Modeling the interannual variability of microbial quality metrics of irrigation water in a Pennsylvania stream

Eun-Mi Hong^{a, b, 1}, Daniel Shelton^a, Yakov A. Pachepsky^a, Won-Ho Nam^c, Cary Coppock^a,
Richard Muirhead^d

Abstract

Knowledge of the microbial quality of irrigation waters is extremely limited. For this reason, the US FDA has promulgated the Produce Rule, mandating the testing of irrigation water sources for many farms. The rule requires the collection and analysis of at least 20 water samples over two to four years to adequately evaluate the quality of water intended for produce irrigation. The objective of this work was to evaluate the effect of interannual weather variability on surface water microbial quality. We used the Soil and Water Assessment Tool model to simulate *E. coli* concentrations in the Little Cove Creek; this is a perennial creek located in an agricultural watershed in south-eastern Pennsylvania. The model performance was evaluated using the US FDA regulatory microbial water quality metrics of geometric mean (GM) and the statistical threshold value (STV). Using the 90-year time series of weather observations, we simulated and randomly sampled the time series of *E. coli* concentrations. We found that weather conditions of a specific year may strongly affect the evaluation of microbial quality and that the long-term assessment of microbial water quality may be quite different from the evaluation based on short-term observations. The variations in microbial concentrations and water quality metrics were affected by location, wetness of the hydrological years, and seasonality, with 15.7 – 70.1% of samples exceeding the regulatory threshold. The results of this work demonstrate the value of using modeling to design and evaluate monitoring protocols to assess the microbial quality of water used for produce irrigation.

Keyword: Food Safety Modernization Act; Irrigation water; Microbial quality; Water quality metrics

1. Introduction

It is currently estimated that approximately 76 million people in the U.S. become ill from foodborne diseases. Over 40% of these cases are linked to fresh produce (Painter et al., 2013). Fresh fruits and vegetables are an increasingly recognized source of foodborne disease outbreaks, not only in many parts of the U.S. but also in several regions of the world (Anderson et al., 2011; Painter et al., 2013; Ceuppens et al., 2014). One of the causes of foodborne diseases is the contamination of fresh produce by pathogens originating from animal waste or wastewater (Oliveira et al., 2012; Liu et al., 2013; Strawn et al., 2013; Holvoet et al., 2014; Hruby et al., 2016). According to many studies, irrigation water can be a significant source of pathogens during the growing season (Wood et al., 2010; Gelting et al., 2011; Pachepsky et al., 2011; Micallef et al., 2012; Park et al., 2013; Ceuppens et al., 2014). Surface water is more vulnerable to microbial contamination than the groundwater due to runoff, direct fecal deposition and growth in bottom sediments (Pagadala et al., 2015). Since measurement of actual pathogen levels is expensive and laborious, the enteric bacterium *Escherichia coli* (*E. coli*) is frequently used as an indicator to evaluate health risks related to water (Edberg et al., 2000; Kay et al., 2004; Chandrasekaran et al., 2015).

To minimize the risks associated with irrigation waters, various government regulators have suggested guidelines and standards for microbial water quality (WHO, 1989; Davies and Mazumder, 2003; Wood et al., 2010; Pachepsky et al., 2011; Edge et al., 2012). In the U.S., the Food Safety Modernization Act (FSMA) has empowered the US Food and Drug Administration (FDA) to promulgate rules to improve the safety of produce. One of the rules developed by the FDA (Produce Rule) specifies regulations for surface irrigation water based on two metrics: the geometric mean (GM) of *E. coli* concentrations and the statistical threshold value (STV) of those concentrations. The GM reflects the central tendency of water quality and its threshold value is 126 CFU *E. coli* per 100 ml. The STV reflects the variability of the water quality caused by

adverse conditions such as extreme precipitation or high water streamflow and represents the concentration at the 90% probability. The threshold value of STV is 410 CFU *E. coli* per 100 ml (USFDA, 2016a; 2016b). The FDA FSMA regulations consider untreated surface water as the most vulnerable (i.e., prone to contamination) since this water is routinely used to grow produce. The US FDA suggests an initial monitoring with at least 20 water samples over the course of two to four consecutive years to determine if the water meets the required criteria for microbial water quality to establish site-specific benchmark values using these water quality metrics. After the initial monitoring, a minimum of five samples per year is required for creating a new time-series dataset of 20 samples. The water cannot be used for irrigation if the microbial water quality metrics exceed these threshold values (Emch and Waite-Cusic, 2016; USFDA, 2016a; 2016b).

Currently, information about the microbial quality of irrigation water is not collected in a systematic way (Pachepsky et al., 2012). Benjamin et al. (2013) found that 14% of water samples from reservoirs and 17% water samples collected in the field contained *E. coli* in concentrations above the Good Agricultural Practices limit of 126 CFU/100 ml. Draper et al. (2016) collected 153 samples from surface water used for irrigation in southeastern Pennsylvania and found that although the concentrations of *E. coli* were low in most samples, 24% had *E. coli* concentrations which exceeded 126 CFU/100 ml. Edge et al. (2012) observed that *E. coli* exceedance rates were significantly higher in streams running through agricultural fields as opposed to streams in predominantly in non-agricultural areas.

It has been previously observed that surface water microbial quality is generally correlated with weather conditions (Ferguson et al., 2003; McCarthy et al., 2012; Strauch et al., 2014; Villarini and Strong, 2014; Ceuppens et al., 2014; Coffey et al., 2016). For example, Coffey et al., (2016) examined the effect of weather parameters on streamflow and microbial water quality under different climate scenarios for the west of Ireland; while Cha et al. (2016) conducted a similar study under synthetic weather scenarios. Increases in amount, intensity or

1 frequency of rainfall can impact microbial transport resulting in higher levels of surface water
2 contamination (Liu et al., 2013; Coffey et al., 2016). Consequently, the results of microbial water
3 quality benchmarking may be affected by weather conditions for a given land use, soil, or water
4 source. To the best of our knowledge, however, little research has yet been published that focuses
5 on the interannual variability of metrics for microbial water. Such analysis currently is possible at
6 the watershed-scale using water quality models.

