**Journal of Environmental Engineering and Science**
Volume 11 Issue JS1

**Rainfall and microbial contamination in Alberta well water**
Valeo, Checkley, He and Neumann

Journal of Environmental Engineering and Science, 2016, **11(1)**, 18–28
http://dx.doi.org/10.1680/jenes.15.00022
**Paper 15.00022**
Received 17/11/2015; accepted 22/06/2016

**Keywords:** groundwater/health & safety/hydrology & water resource

# Rainfall and microbial contamination in Alberta well water

**Caterina Valeo** PhD, PEng
Professor, Department of Mechanical Engineering, University of Victoria, Victoria, BC, Canada (corresponding author: valeo@uvic.ca)

**Sylvia Checkley** DVM PhD
Associate Professor, Department of Ecosystem and Public Health, Faculty of Veterinary Medicine, University of Calgary, Calgary, AB, Canada; Provincial Laboratory for Public Health of Alberta, Calgary, AB, Canada

**Jianxun He** PhD, PEng
Assistant Professor, Department of Civil Engineering, Schulich School of Engineering, University of Calgary, Calgary, AB, Canada

**Norman Neumann** PhD
Professor, School of Public Health, University of Alberta, Edmonton, AB, Canada; Provincial Laboratory for Public Health of Alberta, Edmonton, AB, Canada

**Spatial and seasonal patterns in the positive rates of total coliforms and *Escherichia coli* in Alberta well water were investigated to gain insight into well water microbial contamination. Analysis was conducted in the presence of total coliforms (77 135 tests) and *E. coli* (77 132 tests) in well water from 2004 to 2009 along with monthly estimates of precipitation, all of which were aggregated to 13 zones across Alberta by using Voronoi tessellation. Regression combined with autocorrelation analysis was employed to develop wave functions for data assembled in each zone. Precipitation was found to peak in June or July in all 13 regions. The positive *E. coli* rate was found to peak in June, July or August, but the positive total coliform rate peaked in August, September or October. Spatial statistical analysis revealed a potential association of total coliforms and *E. coli* with precipitation in two heavily populated basins. Spatial density analysis revealed a cluster of positive tests of total coliforms and *E. coli* in a narrow spatial extent in June and July of 2005.**

## Notation

| | |
|---|---|
| $c$ | mean of the function |
| $e_t$ | error at time $t$ |
| $m$ | amplitude of the sine wave function |
| $p$ | period |
| $t$ | time: months |
| $z_t$ | response at time $t$ |
| $\phi$ | initial phase |

## Introduction

Waterborne disease outbreaks can result from microbial contamination of well water (Corkal *et al*., 2004; Raina *et al*., 1999), and the Walkerton tragedy, which occurred in a small community in Ontario in 2000, is one example of how groundwater contamination is an ongoing concern in Canada, particularly in rural regions that largely rely on groundwater as their primary source of potable water. Thus, understanding the temporal and spatial variability in groundwater contamination can assist in developing groundwater sampling, monitoring and management programmes.

Studies exploring the factors influencing waterborne disease outbreaks in Canada include those of Schustre *et al*. (2005), who showed that heavy rainfall was associated with disease outbreaks from drinking water, and Thomas *et al*. (2006), who suggested that extreme rainfall and warmer temperatures were associated with high rates of waterborne disease outbreaks between 1975 and 2001. Studies on surface water quality and the seasonal effects of hydroclimatology have demonstrated that increased overland flow and shallow groundwater infiltration close to the surface would likely lead to an increase in the probability of faecal contamination (Hynds *et al*., 2012). Ueijo *et al*. (2014) found gastrointestinal illness in children increased in weeks with more precipitation (against drier weeks) in municipalities with untreated drinking water from a surface source. Microbial indictors in surface waters have been shown to be dependent on environmental factors and seasonality (Davis *et al*., 2005; He *et al*., 2010, 2015) and will peak after heavy rain events (Chigbu *et al*., 2004; Hyland *et al*., 2003; Mallin *et al*., 2001). Recent studies further confirm these associations, with O'Dwyer *et al*. (2015) using logistic regression to show that antecedent rainfall and high intensity rain events were strongly associated with waterborne outbreaks of *Escherichia coli* regionally – although neither surface or groundwater sources were indicated.

Temporal and spatial variations in the physiochemical quality of groundwater have been investigated in previous works (Bjerg and Christensen, 1992; Fendorf *et al*., 2010), but studies in microbial contamination in groundwater are rare. Waterborne disease outbreaks were strongly associated with extreme precipitation in both surface water and groundwater contamination (Curriero *et al*., 2001), but the authors found a 2-month lag between precipitation and *E. coli* in groundwater. Jean *et al*. (2006) looked at groundwater samples collected with rainfall data to study the associations between rainfall rate and an enterovirus epidemic in Taiwan. Using logistical regression, they found a statistically significant association between rainfall rate and cases of the infection and confirmed that the virus was present in locally used groundwater samples. Hynds *et al*. (2014) conducted a review of enteric pathogen contamination in groundwater for the periods of 1990 and 2013. Their analyses found no significant association between precipitation and occurrence of faecal indicators but did find an association between faecal indicator presence and poorly constructed wells – suggesting a hydro-climatological source of contamination into the well.

The majority of rural residents in the province of Alberta (approximately 19%) rely largely on groundwater wells as source of potable water (Alberta Health, 2014), with the density of domestic wells decreasing northward in the province. While annual trends in waterborne disease infection vary within Alberta and nationally, incidences of waterborne *E. coli* infection were roughly 50% above the national average (between 1985 and 1998) (Alberta Health, 2014). The use of groundwater is expected to intensify in Alberta; thus, effort needs to be directed to safeguarding these resources, which requires an understanding of the spatial and temporal variations in groundwater quality in this region. The objective of this paper is to examine the potential influences of precipitation on two microbial indicators – total coliforms and *E. coli* – in Alberta well water from both a spatial and a temporal (seasonal) perspective.

## Methods

### Sample collection and assay

The Alberta Provincial Laboratory for Public Health (ProvLab) is a government agency tasked with providing laboratory services for medical diagnostic and environmental monitoring functions. In conjunction with environmental monitoring, the laboratory performs a variety of tests that assess the quality of surface water and groundwater for the province. Samples of well water are collected by private well owners and shipped on ice to ProvLab for assay voluntarily. The sample undergoes testing within 24 h of the time of collection; otherwise, the sample is discarded.

