Environ Monit Assess (2013) 185:4551–4569
DOI 10.1007/s10661-012-2888-2

# Analysis of long-term water quality for effective river health monitoring in peri-urban landscapes—a case study of the Hawkesbury–Nepean river system in NSW, Australia

**U. Pinto · B. L. Maheshwari · R. L. Ollerton**

Received: 3 November 2011 / Accepted: 11 September 2012 / Published online: 7 October 2012


**Abstract** The Hawkesbury–Nepean River (HNR) system in South-Eastern Australia is the main source of water supply for the Sydney Metropolitan area and is one of the more complex river systems due to the influence of urbanisation and other activities in the peri-urban landscape through which it flows. The long-term monitoring of river water quality is likely to suffer from data gaps due to funding cuts, changes in priority and related reasons. Nevertheless, we need to assess river health based on the available information. In this study, we demonstrated how the Factor Analysis (FA), Hierarchical Agglomerative Cluster Analysis (HACA) and Trend Analysis (TA) can be applied to evaluate long-term historic data sets. Six water quality parameters, viz., temperature, chlorophyll-a, dissolved oxygen, oxides of nitrogen, suspended solids and reactive silicates, measured at weekly intervals between 1985 and 2008 at 12 monitoring stations located along the 300 km length of the HNR system were evaluated to understand the human and natural influences on the river system in a peri-urban landscape. The application of FA extracted three latent factors which explained more than 70 % of the total variance of the data and related to the 'bio-geographical', 'natural' and 'nutrient pollutant' dimensions of the HNR system. The bio-geographical and nutrient pollution factors more likely related to the direct influence of changes and activities of peri-urban natures and accounted for approximately 50 % of variability in water quality. The application of HACA indicated two major clusters representing clean and polluted zones of the river. On the spatial scale, one cluster was represented by the upper and lower sections of the river (clean zone) and accounted for approximately 158 km of the river. The other cluster was represented by the middle section (polluted zone) with a length of approximately 98 km. Trend Analysis indicated how the point sources influence river water quality on spatio-temporal scales, taking into account the various effects of nutrient and other pollutant loads from sewerage effluents, agriculture and other point and non-point sources along the river and major tributaries of the HNR. Over the past 26 years, water temperature has significantly increased while suspended solids have significantly decreased ($p<0.05$). The analysis of water quality data through FA, HACA and TA helped to characterise the key sections and cluster the key water quality variables of the HNR system. The insights gained from this study have the potential to improve the effectiveness of river health-monitoring programs in terms of cost, time and effort, particularly in a peri-urban context.

U. Pinto (✉) · B. L. Maheshwari
School of Science and Health, Hawkesbury Campus,
Building H3 University of Western Sydney,
Locked Bag 1797,
Penrith, NSW 2751, Australia
e-mail: pintohome@hotmail.com

R. L. Ollerton
School of Computing and Mathematics, Penrith Campus,
Building Y3, University of Western Sydney,
Locked Bag 1797,
Penrith, NSW 2751, Australia

**Keywords** River monitoring · Multivariate analysis · River water quality · Hawkesbury–Nepean River

## Introduction

Worldwide, rivers draining from extensive residential and agricultural areas have become highly polluted over the past few decades and remain a sensitive issue for river management authorities (Simeonov et al. 2003). The condition of a river system depends upon its hydrology and complex interactions between biotic and abiotic interactions which are greatly influenced by the lithological aspects of catchment, climatic variability, land use changes and other activities of anthropogenic nature. In turn, these factors affect a river's surface water quality and quantity in such way that a spatio-temporal variation can be observed in a 'variation gradient' influencing the health of river systems. To understand the spatial and temporal variability of local rivers, many river management authorities focus on measuring water quality parameters as a surrogate measure to assess river health (Nnane et al. 2011). In order to manage river water quality, it is critical that we develop a firm knowledge of the complexity in measured water quality parameters due to pollution originating from anthropogenic factors (spatial scale) and natural factors associated with climate variability (temporal scale) (Alberto et al. 2001; Dillon and Kirchner 1975). For example, nitrogen and mercury are naturally deposited from the atmosphere in river basins through precipitation. Similarly, alkalinity and acidity in surface waters are strongly affected by weathering of bedrocks without human intervention (Ouyang et al. 2006; St-Hilaire et al. 2004).

Rivers often carry a significant amount of pollutant loads originating from natural and anthropogenic sources via one-way flow to the ocean. In particular, rivers in peri-urban landscape have become increasingly vulnerable to a higher level of degradation due to rapid population growth, damming of freshwater for drinking, building of sewage treatment plants (STP) to treat municipal effluent originating mostly from urban townships and extraction of freshwater to grow vegetables (Ford 1999; Buxton et al. 2006). In recent years, the peri-urbanisation and the increased area of impervious surfaces in urban landscapes have led to the generation of large volumes of municipal runoff during rainy periods which end up in rivers, negatively affecting river biota (Pejaman et al. 2009). Treated and untreated sewage effluent, groundwater seepage carrying toxic leachates and nutrient-rich runoff originating from agricultural activities have profoundly altered water quality and the wellbeing of biotic communities in many river basins. As such, water resources in peri-urban landscapes have failed to efficiently provide a number of social and environmental services they used to provide when they were in pristine condition. In order to alleviate river pollution and manage river water quality, it is necessary to clearly understand the interactions of anthropogenic and natural factors, particularly through a well-designed and cost-effective monitoring program. The collected data can then be analysed to evaluate temporal and spatial variations after which informed water management decisions may be implemented.

For this purpose, it is common practice that government agencies collect a number of representative biological, physicochemical and hydro-morphological parameters through a series of monitoring stations along a river at regular intervals. This often yields a large volume of data that is expensive in terms of collection, handling, storage and analysis. The intensity of monitoring has often been questioned when there are budget cuts. Therefore, it is not surprising that in many instances, long-term data are either not available or contain significant gaps. Further, most of the historic data are not collected for the purpose of specific statistical analysis. To make the monitoring sustainable and useful, it is timely to devise analytical methodologies to identify the key water quality parameters which describe the health of a river system and key sections of a river that describe the highest temporal variations in water quality. Previous authors have also highlighted the importance of maintaining long-term records on benchmark stations for the better management of river systems (Askey-Doran et al. 2009).