7
8 The Soil and Water Assessment Tool (SWAT) is a semi-distributed and process-based
9 watershed model which operates with a daily time step. It was developed to investigate the
10 impacts of management on water quality and nonpoint-source pollution. It can simulate changes
11 in water quality over long time periods (Gassman et al., 2007; Arnold et al., 2012). The SWAT
12 bacteria fate and transport subroutine (Sadeghi and Arnold, 2002) simulates grazing land and
13 manure application-associated inputs of bacteria, growth and die-off as affected by temperature,
14 as well as runoff transport to surface water sources affected by rainfall and/or irrigation. The only
15 in-stream process in the original SWAT bacteria module was temperature-dependent bacterial
16 die-off. Kim et al. (2010) modified this bacteria module to include sediment-associated bacteria
17 transport in streams, thus allowing for evaluation of streambed bacteria release and deposition.

18 The objective of this work was to analyze interannual patterns of microbial water quality
19 for irrigation purpose as affected by weather conditions in a small watershed. The watershed-
20 scale water quality model SWAT was used to simulate water flow and *E. coli* fate and transport.

21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65

2. Materials and Methods

2.1 Study Area

This study was conducted in the Little Cove Creek watershed, located in south eastern Pennsylvania (see Fig. S1 in the Supplementary Material). It is surrounded by the Appalachian Mountains and is part of the Chesapeake Bay watershed (Kim et al., 2010). The headwater of the

1 watershed consists of 18.8 ha of forest (93%) and pastures (7%) and extends 7.2 km downstream
2 to the first sampling site (CM2). The middle section of the watershed extends 12.1 km
3 downstream to the second sampling site (CM3); land use includes 18.8 ha of forest (69%), crop
4 agriculture (2%), and pasture (28%). The final section of the watershed extends 17.7 km
5 downstream to the last sampling site (CM7); land use consists of 31.1 ha of forest (62%), crop
6 agriculture (2%), and pasture (36%).
7
8
9
10
11
12
13
14
15
16

17 **2.2 Monitoring data acquisition**

18
19 The *E. coli* concentrations in stream water were monitored weekly at the three monitoring
20 sites (CM2, CM3, and CM7) in 2008-2009. In order to calculate the streamflow at each site, the
21 water depth and velocity were recorded at 30-minute intervals using standard United States
22 Geological Survey (USGS) methods (Rantz et al., 1982). The 30-min streamflow was processed
23 using polynomial regression functions and averaged to a daily streamflow (a day being the
24 minimum time step required for SWAT simulations). The CM7 monitoring site did not have data
25 for the last 20 days of the monitoring period. The stream water samples collected at each
26 monitoring site were transported on ice to the Beltsville Agricultural Research Center (BARC),
27 Beltsville, Maryland, or to the Wilson College Biology Department in Chambersburg,
28 Pennsylvania, where generic *E. coli* bacteria in stream water were quantified within 24 hours as
29 previously described (Kim et al., 2010).
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48

49 **2.3 SWAT model application**

50
51 The SWAT model has inputs for weather, hydrology, soil, crop, nutrients, bacteria and
52 pathogens. The model divides watersheds into multiple sub-watersheds with subdivided
53 hydrologic response units (HRUs), which consist of homogeneous land use, management, and
54 soil characteristics. In this study, the modified SWAT module (Kim et al., 2010) was applied in
55
56
57
58
59
60
61
62
63
64
65

the Little Cove Creek, Pennsylvania. For evaluating model performance of simulated streamflow, the Nash and Sutcliffe model Efficiency (NSE) was applied. The performance of the bacteria module was evaluated based on compliance with FDA regulations: false warning, failure to warn, correct warning, and correct no warning.

SWAT requires spatial data input such as a digital elevation map (DEM), land use, soil data, vegetation data, etc. In this study, the 30-m DEM was obtained from the USGS and used to delineate the watershed boundaries and topography (USGS, 2016b). The land use data were acquired and characterized from the USGS National Land Cover Data 1992 (USGS, 2016a). The major land uses of the study area are forest (72%) and pasture (25%). The major source of bacteria in the watershed was assumed to be from fecal deposition of cattle on pasture lands. Assuming grazing occurred from April through October (approx. 210 days) and a stocking rate of 2.5 cows/ha (based on interviews), the mass of manure (dry weight) spread was estimated to be 22.1 kg/ha/day (ASABE, 2005; Kim et al., 2010). The concentration of *E. coli* in feces/manure was estimated to be 4×10^7 CFU/g (CFU = Colony Forming Unit) per gram dry weight based on the report of the ASABE (2003, 2005). The death rates of *E. coli* in the surface runoff and stream water at 20 °C were assumed to be 0.659 and 0.990 /day, respectively (Crane and Moore's, 1986; Kim et al., 2010). The SWAT Soil Survey Geographic (SSURGO) database of the U.S. Department of Agriculture (USDA) Natural Resources Conservation Service (NRCS) was utilized for describing the soil characteristics (USDA, 2016).

Ninety years of historical meteorological data (precipitation, minimum, and maximum temperatures) from four weather stations near the study area were utilized to evaluate the effect of weather on microbial quality. Daily precipitation and temperature data were obtained from the National Oceanic and Atmospheric Administration (NOAA), the National Centers for Environmental Information (<http://www.ncdc.noaa.gov/oa/ncdc.html>).

2.4. Microbial water quality metrics

The US FDA Produce Rule contains two metrics of microbial water quality: the geometric mean (GM) and the statistical threshold value (STV) of *E. coli* concentrations. The GM threshold value is 126 CFU *E. coli* per 100 ml and the threshold value of STV is 410 CFU *E. coli* per 100 ml (USFDA, 2016a; 2016b).

Leafy green production in PA occurs primarily between April and September (<http://www.almanac.com/gardening/planting-dates/PA/Chambersburg>). Therefore, we considered the growing season to extend from April to September and the non-growing season from October to March. Based on the FDA FSMA, we applied the following two cases to evaluate the metrics of the microbial water quality, 1) randomly collect five samples each year over the whole growing season and calculate GM and STV using a dataset of 20 samples (4-year dataset), or 2) randomly collect five samples each month between April and September, and calculate the GM and STV using a dataset of 20 samples per month. The above random sampling was repeated 1000 times.