The assay of total coliforms and *E. coli* in the database was performed according to standard procedure (APHA, 1992). This procedure is the commercial Colilert defined substrate method (Idexx Laboratories) that may be used for presence/absence (P/A) or quantification testing with Quanti-Tray or Quanti-Tray/2000. This approach differs from traditional media, which provide a nutrient-rich environment that also supports the growth of non-target organisms and, consequently, could lead to false positives. When used for P/A testing, the procedure performs two P/A tests simultaneously on each sample, one for total coliforms and one for *E. coli*, noting that an *E. coli* positive must also result in a positive for total coliforms. All test results are kept on file electronically for approximately 7 years in the agency's Laboratory Information System (LIS).

### Database construction

Total coliform and *E. coli* data for private (and semi-private) well water were extracted from the LIS database for the period of April 2004–March 2009. This resulted in a total of 77 135 tests for total coliforms and 77 132 tests for *E. coli*. Monthly precipitation records were obtained from the Adjusted Historical Canadian Climate Data set provided by Environment Canada. These data were available up to August 2007 (Environment Canada, 2015). Thirteen rain gauge stations that have a complete record of monthly precipitation for the period were found in the study area, and their locations are presented in Figure 1. This paper used Geographic Resources Analysis Support System (Grass) to perform qualitative analysis and some of the quantitative analysis described in the next sections. It was used to create raster maps for data management, spatial modelling and visualisation of the data (Neteler and Mitasova, 2008) by using its animation tool. A Voronoi diagram (Bailey and Gatrell, 1995) was employed to divide the study region into Thiessen polygons based on proximity to a specific point (the rain gauge location) by using a Euclidean distance. The data were then grouped into subsets based on the Thiessen polygons for further analysis. Thiessen polygons for each of the 13 rain gauge stations used are also shown in Figure 1. In each polygon (zone), four monthly time series data sets, (*a*) precipitation, (*b*) number of tests, (*c*) positive total coliform rate and (*d*) positive *E. coli* rate, were assembled. The positive total coliform rate was calculated by

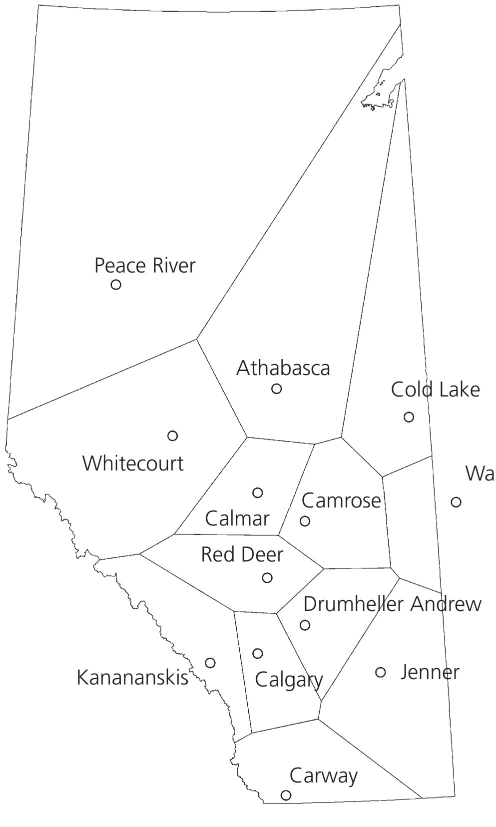


**Figure 1.** Rain gauges used and Thiessen polygons generated using Voronoi diagram

dividing the number of positive tests by the total number of tests. The positive *E. coli* rate was calculated in the same manner. Spatial data analysis was conducted through a variety of methods and temporal variability; specifically, seasonal periodicity was determined using signal processing methods and is described in the following sections.

## Spatial data analysis

### Visualisation of continuous representations of contamination

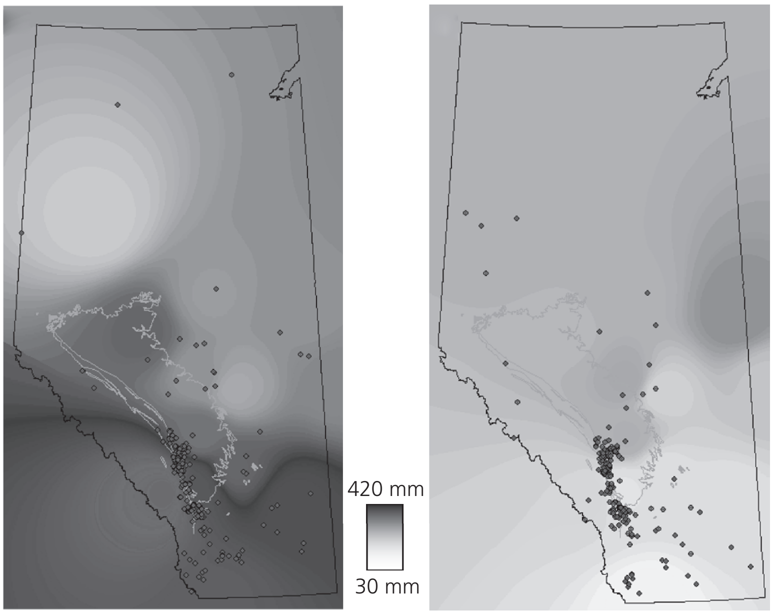
First, analysis by visualising two or more raster maps simultaneously and consequently their potential associations were determined with Grass. A sequence of raster grids was created using the second-order inverse distance-weighted (IDW) interpolation method (Bailey and Gatrell, 1995; Neteler and Mitasova, 2008) for precipitation. Note that each raster map was created using data at a specified time period, thus providing some information on the temporal variations as well.

### Density analysis

To investigate the spatial variation in the variables of interest, a simple and intuitive density analysis in Grass based on quadrats (Bailey and Gatrell, 1995) was conducted. The density analysis divides a rectangle enclosing the province of Alberta into square cells or quadrats. The procedure produces a raster grid in which the value of each cell is the ratio of the number of specimens that tested positive to the total number of specimens tested. Separate rasters were created for total coliforms and *E. coli* at resolutions of 10, 20 and 40 km. Some cells contained no value because no tests were performed in the cells. In addition, some cells had such a small number of specimens that the rate could not be estimated accurately.

