Catchment process affecting drinking water quality, including the significance of rainfall events, using factor analysis and event mean concentrations

Kathy Cinque and Niranjali Jayasuriya

ABSTRACT

To ensure the protection of drinking water an understanding of the catchment processes which can affect water quality is important as it enables targeted catchment management actions to be implemented. In this study factor analysis (FA) and comparing event mean concentrations (EMCs) with baseline values were techniques used to asses the relationships between water quality parameters and linking those parameters to processes within an agricultural drinking water catchment. FA found that 55% of the variance in the water quality data could be explained by the first factor, which was dominated by parameters usually associated with erosion. Inclusion of pathogenic indicators in an additional FA showed that *Enterococcus* and *Clostridium perfringens* (*C. perfringens*) were also related to the erosion factor. Analysis of the EMCs found that most parameters were significantly higher during periods of rainfall runoff. This study shows that the most dominant processes in an agricultural catchment are surface runoff and erosion. It also shows that it is these processes which mobilise pathogenic indicators and are therefore most likely to influence the transport of pathogens. Catchment management efforts need to focus on reducing the effect of these processes on water quality.

Key words | catchment, drinking water quality, event mean concentration, factor analysis

Kathy Cinque (corresponding author)
Melbourne Water,
PO Box 4342,
Melbourne, Victoria 3001,
Australia
Tel.: +61 3 9235 2138,
Fax: +61 3 9235 7226
E-mail: *kathy.cinque@melbournewater.com.au*

Niranjali Jayasuriya
RMIT University,
GPO Box 2476V,
Melbourne, Victoria 3001,
Australia

INTRODUCTION

In the drinking water industry, current best practice in terms of reducing risks to consumers is the use of multiple barrier and preventive risk management. Some examples of common barriers used for the protection of drinking water include detention in large storage reservoirs, treatment/disinfection and a closed distribution system. One of the most critical barriers, however, is catchment management. It is critical because it is usually the first barrier in the system. It is also a preventive measure in that it attempts to control contamination at the source before it enters the water. This preventive approach is emphasised in the Australian Drinking Water Guidelines (ADWG) as it provides a greater surety of the absence of contaminants than does subsequent removal or reduction of contaminants by treatment (NH&MRC 2004).

To ensure that catchment management is effective, it is beneficial to know the processes within the catchment which have the greatest impact on water quality. Improvement works that are known to affect these processes can then be implemented within the catchment with an increased confidence that they will improve water quality. It can be difficult to measure the impact of catchment improvement works such as fencing of streams and gully replanting, and such measurement has often been based on visual inspections rather than on water quality observations (Benham *et al.* 2005). This can be due to the length of time taken for such works to become effective, along with the inherent difficulty in collecting relevant and informative water quality data, especially in terms of pathogenic data. Quantification of the benefit of works for

doi: 10.2166/wh.2010.162

water quality is therefore often challenging and rarely undertaken (Ferguson 2005). By gaining a knowledge of the processes that have the most impact on water quality and targeting those areas, the catchment manager can have confidence that the improvement works will be effective even if not immediately evident in the water quality monitoring data. The ability to show that the on-ground works can have a positive impact on drinking water quality is important for various stakeholders including regulators and the community.

Studies looking at the transport of constituents, and especially pathogens and their indicators, have previously been undertaken at a laboratory or plot scale (Atwill *et al.* 2002; Davies *et al.* 2004) with logistical difficulties often preventing catchment-scale studies. The issue of extrapolating, or up-scaling, results has been identified by Ferguson (2005) as an area where great care should be taken. This study looks at catchment-scale water quality data and draws conclusions about catchment processes affecting water quality. It shows that a water quality monitoring programme can be used to identify the most appropriate catchment management works to improve drinking water quality.

The Tarago Reservoir catchment in Victoria, Australia, was chosen as the study catchment based on a number of factors. It is an agricultural catchment that supplies water for drinking, it has a good quality and extensive data set, both in terms of length and number of parameters and it has a current water quality monitoring programme that includes storm event and pathogenic indicator sampling. Additionally, in an effort to reduce the risks to drinking water quality, a number of catchment management works have been undertaken in the catchment over the past 15 years. These works include: stream frontage management and fencing, introducing whole farm planning initiatives, roadside management and septic tank management, among others. Catchment managers, water authorities and the community all have an interest in the ability of these works to improve water quality.

This paper uses two different techniques, factor analysis (FA) and analysis of event mean concentrations (EMCs), to determine the processes within the Tarago catchment that have the most impact on water quality. The results of these analyses can be used to help set catchment management priorities.

Multivariate statistics, including FA, have been used in various studies assessing different aspects of water quality. In most cases FA is used to evaluate the spatial and temporal changes in water quality and to determine trends. Groundwater analysis has been the focus of a number of studies: for example, delineating the boundaries where groundwater is affected by seawater intrusion (Liu *et al.* 2003) and gaining a better understanding of the processes affecting shallow groundwater in an irrigation district (Ahmed *et al.* 2005). Paul *et al.* (2006) used FA to group different watersheds with similar characteristics to allow the study of the catchments as a group, effectively reducing the amount of sampling and the number of individual studies required. FA will be used in this work for exploratory water quality analysis. It will be used to interpret water quality data and relate it back to processes, environmental or anthropogenic, within the catchment. Siriwardhema (1999) carried out similar work in the Tarago catchment with data from 1974 to 1993. He found that the most significant processes in the catchment were erosion and surface runoff. The current study includes an additional five years of data as well as taking into consideration pathogenic indicators, ensuring that the results are relevant to the implementation of catchment management initiatives from a drinking water perspective.

Evaluation of EMCs is used extensively in urban stormwater engineering (Signor *et al.* 2005). It has been used to characterise and quantify runoff quality as well as evaluate the significance of the first flush phenomenon (Hallberg 2006; Yongjing *et al.* 2009). Dynamic EMCs have been used to explain the variation in pollution levels over the period of a runoff event coming from a paved area (Kim *et al.* 2007). The results were then used to suggest appropriate treatment solutions. In this study EMCs are compared with baseline values in a natural stream in an agricultural catchment to allow an assessment of the effect of rainfall on constituent concentrations. In a predominately pervious catchment, such as an agricultural one, increases in pollutant loads during storm events and the significance of the increase has not been widely researched.