3. Results

3.1 Overview of weather data

Table S1 shows the descriptive statistics of monthly and annual average temperature and total precipitation over 90 years in the study area. The average annual temperature was 10.4 °C, with maximum values occurring in July (22.5 °C) and minimum values in January (-2.1°C). It was typically higher until the early 1930s with the highest value in 1931 (12.5 °C), and then tended to decrease until the late 1970s with the lowest average calculated for 1978 (9.0 °C). More recently, it has increased to 11.9 °C in 2012. The monthly averaged temperature ranged from 9.7 – 22.5 °C during the growing season and -2.1 – 11.5 °C during the non-growing season, with standard deviations ranging between 1.1 – 1.6 °C and 1.5 – 2.7 °C, respectively. The 90-year

average annual precipitation was 899.6 mm with a minimum in February (58.3 mm) and a maximum in May (85.5 mm). The annual precipitation showed a tendency to decrease until the early 1970s and then to increase from the early 1970s to 2012, with the lowest precipitation recorded in 1930 (494.1 mm) and the highest in 1996 (1348.0 mm). The monthly precipitation during the growing season, representing 69% of the annual precipitation, was generally higher than that measured in the non-growing season; however, there were few differences in the monthly average precipitation from month to month. The standard deviation of the monthly precipitation (33.4 – 48.3 mm) and of the daily rainfall intensity (5.28 – 6.05 mm/day) during the growing season was higher than that of the monthly precipitation (29.4 – 49.2 mm) and of the rainfall intensity (4.33 – 5.82 mm/day) during the non-growing season.

3.2 SWAT model performance

3.2.1 Streamflow

In this study, we applied the hydrologic parameters as calibrated by Kim et al. (2010) and evaluated the performance of the hydrology module using the monitored and simulated streamflow data. According to the Kim et al. (2010), twenty-seven hydrologic parameters were selected and utilized for sensitivity analysis, while nine parameters were selected for calibration. Table 1 shows the hydrological primary parameters and the calibrated values. Curve Number II (CN2) was the dominant parameter affecting the surface flow, while ALPHA_BF (baseflow alpha factor) was the dominant parameter affecting subsurface flow (Van Liew et al., 2007). The hydrological characteristics of the Little Cove Creek watershed reflect the fact that soil was highly permeable, resulting in rapid recharge of baseflow and groundwater. Figure 1 shows the temporal variations of the observed and predicted daily streamflow at the three stream monitoring sites in 2008. As shown in Figure 1, despite an overestimation of the streamflow under baseflow conditions at CM7 site, the simulated and observed rainfall streamflow response

were generally consistent. Overall, the SWAT model showed good performance between observed and simulated with 0.502, 0.429, and 0.455 values of Nash and Sutcliffe model Efficiency (NSE) at CM2, CM3, and CM7 respectively.

Fig.1. Temporal variations of the observed (solid line) and predicted (dashed line) daily streamflow at three stream monitoring sites

Table 1 Ranges, values and definitions of hydrological and biological parameters of the SWAT model (after Kim et al., 2010)

Parameters	Range	Applied values	Definition
Hydrology parameters			
CN2	-70 – 50	35 – 46 (-57) ^a	Initial SCS runoff curve number for moisture condition II
SOL_K	-50 – 50	4.6 – 45.9 (-34) ^a	Saturated hydraulic conductivity (mm/hr)
ALPHA_BF	0 – 1	1	Baseflow alpha factor – baseflow recession constant
ESCO	0-1	1	Soil evaporation compensation factor
SOL_AWC	-50 – 50	0.12 – 0.41 (48) ^a	Available water capacity of the soil layer (mm H ₂ O/mm soil)
CANMX	0-15	11	Maximum canopy storage (mm H ₂ O)
Bacteria parameters			
BACT_SWF	0 – 1	0.97	Fraction of manure applied to land areas that has active colony forming units
BACTKDDB	0 – 1	0.36	Bacteria partition coefficient in manure
BIO_MIN	50 – 300	56	Minimum plant biomass for grazing (kg/ha)
FILTERW	0 – 2	1.5	Width of edge-of-field filter strip (m)
PHU_PLT	1,000 – 2000	1,880	Total number of heat units or growing degree days needed to bring plant to maturity (days)
BIO_EAT	10-100	70	Dry weight of manure deposited daily (kg/ha/day)

^a parameters are varied based on the relative change (percentage)

3.2.2 *E. coli* concentrations in the stream

In this study, we utilized the bacteria-related parameters as calibrated by Kim et al. (2010); other parameters were set at the default values of SWAT. Using the modified bacteria module of Kim et al. (2010) provided improved results compared to the original SWAT bacteria module. The most sensitive parameters for *E. coli* simulation were minimum plant biomass for grazing (BIO_MIN), fraction of manure to land areas that had active CFUs and bacteria partition coefficient in manure (BACT_SWF), and total number of growing degree days to bring plant to maturity (PHU_PLT). Table 1 shows the bacteria parameters with ranges, values, and definitions which were used in this study. Figure 2 shows the probability distribution of differences between the logarithms of the simulated and observed *E. coli* concentrations in the stream. Table 2 shows the model performance in terms of compliance with FDA regulations. During the growing season (from April to September), simulation results were generally good, with the exception of CM3 where *E. coli* concentration were underestimated.

Fig.2. Probability distribution of the deviations of the simulated (C_{sim}) from the observed (C_{obs}) *E. coli* concentrations in stream water over the period 2008-2009

Table 2 Model performance in terms of compliance to the FDA regulations

Location	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
CM2												
2008	FAW ^b	FW ^a	FW	CNW	CW	CW	CW	FAW	CW	FAW	FAW	CNW
2009	CW ^c	CNW ^d	CNW	CW	CW	CW	CW	CW	CW	CW	CW	CW
CM3												
2008	FAW	CNW	FW	FW	CW	CW	CW	CW	CW	FAW	FAW	FAW
2009	CW	FW	FW	CW	CW	CW	CW	CW	CW	CW	CW	CW
CM7												
2008	CW	FW	FW	FW	CW	CW	CW	CW	CW	CNW	FAW	FAW
2009	FW	FW	FW	FW	CW	CW	CW	CW	FAW	CW	CW	CW

^aFW : False Warning^bFAW : Failure to Warn^cCW : Correct Warning^dCNW : Correct No Warning

3.3 Interannual variability of *E. coli* concentration as affected by weather conditions

The simulated total streamflow in the growing season averaged over 90 years was 34, 65, and 122 m³/s, respectively at CM2, CM3, and CM7. Figure S2 (see Supplementary Material) shows the interannual variations of the GM and STV *E. coli* during the growing season for the 90 years of available meteorological data. The average GM and STV of *E. coli* concentrations in the watershed increased slightly, along with an increase in the total streamflow, at all three monitoring sites. The companion table S2 (see Supplementary Material) contains means and extreme values of *E. coli* concentrations. While the total streamflow doubled from CM2 to CM3 and from CM3 to CM7, the average GM of *E. coli* concentrations increased only slightly from CM2 to CM3 and there was little change from CM3 to CM7. Extreme GM values at CM7, however, were much wider than at CM3. Average STV values were more than two times larger at CM3 as compared with CM2, and at CM7 as compared with CM3. Maximum STV almost tripled from CM2 to CM3 and from CM3 to CM7.