### Cluster analysis

This paper applied two cluster analysis approaches to identify quantitatively the critical region(s) of the variables of interest. The first approach is the contingency table method, which partitions the data and uses the *p*-value to determine the significance of the partition. This approach requires a presumptive decision on how to partition the data. In addition, in the case of relatively low statistical significance, there is an increased likelihood of the appearance of false clusters if the data are divided repeatedly and the tests are performed many times. The second approach is to use the scan statistical approach within a free software package, SaTScan (Kulldorff, 2009). The scan statistic approach, which is to cluster randomly positioned points, does not need presumptive decision to partition data. SaTScan searches the clusters by the exhaustive enumeration of all partitioning schemes spatially based on predefined geometries and selects the most promising clusters by using a likelihood ratio test. Once a promising partitioning scheme has been found, Monte Carlo simulation is used to determine the significance of the cluster. SaTScan provides several probability models, such as the Bernoulli model, the exponential model and the normal model, when searching for clusters. The Bernoulli model (Kulldorff, 1997) was chosen as the most appropriate for this study, as positive test results constitute *cases* and negative results are *controls*. Two options exist for the geometry of clusters: circles and ellipses. The ellipse geometry, which was selected for the analysis, is a superset of the



(a) (b)

Figure 2. *E. coli* positive results in (a) June of 2005 and (b) July of 2005 shown as filled circles over interpolated rainfall. The irregular enclosed polygon in the lower left is the Paskapoo aquifer

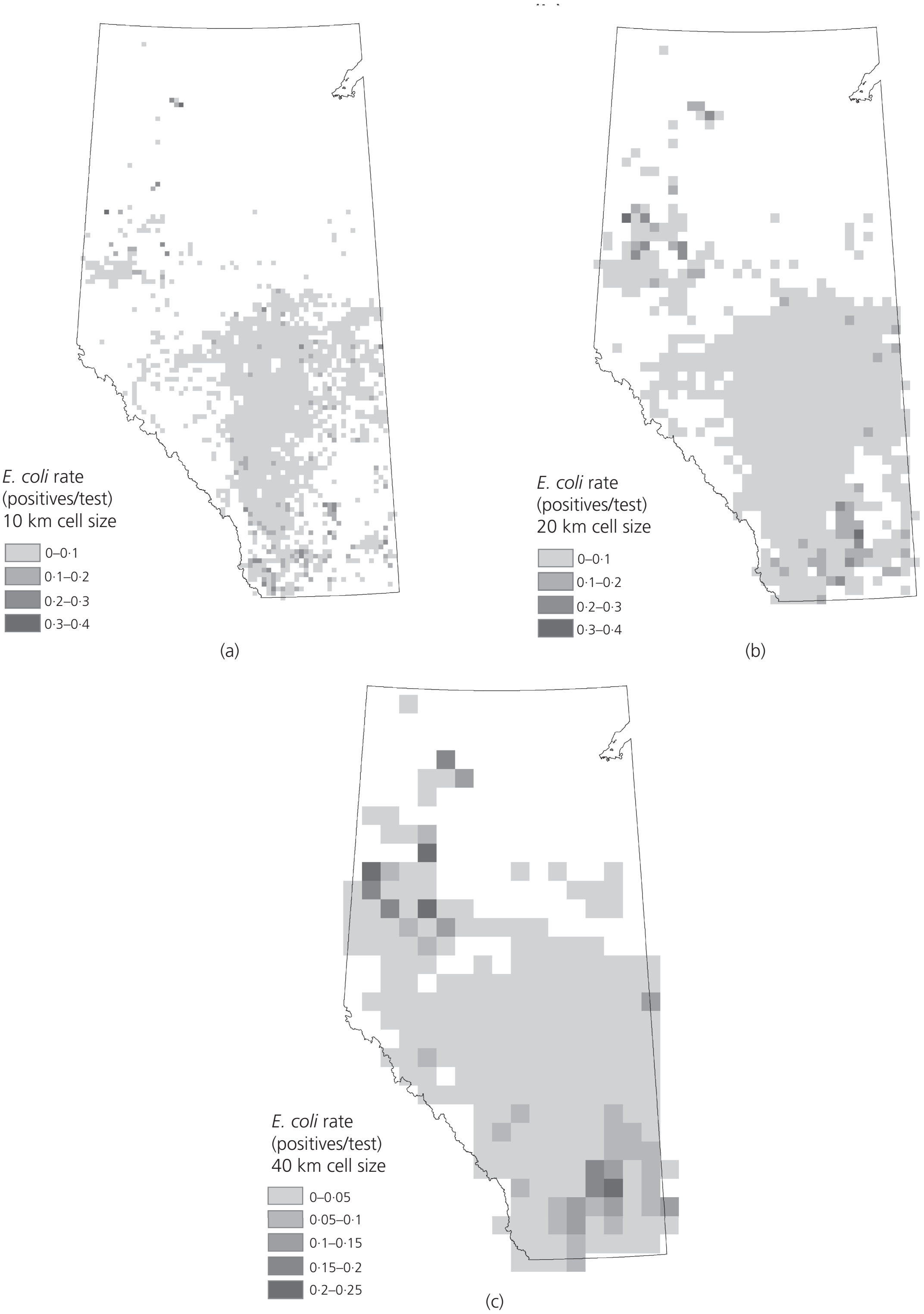


**Figure 3.** Density of *E. coli* rate based on (a) 10-, (b) 20- and (c) 40-km quadrats

circle and requires longer run times as ellipses of varying proportions and orientations are considered in addition to circles.

In addition, the Wilcoxon rank-sum test, which is non-parametric statistical test, was employed to detect if significant difference in medians of two data sets exists. All the statistical analyses were conducted at the significance level of 5%.

### Analysis for seasonal periodicity

A variety of methods exist for the spectral analysis of periodic data. Many of these methods were developed for the field of signal processing (Smith, 1997) and are often not appropriate for the applications of the type in this paper. Thus, for this work, the simplest function – a single sine wave function given by Equation 1, was fit to the observed data

$$z_t = m\sin\left(\frac{2\pi}{p}t + \phi\right) + c + e_t \tag{1}$$

where $z_t$ is response at time $t$, $t$ is time (in months), $m$ is the amplitude of the sine wave function, $p$ is the period, $\phi$ is the initial phase, $c$ is the mean of the function and $e_t$ is the error at time $t$. Autocorrelation (Box and Jenkins, 1976) is used to determine $p$ of the sine wave function. Once $p$ is determined, a combination of trial and error and regression analysis is used to determine $\phi$, $m$ and $c$. Here, $\phi$ is the product of $\pi/6$ and one of 12 possible values (from 0 through 11) corresponding to each of the 12 months. The value of $\phi$ is determined by evaluating all 12 possible functions. The $\phi$ that yields the best performance, is selected. The values of the remaining two parameters, $m$ and $c$, are determined through least-squares regression. The performance of the regression models is evaluated using the coefficient of determination ($R^2$), and the $p$-value is calculated from the $F$ test. The best-fit model has the highest $R^2$ and the lowest $p$-value.