In order to carry out these analyses, an extensive amount of data is required and some of the problems associated with collecting this data are discussed. The paper will endeavour to assess catchment processes such as erosion, direct contamination and groundwater contamination as well as the effect of rainfall runoff. Following the analysis it is hoped that certain catchment management works can be identified that will provide obvious water quality benefits leading to recommendations for improving the management of the catchment.

MATERIALS AND METHODS

Study area

The Tarago Reservoir is a drinking water reservoir located about 100 km east of Melbourne, Victoria, which has a catchment area of 11,400 hectares. Of this area approximately 2,300 ha contributes to direct runoff into the reservoir and the remaining areas drain into three perennial streams, the West Tarago River (with a catchment area of 7,200 ha), East Tarago River (1,300 ha) and Crystal Creek (800 ha), see Figure 1.

The West branch catchment of the Tarago River comprises state-owned forest while the Crystal Creek catchment is a mixture of state and private forest. The East branch catchment is mainly agricultural and rural residential land and is the focus of this study (Figure 1). The East catchment only contributes approximately 25% of the total

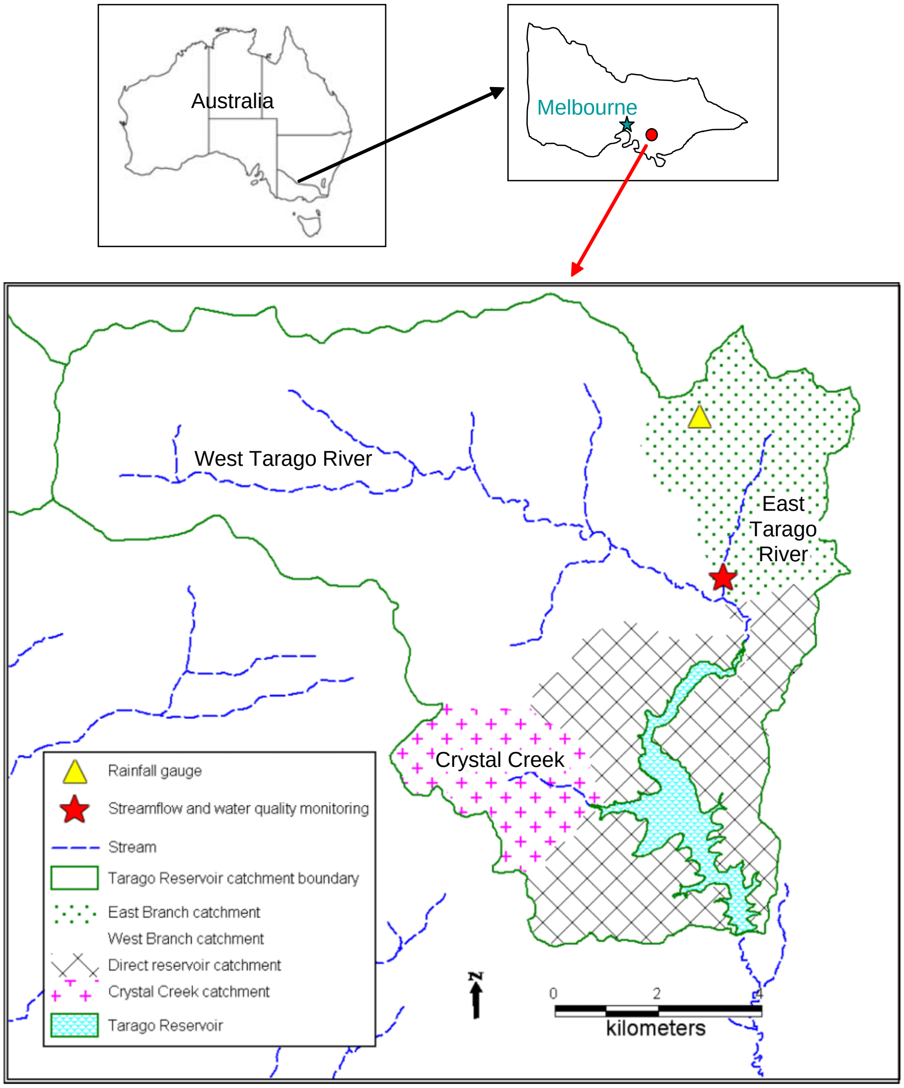

Figure 1 | Tarago Reservoir catchment.

inflow to the reservoir but from a drinking water perspective it represents the majority of the risk. This is due to the rural land uses that it supports in its catchment, which include horticulture, residential, a dairy and grazing. Horticulture mainly consists of annual crops of potatoes and carrots. There are 36 houses in the East catchment and as there is no reticulated sewerage system, all houses have on-site wastewater treatment. The East catchment supports one dairy, comprising over 150 cattle in total, both fully grown heifers and calves. There are also over 150 grazing cattle and one deer farm that is thought to stock up to 300 deer at a time. Land uses that encourage human and/or domestic livestock habitation are known sources of pathogens (Jiang *et al.* 2005) and the greatest risks to consumers of drinking water are pathogenic microorganisms (NH&MRC 2004). The pathogens of most concern in a drinking water catchment are those which have a small infectious dose, are shed in high numbers by an infected host, are environmentally robust and can cause illness in humans: for example, the protozoan pathogens *Cryptosporidium* and *Giardia*.

The Tarago Reservoir was discontinued from use for Melbourne in 1993 because of poor water quality, the lack of appropriate treatment and the availability of more reliable water sources elsewhere. It was reconnected in June 2009 following the construction of a water treatment plant. Owing to the reservoir's use as a drinking water source it is vital that the risks to water quality in the reservoir are understood, managed and most importantly reduced where possible.

Data availability

Routine data collection has been undertaken in the Tarago catchment at the East branch site from 1974 until the present by the water authority, Melbourne Water, and its predecessors. The data set comprises 24 years of data, up to the end of 2007, which includes a period of approximately 10 years when no data was collected because of budgetary constraints. The data set includes almost 40 different parameters, although not all are consistently tested, most likely because of changing priorities. The parameters chosen for analysis were consistently tested throughout the sampling period and are broadly accepted as important indicators of raw water quality. Parameters now known to be important in terms of drinking water, such as dissolved organic carbon, may not have been included in the analysis owing to lack of sufficient data. In terms of pathogenic organisms the problems associated with sampling them directly, which include complexity of the sampling and testing methods, the cost involved and the time required, mean that the data set in this catchment was limited and they are therefore excluded from the analysis. Routine pathogen monitoring is also thought to be of little value in terms of determining the public health risk (Allen *et al.* 2000) because of these sampling difficulties. Monitoring for pathogenic indicators, such as *Escherichia coli*, is standard drinking water quality practice (Haydon 2006).

