Escherichia coli in urban stormwater: explaining their variability

D.T. McCarthy*, V.G. Mitchell*, A. Deletic* and C. Diaper**

*Department of Civil Engineering, Monash University, Clayton, Victoria 3800, Australia
(E-mail: david.mccarthy@eng.monash.edu.au)

**CSIRO, Manufacturing and Infrastructure Technology – Urban Water, Highett, Victoria 3190, Australia

Abstract The development of a model that predicts the levels of microorganisms in urban stormwater will aid in the assessment of health risks when using stormwater for both recreational uses and as an alternate water resource. However, the development of such a model requires an understanding of the dominant processes that influence the behaviour of microorganisms in urban systems. Using simple and multiple regression analyses this paper determines the dominant processes which affect the inter-event variability of the microbial indicator Escherichia coli (E. coli) in four urbanised catchments. The results reveal that a number of antecedent climatic conditions, together with rainfall intensity, can significantly explain the inter-event variation in wet weather E. coli levels.

Keywords Correlation; E. coli; microorganisms; model

Introduction

Understanding the levels of microorganisms in urban stormwater is important for the assessment of health risks relating to both (a) discharging of stormwater into receiving water bodies that are used for recreational purposes (e.g. bathing or other water sports) (Makepeace et al., 1995) and (b) utilisation of urban stormwater as an alternate water resource (McCarthy et al., 2006).

To identify these levels it is impractical to monitor every different microorganism at all urban stormwater sites, and is not possible for urban developments that are yet to be constructed. A more practical way of predicting the microorganism level in urban runoff is to use a modelling tool. The majority of models that predict microorganism levels are based on rural catchments (e.g. Ferguson and Croke, 2005) and, although they can provide important background information, they are not directly transferable to urban systems (Haydon and Deletic, 2006). Thus, there is a requirement for the development of a new model that describes and predicts how microorganisms behave in urban stormwater.

The successful development of an urban microorganism model requires a detailed understanding of how microorganisms behave in an urban environment, and to date, limited information is available to this regard. Furthermore, most of this limited information is built around studies that focus on collecting ‘grab’ samples from large river or estuarine systems (e.g. Ferguson et al., 1996; Atherholt et al., 1998; Eleria and Vogel, 2005). The more relevant studies (i.e. studies which focus on urban stormwater systems) still rely on relationships built around routine ‘grab’ sampling and therefore cannot satisfactorily capture the processes which affect the variation and peaks of microorganisms between and during storm events (e.g. Young and Thackston, 1999; Davies and Bavor, 2000). Moreover, even the most relevant studies (i.e. studies which report event based monitoring of urban stormwater systems) found little correlation between microorganisms and an array of climatic, water quantity and quality parameters.

Although the relationships found in the majority of these types of studies are not directly related to urban systems, it is still useful to draw upon their major conclusions to...
provide insight into what parameters are likely to explain the variation in urban microorganism levels. The following provides a brief overview of the literature available on the relationship between *E. coli* and different climatic, water quality and hydrologic parameters in a variety of different scenarios.

Relevant studies have shown that there is an increase in indicator microorganism concentrations during storm flows (e.g. Davis *et al.*, 1977; Olivieri *et al.*, 1977; McCarthy *et al.*, 2006) indicating a relationship between flow magnitude and microorganism transport. Furthermore, there are studies that show there is a significant relationship between microorganisms and the incidence of rainfall and rainfall intensity (e.g. Gannon, 1988; Young and Thackson, 1999; Davies and Bavor, 2000; Kelsey *et al.*, 2004). These findings suggest that wash-off models, which are commonly used to describe traditional stormwater pollutants, will provide some explanation of the variability of inter-event microorganism levels.

Antecedent climatic conditions have been widely used to help explain the variation of microorganism levels between wet weather events as they help model the ‘build-up’ of microorganisms in catchments. Kelsey *et al.* (2004) showed that antecedent rainfall totals helped explain the level of faecal coliforms (a commonly used indicator organism) in a small urbanised inlet. In contrast, Olivieri *et al.* (1977), in their extensive investigation of microorganisms in urban stormwater, found little correlation between the levels of microorganisms in urban systems and the number of days since the last rainfall. However, this lack of correlation might be due to a different definition of the antecedent dry weather period (ADWP) in the former study as they used rainfall totals and rainfall thresholds to define ADWP whereas Olivieri *et al.* (1977) used the definition of ‘days since previous rainfall’.