In the development of sine wave functions to simulate the four time series, autocorrelation analysis was performed and a correlogram, which is a scatter plot of the sample autocorrelations against the time

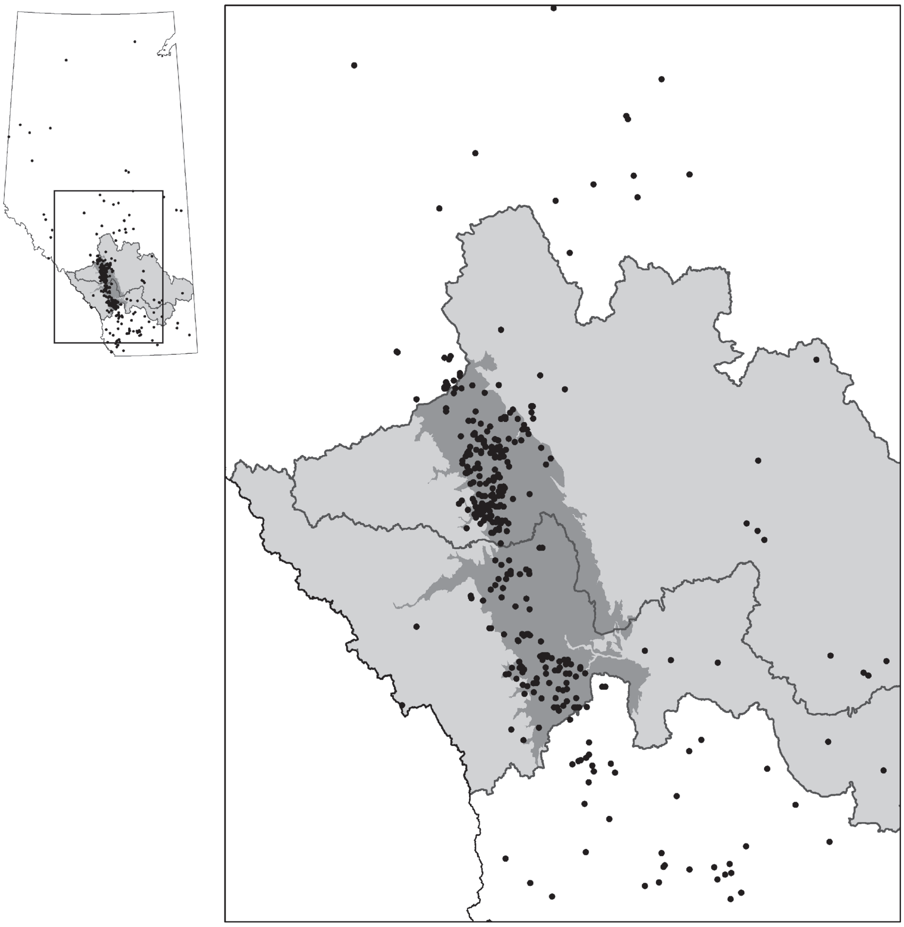


Figure 4. *E. coli* cluster from June and July 2005 and river basins. The dark grey area is the suspected cluster region, the lower grey polygon is the Bow River Basin and the upper grey polygon is the Red Deer River Basin

lags, for each data set in each zone was created to determine the value of $p$. In the analysis, autocorrelations were calculated from a 1-month time lag up to a 30-month time lag. The same approach was used to determine the other parameters, including $m$, $c$ and $\phi$ in Equation 1, for each time series in each zone. An example of a correlogram for the number of tests in the Carway zone is illustrated in Figure S1. The development of the sine wave function for the number of tests at the Carway zone is available in the Supplementary Material and the best-fit equation that results for this zone.

## Results

### Spatial associations between microbial indictors and precipitation

The animation in Grass was used to investigate the number of positive tests of both total coliforms and *E. coli* in relation to precipitation on a monthly basis. Based on the availability of precipitation data, the data from April 2004 to August 2007 inclusive were used in the analyses. Quadrat cells with a test count of six or less were discarded to avoid random errors resulting from small sample sizes. A separate animation was created for each month, including the point locations of any well that tested positive for total coliforms or, alternatively, positive for *E. coli*. The point data for each month and an outline map of Alberta including relevant physical features were overlaid on the corresponding IDW raster grid for interpolated precipitation. Figure 2 shows the point locations of wells that tested positive for *E. coli* during the months of June and July 2005, the Paskapoo aquifer formation in the lower left of the province (Grasby *et al*., 2008) and the interpolated precipitation.

By observing visually the overlaid raster maps for every month over the study time period, the number of positive tests for total coliforms and *E. coli* and precipitation amount were consistently higher in the summer months than in the winter months for all years. The numbers of positive tests for total coliforms and *E. coli* both appeared randomly distributed over the study area, in general, with little evidence of spatial clustering; however, the positive tests appeared concentrated in a crescent-shaped region extending from the southern edge of the Paskapoo Formation in June and July of 2005. The region of the suspected cluster falls within the Red Deer River and Bow River basins in a narrow band falling between elevations of 1000 and 1300 m above mean sea level. However, a region of high rainfall observed upstream of the crescent-shaped region in the month of June 2005 (Figure 2) was not seen in July of 2005.

### Density analysis of microbial indicators

Density analyses were performed using 40-, 20- and 10-km quadrats for both the positive total coliform and *E. coli* rates. In the analyses, the data collected during April 2004 to March 2009 were used. The analyses produced six maps in total. The density maps for the positive *E. coli* rates developed for each quadrat resolution are shown in Figure 3. In the map, some quadrats contain no value because no tests were performed in those quadrats. On the other hand, some quadrats have such a small number of samples that the rates cannot be estimated accurately. As illustrated in Figure 3, and in the maps produced for total coliform densities (not shown here), the rates are, in general, low and in the range of 0·3–0·4 in many quadrats. When the analysis was conducted for each month (results not shown), obvious dense region of positive tests were observed only in the months of June and July of 2005, thus implicating the effect of rainfall on the presence of total coliforms and *E. coli* upstream of the crescent-shaped region in Figure 2.

