75	**0.0005**	7.94	7.74
KNT (as N, mg/L)	0.405	0.592	**8×10^{-6}**	0.386	0.542
NNN (as N, mg/L)	0.805	0.931	0.1637	0.618	0.882
TSS (mg/L)	8.16	75.4	**5×10^{-6}**	4.35	25.0
SRP (as P, µg/L)	14.0	22.0	**4×10^{-5}**	12.2	19.2
SC (µS/cm)	357	316	**0.0005**	355	308
Tcoliform (colonies/100 ml)	9696	36964	**4×10^{-6}**	5802	21065
TP (as P, µg/L)	32.1	103	**1×10^{-6}**	28.5	65.4
turbidity (NTU)	7.34	67.8	**2×10^{-6}**	5.38	25.0
discharge (cfs)	117	451	**6×10^{-7}**	77.3	243
airT (℃)	9.38	6.99	0.1551	n.a.	n.a.
7dP (mm)	22.3	26.8	0.2367	n.a.	21.5
Ecoliload (colonies/s)	6.02×10^{6}	1.71×10^{8}	**9×10^{-5}**	2.16×10^{6}	3.77×10^{8}

Bold: significantly different
n.a.: contains negative number or 0, use arithmetic mean instead

The correlation coefficients between *E. coli* concentration / loading and the other variables are shown in Table 4.3 and Table 4.4. *E. coli* is a good indicator of total coliform ($p<0.001$) under both baseflow and stormflow conditions (Table 4.3 and Table 4.4). For all the data points (n=102) (Table 4.3), we found that *E. coli* concentration and *E. coli* loading are essentially significantly correlated with the same variables. Most notably, neither of them is significantly correlated with any season or any sampling site, but both of them are significantly negatively correlated with baseflow.

However, when we separated the data into two subgroups—baseflow and stormwater (Table 4.4), things changed, especially for baseflow. *E. coli* level became correlated to summer and sampling sites, but not correlated to variables related to pH and seven day antecedent precipitation any more.

Table 4.3 Pearson correlation coefficients between *E. coli* concentration and loading and other variables for all the data points (*n*=102).

	E. coli concentration (MPN/100mL)	*E. coli* loading (MPN/s)
B	-0.41***	-0.38***
W	0.08	0.10
Sp	-0.14	0.003
Su	0.08	-0.14
F	0.04	0.03
FRB	-0.05	-0.04
CSB	0.08	0.16
FCF	-0.04	-0.12
alkalinity	-0.30**	-0.44***
Cl	-0.22*	-0.28**
hardness	-0.30**	-0.46***
pH	-0.27**	-0.22*
KNT	0.74***	0.67***
NNN	-0.13	-0.09
TSS	0.52***	0.59***
SRP	0.58***	0.59***
SC	-0.26**	-0.46***
Tcoliform	0.82***	0.71***
TP	0.61***	0.64***
turbidity	0.61***	0.69***
discharge	0.13	n/a
airT	0.12	-0.02
7dP	0.45***	0.50***

* Significant at 0.05
** Significant at 0.01
*** Significant at 0.001

Table 4.4 Pearson correlation coefficients between *E. coli* concentration / loading and other variables for baseflow conditions (*n*=68), and stormwater conditions (*n*=34).

	E. coli concentration (MPN/100mL)	*E. coli* loading (MPN/s)	*E. coli* concentration (MPN/100mL)	*E. coli* loading (MPN/s)
	Baseflow		Stormwater	
W	-0.11	0.11	-0.02	0.006
Sp	-0.21	-0.02	-0.20	0.03
Su	0.43***	0.16	0.08	-0.25
F	-0.15	-0.23	0.16	0.29
FRB	-0.30*	-0.19	-0.008	-0.06
CSB	0.25*	0.35**	0.05	0.23
FCF	0.04	-0.16	-0.04	-0.18
alkalinity	0.21	-0.10	-0.23	-0.43*
Cl	0.09	0.19	-0.32	-0.40*
hardness	0.19	-0.03	-0.22	-0.45**
pH	-0.18	-0.20	-0.19	-0.14
KNT	0.72***	0.33**	0.77***	0.76***
NNN	-0.06	0.09	-0.52**	-0.41*
TSS	0.68***	0.40***	0.41*	0.52**
SRP	0.27*	0.22	0.66***	0.72***
SC	0.27*	0.07	-0.21	-0.47**
Tcoliform	0.46***	0.60***	0.81***	0.67***
TP	0.64***	0.39**	0.52**	0.58***
turbidity	0.69***	0.46***	0.51**	0.64***
discharge	-0.16	n.a.	-0.07	n.a.
airT	0.26*	-0.07	0.34*	0.09
7dP	-0.02	0.11	0.74***	0.80***

n.a. not independent
* Significant at 0.05
** Significant at 0.01
*** Significant at 0.001

4.4.2 Partial least squares regression (PLS)

The *E. coli* concentration and *E. coli* loading were dependent variables, and all the other variables except for concentration of total coliform were independent variables. Note that fall is the default season category and FCF is the default site category. We used segmented cross validation (Esbensen et al., 2002), i.e. dividing all the data points into segments, using one segment as the test set each time, until each segment having been test set once.

After trial and error, and comprehensively considering R^2, adjusted R^2, root mean square error of prediction (RMSEP) and residual (Esbensen et al., 2002), we found that the models using the log of *E. coli* concentration and *E. coli* loading, and using seven principle components give the best results. In addition, variables whose regression coefficients were < 0.002, were deleted from the corresponding model.

The PLS results for both baseflow condition and stormwater condition are shown in Figure 4.3, and the regression coefficients for each model are shown in Table 4.5. Note these models are based on the logs of *E. coli* concentrations / loadings and autoscaled variables. It can be seen that in baseflow condition (Figure 4.3a-d), it was hard to precisely predict either *E. coli* concentrations/ loadings, as both of the models have relatively small R^2 and adjusted R^2, relatively large RMSEP, and the residuals showed clear systematic patterns. However, in stormwater condition (Figure 4.3e-h) the predicted vs. measured *E. coli* concentrations / loadings are close to 1:1 line, distributed well over the whole range, and the residuals have no clear, systematic pattern. In addition, the predicted *E. coli* concentrations / loadings have relatively large R^2 and adjusted R^2, and relatively small RMSEP (Figure 4.3e, g).