A range of parameters were chosen for this study based on the amount of data available (see Table 1). Water quality samples were taken as grab samples from the monitoring point indicated on Figure 1. The smaller number of data points for the pathogenic indicators are a result of the comprehensive monitoring programme implemented by Melbourne Water, beginning in 2004. Prior to this these parameters were not monitored.

Water quality monitoring during significant rainfall events was undertaken using automated sampling equipment that collected up to 24 samples per event. The programme was based on a system developed by Roser *et al.* (2002). The sampling equipment was located in a secure housing adjacent to the East Branch sampling site and included a pressure sensor to detect stream height, a pump, a logger, 24 × 1 l sample bottles and 24 × 10 l sample bottles. The equipment was set up to automatically trigger and start sampling at a set stream height. The trigger level and sampling interval were determined based on rainfall and hydrological data over a number of years and aimed to capture both the rising and falling limb of the hydrograph. Samples were refrigerated and transported to the laboratory within 24 h of collection. Over the four-year period in which the samplers were deployed they triggered a total of five times.

Flow and rainfall data is also necessary to enable the calculation of storm flow weighted averages and to assess storm flow characteristics. Flow data was obtained from a data logger which recorded hourly flow measurements on the East branch at the monitoring point indicated on Figure 1. Rainfall is collected using a 203 mm tipping bucket

Table 1 | Parameters used in the analysis

Parameter	Units	Collection period	Average frequency	No. data points	Analysed*
pH	–	1974–2007	2 weeks	450	FA
Colour	PtCo	1974–2007	2 weeks	454	FA
Turbidity	NTU	1974–2007	2 weeks	474†	All
Iron	$\mathrm{mg\,l^{-1}}$	1974–2007	2 weeks	398	FA
Manganese	$\mathrm{mg\,l^{-1}}$	1974–2007	2 weeks	363	FA
Nitrate	$\mathrm{mg\,l^{-1}}$	1974–2007	1 month	297†	All
Phosphorus	$\mathrm{mg\,l^{-1}}$	1974–2007	2 weeks	364†	All
Ammonia	$\mathrm{mg\,l^{-1}}$	1989–2007	2 weeks	203	FA
Suspended solids	$\mathrm{mg\,l^{-1}}$	1989–2007	2 weeks	169†	All
TKN	$\mathrm{mg\,l^{-1}}$	1989–2007	2 weeks	231†	All
Total coliforms	orgs/100 ml	2005–2007	2 months	20	FAI
E. coli	orgs/100 ml	2005–2007	1 month	43†	FAI + E
Enterococcus	orgs/100 ml	2005–2007	1 month	50†	FAI + E
C. perfringens	orgs/100 ml	2004–2007	1 month	36	FAI
Nitrite	$\mathrm{mg\,l^{-1}}$	1989–2007	1 month	81†	E
Orthophosphate	$\mathrm{mg\,l^{-1}}$	1989–2007	2 weeks	151†	E

*All = analysed in both FA and events; FA = analysed in FA; FAI = analysed in FA with indicators; E = analysed in events.
†Inclusive of event data.

at a site within the East Tarago catchment, see Figure 1. Both streamflow and rainfall data are collected and quality checked by Melbourne Water.

Before undertaking analysis of any data set it is important that the limitations of that data are explored. Water quality data sets taken over a long period of time are likely to exhibit some of the following problems: missing data, errors, multiple observations, censored data, outliers and changes to measurement or recording technique (Gilbert 1987). Missing data within the selected Tarago data set was not deemed important owing to the large number of observation points (Tabachnick & Fidell 1996). It was found that less than 1% of the data points were multiple observations and these were deleted from the data set. Less than 1% of the data were determined to be outliers; that is, more than three times the standard deviation away from the median. Outliers were individually assessed so that if that point corresponded with high values in other parameters, indicating an event, environmental or anthropogenic, it was considered to be reasonable and retained. However, if it was the only high value for that time stamp it was deleted from the data set. Censored values are defined as those which fell below a level associated with some minimum acceptable level of reliability (Gilliom *et al.* 1984). These values only made up 2% of the total data set. Censored data were replaced with half their detection limit, which is an effective and efficient method of dealing with these values (Zhang *et al.* 2004).

Data analysis

There are many processes within a catchment which may affect water quality, both environmental and anthropogenic. In an attempt to determine the most significant and dominant catchment processes, FA and event data analysis was undertaken.

Factor analysis (FA)

FA is a multivariate statistical method that can be used to reduce the amount of data being used to predict a response. FA summarises the data by means of a linear combination of observed variables with the goal being to determine the smallest number of variables that will explain most of the variance. It attempts to explain the common variance shared by the observed variables and tries to find any underlying factors that are responsible for the interrelationships between observed variables.

FA is particularly beneficial for data sets with a large number of variables as it can reduce the number of variables to those which explain the most variance and therefore contain the most information. When analysing water quality data, FA can also be used as a means to identify the major pollution sources within a catchment by grouping related parameters. Parameters analysed on an ad hoc basis were not included in the FA.

Interpreting the loadings obtained from FA is an important part of the process. The loadings represent the degree to which that variable, or parameter, is influenced by that factor. According to a large-scale study by Liu *et al.* (2003), factor loadings can be classified as strong, medium and weak corresponding to values of >0.75, $0.75 > 0.5$ and $0.5 > 0.3$, respectively. Other researchers, such as Comrey & Lee (1992), indicated a similar interpretation of results: $> 0.71 =$ excellent; $0.7 > 0.63 =$ very good; $0.62 > 0.55 =$ good; $0.54 > 0.45 =$ fair and $0.44 > 0.32 =$ poor. Based on these studies, in this work a factor loading of 0.75 was chosen as indicating that the parameter was a major influence for that factor.