For the analysis of long-term datasets originating from river monitoring programs, multivariate statistical techniques such as Principal Component Analysis (PCA), Factor Analysis (FA), Hierarchical Agglomerative Cluster Analysis (HACA) and Discriminant Function Analysis have been proposed (Simeonov et al. 2003; Vega et al. 1998; Alberto et al. 2001). Of these techniques, PCA and FA have gained more acceptance over other techniques due to their ability to handle large datasets and reduce the dimensionality to a manageable size while keeping as much of the original information as possible. In particular, FA helps to understand the structure of the data set by allowing researchers to identify latent factors (combination of variables that influence conditions of rivers, e.g., pH) unique to a given river system by extracting the most useful groups of measured variables (Fields 2009).

The long-term monitoring of river water quality is likely to suffer from data gaps due to reasons such as funding cuts. Nevertheless, we need to assess river health based on the available information. Therefore, the two main objectives of the present study are (1) to obtain a snapshot of water quality degradation along different sections of the river over the past decade by dealing with limited data collected by the river managing authorities and (2) to understand the major trends in water quality variables and relate them with land use and other changes in the catchment. The present study differs from many other similar studies due to a number of reasons. Firstly, it identifies the major impacts on river waters in the rural–urban fringes of the landscape where anthropogenic impacts are quite different to those in a purely urban river system. Secondly, it takes long-term data variability of the river system into consideration to better understand the magnitude of impacts on river water quality due to peri-urbanisation. Almost all similar studies, which follow the same statistical approach to understand river water quality, examine short-term data sets collected mostly between 1 and 8 years (Ouyang 2005; Shrestha and Kazama 2007; Ouyang et al. 2006; Filik Iscen et al. 2008; Nnane et al. 2011). Finally, this study proposes the usefulness of multivariate techniques, which were previously applied mostly to urban river systems, to reduce the dimensionality of data sets and understand the key sections of a peri-urban river for monitoring purposes.

## The Hawkesbury–Nepean river system

Western Sydney, mostly a peri-urban region, is one of the fastest growing areas in South-Eastern Australia. There are often conflicts and debates among stakeholders, government agencies and the community at large over the sustainable use and management of land and water resources, including the Hawkesbury–Nepean River (HNR) system (33° 34′ 14.72″ S, 151° 20′ 16.36″ E to 34° 11′ 31.59″ S, 150° 43′ 11.57″ E). Due to a range of urban and peri-urban activities, the region presents particular challenges in terms of water quality, quantity and river health management. The HNR system is the main source of water supply for the Sydney Metropolitan area. Almost all major reaches of the river system flow through peri-urban landscapes in the catchment. The catchment of the HNR covers about 22,000 $km^2$, and the total length of the HNR system is approximately 300 km. The HNR system is a combination of two major rivers, the Nepean River (155 km) and the Hawkesbury River (145 km) (Markich and Brown 1998). The Nepean becomes the Hawkesbury River at the Grose River confluence near a rural town of Yarramundi, New South Wales (NSW) (Fig. 1). There are 22 large dams and 15 weirs established along the HNR system. The major dam is Warragamba which holds about $2.031 \times 10^9 m^3$ of freshwater captured from the catchment (Turner and Erskine 2005). The hydrology of HNR system has been profoundly altered from its pristine state, mainly due to land use changes and modifications of riverine physical habitats over the last 50 years (Gavin et al. 1998). Human activities, such as the construction of water storage sites and weirs have considerably reduced flood events of the river, altered natural flow regimes and, most importantly, reduced natural mixing of water in weir pools (Turner and Erskine 2005).

Land use in the HNR catchment includes regions that are heavily peri-urbanised and industrialised and which are important for recreational activities, agricultural activities and tourism (Baginska et al. 2003). By 1998, some 25 % of the catchment was cleared for agriculture (Healthy Rivers Commission 1998) and the remaining forested landscape is now in a constant battle with extensive land clearing to accommodate the growing population in Western Sydney. There are numerous point and diffuse sources of anthropogenic pollution, which primarily originate from urban, agricultural and industrial activities. Point-source pollutions are attributed to STPs, mining activities and discharged industrial effluent, while non-point sources of pollution are related to urban runoff and agricultural activities associated with farms and market gardens. Agricultural runoff contributes approximately 40–50 % of phosphorus loads and 25 % of nitrate loads into the HNR system which are believed to have originated from improved pasture, market gardens and animal farms (Markich and Brown 1998). There are 18 STPs along the HNR discharging significant volumes of treated municipal wastewater into the river (Howard 2009). Gavin et al. (1998) and Simonovski et