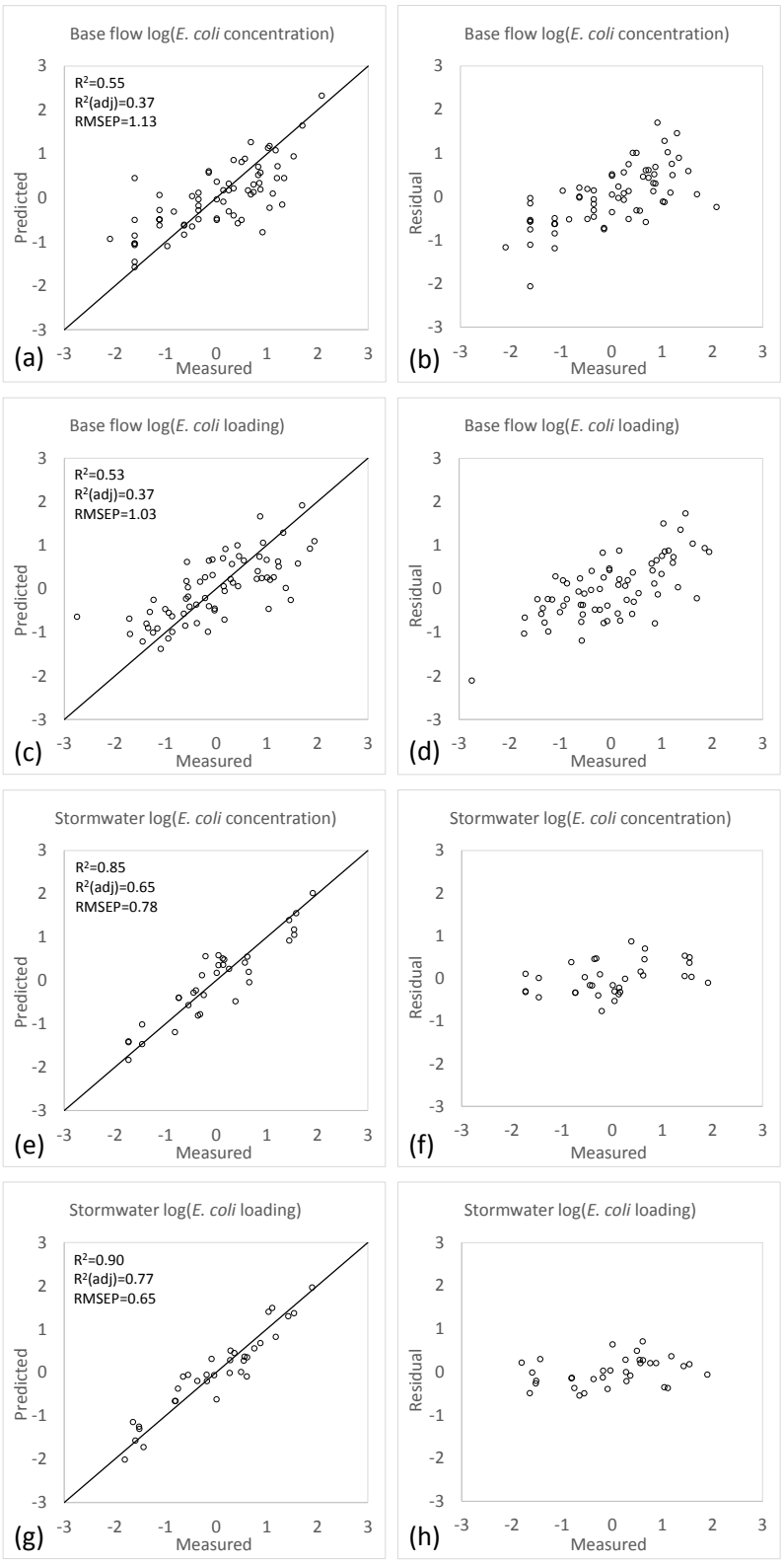

Base flow log(E. coli concentration)
R²=0.55
R²(adj)=0.37
RMSEP=1.13
Predicted
Measured
(a)
Base flow log(E. coli concentration)
Residual
Measured
(b)
Base flow log(E. coli loading)
R²=0.53
R²(adj)=0.37
RMSEP=1.03
Predicted
Measured
(c)
Base flow log(E. coli loading)
Residual
Measured
(d)
Stormwater log(E. coli concentration)
R²=0.85
R²(adj)=0.65
RMSEP=0.78
Predicted
Measured
(e)
Stormwater log(E. coli concentration)
Residual
Measured
(f)
Stormwater log(E. coli loading)
R²=0.90
R²(adj)=0.77
RMSEP=0.65
Predicted
Measured
(g)
Stormwater log(E. coli loading)
Residual
Measured
(h)

Figure 4.3 PLS results: (a) and (b) are for log of *E. coli* concentration in baseflow conditions; (c) and (d) are for log of *E. coli* loading in baseflow conditions; (e) and (f) are for log of *E. coli* concentration in stormwater conditions; and (g) and (h) are for log of *E. coli* loading in stormwater conditions.

Table 4.5 Regression coefficients corresponding to the variables for the four PLS models.

	Baseflow		Stormwater	
	log(*E. coli* concentration)	log(*E. coli* loading)	log(*E. coli* concentration)	log(*E. coli* loading)
W	-0.0072	0.031	0.25	0.31
Sp	-0.13	-0.11	0.11	0.14
Su	0.037	0.082	0.11	0.067
FRB	-0.39	-0.19	-0.24	-0.076
CSB	-0.012	0.22	-0.35	-0.16
alkalinity	0.041	-0.11	-0.13	-0.22
Cl	0.088	Del.	0.19	0.19
hardness	0.035	0.037	-0.012	-0.092
pH	-0.059	-0.033	0.066	0.031
KNT	0.56	0.51	0.76	0.57
NNN	-0.10	0.14	-0.17	-0.033
TSS	-0.24	-0.25	-0.059	-0.021
SRP	-0.0068	Del.	-0.090	-0.10
SC	0.21	0.25	0.17	0.10
TP	-0.0073	0.036	0.068	Del.
turbidity	0.092	0.049	-0.20	-0.14
discharge	0.24	0.51	0.14	0.39
airT	-0.0085	-0.034	0.18	0.13
7dP	-0.17	-0.15	0.33	0.37

Del.: corresponding variable deleted from model due to a less than 0.002 regression coefficient.

4.4.3 Principal component analysis (PCA)

In order to obtain the most important processes controlling the *E. coli* level, we used as few variables as possible to do the PCA analysis: (1) combine the W, Sp, Su and F into one variable,

Season which is 1 for winter, 2 for spring, 3 for summer and 4 for fall; (2) neglect the spatial difference as the three sites are not far from each other and all located at lower reaches of the Fall Creek Watershed (Figure 4.1), and the *E. coli* concentration were not significantly different (p=0.71) at the three sites (Figure 4.3); (3) delete hardness which can be represented by alkalinity; (4) not use Ecoliload, as it is the product of Ecoli and discharge.

Table 4.6 Principal component coefficients and explained variances.

	PC1	PC2	PC3	PC4
B	-0.25	-0.02	0.05	0.20
Season	-0.20	-0.36	0.00	-0.11
alkalinity	-0.32	-0.21	0.17	0.17
Cl	-0.11	-0.20	-0.41	-0.39
EColi	0.21	-0.31	0.30	-0.19
pH	-0.17	-0.09	-0.32	0.56
KNT	0.32	-0.25	0.07	0.03
NNN	0.10	0.40	0.20	0.28
TSS	0.34	-0.18	-0.29	0.21
SRP	0.24	-0.21	0.33	0.12
SC	-0.28	-0.30	-0.01	0.16
TP	0.35	-0.21	-0.21	0.18
turbidity	0.34	-0.20	-0.23	0.20
discharge	0.26	0.22	-0.15	-0.35
airT	-0.12	-0.39	0.08	-0.21
7dP	0.16	-0.12	0.49	0.10
Explained (%)	37.5	18.6	9.0	6.6

For the first principle component (PC1), negative coefficient for the baseflow variable means storm water, which leads to positive coefficients of discharge and those related to sediment erosion and resuspension (total suspended solids, total phosphorus and turbidity), and leads to the dilution of ions (low specific conductance and low chloride) and dilution and neutralization of bases (negative coefficient of alkalinity). The increase of soluble reactive phosphorus may be the result of the increased total phosphorus. And the increase of total Kjeldahl nitrogen, which consisted of ammonia, organic and reduced nitrogen, may be that the organic matter in the top soil got eroded. So, it is possible that PC1 is the runoff and erosion component. As *E. coli* was found to be eroded from the top soil (Chapter 2 and 3) and resuspended from the bottom of the stream together with sediment (Ferguson et al., 2003), the positive PC1 coefficient of *E. coli* concentration is expected.

For the second principle component (PC2), when the air temperature is low, the microbial activity is low, leading to decreased denitrification and increased nitrate-nitrite-nitrogen. So, it is possible that PC2 is the microbial activity component. Low microbial activity leads to less viable *E. coli*. It was reported by Tate (1995) that it is not freezing but thawing that kills the microbes. Therefore, the low microbial activity not only result from the low enzyme activity at low temperature, but can also result from freezing or the dying off of microbes due to thawing. And the relatively large negative coefficient of season further confirms this: winter and spring are the time when temperature is low and freezing and thawing happens. The increase of discharge may due to snowmelt, and the increased discharge dilute the stream water and lead to decreased total Kjeldahl nitrogen and specific conductance.