Other antecedent climatic conditions, especially ones that are known to influence the survival of microorganisms, have also been used to help explain variations between events. Crane and Moore (1986), in their review of modelling bacterial die-off, showed that the survival of microorganisms is affected by many environmental conditions, including: sunlight (irradiance), temperature and desiccation. Eleria and Vogel (2005) used variables such as net radiation and cloud cover in an attempt to model the die-off of faecal coliforms between events but found little explanatory power from these two variables.

Determination of relationships between microorganisms and other common water quality parameters can help identify the major processes and sources of microorganisms. Duncan (1999), in his statistical review of urban stormwater quality, showed that faecal coliforms were significantly correlated with other water quality parameters, including both total phosphorous and turbidity (both having *p* values < 0.01 and correlation coefficients *R* > 0.61). Others have also shown significant correlations between different water quality parameters, including Davis *et al.* (1977) who showed correlations exist between microorganisms and both discharge and suspended solids for stormwater runoff.

Although the relationships found in the above studies have provided some insight into how microorganisms behave in urban systems, the relationships may not be directly transferable to a new urban microorganism model. Therefore, to develop this model more research is required to understand the behaviour of microorganisms in urban stormwater systems. It is the aim of this paper to provide evidence of the dominant processes that influence the way in which a microbial indicator, *Escherichia coli*, behaves in urban stormwater runoff. Specifically, the inter-event behaviour of *E. coli* will be studied using simple and multiple regression analyses.

**Methods**

**Data**

The data used in this paper was collected from the stormwater pipes at four urban catchments in Melbourne, Australia, all of which have separate storm and sanitary sewers.
The catchments ranged in size from 10ha to over 100ha and have various levels of imperviousness (20%–80%) and land uses (see Table 1).

All sites were equipped with Doppler based flow meters and standard 0.2 mm tipping bucket rain gauges, both of which logged at one minute intervals. The samplers were programmed to collect samples using flow based intervals that ensured events were accurately represented. In total, 46 wet weather events (see Table 1) were sampled during a one year period from April 2005 to April 2006.

*E. coli*, the chosen bacterial indicator for this study, were enumerated in a NATA (National Association of Testing Authorities, Australia) approved laboratory using the Colilert® (IDEXX Laboratories) method. All other water quality parameters were analysed using standard analytical techniques, also in a NATA approved laboratory. Climatic data (including: sunshine hours, evaporation, radiation and temperature) were all obtained from nearby weather stations operated by the Bureau of Meteorology (www.bom.gov.au).

**Statistics used**

Standard statistical methods, including simple and multiple regressions, were used to determine significant correlations between the response and explanatory variables. The coefficient of correlation ($R$) was used to indicate if a linear relationship exists for both the simple and multiple regressions. To indicate if a non linear relationship exists the Spearman Rank correlation coefficient ($R_s$) was used for the simple regressions. For the multiple regressions log transformed data were used to determine if non linear relationships exist ($R_l$).

Standard statistical inferences were used to detect significant relationships. Significance levels ($p$-values) were obtained using Student’s $t$-tests for the correlation coefficients ($p$ and $p_l$ for untransformed and log transformed data, respectively). Inferences about the Spearman Rank correlation coefficient were done using a standard critical values table ($p_r$).

**Explanatory and studied variables**

The 18 explanatory variables used in the regression analyses were carefully selected based upon an extensive literature review of the knowledge surrounding the behaviour of microorganisms in water systems. Table 2 gives a brief description of each explanatory variable.