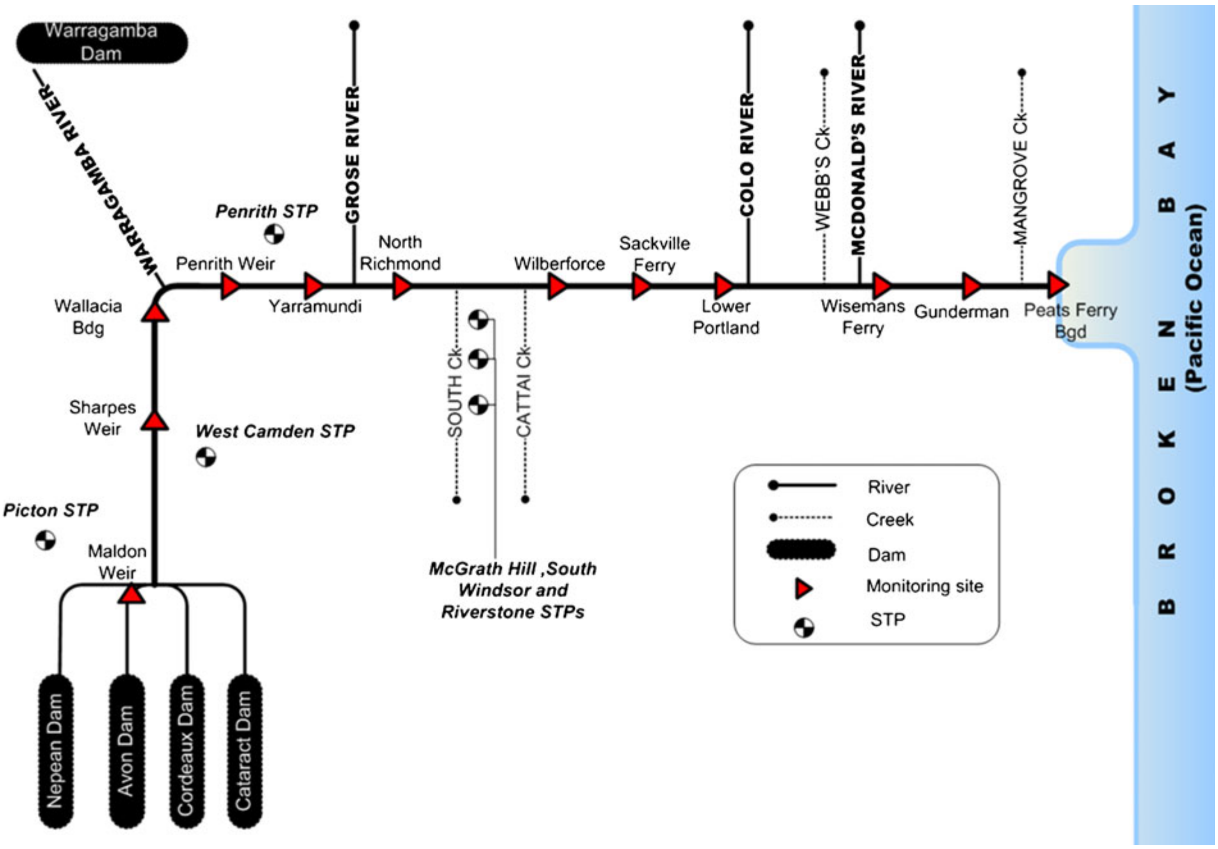



**Fig. 1** Schematic representation of the Hawkesbury–Nepean River and its tributaries

al. (2003) found that the river is less polluted by heavy metals in the upper reaches but concentrations increase near STP discharge points. Increased levels of heavy metals were also observed in oyster tissues in the middle and lower reaches of the estuary as well as at the head waters of the Warragamba River (Gavin et al. 1998; Painuly et al. 2007). Within the HNR catchment, vegetation clearance has been continuously practised over the last 200 years causing increased subsurface and agricultural runoff and sediment loads into the river system (Thoms et al. 2000). The entire length of the HNR system supports a variety of recreational activities for both residents and tourists. Between Windsor (closest monitoring station at North Richmond, see Fig. 1) and Wisemans Ferry, a number of caravan parks and picnic grounds have been established (SPCC 1983). Due to extensive human interactions with the river system over the last 30 years, including dredging and soil extraction from the banks, the width of the Nepean River has increased at many locations. (Turner and Erskine 2005). These are noticeably large and are irreversible changes to the river geomorphology. Due to ongoing human influences, the condition of HNR system has changed considerably from its original state. These influences are well documented in the context of macroinvertebrates, diatoms, fish assemblages, riparian vegetation and fluvial sediment quality and suggest the need for strategic river health management action to improve the river's health (Gavin et al. 1998; Simonovski et al. 2003; Growns and Growns 2001; Growns et al. 2003).

## Data analyses

### *Screening the data set for analysis*

Water quality data from samples collected at weekly intervals were obtained from the Sydney Catchment Authority for a total period of 23 years (1985–2008). Due to changes in priorities of the organisation, funding arrangements and other logistic reasons at different times during this period, the monitoring at some sites was discontinued temporarily or permanently, and some parameters were excluded from the monitoring (i.e., river flow). Therefore, six water quality variables, viz., Temperature (TEMP), Chlorophyll-a (CHL), Dissolved Oxygen (DO), Oxides of Nitrogen (NOx), Suspended Solids (SS) and Reactive Silicate (SIL) were chosen for detailed study due to their continuity of monitoring. The total number of monitoring sites included in this study was 12. Five sites were located on the Nepean River (Maldon Weir, Sharpes Weir, Wallacia, Penrith Weir and Yarramundi) and seven on the Hawkesbury River (North Richmond, Wilberforce, Sackville, Lower Portland, Wisemans Ferry, Gunderman, Peats Ferry) (Table 1). Maldon Weir is the most upstream station and Peats Ferry station is the most downstream station, located close to the mouth of the HNR system. It is noted that short-term (daily or weekly) low-flow or high-flow events are important but excluded from the present analysis as the averaging of other variables at monthly level could mask the effect of variable flow events. Further, this analysis includes stations from headwaters to the river mouth (highly impacted by tide), thus inclusion of river flow in general was not appropriate.

An initial examination of the full dataset revealed that over 23 years of monitoring across 12 stations, the water quality data were often collected on different days of the month. This was probably due to the time required for field sampling and other logistic reasons. We also noted that the distributions of some parameters were positively skewed while others were negatively skewed. In addition, there were statistically extreme readings at various monitoring stations. To overcome these problems, following the approach of Ouyang (2005) and St-Hilaire et al. (2004), we calculated the annual median values of the parameters TEMP, CHL, DO, NOx, SS and SIL and used these values as the basis for analysis in the study. The water quality variables were log-transformed as necessary and subsequently z-scale standardised for FA and HACA to reduce the effects of different units attached to the variables and misclassification that could arise due to differences in magnitude of numerical values. However, untransformed data were used for TA.

**Table 1** Distances of the monitoring stations from ocean (SPCC 1983)

| | Site name | Distance (km) |
|---|---|---|
| Nepean River | Maldon Weir | 256 |
| | Sharpes Weir | 216 |
| | Wallacia Bridge | 183 |
| | Penrith Weir | 162 |
| | Yarramundi Bridge | 143 |
| Hawkesbury River | North Richmond | 140 |
| | Wilberforce | 122 |
| | Sackville Ferry | 97 |
| | Lower Portland | 83 |
| | Wisemans Ferry | [a]64 |
| | Gunderman | [a]49 |
| | Peats Ferry Bridge | 16 |

[a] Distance calculated from Google Maps

## Factor analysis

Factor analysis is a tool, which usually follows a PCA reducing the dimensionality of the water quality data set without loss of embedded information. In FA, components extracted from PCA are rotated according to a mathematically established rule (i.e., varimax, equamax and quarimax) yielding easily interpretable new variables, called varifactors (VFs).

The difference between PCs obtained in PCA and VFs obtained in FA is that PC are linear combinations of observable water quality parameters but VF are unobservable, hypothetical and latent variables (Alberto et al. 2001; Shrestha and Kazama 2007; Chapman 1992). Because of FA, a small number of factors will usually account for approximately the same amount of information as with the much larger set of original observations. To retain the factor loadings that are important, past studies have suggested different cut-off values of significance. For example, Alberto et al. (2001) and Simeonov et al. (2003) suggested 0.7, Ouyang (2005) and Pejaman et al. (2009) suggested 0.75 as the cut-off value of significance to retain the factors. A rather low value of 0.6 and extremely high value of 0.95 was used by Mazlum et al. (1999) and Ouyang et al. (2006) in their studies. However, Liu et al. (2003) provided an easy-to-follow scale-based methodology for this purpose. According to this scale factor loadings above 0.7 are considered as 'strong effects', factor loadings between 0.7 and 0.5 are considered as 'moderate effects' and factor loadings between 0.5 and 0.3 are considered as 'weak effects' (Liu et al. 2003). The factor loadings of the present study were interpreted based on guidelines provided by Liu et al. (2003).