Density analysis for the pooled 5-year data indicated that the positive total coliform and *E. coli* rates were low (in the range of 0–0·1) in the majority of quadrats. When using data in the months of June and July of 2005, the density analysis demonstrated a denser region of positive results, located upstream of the Kananaskis rain gauge where an extreme rainfall event was recorded in June of 2005. The results suggest that ignoring the seasonal/or periodical variation in microbial indicators would even out their spatial variations, which may be very pronounced in a short time period (such as in June and July of 2005 in this study time period). Thus, the density maps created based on a pool of data over a long time period cannot detect variations effectively over a short duration. On the other hand, there is a trade-off between accuracy in estimating rates and the ability to discern spatial variability. A larger number of tests in a quadrat, which can be realised by increasing the size of the quadrat, will tend to yield a more accurate estimate of the rate. However, an increase in quadrat size would smooth out local spatial variations in rates. In contrast, random errors in the estimation of the rates may lead to the creation of false spatial variability when using small quadrats.

| Microbial indicators | | Total number of observations | | Total number of tests |
|---|---|---|---|---|
| Total coliform | | Group 1 | Group 2 | |
| A | Positives | 999 | 3253 | 4252 |
| | Negatives | 1145 | 14 826 | 15 971 |
| | Total | 2144 | 18 079 | 20 223 |
| B | Positives | 1280 | 2972 | 4252 |
| | Negatives | 1519 | 14 452 | 15 971 |
| | Total | 2799 | 17 424 | 20 223 |
| | *E. coli* | Group 1 | Group 2 | |
| A | Positives | 342 | 498 | 840 |
| | Negatives | 1802 | 17 580 | 19 382 |
| | Total | 2144 | 18 078 | 20 222 |
| B | Positives | 434 | 406 | 840 |
| | Negatives | 2365 | 17 017 | 19 382 |
| | Total | 2799 | 17 423 | 20 222 |

Marginal totals from the suspected/or identified cluster region are reported in column for group 1, while the results from their exterior regions are documented in the column for group 2

Table 1. Contingency table for suspected cluster region (A) and its exterior region and identified cluster region (B) by SaTScan and its exterior region for June and July of 2005

### Clustering of microbial indicators

The Wilcoxon rank-sum test was used to assess whether the positive rates of total coliforms and *E. coli* in June and July of 2005 were significantly different from those in June and July for every other year in the time period for all locations in the study region. The test indicated that the positive rates in June and July of 2005 are significantly higher than in other years. Therefore, the cluster analysis was conducted using only the data collected from these 2 months.

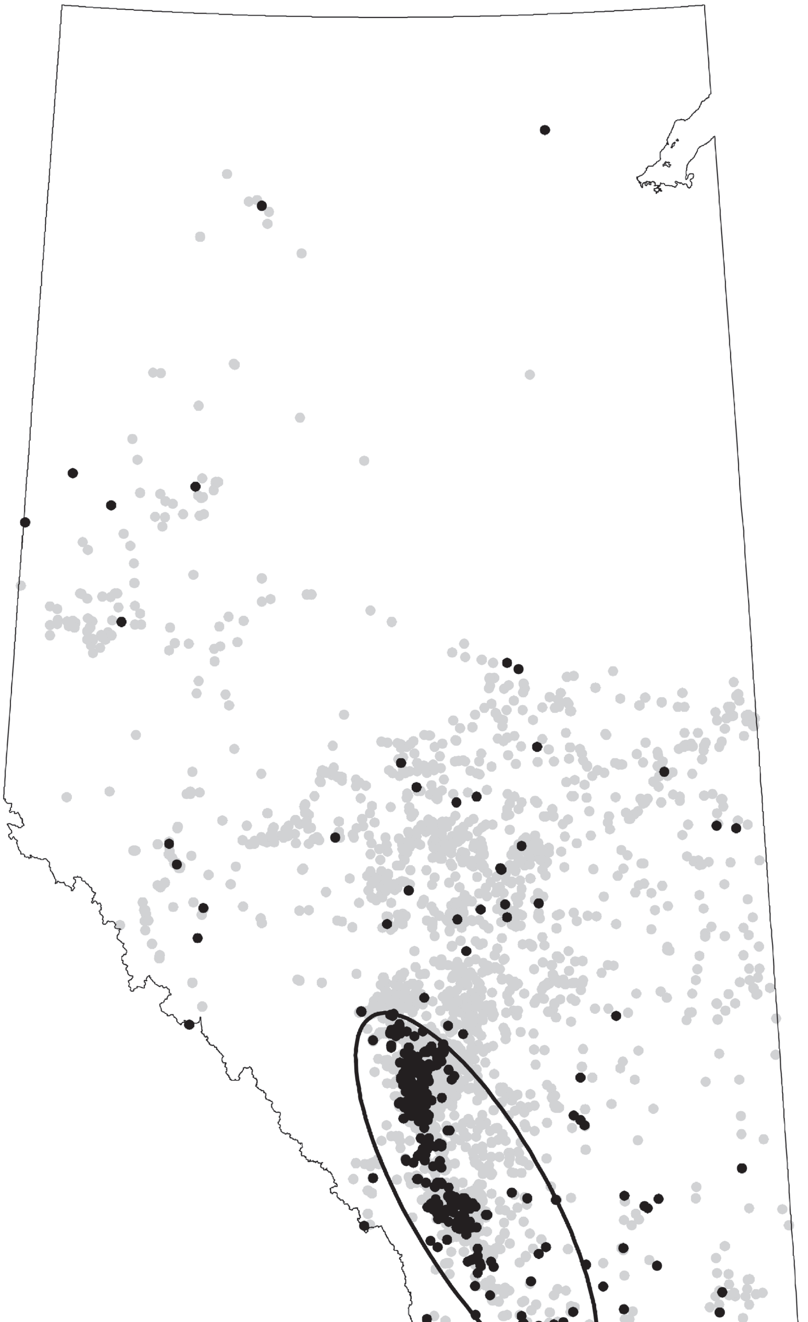


**Figure 5.** *E. coli* cluster from June and July 2005 identified by SaTScan using elliptical geometry. Positives are shown as black dots; negative results are light grey

Figure 4 displays the spatial distribution of the positive *E. coli* tests in June and July of 2005 over the entire study region. The figure indicates a potential cluster region straddling the Bow River and Red Deer River basins. It was also observed that this potential cluster of positive tests seems to fall largely between elevations of 1000 and 1300 m above sea level. Using Grass and a digital elevation model of the region, this suspected cluster region was created as that area bounded by these two elevations across the two basins and is indicated in dark grey in Figure 4. Table 1, the contingency table, provides the marginal totals of positive and negative tests, in and out of the suspected cluster region shown in Figure 3. A Monte Carlo simulation indicated that positive results for the suspected cluster region were significant with a *p*-value less than 0·01. Cluster analysis results using SaTScan are illustrated in Figure 5. In this figure, the positive tests are shown as black circles and negative tests are shown in light grey circles. The ellipse shown in this figure is the region identified as exhibiting significant clustering. Table 1 also provides the marginal totals of positive and negative tests, in and out of the cluster region identified in SaTScan. The calculated *p*-values for all cases are less than 0·01. Similar results were obtained for total coliforms but are not shown.