For the third principle component (PC3), baseflow shows very small coefficient, which means that the large seven day antecedent precipitation does not result in stormwater. The only explanation we could think of is that the sampling can be during stormwater or soon after stormwater. So, it is possible that PC3 is the component of shallow subsurface flow. The relatively large negative coefficients of the variables related to sediment (total suspended solids, total phosphorus and turbidity) further confirmed our inference. And the relatively large positive coefficients of *E. coli* concentration and soluble reactive phosphorus is not hard to explain: the increase of shallow subsurface flow provide more liquid water in the soil, so that those *E. coli* previously strained by soil or adsorbed to soil have more chances to get suspended in liquid water, and those phosphorus previously adsorbed to soil or precipitate out have more chances to dissolve in liquid water, thus both *E. coli* and phosphorus previously fixed in the soil flow into streams. The relatively large coefficients of chloride and pH (7-8.5 in our case) may due to dilution and neutralization, as rainwater has low chloride concentration and less than 7 pH value.

The baseflow and discharge coefficients show that the fourth principle component (PC4) is surely related to baseflow, or groundwater. Thus, the relatively large negative coefficients of *E. coli* concentration and pH, and the relatively large positive coefficients of nitrate-nitrite-nitrogen and sediments were explained. The sources of *E. coli* are at or near the soil surface, so the *E. coli* concentration is low in groundwater. The rain water is generally acidic in this region (Harpold et al., 2010), so the pH increase of the stream should due to the bases in groundwater. For nitrate-nitrite-nitrogen and sediments, the stream water was concentrated by the decrease of discharge. While the low air temperature may suggests that this groundwater component is more important

in colder seasons, i.e. winter, early spring and late fall, when storm turn into stormwater more slowly.

4.5 Discussion

The strategy of applying both PCA and PLS regression provided a useful, complementary and relatively simple way for evaluating datasets with lots of collinearity, both through time and space. The PCA suggested possible underlying biophysical mechanisms driving the variation of the variables. In this set, water quality characteristics, and particularly *E.coli* levels, are driven by differences between stormwater and baseflow conditions interacting with seasonal, temperature-related drivers. PCA did not provide a predictive tool of the actual concentrations / loadings; whereas, the PLS regression was useful, once the baseflow and stormwater data were separated, to create predictive equations of the actual concentrations / loadings.

As has been documented previously, *E. coli* concentrations were highly correlated with sediment concentrations, and also with related variables including, total phosphorus and turbidity. The importance of sediment is consistent with the findings by Vidon et al. (2008) and Gentry et al. (2006), however contrary to Dussart-Baptista et al. (2003). According to previous findings (Chapter 2 and 3), one reason is that *E. coli* and sediment transport can be explained by the same mechanisms. Baseflow has large negative coefficient for PC1, which means stormwater dominates. So we decided to build one model for each flow condition, instead of a general model. Discharge has large coefficient for PC1 too, it is related to stormwater and baseflow. This is consistent with Vidon et al. (2008) who found that *E. coli* concentration at baseflow and high flow conditions are

significantly different and analyzed their data for the two conditions separately. All of these are generally explained by surface erosion and the co-transport of *E.coli* both alone and bound to eroding sediments (Muirhead et al., 2006a; Oliver et al., 2005). It is a common process in agricultural lands, especially where no cover crops or conservation tillage are used and, therefore, contributed to the largest PCA factor.Collins and Rutherford (2004) and Vidon et al. (2008) also found that at high flow, *E. coli* is transported by overland flow during precipitation.

Interestingly, the second largest driver of *E. coli* concentration appears to be a seasonal influence on temperature and microbial activity. Season has large coefficient for PC2, and, for PC2, as *E.coli* concentration and air temperature decrease, nitrate-nitrite-nitrogen increase. It is possible that microbial activity is low at winter and spring due to freeze and thaw (Tate, 1995). Vidon et al. (2008) found that winter/spring and summer/fall constitute two identifiably different conditions for *E. coli*. Brooks et al. (2013) found that variables related to season, like Julian date and month, affect *E. coli*, and that air temperature is important. Whitman et al. (2008) found that season was important too, and, they also found that sunlight and snowmelt affect *E. coli* populations. We found that nitrate-nitrite-nitrogen, a previously ignored factor, is also a good *E. coli* indicator, as it indicates microbial activity, which generally follows seasonal and temperature trends.

For the shallow subsurface flow component, seven day antecedent precipitation is high, and *E. coli* and soluble reactive phosphorus increase while sediments (total suspended solids, total phosphorus, and turbidity) decrease. Seven day antecedent precipitation controls shallow subsurface flow, not just soil moisture. Gentry et al. (2006) and Vidon et al. (2008) found that the seven day antecedent

precipitation is one of the best indicators of *E. coli* under baseflow conditions. To our knowledge, soluble reactive phosphorus is identified to be correlated with *E. coli* level for the first time. It behaves similar to seven day antecedent precipitation (Table 4.5). *E. coli* are prone to straining and sorption (Oliver et al., 2005) and so is phosphorus, especially sorption (McGechan and Lewis, 2002b). So, perhaps the dilution effect of increased soil water content is counter-balanced by the increased mobility through soils. Besides these, pH had large coefficient for shallow subsurface flow component too. It is possible that pH is an indicator of the flow path of water, as pH is lower in rainfall and higher in water underground in the Northeastern US due to acidic atmospheric deposition (Harpold et al., 2010), although people usually do not think pH is useful for predicting *E. coli* population.

For PLS model, we are able to predict *E. coli* levels under stormwater conditions better than under baseflow conditions. However, stormwater conditions (n=34) are significantly higher than baseflow conditions (n=68) for both *E. coli* concentration ($p=1.79\times10^{-5}$) and *E. coli* load ($p=9.48\times10^{-5}$). So, we are good at predicting the hazardous events. Also, a large relative error in predicting small values is perhaps not significant relative to the entire range of *E coli* concentrations / loads.

Our results are consistent with Vidon et al. (2008) in that baseflow has significantly higher pH and specific conductance than stormwater (Table 4.2); and has significantly lower turbidity, *E. coli* concentration, discharge and *E. coli* loading than stormwater (Figure 4.2 and Table 4.2).

For baseflow conditions, we have similar geometric means (Table 4.2) of pH and turbidity as those of Vidon et al. (2008), lower specific conductance and temperature, and have much lower *E. coli* concentration (10 to 20 fold) and much higher discharge (10 to 20 fold), and results in similar *E. coli* loading.

For stormwater (Table 4.2), we have similar geometric means of pH and turbidity as those of Vidon et al. (2008), lower specific conductance and temperature, and have much lower *E. coli* concentration (5 to 8 folds) and much higher discharge (10 to 20 folds), and results in higher *E. coli* loading.

In terms of the Pearson correlation coefficients, our results are again consistent with Vidon et al. (2008) that both *E. coli* concentration and *E. coli* loading are significantly lower at baseflow than stormwater ($p<0.001$ for concentration and $p<0.001$ for loading) (Figure 4.2 and Table 4.2), and the *E. coli* concentration / loading only showed spatial differences at baseflow ($p<0.05$ for concentration and $p<0.01$ for loading) (Table 4.3 and Table 4.4). And the only significant correlation related to season we have found is at baseflow: *E. coli* concentration is significantly higher in summer ($p<0.001$) (Table 4.3 and Table 4.4).

In terms of correlations with other variables, we agree with Vidon et al. (2008) that variables related to pH (pH, alkalinity, and hardness) are significantly correlated with *E. coli* level when

combining all the data points, while not significantly correlated when the data were separated by flow conditions (Table 4.3 and Table 4.4). And we also found that total Kjeldahl nitrogen and soluble reactive phosphorus are significantly correlated with *E. coli* level. Contrary to Vidon et al. (2008) who found that turbidity is not significantly correlated with *E. coli* level during high flow conditions, we found that sediment (total suspended solid, total phosphorus, and turbidity) is always significantly correlated with *E. coli* level whatever the flow condition is.