In the past, FA has been used in conjunction with Analysis of Variance (ANOVA) methods to determine the significance of two simultaneously occurring environmental factors (Simeonov et al. 2003). Similarly, they have been used with multiple regression analysis to understand the contribution of identified sources (by PCA) to the concentration of each parameter, a modelling approach commonly known as 'source apportionment', and for the development of indices of sediment quality (Vega et al. 1998; Simeonov et al. 2003; Shin and Lam 2001). In water research, PCA has been widely used to link multivariate community structure and abiotic gradients in marine environments, to assess seasonal and temporal variation in surface freshwater and to understand groundwater dynamics in multiple locations (Clarke and Ainsworth 1993; Shrestha and Kazama 2007; Winter et al. 2000). The PCA alone can only indicate possible variables related to the highest variance of the data set. To increase the clarity of the analysis, FA should also be used identifying similar variables responsible for differences in overall variances.

We evaluated the annual median values (1994–2008) of each monitoring station using the FA. The Sackville, Gunderman and Peats Ferry sites were removed from the analysis due to large data gaps. Annual median datasets were examined for assumptions of multivariate normality using Anderson–Darling test and Draftsman's plots prior to FA. Keiser-Meyer-Olkin (KMO) and Bartlett's tests were also used to check whether the data were suitable for the FA analysis (Shrestha and Kazama 2007). The KMO test is a measure of sample adequacy in which a value of zero indicates diffusion in the correlation pattern and one indicates relative compactness of correlations. As a rule of thumb, values greater than 0.5 are considered acceptable. If the values are less than 0.5, it is then necessary to consider collecting more data and/or reassessing which variables to include. A significant Bartlett's test ($p<0.05$) indicates that the correlation matrix is not an identity matrix. Furthermore, it indicates that there are significant relationships among the variables used in the matrix and hence the dataset is appropriate for factor analysis (Fields 2009). Factors were extracted using PCA method and were mathematically rotated using a varimax rotation in order to reduce the contribution of variables with minor significance. Factor loadings below 0.1 were suppressed for ease of reading and interpretation in the result (Mazlum et al. 1999). The selection of a varimax rotation was done in accordance with previously published work involving spatial and temporal chemometric evaluations (Alberto et al. 2001; Vega et al. 1998). A schematic diagram of the main steps followed during FA is summarised in Fig. 2.

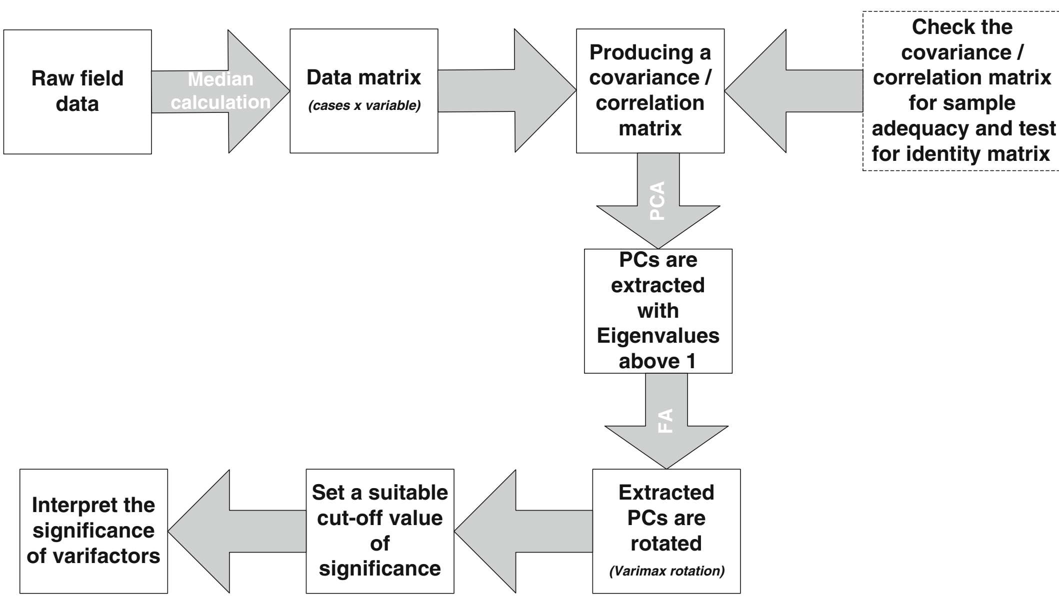



**Fig. 2** PCA and FA methodology

### Cluster analysis

Cluster analysis is potentially a powerful technique to analyse water quality data monitored along river systems. There are two types of cluster analysis available, hierarchical and non-hierarchical. The most widely used method is the HACA—a simple multivariate technique that separates variables into distinct groups based on their natural characteristics. The HACA classification is a non-parametric, unsupervised method and does not depend upon assumptions of normality (Vega et al. 1998). The HACA has been applied to large sets of river data in many parts of the world, e.g., Japan (Fuji River), Argentina (Suquia River) and Iran (Haraz River) for visualizing the differences in spatial and temporal scales and as the first explanatory method (Zhou et al. 2007; Shrestha and Kazama 2007; Alberto et al. 2001; Pejaman et al. 2009).

In this study, we employed HACA using the Squared Euclidean Distance as distance measure and Ward's linkage method to understand similarities among monitoring stations established along the HNR system. The HACA requires a data set without missing values, therefore we only included annual data collected in 1985, 1986, 1996, 1997, 2000 and 2001 (Gunderman site excluded) in the analysis. The use of Squared Euclidean Distance measure has been widely applied for multivariate surface water quality classifications as it is capable of progressively placing a greater weight on variables which are further apart (Alberto et al. 2001; Shrestha and Kazama 2007). Similarly, the Ward's linkage is suitable for this type of work because it uses the ANOVA approach to evaluate the distances between and within cluster variances in an attempt to minimize the sum of squares of any two clusters that can be formed at each step (Shrestha and Kazama 2007).