There were significant differences in microbial water quality under different precipitation and temperature conditions. Cumulative probability distribution functions of *E. coli* concentrations at CM7 for years with different precipitation and temperature are shown in Figure 3. The curve of *E. coli* cumulative probability in the stream for the driest time of the year (between the minimum and the lowest 25% of total precipitation) is located on the left while the curve for the wettest time of the year (between the highest 25% and the maximum of total precipitation) is located far to the right (Fig. 3a). In addition, increases in the daily streamflow during the growing season correlated with increases of GM and STV of *E. coli* concentrations. Generally, the years with a low average streamflow tended to display low GM and STV of *E. coli* concentrations. Despite a slight correction of *E. coli* concentrations with the average temperature during the growing season, there was little difference in *E. coli* concentrations due to temperature variation (Fig. 3b).

(a) Cumulative probability with different precipitation

(b) Cumulative probability with different temperatures

Fig.3. Cumulative probability distribution of *E. coli* concentrations at CM7 as affected by precipitation and temperature conditions (min-1st Q: *E. coli* concentration between minimum and 1st quartile of precipitation or temperature, 1st Q - avg: *E. coli* concentration between 1st quartile and average of precipitation or temperature, avg-3rd Q: *E. coli* concentration between average and 3rd quartile of precipitation or temperature, 3rd Q max: *E. coli* concentration between 3rd quartile and maximum of precipitation or temperature. Two vertical lines are drawn to show the FDA GM threshold value of 126 CFU/100 ml, and the FDA STV threshold value of 410 ml

Figure 4 shows the seasonality pattern of *E. coli* concentrations in the stream water at CM2, CM3, and CM7, with different streamflows: P1 (period between minimum and 10th percentile of daily streamflow), P2 (period between 10th percentile and median of daily streamflow), P3 (period between median and 90th percentile of daily streamflow), and P4 (period between 90th percentile and maximum of daily streamflow). *E. coli* concentrations generally increased with the increase in daily streamflow. The average total monthly streamflow in April was the highest and showed a tendency to decrease in August. Mean monthly rainfall from April to September gradually decreased circa 10% (from 85 to 76 mm), but the percentage of the decrease in stream flow was much larger (Table S1). Although the flow decreased as the growing season progressed from April to June, the average of *E. coli* concentration, particularly with high flow, increased from month to month. Over the growing season, there was a trend of increasing average and standard deviation *E. coli* concentrations from P1 to P2, from P2 to P3, and from P3 to P4. However, there were little monthly differences of minimum *E. coli* concentration with flow quartile and there was little dependence of maximum *E. coli* concentration on the flow quartile except P1 flow (see Table S3 in the Supplementary Material).

(a) CM2

(b) CM3

(c) CM7

- P1** : the period between minimum and 10th percentile of daily streamflow
- P2** : the period between 10th percentile and median of daily streamflow
- P3** : the period between median and 90th percentile of daily streamflow
- P4** : the period between 90th percentile and maximum of daily streamflow

Fig.4. Seasonal pattern of *E. coli* concentrations in the stream at three different sites with different streamflows; box plots the median and whiskers (5th-95th percentiles) with error bars.

3.4 Metrics of the microbial water quality during the growing season

Figure 5 describes the GM and STV characteristics in 20 samples over four consecutive years (n=1000) for each month and for the entire growing season over 90 years (see Table S4 in the Supplementary Material). The average GM was 82 and 121 CFU/100 ml at CM2 and CM3, respectively. These values were lower than the FDA standard of 126 CFU/100 ml, while they were higher at CM7, with 141 CFU/100 ml. The average STV at CM2 was 263 CFU/100 ml, which was lower than the threshold value from the FDA standard of 410 CFU/100 ml. At CM3 and CM7, however, the STV was equal to 511 and 489 CFU/100 ml, respectively, which was comparable to or significantly higher than the FDA threshold.

Fig.5. Monthly geometric mean (GM) and statistical threshold value (STV) variability of *E. coli* concentrations in the stream at the monitored sites over the growing season (GS); box plots the median and whiskers (5th-95th percentiles) with error bars, Horizontal lines are drawn to show the FDA GM threshold value of 126 CFU/100 ml and the FDA STV value of 410 CFU/100 ml.

The average monthly GM for 20 random samples over 4 consecutive years was the highest in April or May at all three observation locations, and it reached the minimum in September. The standard deviation of the monthly GM tended to increase from April through June and decreased in September. The maximum value of the monthly GM was the highest in June. At CM2, the average monthly GM did not exceed the FDA threshold, whereas the average monthly GM from April to June at CM3 and from April to July at CM7 exceeded this threshold

The average STV was the lowest in April and, unlike the GM, did not exceed the FDA regulation. The STV was mostly greater in spring than during the summer, with maximum values in June. At CM2, even the maximum monthly STV except in April exceeded the regulation threshold; however, the average STV during the growing season did not exceed the FDA threshold. The monthly average STV from June to September at CM3 and CM7, respectively, exceeded the threshold values set by the FDA.