Two variables that quantify the level of *E. coli* in stormwater during wet weather events were studied:

1. **EMC** *E. coli* [MPN/100 mL] – the event mean *E. coli* concentration was calculated using a flow weighted average of the discrete *E. coli* concentrations. EMC *E. coli* levels were hypothesised to be correlated with: antecedent climatic variables, stormwater quality variables, $V$, $Q_{max}$, $U_{ave}$, $Q_{ave}$, $P$ and $P_{ave}$; and,

---

**Table 1** Site descriptions and data characteristics used in the analyses

<table>
<thead>
<tr>
<th>Site</th>
<th>Primary Land Use</th>
<th>Area (ha)</th>
<th>Imperviousness (%)</th>
<th>Number of Collected Events</th>
<th>SMC$^1$ E. coli (MPN/100 mL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clayton</td>
<td>Light Industrial</td>
<td>28</td>
<td>80</td>
<td>14</td>
<td>610</td>
</tr>
<tr>
<td>Doncaster</td>
<td>Med Density Res</td>
<td>106</td>
<td>51</td>
<td>9</td>
<td>8 950</td>
</tr>
<tr>
<td>Narre Warren</td>
<td>Rural Residential</td>
<td>10</td>
<td>20</td>
<td>14</td>
<td>29 630</td>
</tr>
<tr>
<td>Richmond</td>
<td>High Density Res</td>
<td>89</td>
<td>74</td>
<td>9</td>
<td>20 050</td>
</tr>
</tbody>
</table>

$^1$SMC is the Site Mean Concentration and is calculated using a volume weighted average of event mean concentrations.
2. **Total E. coli Load** [MPN] – the total E. coli load for each wet weather event was obtained by multiplying the EMC E. coli (in MPN/m³) by the total event volume (V).

The total E. coli load was hypothesised to be correlated with: PIF and antecedent climatic variables.

### Results and discussions

#### Event mean concentrations of E. coli

**Simple regressions.** The E. coli EMCs were significantly correlated with a number of different explanatory parameters but these parameters were not consistent between sites (Table 3). Narre Warren’s EMC of E. coli was well correlated with a common sewage constituent, ammonium, indicating these pollutants have similar processes and sources at this site. For example, high sewage infiltration could explain this relationship. Clayton’s EMC of E. coli is highly correlated with the event mean Total Dissolved Nitrogen (TDN) concentration and suggests that the sources and processes of E. coli and TDN are similar at this site. While sewage contains high levels of TDN, the majority of TDN in sewage is ammonium and since there was no correlation between E. coli and ammonium it is unlikely that the E. coli source at this site is anthropogenic.

Doncaster’s EMC of E. coli was only significantly correlated with the previous day’s total sunshine hours implying that the event mean E. coli levels were dependent on the amount of die-off caused by irradiation on the previous day. Richmond’s EMC of E. coli was most correlated with the average rainfall intensity, indicating that this site’s E. coli levels were well modelled using a simple ‘wash-off’ relationship.

**Multiple regressions.** Although significant simple regressions were obtained, very few had high levels of explanatory power; possibly because E. coli levels depend on not just one, but many, explanatory variables. To capture this dependence a multiple regression analysis (using a maximum of two explanatory variables) was conducted.

The Clayton site’s EMC of E. coli was most highly correlated with the average rainfall intensity and the average antecedent weekly temperature ($R_1 = 0.84, p_1 < 0.01$) while Richmond’s EMC of E. coli was also highly correlated with the average rainfall intensity.
and another antecedent climatic condition (the prior day’s total sunshine hours, $R = 0.98$, $p < 0.01$). Both Doncaster’s and Narre Warren’s E. coli EMC were highly correlated with the average flow intensity and different climatic conditions ($\text{Sun}_1$, $R = 0.81$, $p < 0.05$ and $\text{Evap}_{1}$, $R_l = 0.85$, $p < 0.01$, respectively).

While all of the sites’ highest correlations were with inconsistent pairs, there are some similarities between sites. Both the Clayton and Richmond sites shared average rainfall intensity for their most significant correlations, while both Narre Warren and Doncaster shared average flow intensity. Furthermore, the pair of average rainfall intensity and the previous day’s total sunlight hours was able to significantly explain the variation in E. coli EMCs consistently across all four sites. Using these variables the following regression relationships were found:

- **Clayton**: $\log(\text{E. coli EMC}) = 10.4 \times \text{PI}_{\text{ave}} + 0.02 \times \text{Sun}_1 + 2.1$ ($R_l = 0.68$, $p < 0.01$)
- **Doncaster**: $\log(\text{E. coli EMC}) = 19.6 \times \text{PI}_{\text{ave}} - 0.08 \times \text{Sun}_1 + 3.7$ ($R_l = 0.77$, $p < 0.01$)
- **Narre Warren**: $\log(\text{E. coli EMC}) = 9.8 \times \text{PI}_{\text{ave}} - 0.01 \times \text{Sun}_1 + 3.9$ ($R_l = 0.70$, $p < 0.01$)
- **Richmond**: $\log(\text{E. coli EMC}) = 11.9 \times \text{PI}_{\text{ave}} - 0.05 \times \text{Sun}_1 + 3.8$ ($R_l = 0.94$, $p < 0.01$)

The above regression models have an average correlation coefficient of over 0.77 ($R_l$) indicating that these variables could satisfactorily explain, on average, nearly 60% of the variation in inter-event E. coli levels (using the coefficient of determination). Although the focus of the paper is on the variation of E. coli between events, it is clear that some general trends exist between these coefficients and some site characteristics. These trends include: a) as catchment area increases the PI$_{\text{ave}}$ coefficient increases, b) as catchment area increases the Sun$_1$ coefficients decrease, and c) as the E. coli SMC increases the intercept values increase. Future work by the authors is aimed at using data from more sites to validate the aforementioned trends.
It may be concluded that the parameters which represent the ‘build-up/die-off’ processes (e.g. antecedent climatic conditions) and the ‘wash-off’ processes (e.g. average rainfall intensity) should be included when modelling the inter-event variation of *E. coli* levels in urban stormwater.

**Total *E. coli* load**

*Simple regressions.* Table 4 shows that significant correlations were obtained between the total *E. coli* load and the parameter PIF$_x$ (which represents the sum of the rainfall intensity to a power $x$), with PIF$_2$ providing the most accurate relationship (on average). PIF$_2$, which is an indicator of the total raindrop energy over the event, has been shown to provide good explanation for stormwater sediment transport (*Francey et al.*, 2005). *E. coli* is known to behave as particulate matter (as they have diameters > 0.45 μm), or combines with particulate matter in the environment, which clarifies why PIF$_2$ is a good parameter for describing the inter-event variation of *E. coli* levels.

*Multiple regressions.* Using the outcomes of the simple regression analysis, a multiple regression analysis was conducted between PIF$_2$ and a number of other explanatory variables. Table 5 shows only the correlations which were consistently significant across all four sites.

Four antecedent climatic conditions were significantly correlated with the total load of *E. coli* at each site when paired with PIF$_2$. On average, across all four sites, the previous day’s rainfall showed the most promising results with a correlation coefficient ($R_l$) of 0.90. Rain$_x$ is commonly used to model the amount of material available for transport during storm events. The previous day’s rainfall total is the most significant in this approach and this implies that the ‘memory’ of the sites’ accumulated load is less than...

---

**Table 4** Results of the simple regression analysis using total *E. coli* Load as the response variable and PIF$_x$ as the explanatory variable. (**Bold** font: $p < 0.01$, normal font: $p < 0.05$ and ‘–’: $p > 0.05$)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Clayton</th>
<th>Doncaster</th>
<th>Narre Warren</th>
<th>Richmond</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$R$</td>
<td>$R_l$</td>
<td>$R$</td>
<td>$R_l$</td>
</tr>
<tr>
<td>PIF$_1$</td>
<td>0.76</td>
<td>0.87</td>
<td>–</td>
<td>0.67</td>
</tr>
<tr>
<td>PIF$_2$</td>
<td>0.87</td>
<td>0.89</td>
<td>–</td>
<td>0.72</td>
</tr>
<tr>
<td>PIF$_3$</td>
<td>0.76</td>
<td>0.91</td>
<td>–</td>
<td>0.73</td>
</tr>
</tbody>
</table>

**Table 5** Results of the multiple regression analysis using *E. coli* Load as the response variable showing only the results which were consistently significant across all four sites (**Bold** font: $p < 0.01$, normal font: $p < 0.05$ and ‘–’: $p > 0.05$)