### Trend analysis

Trend analysis is an exploratory data visualisation technique useful in identifying emerging patterns of water quality in historic data records. Burt et al. (2010) extensively used different types of TA plots to understand seasonal cycles, episodic responses and long-term trends of nitrate concentration in UK rivers. To test the statistical significance of trends in the time-series data, the Mann–Kendall test (MK) is often used because in reality, the long-term water quality data originating from large river systems does not follow conventional probability distributions (i.e., normal or lognormal distributions) and contains missing data points (Mann 1945; Lettenmaier et al. 1991; Yue et al. 2002). The null hypothesis for the MK statistic is that there is no significant trend in each variable considered in the time series.

For the TA, we used annual median values of all sites between 1991 and 2008 (NOx—total of 18 years, CHL—total of 17 years). However, we included slightly larger data records for the MK test collected between 1985 and 2009. The serial independence was checked with the Dublin–Watson test statistic prior to conducting the MK test. Dublin–Watson test (value= 2) indicated that serial correlation is not a problem to yield a valid MK test result (Deepesh and Madan 2012). The TA plots of variables were created using Origin™ 8.1 to visualize overall fluctuations of river water quality variables on spatial and temporal scales and the MK statistic was calculated using an Excel™ macro developed for this purpose (Anders 2010).

## Results and discussion

### Descriptive statistics of the data set

The summary statistics of the original data set indicating the number of data points, range, minimum, maximum, mean, standard error, standard deviation and variance are shown in Table 2. The mean temperature gradually increases across the stations towards the mouth of the river with the highest recorded at Gunderman site (20.4 °C). The mean values of CHL levels gradually increase between Penrith Weir and Sackville Ferry but then decrease towards the river mouth. The mean value of CHL was highest at Sackville Ferry (31.5 mg/L) and lowest at Peats Ferry Bridge (2.88 mg/L). The mean DO remained relatively constant across the stations. Interestingly, the mean SS steadily increased towards the mouth of the river with the lowest recorded at the most upstream station (Maldon Weir) and the highest at the most downstream station (Peats Ferry Bridge). Mean values for NOx indicated two clear peaks, one at Sharpes Weir and other at Wilberforce. From Lower Portland onwards, the mean NOx levels gradually declined towards the mouth of the river. For SIL, the mean values followed a similar trend with two peaks, at Penrith Weir and North Richmond.

**Table 2** Descriptive statistics of river data

| | | Descriptive statistics | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | $N$ | Range | Min. | Max. | Mean | Std. err. | Std. div. | Variance |
| Maldon Weir | TEMP(0C) | 650 | 27.08 | 4.42 | 31.50 | 17.83 | 0.20 | 5.04 | 25.36 |
| | CHL (ug/L) | 668 | 77.20 | 0.00 | 77.20 | 6.63 | 0.35 | 9.13 | 83.35 |
| | DO (mg/L) | 642 | 15.40 | 3.40 | 18.80 | 8.87 | 0.08 | 2.08 | 4.34 |
| | SS (mg/L) | 538 | 121.00 | 1.00 | 122.00 | 3.17 | 0.27 | 6.38 | 40.65 |
| | NOx (mg/L) | 666 | 1.63 | 0.00 | 1.63 | 0.28 | 0.01 | 0.26 | 0.07 |
| | SIL (mg/L) | 643 | 4.90 | 0.10 | 5.00 | 1.76 | 0.04 | 0.93 | 0.86 |
| | Total years | | 24 | | | Missing years | | 0 | |
| Sharpes Weir | TEMP(0C) | 665 | 22.80 | 7.00 | 29.80 | 19.15 | 0.20 | 5.20 | 27.01 |
| | CHL (ug/L) | 707 | 101.40 | 0.20 | 101.60 | 13.36 | 0.55 | 14.54 | 211.33 |
| | DO (mg/L) | 651 | 14.30 | 3.20 | 17.50 | 9.05 | 0.07 | 1.85 | 3.43 |
| | SS (mg/L) | 559 | 87.50 | 0.50 | 88.00 | 5.28 | 0.33 | 7.84 | 61.41 |
| | NOx (mg/L) | 693 | 5.89 | 0.01 | 5.90 | 1.50 | 0.04 | 1.14 | 1.31 |
| | SIL (mg/L) | 666 | 6.14 | 0.10 | 6.24 | 1.92 | 0.04 | 1.08 | 1.16 |
| | Total years | | 24 | | | Missing years | | 0 | |
| Wallacia Bridge | TEMP(0C) | 1013 | 24.00 | 7.50 | 31.50 | 19.53 | 0.17 | 5.38 | 29.00 |
| | CHL (ug/L) | 1037 | 155.43 | 0.20 | 155.63 | 10.65 | 0.37 | 11.93 | 142.24 |
| | DO (mg/L) | 989 | 12.10 | 3.80 | 15.90 | 8.57 | 0.06 | 1.77 | 3.14 |
| | SS (mg/L) | 586 | 488.00 | 1.00 | 489.00 | 9.65 | 1.33 | 32.21 | 1037.51 |
| | NOx (mg/L) | 1036 | 3.42 | 0.00 | 3.42 | 0.33 | 0.01 | 0.34 | 0.11 |
| | SIL (mg/L) | 895 | 23.90 | 0.10 | 24.00 | 1.78 | 0.04 | 1.34 | 1.80 |
| | Total years | | 24 | | | Missing years | | 0 | |
| Penrith Weir | TEMP(C) | 813 | 25.70 | 8.00 | 33.70 | 19.18 | 0.18 | 5.05 | 25.53 |
| | CHL (ug/L) | 821 | 35.98 | 0.00 | 35.98 | 4.63 | 0.13 | 3.86 | 14.92 |
| | DO (mg/L) | 806 | 12.10 | 1.50 | 13.60 | 8.84 | 0.06 | 1.60 | 2.56 |
| | SS (mg/L) | 569 | 259.00 | 0.00 | 259.00 | 7.86 | 1.09 | 26.12 | 682.22 |
| | NOx (mg/L) | 854 | 1.80 | 0.00 | 1.80 | 0.16 | 0.01 | 0.17 | 0.03 |
| | SIL (mg/L) | 796 | 5.45 | 0.10 | 5.55 | 2.49 | 0.04 | 1.13 | 1.28 |
| | Total years | | 24 | | | Missing years | | 0 | |
| Yarramundi Bridge | TEMP(0C) | 764 | 20.70 | 8.90 | 29.60 | 19.23 | 0.18 | 4.98 | 24.78 |
| | CHL (ug/L) | 805 | 111.62 | 0.40 | 112.02 | 11.56 | 0.48 | 13.55 | 183.68 |
| | DO (mg/L) | 756 | 12.40 | 2.50 | 14.90 | 8.26 | 0.07 | 1.85 | 3.41 |
| | SS (mg/L) | 572 | 120.00 | 1.00 | 121.00 | 5.49 | 0.36 | 8.62 | 74.30 |
| | NOx (mg/L) | 803 | 2.29 | 0.01 | 2.30 | 0.50 | 0.01 | 0.36 | 0.13 |
| | SIL (mg/L) | 725 | 5.39 | 0.01 | 5.40 | 1.88 | 0.05 | 1.28 | 1.65 |
| | Total years | | 24 | | | Missing years | | 0 | |
| North Richmond | TEMP(0C) | 978 | 21.90 | 8.30 | 30.20 | 19.56 | 0.17 | 5.36 | 28.74 |
| | CHL (ug/L) | 1025 | 118.98 | 0.24 | 119.22 | 13.62 | 0.45 | 14.51 | 210.57 |
| | DO (mg/L) | 961 | 10.20 | 4.40 | 14.60 | 8.98 | 0.05 | 1.64 | 2.68 |
| | SS (mg/L) | 551 | 189.00 | 1.00 | 190.00 | 5.40 | 0.51 | 11.94 | 142.53 |
| | NOx (mg/L) | 970 | 4.22 | 0.00 | 4.22 | 0.32 | 0.01 | 0.27 | 0.07 |
| | SIL (mg/L) | 933 | 6.27 | 0.03 | 6.30 | 2.60 | 0.04 | 1.28 | 1.65 |
| | Total years | | 24 | | | Missing years | | 0 | |
| Wilberforce | TEMP(0C) | 363 | 20.90 | 10.30 | 31.20 | 19.67 | 0.27 | 5.14 | 26.40 |