In conclusion, we identified more variables that can indicate the level of *E. coli* in a stream, found four important processes driving the *E. coli* fate and transport in a watershed (runoff and erosion, microbial activity, shallow subsurface flow and groundwater), and built a PLS model that can predict *E. coli* level very well at stormwater conditions.

CHAPTER 5 CONCLUSIONS

It was found that under the rainfall and soil conditions set by the experiments the transport of *E. coli* from soil into overland flow is equally well described by Gao solute transport model and Hairsine-Rose soil erosion model. Thus we proposed the microbial solute-particle duality, in which microbes can be modeled as non-diffusing solutes. And the inner relations of the two models were revealed, which provide more point of views to look at solute and particles.

There was substantially more *E. coli* transferred from the sand column than from the sand-clay, because the raindrops penetrated deeper into the soil-media and the proportion of soil water in the total ejected mass is larger. The primary role of clay appears to be changing the soil texture, making the soil harder to penetrate and containing less soil water, which reduces *E. coli* transport into overland flow. Thus, soil texture plays an important role in the co-transport of mineral colloid and bacteria from soil into overland flow under raindrop impact.

Since *E. coli* showed particle-like transport and our results suggest that most of the *E. coli* were not attached to sand particles, we suggest that the best management practices designed for fine sediment retention are also suitable for *E. coli* retention.

Applying our findings at soil column scale to *E. coli* level explanation and prediction at watershed scale, we identified more variables that can indicate the level of *E. coli* in a stream, found four important processes driving the *E. coli* fate and transport in a watershed (runoff and erosion,

microbial activity, shallow subsurface flow and groundwater), and built a PLS model that can predict *E. coli* level very well at stormwater conditions.

In the future, more studies linking discovers of waterborne pathogen transport at different scales are highly encouraged.

APPENDIX A

Analytical solution to the simplified no diffusion Gao solute model

Inserting Eqs. 2-15 into Eq. 2-13 and solving Eq. 2-13,

$$c \cdot C_e = \exp(-\frac{ap}{\rho_b d_e} t) \tag{A1}$$

$$\because C_e = C_o, \text{ at } t = 0 \text{ (Eq. 2-16)}$$

$$\therefore C_e = C_o \cdot \exp(-\frac{ap}{\rho_b d_e} t) \tag{A2}$$

Inserting Eqs. 2-15 into Eq. 2-14 and solving Eq. 2-14,

$$C_w = (\int \frac{ap\theta}{\rho_b d_w} C_e \cdot \exp\left(\int \frac{p}{d_w} dt\right) dt + c_2) \cdot \exp(-\int \frac{p}{d_w} dt) \tag{A3}$$

Plugging in Eq. A2,

$$C_w = (\frac{ap\theta}{\rho_b d_w\left(\frac{p}{d_w} - \frac{ap}{\rho_b d_e}\right)} \cdot C_o \cdot \exp(\left(\frac{p}{d_w} - \frac{ap}{\rho_b d_e}\right) t) + c_2) \cdot \exp(-\frac{p}{d_w} t) \tag{A4}$$

$$\because C_w = 0 \text{ at } t = 0 \text{ (Eq. 2-16)}$$

$$\therefore C_w = \frac{ap\theta}{\rho_b d_w\left(\frac{p}{d_w} - \frac{ap}{\rho_b d_e}\right)} \cdot C_o \cdot (\exp(\left(\frac{p}{d_w} - \frac{ap}{\rho_b d_e}\right) t) - 1) \cdot \exp(-\frac{p}{d_w} t) \tag{A5}$$

Supplemental materials

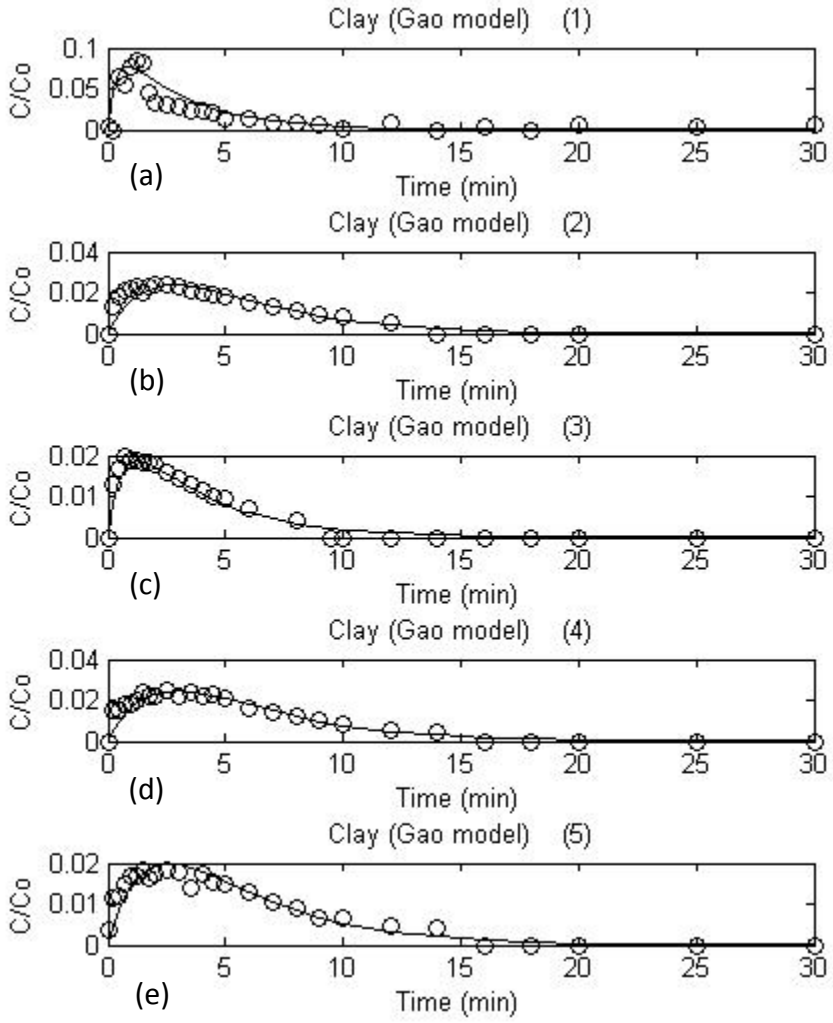


Figure A1. Measured relative concentration of clay (circles) fitted by Gao model.