Event analysis using event mean concentrations (EMCs)

When assessing events it is not just the maximum or average contaminant concentration that is significant but rather the total storm flow weighted average. This is calculated using the EMC, which is defined as the total storm load (mass) divided by the total runoff volume, see Equation (1):

$$\mathrm{EMC}_j = \frac{\sum_{i=1}^{n} Q_{ij} C_{ij}}{\sum_{i=1}^{n} Q_{ij}} \tag{1}$$

where, EMC_j is the EMC of the jth event and Q_{ij} and C_{ij} are the ith flowrate and concentration pair measured during the jth event.

This equation is more correctly displayed iteratively but the difference in values obtained from each method is negligible when compared with other measurement uncertainties (Signor *et al.* 2005).

A *t*-test was used to determine whether the EMCs were statistically significantly different from the mean concentrations during baseflow conditions. A significance level ($\propto$) of 0.05 was chosen as the basis for determining whether the means were statistically different. This is the most commonly used significance level in scientific research (Varkevisser *et al.* 2003).

RESULTS AND DISCUSSION

Factor analysis (FA)

The data set used, shown in Table 1, has over 3,500 data points, which is more than adequate given similar FA carried out by Boyacioglu (2006) and Shrestha & Kazama (2007) who used less than 500 and less than 2,000 data points, respectively.

Undertaking FA requires a correlation matrix to be calculated to determine the factorability of the data, or the amount of intercorrelation between parameters. Tabachnick & Fidell (1996) state that, for a correlation matrix without any correlations over 0.3, FA should be reconsidered. A sufficient number of significant correlations indicate that there may be some underlying processes affecting several parameters and that undertaking FA could successfully reduce the dimensionality of the original data set. Table 2 shows the correlations between parameters and shows that over a third of the correlations were greater than 0.3. This was deemed to be a sufficient number to indicate the possibility of an underlying process and therefore FA could be carried out.

The next step in FA involves determining the appropriate number of factors that need to be extracted in order to explain most, or a sufficient amount, of the variance in the data set. This can be done using a scree test which is a visual test where the eigenvalues, which measure the significance of the factor, are plotted against the factors. The number of factors selected corresponds to the point at which the graph becomes approximately horizontal or at which the eigenvalues go below one. Figure 2 shows the scree plot for this data.

From analysis of this graph it is reasonable to extract three factors for the East catchment as after this point the eigenvalues are clearly below one. Three factors should be enough to explain the majority of the variance and

Table 2 | Correlation matrix for water quality parameters in the East Tarago catchment

	pH –	**Colour (PtCo)**	**Turbidity (NTU)**	**Iron (mg l^{-1})**	**Manganese (mg l^{-1})**	**Nitrate (mg l^{-1})**	**Phosphorus (mg l^{-1})**	**Ammonia (mg l^{-1})**	**Suspended solids (mg l^{-1})**	**TKN (mg l^{-1})**	**Total coliforms (orgs/100 ml)**	***E. coli* (orgs/100 ml)**	***Enterococcus* (orgs/100 ml)**	***C. perfringens* (orgs/100 ml)**
pH	1.0000	0.0889	0.0004	0.0349	0.0019	0.0000	0.0026	0.0773	0.0104	0.0002	0.0086	0.0112	0.0000	0.0036
Colour		1.0000	0.2948	0.0511	0.1763	0.0042	0.1372	0.0012	0.6943	0.3878	0.0028	0.0205	0.0856	0.0002
Turbidity			1.0000	0.4259	0.3690	0.0104	0.4802	0.0112	0.7646	0.7566	0.1443	0.1668	0.3433	0.1293
Iron				1.0000	0.5383	0.0017	0.3194	0.0408	0.4266	0.4251	0.1720	0.2057	0.5196	0.1369
Manganese					1.0000	0.0275	0.5900	0.0383	0.7707	0.6552	0.4113	0.4713	0.7804	0.3862
Nitrate						1.0000	0.0020	0.0227	0.0089	0.0114	0.0444	0.4725	0.2338	0.0146
Phosphorus							1.0000	0.0001	0.5688	0.4066	0.0490	0.1994	0.4498	0.2776
Ammonia								1.0000	0.0000	0.0110	0.0729	0.0546	0.0404	0.0317
Suspended solids									1.0000	0.8108	0.1589	0.1250	0.2433	0.1156
TKN										1.0000	0.1562	0.2627	0.2750	0.3273
Total coliforms											1.0000	0.2214	0.1260	0.0214
E. coli												1.0000	0.6682	0.3475
Enterococcus													1.0000	0.3066
C. perfringens														1.0000

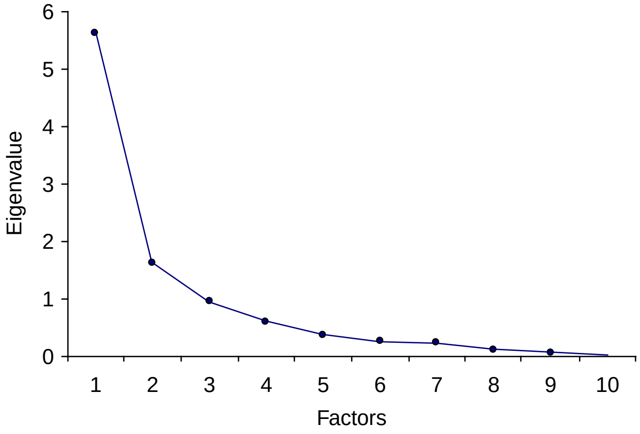

Figure 2 | Scree plot for the East Tarago catchment.

assess the processes in the catchment affecting water quality. Table 3 shows the amount of variance explained by each of the first three factors as well as the total cumulative variance.

Over 80% of the variance in the data is explained with the first three factors. Additionally the first factor explains over 55% of the data variance and so this will represent the dominant transport processes within the catchment.

In order to find the best solution when undertaking FA, rotation techniques may be implemented. This improves the interpretability and scientific utility of the solution but it does not improve the quality of the results (Tabachnick & Fidell 1996). There are many rotational techniques but by far the most popular is the varimax method devised by Kaiser (1958). Varimax rotation maximises the variance of the squared loadings for each factor and polarises the loadings so they are either high or low thus making it easier to identify factors with specific observed parameters (Marcoulides & Hershberger 1997). Table 4 shows the varimax rotated factor loadings for the East Tarago catchment, which were obtained using the computer program MINITAB®.