**Table 2** (continued)

| | | Descriptive statistics | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | *N* | Range | Min. | Max. | Mean | Std. err. | Std. div. | Variance |
| | CHL (ug/L) | 383 | 81.53 | 0.79 | 82.32 | 20.60 | 0.70 | 13.79 | 190.04 |
| | DO (mg/L) | 359 | 10.20 | 3.90 | 14.10 | 8.31 | 0.10 | 1.89 | 3.56 |
| | SS (mg/L) | 337 | 241.00 | 1.00 | 242.00 | 17.73 | 1.02 | 18.67 | 348.60 |
| | NOx (mg/L) | 366 | 4.29 | 0.01 | 4.30 | 1.15 | 0.05 | 0.87 | 0.76 |
| | SIL (mg/L) | 346 | 6.18 | 0.02 | 6.20 | 1.90 | 0.10 | 1.78 | 3.15 |
| | Total years | | 24 | | | Missing years | | 0 | |
| Sackville Ferry | TEMP(0C) | 439 | 18.50 | 10.80 | 29.30 | 20.05 | 0.24 | 5.02 | 25.18 |
| | CHL (ug/L) | 458 | 191.20 | 1.00 | 192.20 | 31.51 | 1.30 | 27.78 | 771.64 |
| | DO (mg/L) | 431 | 12.50 | 1.80 | 14.30 | 8.68 | 0.09 | 1.84 | 3.37 |
| | SS (mg/L) | 266 | 174.00 | 1.00 | 175.00 | 17.03 | 1.17 | 19.01 | 361.23 |
| | NOx (mg/L) | 436 | 3.18 | 0.00 | 3.18 | 0.67 | 0.03 | 0.57 | 0.32 |
| | SIL (mg/L) | 377 | 6.00 | 0.00 | 6.00 | 1.36 | 0.09 | 1.70 | 2.88 |
| | Total years | | 17 | | | Missing years | 2002, 2003, 2004, 2005, 2006, 2007, 2008 | | |
| Lower Portland | TEMP(0C) | 319 | 18.40 | 10.60 | 29.00 | 20.27 | 0.28 | 5.02 | 25.18 |
| | CHL (ug/L) | 337 | 252.60 | 0.50 | 253.10 | 19.59 | 1.00 | 18.35 | 336.68 |
| | DO (mg/L) | 317 | 9.40 | 4.00 | 13.40 | 8.49 | 0.09 | 1.62 | 2.63 |
| | SS (mg/L) | 304 | 100.00 | 2.00 | 102.00 | 9.99 | 0.54 | 9.34 | 87.29 |
| | NOx (mg/L) | 334 | 1.99 | 0.01 | 2.00 | 0.25 | 0.02 | 0.28 | 0.08 |
| | SIL (mg/L) | 286 | 5.90 | 0.10 | 6.00 | 1.63 | 0.09 | 1.51 | 2.28 |
| | Total years | | | 24 | | Missing years | | 0 | |
| Wiseman's Ferry | TEMP(0C) | 425 | 17.80 | 11.00 | 28.80 | 20.01 | 0.23 | 4.79 | 22.91 |
| | CHL (ug/L) | 433 | 52.68 | 0.50 | 53.18 | 7.92 | 0.35 | 7.38 | 54.42 |
| | DO (mg/L) | 420 | 10.80 | 4.30 | 15.10 | 7.90 | 0.07 | 1.47 | 2.15 |
| | SS (mg/L) | 401 | 149.00 | 1.00 | 150.00 | 15.51 | 0.85 | 16.96 | 287.60 |
| | NOx (mg/L) | 424 | 1.07 | 0.01 | 1.08 | 0.22 | 0.01 | 0.19 | 0.04 |
| | SIL (mg/L) | 390 | 6.20 | 0.10 | 6.30 | 1.62 | 0.08 | 1.48 | 2.20 |
| | Total years | | | 24 | | Missing years | | 0 | |
| Gunderman | TEMP(0C) | 255 | 16.40 | 12.00 | 28.40 | 20.41 | 0.28 | 4.55 | 20.67 |
| | CHL (ug/L) | 246 | 40.38 | 0.30 | 40.68 | 4.86 | 0.33 | 5.13 | 26.27 |
| | DO (mg/L) | 247 | 8.30 | 4.10 | 12.40 | 7.33 | 0.08 | 1.31 | 1.72 |
| | SS (mg/L) | 235 | 177.00 | 1.00 | 178.00 | 18.74 | 1.59 | 24.37 | 594.09 |
| | NOx (mg/L) | 261 | 0.90 | 0.01 | 0.91 | 0.20 | 0.01 | 0.16 | 0.03 |
| | SIL (mg/L) | 218 | 5.40 | 0.10 | 5.50 | 1.63 | 0.09 | 1.40 | 1.96 |
| | Total years | | 16 | | | Missing years | 2002, 2003, 2004, 2005, 2006, 2007, 2008 | | |
| Peats Ferry Bridge | TEMP(0C) | 225 | 16.50 | 11.00 | 27.50 | 19.85 | 0.27 | 4.03 | 16.23 |
| | CHL (ug/L) | 230 | 12.13 | 0.06 | 12.19 | 2.88 | 0.13 | 1.98 | 3.92 |
| | DO (mg/L) | 221 | 9.70 | 4.80 | 14.50 | 7.11 | 0.08 | 1.18 | 1.38 |
| | SS (mg/L) | 211 | 400.00 | 1.00 | 401.00 | 32.74 | 4.06 | 58.97 | 3477.24 |
| | NOx (mg/L) | 225 | 0.66 | 0.00 | 0.66 | 0.06 | 0.01 | 0.08 | 0.01 |
| | SIL (mg/L) | 204 | 3.00 | 0.10 | 3.10 | 0.77 | 0.03 | 0.48 | 0.23 |
| | Total years | | 16 | | | Missing years | 2002, 2003, 2004, 2005, 2006, 2007, 2008 | | |