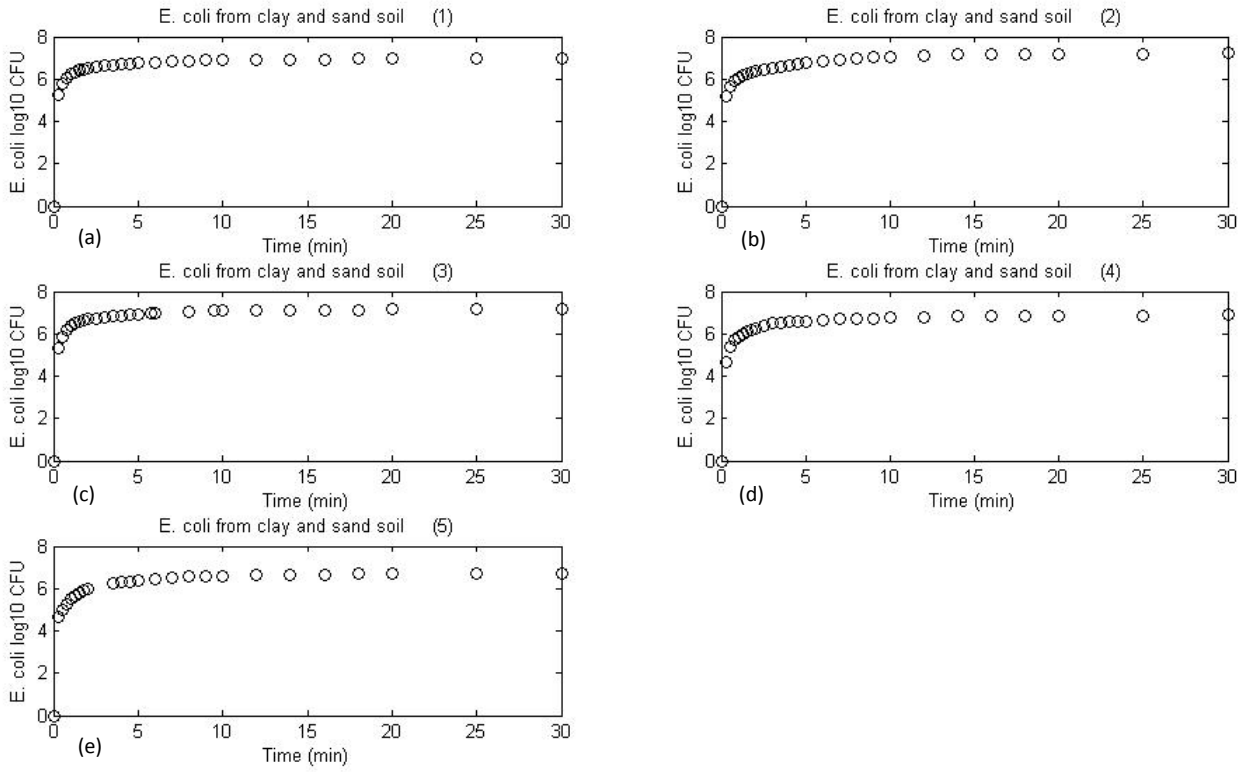
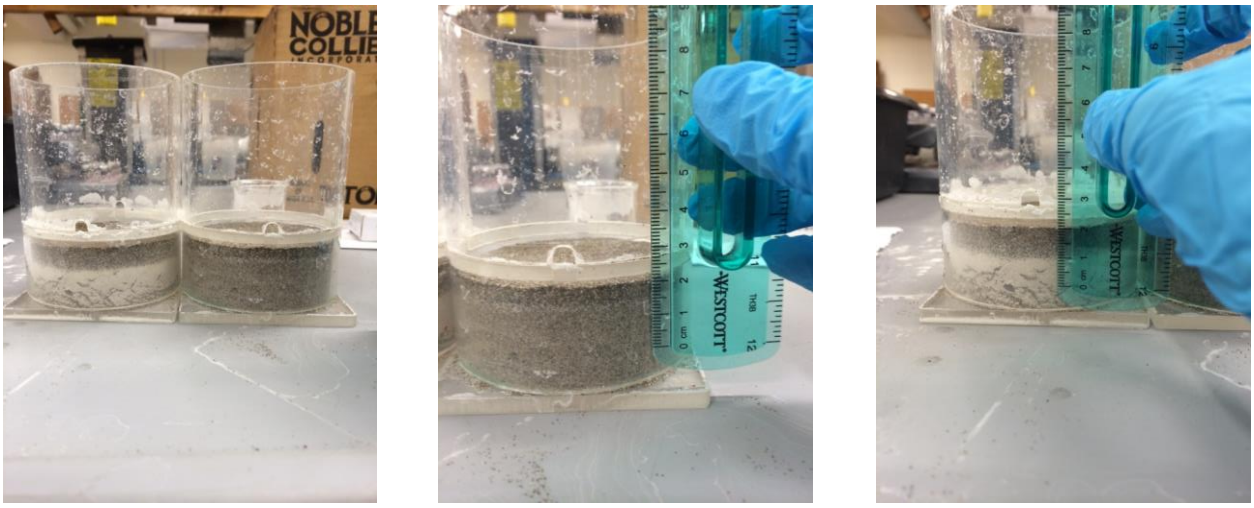


Figure A2. Cumulative load of *E. coli* from clay and sand soil. The index means from which run of the experiment.

APPENDIX B

The exchange layer for the clay-sand mixture was very distinct, because the underlying soil was lighter due to the white clay (Figure B1a and B1c). Surprisingly, the exchange layer for the pure sand was visually distinguishable because a white precipitate formed in the soil below the exchange layer (Figure B1a and B1b).



(a) (b) (c)

Figure B1. Soil columns air dried after rainfall experiment to measure shield layer depth and mass per unit area: (a) two kinds of soil columns side by side, (b) the air dried no clay soil, and (c) the air dried clay-sand mixture soil.

We also measured pH and conductivity of the soil water and designed a small experiment to determine what the white precipitate was. We made three soil columns, (1) add only 1.08% KCl (with the ratio of 80 ml in 225g sand), (2) add 1.08% KCl and TSB (with the ratio of 80 ml KCl and 2 ml TSB in 225g sand), and (3) add 1.08% KCl and *E. coli*-TSB solution (with the ratio of 80 ml KCl and 2 ml *E. coli*-TSB solution in 225g sand).

Table B1. pH and conductivity.

	KCl (1.08%, 16.4ml)		KCl (1.08%, 16ml)+TSB (30g in 1L DI water, 0.4ml)	
	Liquid	Mix with sand	Liquid	Mix with sand
pH	6.15	7.45	7.16	7.31
Conductivity (ms/cm)	18.6	17.7	17.1	17.5

After drying, the soil with only KCl and the soil with KCl and TSB only showed white precipitate on the top of the soil but no white precipitate in the column. However, the soil with KCl, TSB and *E. coli* cells showed while precipitate along the column. So, it is possible that the white precipitate is mainly composed of dead *E. coli* cells, and maybe a little dried TSB.

We initially applied the directly measured d_e and C_o values to the Gao model and got seemingly random and physically unrealistic results (Figure B2).

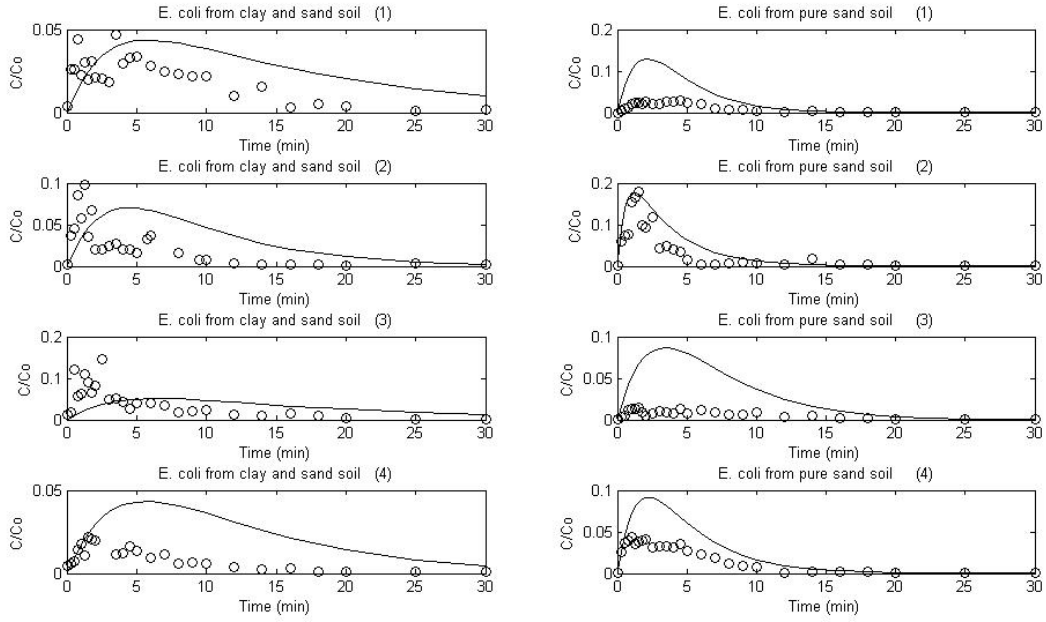


Figure B2. Gao solute model for relative concentration of *E. coli* from clay and sand soil (left column) and from pure sand soil (right column), using measured *de* and *Co*.

BIBLIOGRAPHY

Abdel-Salam, A., Chrysikopoulos, C., 1995. Modeling of colloid and colloid-facilitated contaminant transport in a two-dimensional fracture with spatially variable aperture. Transp. Porous Media 20(3): 197-221.