Table 3 | Factor variability for the East Tarago catchment

Factor	Variability (%)
1	55.6
2	16.1
3	10.5
Cumulative total variability	82.2

Table 4 | Varimax rotations for the East Tarago catchment

Parameter	Units	Factor 1	Factor 2	Factor 3
pH	–	−0.015	*−0.833*	0.173
Colour	PtCo	*0.876*	0.087	−0.001
Turbidity	NTU	*0.942*	−0.199	0.014
Iron	$mg\,l^{-1}$	*0.854*	−0.372	0.117
Manganese	$mg\,l^{-1}$	*0.875*	−0.097	0.145
Nitrate	$mg\,l^{-1}$	−0.074	0.091	*−0.976*
Phosphorus	$mg\,l^{-1}$	*0.853*	0.003	−0.053
Ammonia	$mg\,l^{-1}$	−0.06	*0.834*	0.060
Suspended solids	$mg\,l^{-1}$	*0.949*	0.106	−0.050
TKN	$mg\,l^{-1}$	*0.879*	0.060	0.162

Values in italic indicate influential parameters for that factor (i.e. values with a loading over 0.75).

It is evident from the dominant parameters for factor 1 that it represents surface runoff and erosion. Parameters such as turbidity, suspended solids and colour are strongly associated with sediment and the dislodging of soil from the surface. Iron and manganese, which are also influential in the erosion factor, are most likely the result of the movement of the mineral-rich clay soils within the East Tarago catchment (Siriwardhema 1999). The remaining influential parameters in the erosion factor can be related back to a combination of both soil type and land use. Both total Kjeldahl nitrogen (TKN) and phosphorus had high loadings for factor 1, most likely as a result of the agricultural land use. TKN can be an indication of disposal of sewage, fertiliser use or decaying plant matter (United States Geological Survey 2004). Its inclusion in the erosion factor could be an indication that the septic systems within the catchment are overflowing and impacting surface water quality, which may require more investigation and possible remediation works. The inclusion of phosphorus could possibly be due to fertiliser application or the presence of animals in the catchment. The most likely reason, however, for it being included in the erosion factor is that the soils in the East Tarago catchment have naturally high phosphorus levels (Swan & Volum 1984) and as the land is cleared it is prone to erosion. This supports the link between phosphorus and more recognised indicators of erosion such as turbidity and suspended solids.

Both ammonia and pH have high loadings in factor 2 and they are negatively correlated. The soluble nature of ammonia in water means that the likely transport mechanism will be in solution as surface runoff or as groundwater. The delivery of ammonia to the stream is not related to erosion or the movement of sediments and therefore not influential in factor 1. The negative correlation between the two parameters is due to the fact that as pH decreases the solubility of ammonia increases.

The third factor relates to nitrate. Nitrate, like ammonia, is soluble and this factor is therefore again indicating the influence of groundwater or surface runoff on the water quality. Nitrate has been shown to partition into soil pores and is flushed out when the soil moisture increases (Rosenzweig *et al.* 2008). This indicates that the catchment needs to be partially saturated before nitrate will be released. As with factor 2, this factor is not related to erosion or to particle movement. Ammonia and nitrate influencing two different factors suggests that their transport mechanisms are also unique from each other.

The overall ability of FA to reduce the number of parameters needed to explain the majority of the data in this case was weak. That is, to explain 82% of the variance across three factors all 10 parameters were needed. FA was, however, successful in identifying the most significant processes within the catchment.

FA with the inclusion of pathogenic indicators

Including the pathogenic indicators in a FA was undertaken to determine the processes that are most likely to be related to pathogen movement. This was done as a separate piece of work because of the shorter time period of data collection. There were 19 days over a 3-year period where all parameters were collected and with 14 parameters this resulted in 266 data points; this was still deemed as an acceptable number given the work by Boyacioglu (2006) who separated their limited data points into two groups, one of 180 and one of 350. Varimax rotation with three factors extracted was undertaken (see Table 5).

This analysis showed that the indicators *Clostridium perfringens* (*C. perfringens*) and *Enterococcus* were related to factor 1, which is the surface runoff and erosion factor. This suggests that these indicators are either attached to soil particles (as suggested by Mallin *et al.* 2000) or behave in a similar way. Factor 2 was dominated by total coliforms, which are commonly but incorrectly used as an indicator of faecal pollution. Total coliforms can occur naturally; most have an environmental origin either as plant pathogens or as normal inhabitants of soil and water (Stevens *et al.* 2003). In this catchment total coliforms are not associated with factor 1 and, therefore, neither are they associated with surface runoff or *E. coli* (a type of faecal coliform found exclusively in the gut of humans and warm-blooded animals), which indicates that they are environmental rather than faecal in origin. It is interesting that *E. coli* is not a major influence with any of the factors, even those that include the other two faecal indicators. Both *Enterococcus* and *C. perfringens* can survive longer in the environment than *E. coli* (Medema *et al.* 1997) and this may be why they are more influential in factor 1. *E. coli* does have a loading of 0.622 for factor 1, which is high but not influential when judged on the criteria set out at the beginning of the analysis. This result may simply be due to the fact that there was an insufficient amount of data available (116 data points) to give a significant conclusion. In this case 11 out of the 14 parameters were needed to explain 79.4% of the variance in the data set.

Table 5 | East catchment factor loadings including indicators

Parameter	**Units**	**Factor 1**	**Factor 2**	**Factor 3**
pH	–	−0.14	0.01	*0.772*
Colour	PtCo	0.301	0.604	−0.582
Turbidity	NTU	*0.798*	0.377	−0.399
Iron	$mg\,l^{-1}$	*0.817*	0.433	−0.306
Manganese	$mg\,l^{-1}$	*0.86*	0.379	−0.266
Nitrate	$mg\,l^{-1}$	−0.127	−0.384	*−0.784*
Phosphorus	$mg\,l^{-1}$	*0.881*	0.107	−0.137
Ammonia	$mg\,l^{-1}$	−0.048	−0.658	−0.097
Suspended solids	$mg\,l^{-1}$	*0.768*	0.392	−0.464
TKN	$mg\,l^{-1}$	*0.812*	0.403	−0.393
Total coliforms	orgs/100 ml	0.303	*0.784*	0.145
E. coli	orgs/100 ml	0.622	0.241	0.279
Enterococcus	orgs/100 ml	*0.897*	0.119	0.087
C. perfringens	orgs/100 ml	*0.888*	−0.129	0.125
Variability (%)		44.8	17.5	17.1

Values in italic indicate an influential parameter for that factor.