Factor analysis

In this study, FA was employed to understand the compositional patterns between environmental variables in the HNR system and possibly identify factors affecting each variable. The KMO was 0.473 and Bartlett's test was significant ($p$<0.05). The low value of KMO was probably due to a low number of variables used in the analysis. We have taken into consideration the low values indicated by KMO when interpreting our results. The correlation matrix of the water quality variables is provided in Table 3. Most correlations between parameters tended to be weak with the highest correlation between SS and CHL (0.661). A low correlation (0.207) was recorded for NOx and CHL, indicating that CHL did not fully depend upon NOx in the HNR system.

The PCA extracted three components with eigenvalues above 1 capturing 74 % of the total (Table 4) and the FA yielded three VFs with weak, moderate and strong factor loadings (Table 5). We reported only on the strong factor loadings in this study. Veri-factor 1 explained 29 % of the total variance and indicates strong positive loadings on CHL (0.794) and SS (0.702) and a negative loading on SIL (−0.785). Veri-factor 2 accounted for 25 % of the total variance and recorded loadings on DO of −0.793 and on TEMP of 0.750. VF3 accounted for 19 % of the total variance and had the highest positive loading on NOx (0.894). It is also interesting to note that VF1 represents the bio-geographical factor of water quality, VF2 the natural factor and VF3 the nutrient pollution factor. The bio-geographical and nutrient pollution factors more likely related to the direct influence of changes and activities of peri-urban natures and account for about 50 % of variability in water quality.

The three variables CHL, SIL, and SS picked up by VF1 were considered further. Here, VF1 is indicative of CHL, which is a biological measure of water quality, SIL which is a measure of catchment geology and SS which is a measure of total organic and inorganic suspended matter. Suspended solids usually consist of particulates with a diameter less than 62 μm (Waters 1995). In many instances, SS are related to soil leaching and soil erosion (Simeonov et al. 2003; Shrestha and Kazama 2007). On the other hand, SIL are derived from the weathering of silicate rocks (Semhi et al. 2000). However, the factor loadings are opposite for SIL and SS suggesting the origin and existence of SIL in the HNR system are not similar to those of SS. Further, SIL did not appear to contribute significantly to the increase in CHL levels. The relationship between CHL and SS contradicts the fact that increased SS reduce the light penetration into the water column thereby reducing the growth of chlorophyll producing flora. Instead, suspended solids appear to have had an additive effect on chlorophyll producing flora in the HNR system. Nutrients flocculated around suspended solid particles may be supporting the increase in chlorophyll producing flora (Heathwaite 1994). Further to this the negative loading of SIL in connection with a positive loading of CHL may be an indication of silicate uptake by diatoms during phytoplankton blooms. However, further research is warranted to understand such nutrient dynamics.

Two variables (DO and TEMP) with high absolute factor loadings on VF2 are indicative of oxygen desaturation processes in the water column. The inverse relationship obtained by VF2 reflects warmer water holding less oxygen as it saturates with gas faster than colder water. Shrestha and Kazama (2007) obtained similar results (0.889 for temperature and −0.885 for DO under VF2) for the Fuji River in Japan.

**Table 3** The values of the correlation matrix of the six physicochemical variables

| Correlation matrix[a] | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | TEMP | CHL | DO | SS | NOx | SIL |
| Correlation | TEMP | 1.000 | 0.289 | −0.296 | 0.290 | −0.080 | −0.042 |
| | CHL | 0.289 | 1.000 | −0.038 | 0.661 | 0.204 | −0.316 |
| | DO | −0.296 | −0.038 | 1.000 | −0.442 | 0.065 | 0.119 |
| | SS | 0.290 | 0.661 | −0.442 | 1.000 | 0.051 | −0.313 |
| | NOx | −0.080 | 0.204 | 0.065 | 0.051 | 1.000 | 0.117 |
| | SIL | −0.042 | −0.316 | 0.119 | −0.313 | 0.117 | 1.000 |

[a]Determinant=0.252

**Table 4** Eigenvalues for the first three PCs for the water quality data set of HNR

| Total variance explained | | | |
|---|---|---|---|
| Component | Rotation sums of squared loadings | | |
| | Total | Percent of variance | Cumulative % |
| 1 | 1.756 | 29.270 | 29.270 |
| 2 | 1.523 | 25.384 | 54.654 |
| 3 | 1.147 | 19.124 | 73.778 |

Extraction method: principal component analysis

Variables with a large loading on VF3 relate to effects of nutrient pollutants in the river. This factor has been reported in several previous studies for highly urbanized river systems including Aliakmon, Axios, Gallikos, Loudias and Strymon Rivers (Simeonov et al. 2003) and Fuji River (Shrestha and Kazama 2007). The raw data for NOx included in the analysis contained both nitrate and nitrite oxides. The increased levels of NOx were probably due to point and non-point-source pollution from urban and agricultural areas and orchards in the catchment. Nitrogenous fertilizers applied for crops undergo nitrification process and $NO_3$–N reaches river waters through groundwater seepage and runoff (Shrestha and Kazama 2007). The nutrient factor in the present study only accounts for 19 % of the total variance and was captured by the third PC. However, compared to the first (29 %) and second (25 %) components, the effect of the third component (19 %) is relatively low in the total variance. Nevertheless, it indicates the prevalence of nutrient pollution in the HNR.