3.5 Exceedance rates for microbial water quality

Table 3 shows the exceedance rate for the five random samples collected for four consecutive years (total 20 samples, n=1000), based on the FDA regulation. According to model simulations, the monthly *E. coli* concentrations were above the GM or STV threshold level 8.6-34.3%, 52.3-89.4%, and 55.9-99.5% of the time at CM2, CM3, and CM7 respectively during the growing season for the 90 years of available meteorological data (Table 3). Since the CM2 headwaters site had lower levels of grazing-related *E. coli* contamination, it was least likely to exceed the FDA threshold for GM or STV. With the exception of GMs for April and May, exceedance rates were consistently less than 20%. Exceedance rates for CM3 and CM7, both of which included grazing cattle in the watershed, were generally higher. Exceedance rates for GMs were high in April and May, and then decreased through the remainder of the growing season. In comparison, exceedance rates for STVs were low in April (approx. 1-5%), increased dramatically

through June (approx. 71-80%), and then decreased gradually through the remainder of the growing season.

Table 3 Exceedance rates (%) for microbial water quality during each month of the growing season and over the entire growing season

Location & Microbial water quality metric		Exceedance rates (%)						
		Apr	May	Jun	Jul	Aug	Sep	Growing season
CM2								
GM ^a		8.6	8.1	1.6	0.3	0.3	0.1	0.7
STV ^b		0	3.8	19.8	18.6	20.6	16.0	12.4
GM and STV		0	9.8	13	1.9	1.9	0.5	2.6
Total		8.6	21.7	34.4	20.8	22.8	16.6	15.7
CM3								
GM		88.2	60.9	12.3	2.1	0.1	0.1	13.9
STV		0	0.1	7.1	31.5	43.3	42.7	14.9
GM and STV		0	28.4	61.7	24	12.7	9.5	26.6
Total		88.2	89.4	81.1	57.6	56.1	52.3	55.4
CM7								
GM		96.0	64.0	27.1	19.3	8.9	6.8	26.0
STV		0	0.3	10.5	31.4	40.5	41.9	8.3
GM and STV		2.3	35.2	59.5	20.6	9.2	7.2	35.8
Total		98.3	99.5	97.1	71.3	58.6	55.9	70.1

^a GM= Geometric Mean

^b STV= Statistic Threshold Value

There were marked differences between dynamics of exceedance rates at the three monitoring sites. At CM2, the exceedance rate during the growing season was 15.7%. STVs were higher than the FDA threshold value in 12.4% of the cases. More precisely, 8.6% of the samples in April exceeded the GM limit, but *E. coli* concentrations in the stream satisfied the threshold STV. In June, the exceedance rate for microbial water quality was the highest with 19.8% for the STV and 13.0% for both the STV and GM limit. From July to September, the exceedance rate was similar to that estimated between April and June, however, in most cases this was due to STVs above the threshold (16.6 – 22.8%). At CM3, the exceedance rate during the growing season was 55.4%, while it reached up to 88.2% in April and 60.9% in May for the GM limit, even though it complied with the STV. The exceedance rate in June was 81.1%, which was mostly caused by the STV exceeding the limit. This rate decreased from 67.8 to 52.2% between July and September. The exceedance rate from July to September was primarily due to STVs. At CM7, the downstream watershed with approximately 36% of pasture, the exceedance rate during the growing season was 70.1%. Over 90% of samples at this site did not comply with the water quality standards from April to June. Most measurements in April and May exceeded the GM limit of *E. coli* concentrations, and 59.5% of samples in June did not satisfy the regulations of both the GM and STV thresholds. From July to September, the exceedance rate was lower than in spring season (55.9 – 71.3%), but in most cases, thresholds regulations were exceeded due to high STVs.

4. Discussion

The modified watershed-scale water quality model SWAT was applied to evaluate the long-term outlook for microbial water quality compliance with FDA irrigation water standards. The use of microbial water quality compliance metrics as a measure of model performance distinguishes this work from earlier microbial water quality modeling studies where traditional model accuracy measures like root-mean-squared error or Nash-Sutcliffe efficiency were utilized (Parajuli et al., 2009; Cho et al., 2012). The microbial water quality compliance metrics, GM of concentrations and 90% quartile (STV), are two statistics derived from the statistical distribution of logarithms of *E. coli* concentrations. Using the comparison of statistical distribution of simulated and measured concentrations to evaluate model performance was proposed by Iudicello and Chin (2013). However, rather than using the entire concentration distributions, we compare two quantiles of these distributions with regulation thresholds and compute frequencies of threshold exceedance.

We found a good correspondence between simulated and measured GM and STV threshold exceedance rates (Table 2). Small numbers of false positives (simulated values) and false negatives (measured values) were found in April and September. The compliance rates during most of growing season were accurately predicted.

The interannual variations in microbial water quality metrics demonstrate that there is substantial spatio-temporal variability. Significant differences in compliance were observed at the three monitoring locations, even though they were separated by relatively small distances (7-10 km). Compliance was impacted by monitoring location, the wetness of the hydrological years, and the seasonality. Monitoring sites along the creek were chosen to reflect different land uses in the watershed; in particular the transition from forest (CM2) to grazing agriculture (CM3 and CM7). The deterioration in microbial water quality was a function of the increased input of *E. coli* from grazing cattle due to direct fecal deposition and/or runoff. The differences in monthly

1 median GM values at each site (CM2, CM3, and CM7) were statistically significant ($P < 0.001$).
2
3 Significant changes of microbial water quality over relatively small distances have been
4
5 previously documented (Stocker et al., 2016). Such variation indicates microbial water quality
6
7 estimations can vary dramatically as a function of changes in land use and/or point source
8
9 contamination.
10

11
12 Wetness/rainfall also controls microbial water quality (Fig. 3a and 4, Chandrasekaran et
13
14 al., 2015). Intense precipitation increases surface runoff and enhanced transport of indicator
15
16 bacteria from pasture to streams (Castillo et al., 2004; Donnison and Ross, 2009; Liu et al., 2013).
17
18 The concentrations of bacteria in watersheds have been previously observed to increase
19
20 following heavy rainfall (Goyal et al., 1977; Kistemann et al., 2002; Cooley et al., 2007;
21
22 Hodgson et al., 2016). Consequently, the probability of non-compliance with water quality
23
24 standards is higher. However, since the need for irrigation is reduced, or eliminated, when there
25
26 is adequate rainfall, the issue of non-compliance may be a moot point. This illustrates the need to
27
28 couple a watershed-scale microbial water quality model with a reliable crop produce model to
29
30 predict the demand for irrigation. This presents an interesting avenue for future research, and
31
32 requires introduction of socioeconomic information on the strategy that farmers use to decide on
33
34 applying irrigation. We note that indicator bacteria concentrations at high flows can be a
35
36 prominent influence of irrigation water quality if the water is not immediately used in field
37
38 irrigation, e.g. is used to replenish reservoirs of water for future use or directed to greenhouses
39
40 (Holvoet et al., 2014).
41
42
43
44
45
46
47