The animations indicated seasonal variations in precipitation and the number of positive tests of total coliforms and *E. coli*. The results also demonstrated that the positive tests for total coliforms and *E. coli* in June and July of 2005 were observed concurrently with extreme rainfall recorded in June of 2005 in the Red Deer River and Bow River basins. The monthly precipitation recorded at the Kananaskis rain gauge station in June of 2005 was 422 mm, which is more than four times higher than the monthly average at this station during the 5-year study period. The high precipitation in June 2005 appeared to result in high positive tests of both total coliforms and *E. coli* in the month of June and even a month later in July 2005. These results might suggest the prolonged effects of rainfall (2 or more than 2 months) on microbial contamination in groundwater in this region. Both the suspected cluster shown in Figure 4 and the cluster identified by SaTScan also demonstrate that in June and July of 2005, many positive tests of total coliforms and *E. coli* were located in a region that received extreme amounts of rainfall in June of 2005. The results from the contingency table along with calculated *p*-values confirmed statistically that the number of positive tests of total coliforms and *E. coli* in both the suspected and identified cluster regions were significantly higher than those in their corresponding exterior regions. In addition, the elliptical cluster region identified by SaTScan appears to be slightly more significant than the suspected clustering region bounded by elevation for both total coliforms and *E. coli*.

### Seasonal periodicity

Close visual inspection (data not shown) of the total number of tests each month revealed substantial seasonal variation in all 13 zones, with a higher number of tests occurring during the summer months. The number of positive total coliform and *E. coli* tests also exhibited seasonal variations. Similarly, the positive total coliform and *E. coli* rates also revealed evidence of seasonal periodicity in the form of repeated peaks and valleys spaced approximately 12 months apart.

The analysis described previously for the number of tests was also performed for precipitation, positive total coliform and *E. coli* rates for all 13 zones with the peak month calculated and shown in Table 2. Throughout the province of Alberta, precipitation peaked in either June or July. The number of tests undertaken in 1 month also reached a peak in July at all zones except at Calgary, where the peak occurred in June. The positive total coliform rate had peak months ranging from August to October, with the majority (9 of 13) occurring in September. The positive *E. coli* rate peaked 1 or 2 months earlier than the positive total coliform rate at all zones

| Zone | Number of tests | Precipitation peak | Number of tests peak | Positive *E. coli* rate peak | Positive total coliform rate peak |
|---|---|---|---|---|---|
| Athabasca | 3565 | July | July | August | September |
| Calgary | 13 167 | June | June | July | August |
| Calmar | 10 689 | July | July | July | September |
| Camrose | 5754 | July | July | August | September |
| Carway | 6929 | June | July | August | September |
| Cold Lake | 3296 | July | July | August | October |
| Drumheller | 2820 | June | July | August | September |
| Jenner | 2771 | June | July | August | September |
| Kananaskis | 5788 | July | July | July | August |
| Peace River | 3750 | June | July | August | September |
| Red Deer | 12 906 | July | July | July | September |
| Waseca | 1628 | July | July | July | September |
| Whitecourt | 7200 | July | July | June | October |

**Table 2.** Peak months calculated from developed spectral models at all 13 zones in Alberta

except at Whitecourt, where the peak occurred 3 months earlier than the positive total coliform rate. In addition, the positive *E. coli* rate always peaked after 0–2 months within the peak for precipitation at most zones, whereas it peaked earlier than precipitation at Whitecourt. Overall, the precipitation peaked in June–July, followed by the peak of positive *E. coli* rate in June–August, followed by the peak of the positive total coliform rate in August–September in the study region.

Figure 6 shows graphically the observed and predicted positive total coliform and *E. coli* rates by the developed sine wave functions using the pooled data from all 13 zones. It is not surprising that the positive *E. coli* rate is always lower than the positive total coliform rate as *E. coli* is a subset of total coliforms. The developed sine wave functions predict quite well not only the seasonal patterns of positive total coliform and *E. coli* rates quite well but also their magnitudes. In addition, the figure also demonstrates clearly that there is a consistent time lag between the periodicity of the positive *E. coli* rate and the positive total coliform rate and that the peak of the positive *E. coli* rate always appears before the peak of the positive total coliform rate in the observations. All these were captured quite well by the developed sine wave functions. It can been seen from Figure 6 that the observed positive total coliform and *E. coli* rates are relatively high from months 15 to 21, which