Aguilera, P.A., Garrido Frenich, A., Castro, H., Martinez Vidal, J.L., 2000. PLS and PCR methods in the assessment of coastal water quality. Environ. Monit. Assess. 62(2): 193-204.

Al-Holy, M.A., Lin, M., Cavinato, A.G., Rasco, B.A., 2006. The use of Fourier transform infrared spectroscopy to differentiate *Escherichia coli* O157:H7 from other bacteria inoculated into apple juice. Food Microbiology 23(2): 162-168.

Al-Qadiri, H.M., Lin, M., Cavinato, A.G., Rasco, B.A., 2006. Fourier transform infrared spectroscopy, detection and identification of *Escherichia coli* O157:H7 and *Alicyclobacillus* strains in apple juice. Int. J. Food Microbiol. 111(1): 73-80.

Barton, J.W., Ford, R.M., 1995. Determination of effective transport coefficients for bacterial migration in sand columns. Appl. Environ. Microbiol. 61(9): 3329-3335.

Barton, J.W., Ford, R.M., 1997. Mathematical model for characterization of bacterial migration through sand cores. Biotechnol. Bioeng. 53(5): 487-496.

Basant, N., Gupta, S., Malik, A., Singh, K.P., 2010. Linear and nonlinear modeling for simultaneous prediction of dissolved oxygen and biochemical oxygen demand of the surface water — A case study. Chemom. Intell. Lab. Syst. 104(2): 172-180.

Bech, T.B., Rosenbom, A.E., Kjaer, J., Amin, M.G.M., Olsen, P., Jacobsen, C.S., 2014. Factors influencing the survival and leaching of tetracycline-resistant bacteria and *Escherichia coli* through structured agricultural fields. Agric. Ecosyst. Environ. 195: 10-17.

Bekhit, H.M., El-Kordy, M.A., Hassan, A.E., 2009. Contaminant transport in groundwater in the presence of colloids and bacteria: Model development and verification. J. Contam. Hydrol. 108(3–4): 152-167.

Benham, B., Baffaut, C., Zeckoski, R., Mankin, K., Pachepsky, Y., Sadeghi, A., Brannan, K., Soupir, M., Habersack, M., 2006. Modeling bacteria fate and transport in watersheds to support TMDLs. Trans. ASAE 49(4): 987-1002.

Beversdorf, L.J., Bornstein‐Forst, S.M., McLellan, S.L., 2007. The potential for beach sand to serve as a reservoir for *Escherichia coli* and the physical influences on cell die‐off. J. Appl. Microbiol. 102(5): 1372-1381.

Brooks, W.R., Fienen, M.N., Corsi, S.R., 2013. Partial least squares for efficient models of fecal indicator bacteria on Great Lakes beaches. J. Environ. Manag. 114(0): 470-475.

Buddemeier, R.W., Hunt, J.R., 1988. Transport of colloidal contaminants in groundwater: radionuclide migration at the Nevada test Site. Appl. Geochem. 3(5): 535-548.

Carroll, S.P., Dawes, L., Hargreaves, M., Goonetilleke, A., 2009. Faecal pollution source identification in an urbanising catchment using antibiotic resistance profiling, discriminant analysis and partial least squares regression. Water Res. 43(5): 1237-1246.

Chen, G., 2012. *S. typhimurium* and *E. coli* O157:H7 retention and transport in agricultural soil during irrigation practices. Eur. J. Soil Sci. 63(2): 239-248.

Chen, G., Flury, M., Harsh, J.B., Lichtner, P.C., 2005. Colloid-facilitated transport of cesium in variably saturated hanford sediments. Environ. Sci. Technol. 39(10): 3435-3442.

Chrysikopoulos, C.V., Sim, Y., 1996. One-dimensional virus transport in homogeneous porous media with time-dependent distribution coefficient. J. Hydrol. 185(1–4): 199-219.

Collins, R., Rutherford, K., 2004. Modelling bacterial water quality in streams draining pastoral land. Water Res. 38(3): 700-712.

Community Science Institute, 2014. CSI database. http://communityscience.org (July 1, 2014).

Corapcioglu, M.Y., Jiang, S., 1993. Colloid-facilitated groundwater contaminant transport. Water Resour. Res. 29(7): 2215-2226.

de Sousa Marques, A., Nicácio, J.T.N., Cidral, T.A., de Melo, M.C.N., de Lima, K.M.G., 2013. The use of near infrared spectroscopy and multivariate techniques to differentiate *Escherichia coli* and *Salmonella Enteritidis* inoculated into pulp juice. J. Microbiol. Methods 93(2): 90-94.

Dussart-Baptista, L., Massei, N., Dupont, J.-P., Jouenne, T., 2003. Transfer of bacteria-contaminated particles in a karst aquifer: evolution of contaminated materials from a sinkhole to a spring. Journal of Hydrology 284(1): 285-295.

Dwivedi, D., Mohanty, B.P., Lesikar, B.J., 2013. Estimating *Escherichia coli* loads in streams based on various physical, chemical, and biological factors. Water Resour. Res. 49(5): 2896-2906.

Esbensen, K.H., Guyot, D., Westad, F., Houmoller, L.P., 2002. Multivariate data analysis-In practice: An introduction to multivariate data analysis and experimental design. CAMO Process AS, Oslo, Norway.

Falbo, K., Schneider, R.L., Buckley, D.H., Walter, M.T., Bergholz, P.W., Buchanan, B.P., 2013. Roadside ditches as conduits of fecal indicator organisms and sediment: Implications for water quality management. J. Environ. Manag. 128: 1050-1059.

Ferguson, C., Davies, C., Kaucner, C., Ashbolt, N., Krogh, M., Rodehutskors, J., Deere, D., 2007. Field scale quantification of microbial transport from bovine faeces under simulated rainfall events. J. Water Health 5(1): 83-95.

Ferguson, C., Husman, A.M.d.R., Altavilla, N., Deere, D., Ashbolt, N., 2003. Fate and transport of surface water pathogens in watersheds.

Gao, B., Walter, M.T., Steenhuis, T.S., Hogarth, W.L., Parlange, J.-Y., 2004. Rainfall induced chemical transport from soil to runoff: Theory and experiments. J. Hydrol. 295(1): 291-304.

Gao, B., Walter, M.T., Steenhuis, T.S., Parlange, J.-Y., Nakano, K., Rose, C.W., Hogarth, W.L., 2003. Investigating ponding depth and soil detachability for a mechanistic erosion model using a simple experiment. J. Hydrol. 277(1): 116-124.

Gao, B., Walter, M.T., Steenhuis, T.S., Parlange, J.-Y., Richards, B.K., Hogarth, W.L., Rose, C.W., 2005. Investigating raindrop effects on transport of sediment and non-sorbed chemicals from soil to surface runoff. J. Hydrol. 308(1): 313-320.

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

Gill, C.O., Delacy, K.M., 1991. Growth of *Escherichia coli* and *Salmonella typhimurium* on high-pH beef packed under vacuum or carbon dioxide. Int. J. Food Microbiol. 13(1): 21-30.

Gray, E., 1975. Survival of *Escherichia coli* in stream water in relation to carbon dioxide and plant photosynthesis. J. Appl. Bacteriol. 39(1): 47-54.

Grolimund, D., Borkovec, M., 2005. Colloid-facilitated transport of strongly sorbing contaminants in natural porous media: mathematical modeling and laboratory column experiments. Environ. Sci. Technol. 39(17): 6378-6386.

Guber, A.K., Shelton, D.R., Pachepsky, Y.A., 2005a. Effect of manure on attachment to soil. J. Environ. Qual. 34(6): 2086-2090.

Guber, A.K., Shelton, D.R., Pachepsky, Y.A., 2005b. Transport and retention of manure-borne coliforms in soil. Vadose Zone J. 4(3): 828-837.