Event analysis

Overall, five events were sampled and a summary of each, along with the EMCs, is provided in Table 6. Half the parameters that were analysed in the FA were not measured during events. However two additional parameters, nitrite and orthophosphate, were.

Each event is unique in terms of the environmental conditions such as rainfall intensity and preceding dry period and it is important to remember this when assessing and comparing concentrations generated during events. It is also important to consider the sampling regime, as the stage of the hydrograph where the samples are taken can affect concentrations. Each event's sampling regime was different owing to storm size, both peak and duration, and sampling equipment performance. Figure 3 shows the sampling regime during event 3, which is ideal as it captures the peak as well as the rising and falling limbs of the hydrograph.

Capturing a storm in this way could provide an indication as to whether there is a flush of contaminants from the catchment, as is reported in urban storm water literature (Lee *et al.* 2002; Taebi & Droste 2004), or whether once mobilised there is a seemingly endless supply of pollutants, as reported by Roser *et al.* (2002). This sampling regime is similar to the sampling regime in event 1. Events 2 and 4 were, however, sampled mostly on the falling limb of the hydrograph whereas event 5 was sampled on the rising limb.

Comparison of the EMCs with the average concentration from baseline sample data yielded some interesting results. As expected, most parameters showed a statistically significant ($\propto < 0.05$) increase in concentration during storm events (see Table 7). There were some, however, that were not statistically significantly different.

Both nitrate and nitrite were not significantly different from baseline values in the three events that had a smaller total rainfall and smaller intensities, namely events 2, 4 and 5. This observation may signify that the majority of the nitrogen reaching the stream is doing so via the groundwater and is not affected by rainfall unless the intensity is high. This observation is also consistent with the findings of the FA that nitrogen is not related to erosion and has a unique transport mechanism.

Enterococcus concentrations are not significantly different from baseline values in events 2 and 4, whereas for events 1, 3 and 5 they were significantly different. Of the five events, events 2 and 4 have the two lowest rainfall totals and the lowest rainfall intensities. It could therefore be hypothesised that this indicator needs a high rainfall intensity to be mobilised. Events 2 and 4 are also preceded

Table 6 | Event details and EMCs

Event details	**Units**		**Event 1**	**Event 2**	**Event 3**	**Event 4**	**Event 5**
Rainfall total	mm		54.21	15.47	24.64	13.81	17.52
Event intensity	mm/h		1.47	0.77	1.23	0.63	1.03
Preceding dry period	days		12	2	8	7	17
Number of samples	–		22	24	24	24	24
Parameters		**Average baseline conc.**	**EMCs**				
Turbidity	NTU	18.11	179.57	87.33	297.86	235.05	99.00
Nitrate	$mg\,l^{-1}$	0.980	1.220	0.916	2.124	0.954	0.862
Phosphorus	$mg\,l^{-1}$	0.07	0.467	0.194	0.378	0.503	0.338
Suspended solids	$mg\,l^{-1}$	36.29	348.89	113.37	463.47	398.99	175.88
TKN	$mg\,l^{-1}$	0.51	5.73	2.23	4.40	3.73	12.39
E. coli	orgs/100 ml	412.5	2215.86	1038.37	1292.72	1260.90	3434.50
Enterococcus	orgs/100 ml	355.03	6810.21	464.49	5415.45	482.82	1462.28
Nitrite	$mg\,l^{-1}$	0.002	0.010	0.012	0.005	0.001	0.001
Orthophosphate	$mg\,l^{-1}$	0.017	0.110	0.014	0.002	0.002	0.026

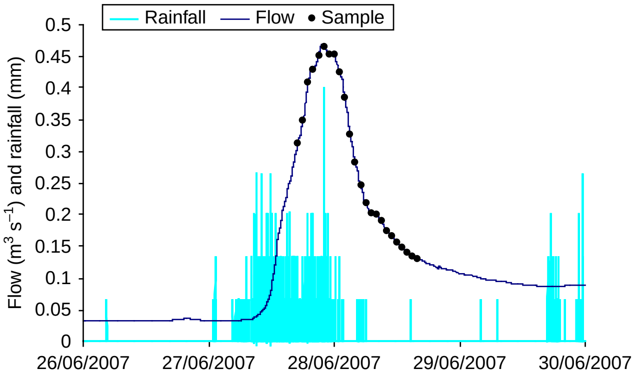

Figure 3 | Time versus concentration, flow and rainfall showing the sampling regime.

by a low number of days without rainfall which could indicate that *Enterococcus* did not have time to build up in the catchment. It could also be an indication that the level of *Enterococcus* in faeces is lower than the level of *E. coli*, as reported by Moriarty *et al.* (2008). The most likely reason for the lack of significant difference between baseline values and event values in events 2 and 4 is, however, the fact that these two events were sampled only on the falling limb of the hydrograph. This indicates that *Enterococcus* is impacted by a first flush, which means that the majority of the pollutants are transported in the initial part of the storm (Teabi & Droste 2004). It also indicates that there is a limited supply of this particular indicator in the catchment and it therefore needs time to build up in the catchment before it is mobilised. The lack of a similar finding for *E. coli* in these events suggests that the first flush is unique to *Enterococcus*. Therefore, monitoring *Enterococcus* to indicate pathogen movement, in this catchment, may underestimate the risk to public health as it is less readily available than other parameters.

It is interesting to assess the *E. coli* results in the two events samples on the falling limb, events 2 and 4. The event values show statistically significantly higher concentrations compared with baseline values in the falling limb of the hydrograph, suggesting an unlimited supply of this indicator. This is similar to findings reported by Roser *et al.* (2002).

In addition to the above, the first flush concept says that a certain storm event will produce more or less pollutant load to the receiving water depending on the intensity of the rainfall (Taebi & Droste 2004). For the five events in the East catchment that were sampled, the rainfall patterns and intensities varied significantly both in total volume and timing, making comparison difficult. However, the results for *Enterococcus*, as discussed above, seem to indicate a first flush effect. It would be necessary to sample a number of additional events if conclusions about parameter concentration were to be drawn based on rainfall intensity specifically.