<table>
<thead>
<tr>
<th>Variable 1</th>
<th>Variable 2</th>
<th>Clayton</th>
<th>Doncaster</th>
<th>Narre Warren</th>
<th>Richmond</th>
</tr>
</thead>
<tbody>
<tr>
<td>PIF$_2$</td>
<td>Evap$_7$</td>
<td>0.92</td>
<td>0.96</td>
<td>–</td>
<td>0.82</td>
</tr>
<tr>
<td>PIF$_2$</td>
<td>Evap$_{14}$</td>
<td>0.92</td>
<td>0.95</td>
<td>–</td>
<td>0.81</td>
</tr>
<tr>
<td>PIF$_2$</td>
<td>Evap$_{28}$</td>
<td>0.94</td>
<td>0.95</td>
<td>–</td>
<td>0.91</td>
</tr>
<tr>
<td>PIF$_2$</td>
<td>Rad$_0$</td>
<td>0.88</td>
<td>0.94</td>
<td>–</td>
<td>0.81</td>
</tr>
<tr>
<td>PIF$_2$</td>
<td>Rad$_1$</td>
<td>0.89</td>
<td>0.94</td>
<td>–</td>
<td>0.83</td>
</tr>
<tr>
<td>PIF$_2$</td>
<td>Rad$_7$</td>
<td>0.89</td>
<td>0.95</td>
<td>–</td>
<td>0.84</td>
</tr>
<tr>
<td>PIF$_2$</td>
<td>Rad$_{14}$</td>
<td>0.90</td>
<td>0.95</td>
<td>–</td>
<td>0.87</td>
</tr>
<tr>
<td>PIF$_2$</td>
<td>Rad$_{28}$</td>
<td>0.92</td>
<td>0.95</td>
<td>–</td>
<td>0.89</td>
</tr>
<tr>
<td>PIF$_2$</td>
<td>Rain$_1$</td>
<td>0.87</td>
<td>0.94</td>
<td>–</td>
<td>0.92</td>
</tr>
<tr>
<td>PIF$_2$</td>
<td>Temp$_0$</td>
<td>0.89</td>
<td>0.95</td>
<td>–</td>
<td>0.83</td>
</tr>
</tbody>
</table>
one day (i.e. the build-up of \textit{E. coli} might follow a function that has an asymptote after one day of antecedent conditions).

The remaining parameters listed in Table 5 are all climatic conditions that represent different factors known to effect the survival of \textit{E. coli}. For example, a) antecedent evaporation could be representing the desiccation effects on the survival of \textit{E. coli}, b) antecedent radiation might be representing the die-off of \textit{E. coli} due to irradiation, and c) antecedent average temperatures could be modelling the effects that temperature has on the survival of \textit{E. coli}.

**Recommendations for a new microorganism model**

The significant relationships found in the preceding sections will be used to create a model which is capable of predicting the levels of microorganisms in urban stormwater systems. There are several recommendations for this new model that follow from the above results and discussions, including:

- simple regressions were not able to consistently explain \textit{E. coli} EMCs in urban stormwater and hence the model requires more than just one explanatory variable; the model should include conceptual representations of build-up, die-off and wash-off processes;
- inter-event \textit{E. coli} levels were related to antecedent climatic conditions and it is suggested that antecedent climatic conditions are used to model both the microorganism build-up, die-off and wash-off processes;
- rainfall intensity and PIF\textsubscript{x} had high predictive power for the variation of inter-event \textit{E. coli} levels and therefore it is recommended that one of these variables be included in the model to represent the microorganism wash-off process; and,
- although many significant linear correlations were determined, the majority of the high powered regressions were obtained using non linear correlation coefficients and therefore the model should take this non linearity into account.

While this study developed relationships that explain the variation of \textit{E. coli} between wet weather events, there has been no attempt at explaining: a) the variation of \textit{E. coli} levels during events, or b) the variation of \textit{E. coli} levels between different sites. Current work by the authors will help elucidate these areas.

**Conclusions**

The results of these correlation analyses helped identify the most influential factors that explain the level of \textit{E. coli} in urban stormwater systems. The most important variables that explain the variation of \textit{E. coli} between wet weather events were antecedent climatic conditions and rainfall intensity. The significance of each parameter was presented and discussed in relation to a new urban microorganism model. Future work endeavours to use the findings of this research to create a model that predicts microorganism levels in urban stormwater.

**References**