Guiné, V., Martins, J., Gaudet, J., 2003. Facilitated transport of heavy metals by bacterial colloids in sand columns, J. Phys. IV Fr. EDP sciences, pp. 593-596.

Hairsine, P.B., Rose, C.W., 1991. Rainfall detachment and deposition: Sediment transport in the absence of flow-driven processes. Soil Sci. Soc. Am. J. 55(2): 320-324.

Harpold, A.A., Burns, D.A., Walter, T., Shaw, S.B., Steenhuis, T.S., 2010. Relating hydrogeomorphic properties to stream buffering chemistry in the Neversink River watershed, New York State, USA. Hydrological Processes 24(26): 3759-3771.

Hedges, A., 2002. Estimating the precision of serial dilutions and viable bacterial counts. Intl. J. Food Microbiol. 76(3): 207-214.

Heilig, A., DeBruyn, D., Walter, M.T., Rose, C.W., Parlange, J.-Y., Steenhuis, T.S., Sander, G.C., Hairsine, P.B., Hogarth, W.L., Walker, L.P., 2001. Testing a mechanistic soil erosion model with a simple experiment. J. Hydrol. 244(1): 9-16.

Hwang, C.-A., Sheen, S., Juneja, V., Hwang, C.-F., Yin, T.-C., Chang, N.-Y., 2014. The influence of acid stress on the growth of *Listeria monocytogenes* and *Escherichia coli* O157:H7 on cooked ham. Food Control. 37(0): 245-250.

Ibaraki, M., Sudicky, E.A., 1995. Colloid-facilitated contaminant transport in discretely fractured porous media: 1. Numerical formulation and sensitivity analysis. Water Resour. Res. 31(12): 2945-2960.

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

Jiang, G., Noonan, M.J., Buchan, G.D., Smith, N., 2007. Transport of *Escherichia coli* through variably saturated sand columns and modeling approaches. J. Contam. Hydrol. 93(1): 2-20.

Kanti Sen, T., Khilar, K.C., 2006. Review on subsurface colloids and colloid-associated contaminant transport in saturated porous media. Adv. Colloid Interface Sci. 119(2–3): 71-96.

Kim, S.-B., Corapcioglu, M.Y., 2002a. Contaminant transport in dual-porosity media with dissolved organic matter and bacteria present as mobile colloids. J. Contam. Hydrol. 59(3–4): 267-289.

Kim, S.-B., Corapcioglu, M.Y., 2002b. Contaminant transport in riverbank filtration in the presence of dissolved organic matter and bacteria: A kinetic approach. J. Hydrol. 266(3–4): 269-283.

Kim, S.-B., Yavuz Corapcioglu, M., Kim, D.-J., 2003. Effect of dissolved organic matter and bacteria on contaminant transport in riverbank filtration. J. Contam. Hydrol. 66(1–2): 1-23.

Kistemann, T., Claßen, T., Koch, C., Dangendorf, F., Fischeder, R., Gebel, J., Vacata, V., Exner, M., 2002. Microbial load of drinking water reservoir tributaries during extreme rainfall and runoff. Appl. Environ. Microbiol. 68(5): 2188-2197.

Kouznetsov, M.Y., Roodsari, R., Pachepsky, Y.A., Shelton, D.R., Sadeghi, A.M., Shirmohammadi, A., Starr, J.L., 2007. Modeling manure-borne bromide and fecal coliform transport with runoff and infiltration at a hillslope. J. Environ. Manag. 84(3): 336-346.

McGechan, M.B., Lewis, D.R., 2002a. Transport of particulate and colloid-sorbed contaminants through soil, part 1: General principles. Biosyst. Eng. 83(3): 255-273.

McGechan, M.B., Lewis, D.R., 2002b. Transport of Particulate and Colloid-sorbed Contaminants through Soil, Part 1: General Principles. Biosystems Engineering 83(3): 255-273.

Meays, C.L., Broersma, K., Nordin, R., Mazumder, A., Samadpour, M., 2006. Diurnal variability in concentrations and sources of *Escherichia coli* in three streams. Can. J. Microbiol. 52(11): 1130-5.

Muirhead, R.W., Collins, R.P., Bremer, P.J., 2005. Erosion and subsequent transport state of *Escherichia coli* from cowpats. Appl. Environ. Microbiol. 71(6): 2875-2879.

Muirhead, R.W., Collins, R.P., Bremer, P.J., 2006a. The association of *E. coli* and soil particles in overland flow. Water Sci. & Technol. 54(3): 153-159.

Muirhead, R.W., Collins, R.P., Bremer, P.J., 2006b. Interaction of *Escherichia coli* and soil particles in runoff. Appl. and Environ. Microbiol. 72(5): 3406-3411.

Muirhead, R.W., Collins, R.P., Bremer, P.J., 2006c. Numbers and transported state of *Escherichia coli* in runoff direct from fresh cowpats under simulated rainfall. Lett. Appl. Microbiol. 42(2): 83-87.

Muirhead, R.W., Littlejohn, R.P., 2009. Die ‐ off of *Escherichia coli* in intact and disrupted cowpats. Soil Use and Manag. 25(4): 389-394.

Neidhardt, F.C., Ingraham, J.L., Schaechter, M., 1990. Physiology of the Bacterial Cell: A Molecular Approach. Sinauer Associates, Inc., Sunderland, Massachusetts, pp. 8.

Nevers, M.B., Whitman, R.L., 2005. Nowcast modeling of *Escherichia coli* concentrations at multiple urban beaches of southern Lake Michigan. Water Res. 39(20): 5250-5260.

NOAA, 2014. NOAA National Centers for Environmental Information: Climate Data Online. http://gis.ncdc.noaa.gov/map/viewer/#app=cdo&cfg=lcd&theme=lcd

Oliver, D.M., Clegg, C.D., Haygarth, P.M., Heathwaite, A.L., 2005. Assessing the potential for pathogen transfer from grassland soils to surface waters. Adv. Agron. 85: 125-180.

Oliver, D.M., Clegg, C.D., Heathwaite, A.L., Haygarth, P.M., 2007. Preferential attachment of *Escherichia coli* to different particle size fractions of an agricultural grassland soil. Water, Air, and Soil Pollut. 185(1-4): 369-375.

Olson, M.S., Ford, R.M., Smith, J.A., Fernandez, E.J., 2005. Analysis of column tortuosity for $MnCl_2$ and bacterial diffusion using magnetic resonance imaging. Environ. Sci. Technol 39(1): 149-154.

Ortiz-Estarelles, O., Martin-Biosca, Y., Medina-Hernandez, M.J., Sagrado, S., Bonet-Domingo, E., 2001a. Multivariate data analysis of quality parameters in drinking water. Analyst 126(1): 91-96.

Ortiz-Estarelles, O., Martín-Biosca, Y., Medina-Hernández, M.J., Sagrado, S., Bonet-Domingo, E., 2001b. On the internal multivariate quality control of analytical laboratories. A case study: the quality of drinking water. Chemom. Intell. Lab. Syst. 56(2): 93-103.

Ouyang, Y., Shinde, D., Mansell, R.S., Harris, W., 1996. Colloid ‐ enhanced transport of chemicals in subsurface environments: A review. Crit. Rev. Environ. Sci. Technol. 26(2): 189-204.

Pang, L., Close, M., Noonan, M., Flintoft, M., Van den Brink, P., 2005. A laboratory study of bacteria-facilitated cadmium transport in alluvial gravel aquifer media. J. Environ. Qual. 34(1): 237-247.

Pang, L., Šimůnek, J., 2006. Evaluation of bacteria-facilitated cadmium transport in gravel columns using the HYDRUS colloid-facilitated solute transport model. Water Resour. Res. 42(12): W12S10.

Pettibone, G.W., Irvine, K.N., 1996. Levels and Sources of Indicator Bacteria Associated with the Buffalo River "Area of Concern," Buffalo, New York. J. Gt. Lakes Res. 22(4): 896-905.

Powelson, D.K., Mills, A.L., 2001. Transport of *Escherichia coli* in sand columns with constant and changing water contents. J. Environ. Qual. 30(1): 238-245.