The concentration of orthophosphate in the majority of events is not significantly affected by rainfall runoff. This suggests that, like nitrate and ammonia, it has a unique transport mechanism, possibly related to groundwater.

It is clear, from the five events sampled, that the concentrations of pollutants increase significantly during rainfall and therefore these times are when the risk to water quality is the greatest. This work highlights the importance of monitoring rainfall events in a drinking water catchment.

Table 7 | *t*-test statistics

	Units	Event 1	Event 2	Event 3	Event 4	Event 5
Turbidity	NTU	*8.31*	*5.89*	*6.75*	*54.67*	*11.85*
Nitrate	$\mathrm{mg\,l^{-1}}$	*5.78*	− 1.08	*15.24*	− 0.89	− 4.19
Phosphorus	$\mathrm{mg\,l^{-1}}$	*5.21*	*4.54*	*7.86*	*21.47*	*10.31*
Suspended solids	$\mathrm{mg\,l^{-1}}$	*5.11*	*4.04*	*6.82*	*20.52*	*9.90*
Total Kjeldahl nitrogen	$\mathrm{mg\,l^{-1}}$	*7.68*	*3.49*	*8.93*	*29.95*	*13.34*
E. coli	orgs/100 ml	*6.39*	*2.53*	*3.28*	*3.62*	*6.49*
Enterococcus	orgs/100 ml	*8.82*	0.49	*4.26*	0.65	*4.63*
Nitrite	$\mathrm{mg\,l^{-1}}$	*5.73*	1.32	*7.37*	− 5.58	− 5.37
Orthophosphate	$\mathrm{mg\,l^{-1}}$	− 3.66	− 1.84	− 10.21	− 10.21	*2.47*

Values in italic indicate that the mean concentration during a storm is significantly higher than that seen during baseflow conditions.

It also shows that the analysis of events is not always simple given the amount of variables that characterise an event and the often randomness of the sampling regime. Given these facts it is important in a drinking water supply catchment to monitor as many events as possible in order to gain a good understanding of the catchment and the risks to water quality.

CONCLUSIONS

Based on the statistical analysis reported above, the most dominant process in the East catchment affecting water quality is surface runoff and erosion. It follows that in high runoff events, parameters associated with these processes, including turbidity, suspended solids, TKN, phosphorus and *Enterococcus*, are even more dominant. This work supports the idea that the highest risk period in a catchment supplying drinking water is during storm events and, in order to adequately assess the risk, this is when monitoring should be undertaken.

Interestingly, an increase in *Enterococcus* during the events that were only sampled on the falling limb of the hydrograph was not evident, despite an increase in most other parameters. The flushing of *Enterococcus* from the catchment indicates a limited supply and may suggest that *Enterococcus* needs time to build up in the catchment prior to a significant rainfall event before it is detectable in the stream. In this catchment *Enterococcus* needs to be part of a range of indicators monitored for as they may not be as readily available in the environment as other indicators and may therefore underestimate the pathogenic risk.

The results of this study support the view that total coliforms are inadequate as indicators of pathogens as they were not related to surface runoff as were the other three indicators. As a consequence they should not be used as evidence for improvement in stream water quality from a public health perspective. Additionally *E. coli* may not be the best indicator in this catchment as it was not related to the erosion factor as strongly as *C. perfringens* or *Enterococcus*. It is important that these latter two organisms are included in any routine and event-based sampling programmes as they are more likely to be good indicators of pathogens.

Nitrate and nitrite concentrations are not affected by rainfall or by erosion and if a reduction in this parameter in stream is thought to be necessary, for example if there was an algal problem in Tarago Reservoir and it was nitrogen limited, management techniques other than on-ground physical barriers to flow may be required. The actual transport process would need to be determined to ensure that the alternative techniques were successful.

The results from the two analyses undertaken on the available data show that for the Tarago East catchment, surface runoff and erosion are the dominant processes affecting water quality and that rainfall has a significant impact on this process. It also showed that this process mobilises pathogenic indicators. Catchment management efforts, therefore, need to focus on lessening the effect of these processes in order to reduce risks and protect drinking water quality. This could be achieved through the implementation of stream frontage plantings, fencing off stream banks and stabilisation of slopes.

Monitoring and analysing more events will help confirm the conclusions reached in this study. It will also allow for events of higher or lower intensity to be monitored, enabling the effect of storm characteristics on contaminant concentrations found in streams to be determined.

REFERENCES

Ahmed, S. M., Hussain, M. & Abderrahman, W. 2005 Using multivariate factor analysis to assess surface/logged water quality and source of contamination at a large irrigation project at Al-Fadhli, Eastern Province, Saudi Arabia. *Bull. Eng. Geol. Environ.* **64**, 319–327.

Allen, M. J., Clancy, J. L. & Rice, E. W. 2000 The plain hard truth about pathogen monitoring. *J. Am. Water Works Assoc.* **92**(9), 64–76.

Atwill, E. R., Hou, L., Karle, B. M., Harter, T., Tate, K. E. & Dahlgren, R. A. 2002 Transport of *Cryptosporidium parvum* oocysts through vegetated buffer strips and estimated filtration efficiency. *Appl. Environ. Microbiol.* **68**(11), 5517–5527.

Benham, B. L., Robbins, J. H., Brannan, K. M., Mostaghimi, S., Diliaha, T. A., Pease, J. W. & Smith, E. P. 2005 Development of survey-like assessment tools to assess agricultural best management practice performance potential. *J. Soil Water Conserv.* **60**(5), 251–259.

Boyacioglu, H. 2006 Surface water quality assessment using factor analysis. *Water SA* **32**(3), 389–393.

Comrey, A. L. & Lee, H. B. 1992 *A First Course in Factor Analysis*. Lawrence Erlbaum Associates, New Jersey.

Davies, C. M., Ferguson, C. M., Kaucner, C., Krogh, M., Altavilla, N., Deere, D. & Ashbolt, N. J. 2004 Dispersion and transport of *Cryptosporidium* oocysts from faecal pats under simulated rainfall events. *Appl. Environ. Microbiol.* **70**(2), 1151–1159.