48
49 Interannual patterns were found for both microbial water quality metrics. In April and
50
51 May, the GM of the stream *E. coli* concentrations was higher than in other months and there were
52
53 little differences among *E. coli* concentrations during periods of different flow rates. At CM3 and
54
55 CM7 sampling sites, the exceedance rate as determined by GM was approximately 90% in April
56
57 and 60% in May. The two microbial water quality metrics had different interannual patterns of
58
59
60
61
62
63
64
65

1 exceedance. In April, microbial water quality was worse than in the following months and was
2 independent of flow rate. Management decisions to increase the time between irrigation and
3 harvest, may be different for April as opposed to June. The exceedance rate peaked in June
4 including high rates of both GM and STV as opposed to the spring when only GM exceeded the
5 regulatory threshold. *E. coli* concentrations in June were the highest as compared to other months,
6 even though the June stream flow rates were smaller than spring stream flow rates. This could be
7 caused by the interplay of greater microbial loading and the lower dilution associated with a low
8 flow rate, which may yield high microbial concentrations (Coffey et al., 2016). Overall, June was
9 the most vulnerable month for non-compliance with water quality standards. From July to
10 September, the exceedance rate was smaller than between April and May, and in most cases was
11 affected primarily by non-compliance of STVs. The interannual pattern of the microbial water
12 quality may require accounting for seasonality in produce crop management.

13
14 It has to be noted that when we collected 20 random samples over four consecutive years
15 for simulations, we did not distinguish between days of rainfall, antecedent rainfall, rainfall
16 recessions, or baseflow. The irrigation water requirement can be different for these hydrological
17 situations and also differ between crop species, rooting depth and crop growth stage and water
18 availability (Hong et al., 2016). The methodology of this study can be applied for specific crops
19 and hydrological situations when the crop module in SWAT is developed and tuned to this crop
20 and can be applied for irrigation strategy of a specific agricultural water management.

21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65

5. Conclusions

66 This study aimed to analyze patterns of the microbial quality of a stream water body
67 using the SWAT watershed model. Watershed-scale modeling generally provided reliable
68 information on compliance of waters to water quality regulatory requirements. Our modeling
69 results showed that substantial interannual variability in microbial water quality metrics can be

1 expected due to differences in location, weather pattern, and seasonality. *E. coli* concentrations
2 increased following heavy rainfall and in particular during wet growing seasons more so than dry
3 growing seasons. High *E. coli* GMs were the dominant reason in non-compliance with water
4 quality standards in the spring (April and May), while high STV were the dominant reason in
5 non-compliance in the summer/fall (July to September). June was determined to be the most
6 critical month for water quality management since in which both high *E. coli* GM and high *E.*
7 *coli* STV were reasons of non-compliance.
8
9

10 According to our findings, the results of the evaluation of the microbial water quality
11 depend on the sampling and irrigation schedule. The results are useful in determining the role and
12 the impact of these factors; and this opens an interesting perspective for future research. It is
13 necessary to take into account the exact location and timing of the sample collection in the
14 benchmarking process of irrigation water quality to develop efficient practices of characterizing
15 the microbial quality of irrigation waters.
16
17

18 **Acknowledgments**

19 This research was supported in part by an appointment to the Agricultural Research Service
20 (ARS) Research Participation Program administered by the Oak Ridge Institute for Science and
21 Education (ORISE) through an interagency agreement between the U.S. Department of Energy
22 (DOE) and the U.S. Department of Agriculture (USDA). ORISE is managed by ORAU under
23 DOE contract number DE-AC05-06OR23100.
24
25

26 **References**

27 Anderson, M., Jaykus, L.A., Beaulieu, S., Dennis, S., 2011. Pathogen–produce pair attribution risk
28 ranking tool to prioritize fresh produce commodity and pathogen combinations for further evaluation
29 (P³ARRT). *Food Control* 22(12): 1865–1872.
30
31

- Arnold, J.G., Moriasi, D.N., Gassman, P.W., Abbaspour, K.C., White, M.J., Srinivasan, R., Santhi, C., Harmel, R.D., van Griensven, A., Van Liew, M.W., Kannan, N., Jha, M.K., 2012. SWAT: Model Use, Calibration, and Validation. *Transactions of the ASABE*. Vol. 55(4): 1491-1508.
- ASABE (American Society of Agricultural and Biological Engineers), 2003. Manure Production and Characteristics. ASAE D384.1. February 2003. ASABE (American Society of Agricultural and Biological Engineers), 2005. Manure Production and Characteristics. ASAE D384.2. March 2005.
- Benjamin, L., Atwill, E.R., Jay-Russell, M.J., Cooley, M., Carychao, D., Gorski, L., Mandrell, R.E., 2013. Occurrence of generic *Escherichia coli*, *E. coli* O157 and *Salmonella* spp. in water and sediment from leafy green produce farms and streams on the Central California coast. *Int. J. Food Microbiol.* 165(1): 65-76.
- Castillo, A., Mercado, I., Lucia, L.M., Martinez-Ruiz, Y., Ponce de Leon, J., Murano, E.A., Acuff, G.R., 2004. Salmonella contamination during production of cantaloupe: a binational study. *J. Food Prot.* 67(4), 713–720.
- Ceuppens, S., Hessel, C.T., de Quadros Rodrigues, R., Bartz, S., Tondo, E.C., Uyttendaele, M., 2014. Microbiological quality and safety assessment of lettuce production in Brazil. *Int. J. Food Microbiol.* 181(2): 67-76.
- Cha, Y., Park, M.H., Lee, S.H., Kim, J.H., Cho, K.H., 2016. Modeling spatiotemporal bacterial variability with meteorological and watershed land-use characteristics. *Water Res.* 100:306-315.
- Chandrasekaran, R., Hamilton, M.J., Wang, P., Staley, C., Matteson, S., Dirr, A., Sadowsky, M.J., 2015. Geographic isolation of *Escherichia coli* genotypes in sediments and water of the Seven Mile Creek — A constructed riverine watershed. *Sci. Total Environ.* 538: 78-85.
- Coffey, R., Benham, B., Wolfe, M.L., Dorai-Raj, S., Bhreathnach, N., O’Flaherty, V., Cormican, M., Cummins, E., 2016. Sensitivity of streamflow and microbial water quality to future climate and land use change in the West of Ireland. *Reg. Environ. Chang.*: 1-18.
- Cooley, M., Carychao, D., Crawford-Miksza, L., Jay, M.T., Myers, C., Rose, C., Keys, C., Farrar, J., Mandrell, R.E., 2007. Incidence and tracking of *Escherichia coli* O157:H7 in a major produce production region in California. *PLoS One* 2(11): e1159.