Reynolds, P.J., Sharma, P., Jenneman, G.E., McInerney, M.J., 1989. Mechanisms of microbial movement in subsurface materials. Appl. and Environ. Microbiol. 55(9): 2280-2286.

Roodsari, R.M., Shelton, D.R., Shirmohammadi, A., Pachepsky, Y.A., Sadeghi, A.M., Starr, J.L., 2005. Fecal coliform transport as affected by surface condition. Trans. ASAE 48(3): 1055-1061.

Roy, S.B., Dzombak, D.A., 1997. Chemical factors influencing colloid-facilitated transport of contaminants in porous media. Environ. Sci. Technol. 31(3): 656-664.

Runkel, R.L., Crawford, C.G., Cohn, T.A., 2004. Load Estimator (LOADEST): A Fortran program for estimating constituent loads in streams and rivers. US Department of the Interior, US Geological Survey.

Safadoust, A., Mahboubi, A.A., Mosaddeghi, M.R., Gharabaghi, B., Unc, A., Voroney, P., Heydari, A., 2012a. Effect of regenerated soil structure on unsaturated transport of *Escherichia coli* and bromide. J. of Hydrol. 430–431: 80-90.

Safadoust, A., Mahboubi, A.A., Mosaddeghi, M.R., Gharabaghi, B., Voroney, P., Unc, A., Khodakaramian, G., 2012b. Significance of physical weathering of two-texturally different soils for the saturated transport of *Escherichia coli* and bromide. J. of Environ. Manag. 107: 147-158.

Sagdic, O., Ozturk, I., 2013. Kinetic Modeling of *Escherichia coli* O157:H7 Growth in Rainbow Trout Fillets as Affected by Oregano and Thyme Essential Oils and Different Packing Treatments. Int. J. Food Prop. 17(2): 371-385.

Salleh-Mack, S.Z., Roberts, J.S., 2007. Ultrasound pasteurization: The effects of temperature, soluble solids, organic acids and pH on the inactivation of *Escherichia coli* ATCC 25922. Ultrason. Sonochem. 14(3): 323-329.

Sander, G.C., Hairsine, P.B., Rose, C.W., Cassidy, D., Parlange, J.-Y., Hogarth, W.L., Lisle, I.G., 1996. Unsteady soil erosion model, analytical solutions and comparison with experimental results. J. Hydrol. 178(1): 351-367.

Sauer, A., Moraru, C.I., 2009. Inactivation of *Escherichia coli* ATCC 25922 and *Escherichia coli* O157: H7 in apple juice and apple cider, using pulsed light treatment. J. of Food Prot. 72(5): 937-944.

Schäfer, A., Ustohal, P., Harms, H., Stauffer, F., Dracos, T., Zehnder, A.J.B., 1998. Transport of bacteria in unsaturated porous media. J. Contam. Hydrol. 33(1): 149-169.

Sherwood, J.L., Sung, J.C., Ford, R.M., Fernandez, E.J., Maneval, J.E., Smith, J.A., 2003. Analysis of bacterial random motility in a porous medium using magnetic resonance imaging and immunomagnetic labeling. Environ. Sci. Technol. 37(4): 781-785.

Simon, R.D., Makarewicz, J.C., 2009a. Impacts of manure management practices on stream microbial loading into Conesus Lake, NY. Journal of Great Lakes Research 35, Supplement 1(0): 66-75.

Simon, R.D., Makarewicz, J.C., 2009b. Storm water events in a small agricultural watershed: Characterization and evaluation of improvements in stream water microbiology following implementation of Best Management Practices. Journal of Great Lakes Research 35, Supplement 1(0): 76-82.

Šimůnek, J., He, C., Pang, L., Bradford, S.A., 2006. Colloid-facilitated solute transport in variably saturated porous media. Vadose Zone J. 5(3): 1035-1047.

Singh, K.P., Malik, A., Basant, N., Saxena, P., 2007. Multi-way partial least squares modeling of water quality data. Analytica chimica acta 584(2): 385-396.

Siripatrawan, U., Makino, Y., Kawagoe, Y., Oshita, S., 2011. Rapid detection of *Escherichia coli* contamination in packaged fresh spinach using hyperspectral imaging. Talanta 85(1): 276-281.

Smith, M.S., Thomas, G.W., White, R.E., Ritonga, D., 1985. Transport of *Escherichia coli* through intact and disturbed soil columns. J. Environ. Qual. 14(1): 87-91.

Smith, P.A., Degueldre, C., 1993. Colloid-facilitated transport of radionuclides through fractured media. J. Contam. Hydrol. 13(1–4): 143-166.

Tate, R.L., 1995. Soil Microbiology. John Wiley & Sons, Inc., New York, USA.

Traister, E., Anisfeld, S.C., 2006. Variability of indicator bacteria at different time scales in the upper Hoosic River watershed. Environmental science & technology 40(16): 4990-4995.

United States Environmental Protection Agency (US EPA), 2010. National Summary of Impaired Waters and TMDL Information. http://iaspub.epa.gov/waters10/attains_nation_cy.control?p_report_type=T (May 19, 2015).

USGS, 2007. USGS Instantaneous Data Archive - IDA. http://ida.water.usgs.gov/ida/available_records.cfm (July 22, 2014).

USGS, 2014. USGS Surface-Water Historical Instantaneous Data for the Nation. http://nwis.waterdata.usgs.gov/nwis/uv? (July 20, 2014).

Vasiliadou, I.A., Chrysikopoulos, C.V., 2011a. Cotransport of *Pseudomonas putida* and kaolinite particles through water-saturated columns packed with glass beads. Water Resour. Res. 47(2): W02543.

Vasiliadou, I.A., Chrysikopoulos, C.V., 2011b. Cotransport of Pseudomonas putida and kaolinite particles through water-saturated columns packed with glass beads. Water Resources Research 47(2): W02543.

Vidon, P., Tedesco, L.P., Wilson, J., Campbell, M.A., Casey, L.R., Gray, M., 2008. Direct and indirect hydrological controls on *E. coli* concentration and loading in Midwestern streams. J. Environ. Qual. 37(5): 1761-8.

Walker, F.R., Stedinger, J.R., 1999. Fate and transport model of *Cryptosporidium*. J. Environ. Eng. 125(4): 325-333.

Walter, M.T., Gao, B., Parlange, J.-Y., 2007. Modeling soil solute release into runoff with infiltration. J. Hydrol. 347(3): 430-437.

Wang, C., Parlange, J.-Y., Rasmussen, E.W., Wang, X., Chen, M., Walter, M.T., 2015a. Modeling transport of *Escherichia coli* from soil into overland flow under raindrop impact: Solute or particle? J. Hydrol. (submitted for publication).

Wang, C., Parlange, J.-Y., Schneider, R.L., Rasmussen, E.W., Wang, X., Chen, M., Walter, M.T., 2015b. Transport of *Escherichia coli* under raindrop impact: The role of clay. J. Hydrol. (submitted for publication).

Wang, C., Walter, M.T., Parlange, J.-Y., 2013. Modeling simple experiments of biochar erosion from soil. J. Hydrol. 499: 140-145.

Whitman, R.L., Przybyla-Kelly, K., Shively, D.A., Nevers, M.B., Byappanahalli, M.N., 2008. Sunlight, season, snowmelt, storm, and source affect *E. coli* populations in an artificially ponded stream. Sci. Total Environ. 390(2–3): 448-455.

Wilkinson, J., Jenkins, A., Wyer, M., Kay, D., 1995. Modelling faecal coliform dynamics in streams and rivers. Water Res. 29(3): 847-855.

World Health Organization (WHO), 2011. Guidelines for drinking-water quality. http://www.who.int/water_sanitation_health/dwq/guidelines/en/ (May 11, 2015).

Zyman, J., Sorber, C.A., 1988. Influence of simulated rainfall on the transport and survival of selected indicator organisms in sludge-amended soils. J. Water Pollut. Control. Fed. 60: 2105-2110.