Ferguson, C. 2005 *Deterministic models of microbial sources, fate and transport: a quantitative tool for pathogen catchment budgeting*. PhD Thesis, University of New South Wales, Sydney.

Gilbert, R. O. 1987 *Statistical Methods for Environmental Pollution Monitoring*. Van Nostrand Reinhold, New York.

Gilliom, R. J., Hirsch, R. M. & Gilroy, E. J. 1984 Effect of censoring trace-level water-quality data on trend-detection. *Environ. Sci. Technol.* **18**(7), 530–535.

Hallberg, M. 2006 Assessment of suspended solids concentration in highway runoff and its treatment implication. *Environ. Technol.* **27**(9), 945–950.

Haydon, S. R. 2006 *A simplified process-based model for predicting pathogen transport in catchments*. PhD Thesis, Monash University, Melbourne.

Jiang, J., Alderisio, K. & Xiao, L. 2005 Distribution of *Cryptosporidium* genotypes in storm event water samples from three watersheds in New York. *Appl. Environ. Microbiol.* **71**(8), 4446–4454.

Kaiser, H. F. 1958 The varimax criterion for analytic rotation in factor analysis. *Psychometrika* **23**(3), 187–200.

Kim, L. H., Ko, S. O., Jeong, S. & Yoon, J. 2007 Characteristics of washed-off pollutants and dynamic EMCs in parking lots and bridges during a storm. *Sci. Total Environ.* **376**(1–3), 178–184.

Lee, J. H., Bang, K. W., Ketchum, L. H., Choe, J. S. & Yu, M. J. 2002 First flush analysis of urban storm runoff. *Sci. Total Environ.* **293**(1–3), 163–175.

Liu, C. W., Lin, K. H. & Kuo, Y. M. 2003 Application of factor analysis in the assessment of groundwater quality in a blackfoot disease area in Taiwan. *Sci. Total Environ.* **313**(1–3), 77–89.

Mallin, M. A., Williams, K. E., Cartier Esham, E. & Lowe, R. P. 2000 Effect of human development on bacteriological water quality in coastal watersheds. *Ecol. Appl.* **10**(4), 1047–1056.

Marcoulides, G. A. & Hershberger, S. L. 1997 *Multivariate Statistical Methods*. Lawrence Erlbaum Associates, Mahwah, New Jersey.

Medema, G. J., Bahar, M. & Schets, F. M. 1997 Survival of *Cryptosporidium parvum*, *Escherichia coli*, faecal enterococci and *Clostridium perfringens* in river water: influence of temperature and autochthonous microorganisms. *Water Sci. Technol.* **35**(11–12), 249–252.

Moriarty, E. M., Sinton, L. W., Mackenzie, M. L., Karki, N. & Wood, D. R. 2008 A survey of enteric bacteria and protozoans in fresh bovine faeces on New Zealand diary farms. *J. Appl. Microbiol.* **105**(6), 2015–2025.

NH&MRC (National Health and Medical Research Council) 2004 *Australian Drinking Water Guidelines (ADWG)*. National Health and Medical Research Council and the Natural Resource Management Ministerial Council, Commonwealth of Australia, Canberra.

Paul, S., Srinivasan, R., Sanabria, J., Haan, P. K., Mukhtar, S. & Neimann, K. 2006 Groupwise modelling study of bacterially impaired watersheds in Texas: clustering analysis. *J. Am. Water Resour. Assoc.* **42**(4), 1017–1031.

Rosenzweig, B. R., Moon, H. S., Smith, J. A., Baeck, M. L. & Jaffe, P. R. 2008 Variation in the instream dissolved inorganic nitrogen response between and within rainstorm events in an urban watershed. *J. Environ. Sci. Health A* **43**(11), 1223–1233.

Roser, D. J., Skinner, J., LeMaitre, C., Marshall, L., Baldwin, J., Bilington, K., Kotz, S., Clarkson, K. & Ashbolt, N. J. 2002 Automated event sampling for microbiological and related analytes in remote sites: a comprehensive system. *Water Sci. Technol.: Water Suppl.* **2**(3), 123–130.

Shrestha, S. & Kazama, F. 2007 Assessment of surface water quality using multivariate statistical techniques: a case study of the Fuji river basin, Japan. *Environ. Model. Softw.* **22**(4), 464–475.

Signor, R. S., Roser, D. J., Ashbolt, N. J. & Ball, J. E. 2005 Quantifying the impact of runoff events on microbiological contaminant concentrations entering surface drinking source waters. *J. Water Health* **3**(4), 453–468.

Siriwardhema, V. K. 1999 *Water pollution control in multiple landuse water supply catchments*. PhD Thesis, Royal Melbourne Institute of Technology, Melbourne.

Stevens, M., Ashbolt, N. & Cuncliffe, D. 2003 *Review of Coliforms; As Microbial Indicators of Drinking Water Quality*. National Health and Medical Research Council, Canberra.

Swan, I. R. & Volum, A. G. 1984 *Assessment of Agricultural Quality of Land in Gippsland*. Department of Agriculture, Melbourne.

Tabachnick, B. G. & Fidell, L. S. 1996 *Using Multivariate Statistics*. Harper Collins College, New York.

Taebi, A. & Droste, R. L. 2004 First flush pollution load of urban stormwater runoff. *J. Environ. Eng.* **3**(2), 301–309.

USGS (United States Geological Survey) 2004 Chesapeake Bay River Input Monitoring Program, Water Chemistry at the RIM Stations, http://va.water.usgs.gov/chesbay/RIMP/waterchem.html, accessed 18 June 2009.

Varkevisser, C. M., Pathmanathan, I. & Bronwless, A. 2003 *Designing and Conducting Health Systems Research Projects: Data Analysis and Report Writing*, vol. 2. The International Development Research Centre, Canada.

Yongjing, T., Tian, L., Shaoming, H., Aicheng, Z. & Dan, Z. 2009 Characteristics of urban non-point source pollution in Suzhou City, paper presented at the *Environmental Science and Information Application Technology, International Conference*, 4–5 July 2009, vol. 2, pp. 359–362.

Zhang, Z., Lennox, W. C. & Panu, U. S. 2004 Effect of percent non-detects on estimation bias in censored distributions. *J. Hydrol.* **297**(1–4), 74–94.

First received 26 November 2009; accepted in revised form 11 March 2010. Available online 22 April 2010