- Crane, S.R., Moore, J.A., 1986. Modeling enteric bacterial die-off – a review. *Water Air Soil Pollut.* 27(3): 411–439.
- Davis, J.M., Mazumder, A., 2003. Health and environmental policy issues in Canada: the role of watershed management in sustaining clean drinking water quality at surface sources. *J. Environ. Manage.* 68: 273-286.
- Donnison, A., Ross, C., 2009. Survival and retention of *Escherichia coli* O157:H7 and *Campylobacter* in contrasting soils from the Toenepi catchment. *New Zeal J Agr Res* 52: 133-144.
- Draper, A.D., Doores, S., Gourama, H., and LaBorde, L.F., 2016. Microbial survey of Pennsylvania surface water used for irrigating produce crops. *Journal of Food Protection* 79(6): 902-912.
- Edberg, S.C., Rice, E.W., Karlin, R.J., Allen, M.J., 2000. *Escherichia coli*: the best biological drinking water indicator for public health protection. *J. Appl. Microbiol* 88: 106S-116S.
- Edge, T.A., El-Shaarawi, A., Gannon, V., Jokinen, C., Kent, R., Khan, I.U., Koning, W., Lapen, D., Miller, J., Neumann, N., Phillips, R., Robertson, W., Schreier, H., Scott, A., Shtepani, I., Topp, E., Wilkes, G., van Bochove, E., 2012. Investigation of an *Escherichia coli* environmental benchmark for waterborne pathogens in agricultural watersheds in Canada. *J. Environ. Qual.* 41(1):21-30.
- Emch, A.W., White-Cusic, J.G., 2016. Conventional curing practices reduce generic *Escherichia coli* and *Salmonella spp.* on dry bulb onions produced with contaminated irrigation water, *Food Microbiol.* 53: 41-47.
- Ferguson, C., de Roda Husman, A.M., Altavilla, N., Deere, D., Ashbolt, N., 2003. Fate and transport of surface water pathogens in watersheds. *Crit. Rev. Env. Sci. Tec.* 33 (3), 299-361.
- Gassman, P.W., Reyes, M.R., Green, C.H., Arnold, J.G., 2007. The Soil and Water Assessment Tool: Historical Development, Applications, and Future Research Directions. *T. ASABE*. Vol. 50(4): 1211-1250.
- Gelting, R.J., Baloch, M.A., Zarate-Bermudez, M.A., Selman, C., 2011. Irrigation water issues potentially related to the 2006 multistate *E. coli* O157:H7 outbreak associated with spinach. *Agric. Water Manage.* 98(9): 1395–1402.
- Goyal, S.M., Cerba, C.P., Melnick, J.L., 1977. Occurrence and distribution of bacterial indicators and pathogens in canal communities along the Texas coast. *Appl. Environ. Microbiol.* 24(2): 139-149.

- Hodgson, C.J., Oliver, D.M., Fish, R.D., Bulmer, N.M., Heathwaite, A.L., Winter, M., Chadwick, D.R.,
2016. Seasonal persistence of fecal indicator organisms in soil following dairy slurry application to
land by surface broadcasting and shallow injection. *J. Environ. Manage.* In press.
- Holvoet, K., Sampers, I., Seynnaeve, M., Uyttendaele, M., 2014. Relationships among hygiene indicators
and enteric pathogens in irrigation water, soil and lettuce and the impact of climatic conditions on
contamination in the lettuce primary production. *Int. J. Food Microbiol.* 171: 21-31.
- Hong, E.M., Nam, W.H., Choi, J.Y., Pachepsky, Y.A., 2016. Projected irrigation requirements for upland
crops using soil moisture model under climate change in South Korea. *Agric. Water Manage.* 165:
163-180.
- Hruby, C.E., Soupir, M.L., Moorman, T.B., Shelley, M., Kanwar, R.S., 2016. Effects of tillage and
poultry manure application rates on *Salmonella* and fecal indicator bacteria concentrations in tiles
draining Des Moines Lobe Soils. *J. Environ. Manage.* 171: 60-69.
- Iudicello, J.J., Chin, D.A., 2013. Multimodel, multiple watershed examination of in-stream bacteria
modeling. *J. Environ. Qual.* 139(5): 719-727.
- Kay, D., Bartram, J., Prüss, A., Ashbolt, N., Wyer, M.D., Fleisher, J.M., Fewtrell, L., Rogers, A.I., Rees,
G., 2004. Derivation of numerical values for the World Health Organization guidelines for
recreational waters. *Water Res.* 38: 1296-1304.
- Kim, J.W., Pachepsky, Y.A., Shelton, D.R., Coppock, C., 2010. Effect of streambed bacteria release on *E.*
coli concentrations: Monitoring and modeling with the modified SWAT. *Ecol. Model.* 221: 1592-
1604.
- Kistemann, T., Claßen, T., Koch, C., Dangendorf, F., Fischeder, R., Gebel, J., Vacata, V., and Exner, M.,
2002. Microbial load of drinking water reservoir tributaries during extreme rainfall and runoff. *Appl.*
Environ. Microbiol. 68(5): 2188-2197.
- Liu, C., Hofstra, N., Franz, E., 2013. Impacts of climate change on the microbial safety of pre-harvest
leafy green vegetables as indicated by *Escherichia coli* O157 and *Salmonella* spp. *Int. J. Food*
Microbiol. 163(2-3): 119-128.
- McCarthy, D.T., Hathaway, J.M., Hunt, W.F., Deletic, A., 2012. Intra-event variability of *Escherichia*
coli and total suspended solids in urban stormwater runoff. *Water Res.* 46(20): 6661-6670.