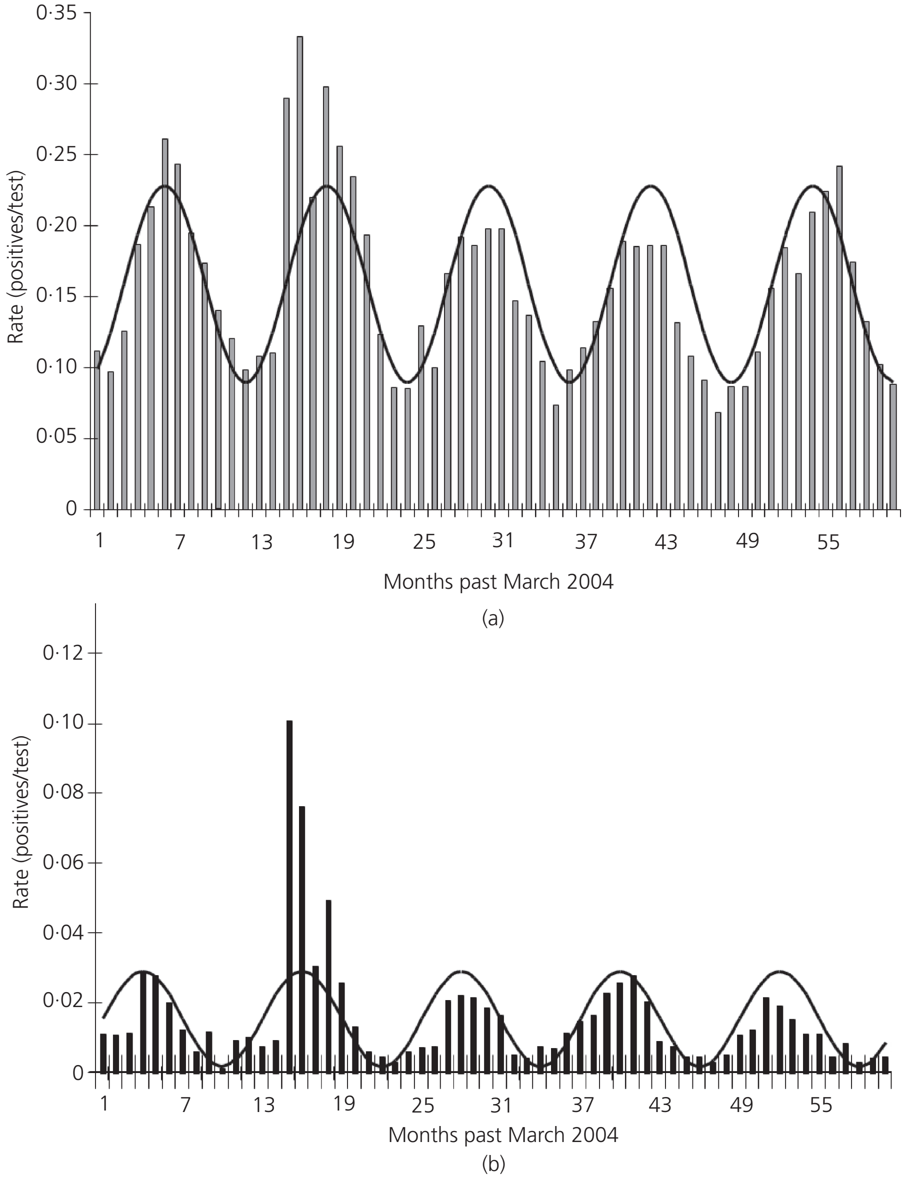


**Figure 6.** Observed (shaded bars) and predicted (solid lines) positive (a) total coliform and (b) *E. coli* rates using single sine wave functions

correspond to the period of June to December of 2005. The sine waves functions underpredicted the observations during this time period.

It has been acknowledged that *E. coli* does not survive well in non-host environments for extended periods of time other than in tropic climatic regions (Winfield and Groisman, 2003). Some coliforms are known to survive and grow in the environment. In laboratory experiments, the growth of coliforms has been observed, while the growth of *E. coli* was not detected (Geldreich, 1967). The time lag between the seasonal variations in positive total coliform and *E. coli* rates might be explained by the fact that there are differences in their persistence, fate and regrowth capability in the environment. Several studies in the literature cited earlier note the strong influence of temperature on contamination levels. The precipitation accessed from Environment Canada was total precipitation, which includes snowmelt. *E. coli* and total coliforms are not expected to grow or thrive in cold, winter months in Alberta, nor are hydrological processes in sub-zero conditions similar to those of above-zero conditions. While snow accumulation in winter months does affect surface run-off during freshet and groundwater recharge within the same year, the results observed in this work can be used only to suggest associations that are really valid only in above-zero temperature conditions within Alberta.

The wells in Alberta pull from various aquifers in the region, all with varying characteristics and, therefore, varying well depth. Precipitation patterns are not necessarily correlated with aquifer formation, and the aquifers sustaining the wells vary extensively across the province; yet the results in this study were similar across the aquifers. This may suggest that shallow groundwater processes (that is, subsurface run-off processes), which are dependent on shallow soil and topographic characteristics, may have a greater influence on the observations seen in this paper than deeper geological formations. Compromised well integrity (such as in poorly capped wells) can create vulnerability to shallow water pathways. Surface water flow pathways that are Hortonian, interflow, or saturation excess overland flow (Valeo and Moin, 2001) could all lead potentially to contamination of a well from contaminated stormflow. Because the Thiessen polygons were generated based on proximity to a rain gauge, it is distinctly possible that a well in any given location is affected by groundwater and surface water hydrological pathways that are outside the Thiessen polygon attributed to that well. Therefore, future studies should involve a tessellation scheme based on watershed delineating zones of surface or subsurface drainage.

When considering the results of this study, it is important to bear in mind that water samples were submitted voluntarily by the wells' owners. A sample might be submitted given the suspicion that the water may be contaminated due to unusual taste, odour, colour or some other factors. Therefore, the positive total coliform and *E. coli* rates and the number of tests do not reflect necessarily those that would be observed with randomised or systematic testing. For this reason, the rates of positive results could be higher than would be expected when conducting random sampling. In addition, there are limitations with the methods used for the assay of the bacteria in the Colilert method, with the collection of precipitation with rain gauges and with other methods/equipment used in constructing the database.

## Conclusions

Spatial statistical analyses and tests of seasonal periodicity conducted in this study on precipitation and positive test rates of total coliform and *E. coli* suggest increasing microbial contamination after heavy rain events in shallow groundwater wells, which are a primary drinking water source in the rural area of Alberta. The analysis revealed the following.

- A cluster region prone to the impacts of extreme rain events was identified in the Red Deer River Basin.
- Seasonal periodicity analysis for all variables investigated showed a periodicity of 12 months at all zones in Alberta.
- After precipitation peaks, the positive *E. coli* rate peaks, and then the positive total coliform rate peaks.
- When attempting to correlate the positive test rates with precipitation, very strong correlation coefficients did not result, because of the existence of time lags among their seasonal variations. This is likely because of a lack of attention to temperature as a variable affecting microbial growth in the wells. Thus, conducting the analysis by using temperature is recommended.

The results suggest a need to develop well water-monitoring programmes that consider precipitation variability in the province because precipitation and surface run-off appear to be important factors influencing groundwater contamination. Potentially, increasing the sampling frequency and locations after extreme events to detect microbial contamination early may be a recommendation for this monitoring programme. On the other hand, further understanding of what magnitude of rain events result in microbial contamination of well water and to what degree (which may be location specific) is needed for developing efficient groundwater monitoring and management.

## Acknowledgements

The authors would like to thank Dr James Davidson, former Calgary Health Region employee, for his contribution to this manuscript.

### REFERENCES

Alberta Health (2014) *Domestic Well Water Quality in Alberta, Fact Sheets*. Alberta Health, Edmonton, AB, Canada. p. 63.

APHA (American Public Health Association) (1992) *Standard Methods for Examination of Water and Wastewater*, 18th edn. APHA, Washington, DC, USA.

Bailey T and Gatrell A (1995) *Interactive Spatial Data Analysis*. Longman Scientific & Technical, Harlow, UK.

Bjerg P and Christensen T (1992) Spatial and temporal small-scale variation in groundwater quality of a shallow sandy aquifer. *Journal of Hydrology* **131(1–4)**: 133–149.

Box G and Jenkins G (1976) *Time Series Analysis*, 4th edn. Wiley, Hoboken, NJ, USA.

Chigbu P, Gordon S and Strange T (2004) Influence of inter-annual variations in climatic factors on fecal coliform levels in Mississippi Sound. *Water Research* **38(20)**: 4341–4352, http://dx.doi.org/10.1016/j.watres.2004.08.019.

Corkal D, Schutzman W and Hilliard C (2004) Rural water safety from the source to the on-farm tap. *Journal of Toxicology and Environmental Health, Part A* **67(20–22)**: 1619–1642, http://dx.doi.org/10.1080/15287390490491918.

Curriero F, Patz J, Rose J and Lele S (2001) The association between extreme precipitation and waterborne disease outbreaks in the United States, 1948–1994. *American Journal of Public Health* **91(8)**: 1194–1199.

Davis K, Anderson M and Yates M (2005) Distribution of indicator bacteria in Canyon Lake, California. *Water Research* **39(7)**: 1277–1288, http://dx.doi.org/10.1016/j.watres.2005.01.011.

Environment Canada (2015) *AHCCD/DCCAH – Welcome*. See http://www.cccma.ec.gc.ca/hccd/ (accessed 30/07/2015).

Fendorf S, Michael H and van Geen A (2010) Spatial and temporal variations of groundwater arsenic in South and Southeast Asia. *Science* **328(5982)**: 1123–1127, http://dx.doi.org/10.1126/science.1172974.

Geldreich E (1967) *Fecal Coliform Concepts in Stream Pollution*. Bureau of Sanitary Engineering, California State Department of Public Health, Berkeley, CA, USA.

Grasby S, Chen Z, Hamblin A, Wozniak P and Sweet A (2008) Regional characterization of the Paskapoo bedrock aquifer system, southern Alberta. *Canadian Journal of Earth Sciences* **45(12)**: 1501–1516, http://dx.doi.org/10.1139/E08-069.

He J, Valeo C, Chu A and Neumann N (2010) Characteristics of suspended solids, microorganisms, and chemical water quality of event-based stormwater runoff from an urban residential area. *Water Environment Research* **82(12)**: 2333–2345, http://dx.doi.org/10.2175/106143010X12681059117058.

He J, Valeo C and Chu A (2015) Variation in water quality of a stormwater pond from diurnal thermal stratification. *Journal of Water Resources and Hydraulic Engineering* **4(1–4)**: 189–198, http://dx.doi.org/10.5963/JWRHE0402008.

Hyland R, Byrne J, Selinger B *et al.* (2003) Spatial and temporal distribution of fecal indicator bacteria within the Oldman River basin of Southern Alberta, Canada. *Water Quality Research Journal of Canada* **38(1)**: 15–32.

Hynds P, Misstear B and Gill L (2012) Development of a microbial contamination susceptibility model for private domestic groundwater sources. *Water Resources Research* **48(12)**: W12504, http://dx.doi.org/10.1029/2012WR012492.

Hynds P, Thomas M and Pintar K (2014) Contamination of groundwater systems in the US and Canada by enteric pathogens, 1990–2013: a review and pooled-analysis. *PLoS ONE* **9(5)**: e93301, http://dx.doi.org/10.1371/journal.pone.0093301.

Jean J, Guo H, Chen S *et al.* (2006) The association between rainfall rate and occurrence of an enterovirus epidemic due to a contaminated well. *Journal of Applied Microbiology* **101(6)**: 1224–1231, http://dx.doi.org/10.1111/j.1365-2672.2006.03025.x.

Kulldorff M (1997) A spatial scan statistic. *Communications in Statistics – Theory and Methods* **26(6)**: 1481–1496.

Kulldorff M (2009) *SaTScanTM User Guide*, 8th edn. Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, MA, USA.

Mallin M, Ensign S, McIver M, Shank G and Fowler P (2001) Demographic, landscape, and meteorological factors controlling the microbial pollution of coastal waters. *Hydrobiologia* **460(1/3)**: 185–193.

Neteler M and Mitasova H (2008) *Open Source GIS: a GRASS GIS Approach*, 3rd edn. Springer, New York, NY, USA.

O'Dwyer J, Morris Downes M and Adley C (2015) The impact of meteorology on the occurrence of waterborne outbreaks of vero cytotoxin-producing *Escherichia coli* (VTEC): a logistic regression approach. *Journal of Water and Health* **14(1)**: 39–46, http://dx.doi.org/10.2166/wh.2015.016.

Raina P, Pollari F, Teare G *et al.* (1999) The relationship between *E. coli* indicator bacteria in well-water and gastrointestinal illnesses in rural families. *Canadian Journal of Public Health* **90(3)**: 172–175.

Schustre C, Ellis A, Robertson W *et al.* (2005) Infectious disease outbreaks related to drinking water in Canada 1974–2001. *Canadian Journal of Public Health* **96(4)**: 254–258.

Smith S (1997) *The Scientist and Engineer's Guide to Digital Signal Processing*. California Technical Pub, San Diego, CA, USA.

Thomas K, Charron D, Waltner-Toews D *et al.* (2006) A role of high impact weather events in waterborne disease outbreaks in Canada, 1975–2001. *International Journal of Environmental Health Research* **16(3)**: 167–180, http://dx.doi.org/10.1080/09603120600641326.

Uejio C, Yale S, Malecki K *et al.* (2014) Drinking water systems, hydrology, and childhood gastrointestinal illness in Central and Northern Wisconsin. *American Journal of Public Health* **104(4)**: 639–646, http://dx.doi.org/10.2105/AJPH.2013.301659.

Valeo C and Moin S (2001) Hortonian and variable source area modeling in urbanizing basins. *Journal of Hydrologic Engineering* **6(4)**: 328–335, http://dx.doi.org/10.1061/(ASCE)1084-0699(2001)6:4(328).

Winfield M and Groisman E (2003) Role of nonhost environments in the lifestyles of *Salmonella* and *Escherichia coli*. *Applied and Environmental Microbiology* **69(7)**: 3687–3694, http://dx.doi.org/10.1128/AEM.69.7.3687-3694.2003.

---

**WHAT DO YOU THINK?**

To discuss this paper, please submit up to 500 words to the editor at journals@ice.org.uk. Your contribution will be forwarded to the author(s) for a reply and, if considered appropriate by the editorial panel, will be published as a discussion in a future issue of the journal.