Micallef, S.A., Rosenberg Goldstein, R.E., George, A., Kleinfelter, L., Boyer, M.S., McLaughlin, C.R., Estrin, A., Ewing, L., Jean-Gilles Beaubrun, J., Hanes, D.E., Kothary, M.H., Tall, B.D., Razeq, J.H., Joseph, S.W., Sapkota, A.R., 2012. Occurrence and antibiotic resistance of multiple *Salmonella* serotypes recovered from water, sediment and soil on mid Atlantic tomato farms. *Environ. Res.* 114: 31–39.

Pachepsky, Y., Morrow, J., Guber, A., Shelton, D., Rowland, R., Davies, G., 2012. Effect of biofilm in irrigation pipes on microbial quality of irrigation water. *Appl. Microbiol.* 54(3): 217-224.

Pachepsky, Y., Shelton, D.R., McLain, J.E.T., Patel, J., Mandrell, R.E., 2011. Irrigation waters as a source of pathogenic microorganisms in produce. A review. *Adv. Agron.* 113:73-138.

Pagadala, S., Marine, S.C., Micallef, S.A., Wang, F., Pahl, D.M., Melendez, M.V., Kline, W.L., Oni, R.A., Walsh, C.S., Everts, C.S., Buchana. C.S., 2015. Assessment of region, farming system, irrigation source and sampling time as food safety risk factors for tomatoes. *Int. J. Food Microbiol.* 196: 98-108.

Painter, J.A., Hoekstra, R.M., Ayers, T., Tauxe, R.V., Braden, C.R., Angulo, F.J., Griffin, P.M., 2013. Attribution of foodborne illnesses, hospitalizations, and deaths to food commodities by using outbreak data, United States, 1998–2008. *Emerg. Infect. Diseases.* 19(3): 407-415.

Parajuli, P.B., Mankin, K.R., Barnes, P.L., 2009. Source specific fecal bacteria modeling using soil and water assessment tool model. *Bioresour. Technol.* 100(2): 953–963.

Park, S., Navratil, S., Gregory, A., Bauer, A., Srinath, I., Jun, M., Szonyi, B., Nightingale, K., Anciso, J., Ivanek, R., 2013. Generic *Escherichia coli* contamination of spinach at the preharvest level: the role of farm management and environmental factors. *Appl. Environ. Microbiol.* 79(14): 4347–4358.

Rantz, S.E. et al., 1982. Measurement and computation of streamflow: volume 1. Measurement of Stage and Discharge. Geological Survey Water-Supply paper 2175.

Sadeghi, A.M., Arnold, J.G., 2002. A SWAT/microbial sub-model for predicting pathogen loadings in surface and groundwater at watershed and basin scales. In: Total Maximum Daily Load (TMDL) Environmental Regulations: Proceedings of the March 11–13, 2002 Conference, (Fort Worth, Texas, USA). ASAE Publication No. 701P0102. ASAE, St. Joseph, MI.

1 Stocker, M.D., Rodriguez-Valentín, J.G., Pachepsky, Y.A., Shelton, D.R., 2016. Spatial and temporal
2 variation of fecal indicator organisms in two creeks in Beltsville, Maryland. *Water Qual. Res. J. Can.*
3
4 51(2): 167-179.
5
6 Strauch, A.M., Mackenzie, R.A., Bruland, G.L., Ill, R.T., Giardina, C.P., 2014. Climate change and land
7 use drivers of fecal bacteria in Tropical Hawaiian Rivers. *J. Environ. Qual.* 43(4): 1475-1483.
8
9 Strawn, L.K., Fortes, E.D., Bihn, E.A., Nightingale, K.K., Gröhn, Y.T., Worobo, R.W., Wiedmann, M.,
10 and Bergholz, P.W., 2013. Landscape and meteorological factors affecting prevalence of three food-
11 borne pathogens in fruit and vegetable farms. *Appl. Environ. Microbiol.* 79(2), 588–600.
12
13 USDA (U.S. Department of Agriculture), 2016. Natural Resources Conservation Service (NRCS). Soil
14 Survey Geographic (STATSGO) database. Available at
15 <http://websoilsurvey.sc.egov.usda.gov/App/WebSoilSurvey.aspx> as of last access date: August 25
16 2016.
17
18 USFDA (U.S. Food and Drug Administration), 2016a. FDA Food Safety Modernization Act (FSMA).
19 Available at <http://www.fda.gov/Food/GuidanceRegulation/FSMA/default.htm> as of last access date:
20 August 25 2016.
21
22 USFDA (U.S. Food and Drug Administration), 2016b. Federal Register: Standards for the growing,
23 harvesting, packing, and holding of produce for human consumption; proposed rule. Available at
24 <https://www.regulations.gov/#!documentDetail;D=FDA-2011-N-0921-0973> as of August 25 2016.
25
26 USGS (U.S. Geological Survey), 2016a. Earth Resources Observation and Science (EROS) Center.
27 National Land Cover Data 1992 (NLCD 92). Available at http://www.mrlc.gov/nlcd92_data.php as
28 of last access date: August 25 2016.
29
30 USGS (U.S. Geological Survey), 2016b. Earth Resources Observation and Science (EROS) Center. USGS
31 National Elevation Dataset. Available at <http://eros.usgs.gov/elevation-products> as of last access
32 date: August 25 2016.
33
34 Van Liew, M., Veith, T.L., Bosch, D.D., Arnold, J.G., 2007. Suitability of SWAT for the conservation
35 effects assessment project: Comparison on USDA agricultural research service watersheds. *J.*
36 *Hydrol. Eng.* 12 (2): 173–189.
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Villarini, G., Strong, A., 2014. Roles of climate and agricultural practices in discharge changes in an agricultural watershed in Iowa. *Agric. Ecosyst. Environ.* 188: 204-211.

WHO, 1989. Health guidelines for the use of wastewater in agriculture and aquaculture. Report of a WHO Scientific Group, Technical Report Series No. 778, WHO, Geneva

Wood, J.D., Bezanson, G.S., Gordon, R.J., Jamieson, R., 2010. Population dynamics of *Escherichia coli* inoculated by irrigation into the phyllosphere of spinach grown under commercial production conditions. *Int. J. Food Microbiol.* 143(3): 198-204.