The FA in this study seemingly suffered from the limited number of complete data sets, resulting from short lengths of continuous monitoring and reduced numbers of parameters in a number of occasions. As such, the data shortage attributed to the time scale, special distances and number of water quality variables considered in the analysis has some impact on the outcome of FA. This problem was also reflected in the KMO pre-test. More frequent and systematic collection of river data is required to obtain additional VFs that would aid in identifying other variables affecting the health of the HNR system. Factor analysis in evaluating the surface water quality usually achieves two targets. Firstly, it reduces the number of variables to a manageable size and secondly it indicates variables important for a larger proportion of water quality variability in the river system. In this study, application of FA did not significantly reduce the number of variables. However, it revealed three underlying factors that governed the health of the HNR system in the past decade.

**Table 5** Rotated factor loadings for extracted principal components

| | Rotated component matrix[a] | | |
|---|---|---|---|
| | VF1 | VF2 | VF3 |
| CHL | 0.794 (Strong) | 0.191 | 0.379 |
| SIL | −0.785 (Strong) | 0.120 | 0.395 |
| SS | 0.702 (Strong) | 0.518 | 0.147 |
| DO | | −0.793 (Strong) | 0.158 |
| TEMP | | 0.750 (Strong) | |
| NOx | | −0.111 | 0.894 (Strong) |

## Hierarchical agglomerative cluster analysis

We employed HACA to understand the similarities and dissimilarities in monitoring sites giving particular emphasis to the spatial variability of the river. For increased clarity, the distinct clusters obtained from the dendrograms are relisted and presented in Table 6. The Gunderman site was removed for the year 2001 data column due to the lack of continuity in available data. For the six-year period (1985–1986, 1996–1997, 2000–2001) considered in this analysis HACA identified two broad clusters. Most upstream and downstream stations were included in one cluster while stations belonging to the middle reach were included in the second cluster. There is a slight mismatch in this pattern in 1985 and 1996 however the overall broad clusters remained the same.

HACA clustered the 12 monitoring stations into two distinct groups based on six water quality variables. Overall, this pattern was consistent for the six-year (non-consecutive) period. We noted that clusters observed in the present study did not belong to the broader areas of the HNR system as predicted by previous studies (Markich and Brown 1998; Diamond 2004). The geology of catchment often influences the quality of the runoff and groundwater that contributes to the flow in the river. Diamond (2004) identified three main types of landscapes of the HNR catchment—the upper one-third having many poorly accessible gorges,

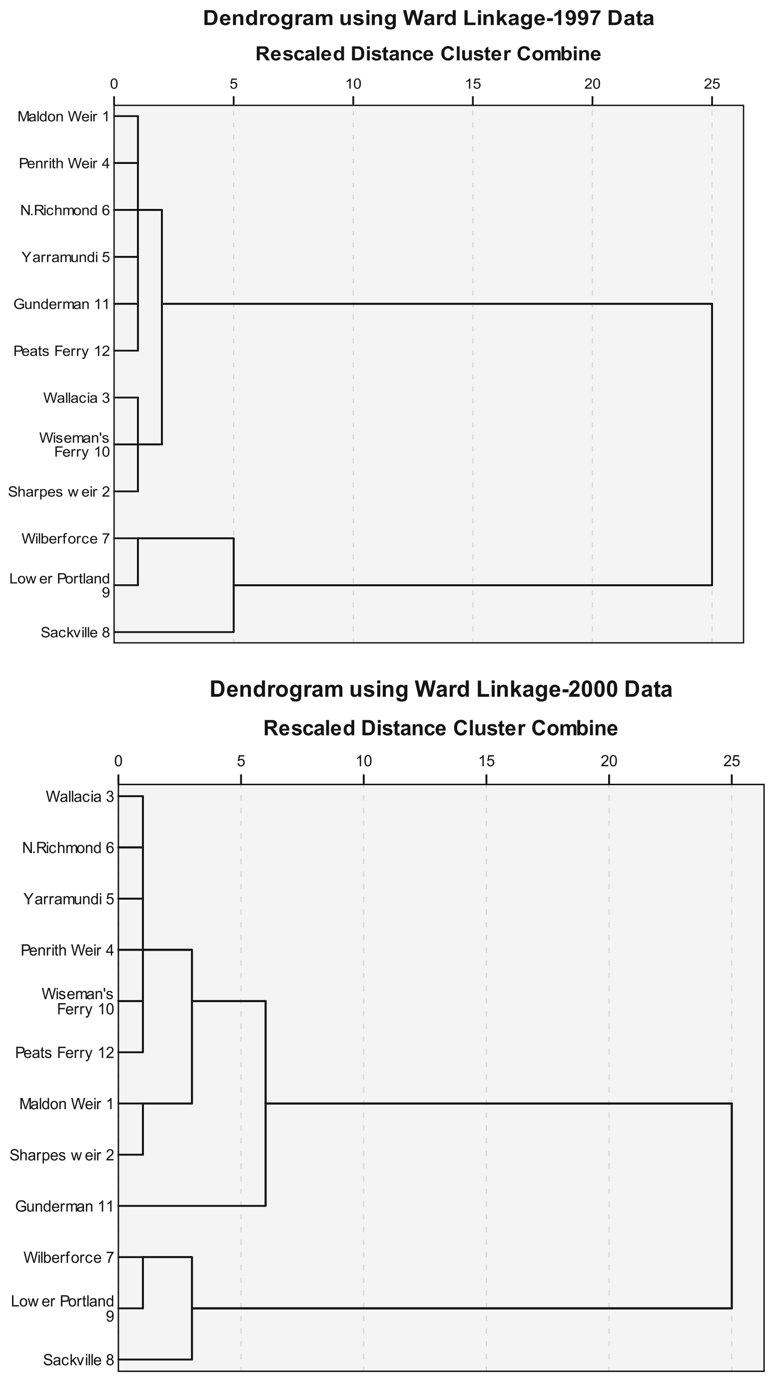
**Table 6** Two clusters obtained from dendrograms at 25-rescaled distance cluster combine level (non-highlighted—cluster 1, highlighted—cluster 2)

| 1985 | 1986 | 1996 | 1997 | 2000 | 2001 |
|---|---|---|---|---|---|
| Maldon Weir | Maldon Weir | Maldon Weir | Maldon Weir | Maldon Weir | Maldon Weir |
| Sharpes Weir | Sharpes Weir | **Sharpes Weir** | Sharpes Weir | Sharpes Weir | Sharpes Weir |
| Wallacia | Wallacia | Wallacia | Wallacia | Wallacia | Wallacia |
| Penrith Weir | Penrith Weir | Penrith Weir | Penrith Weir | Penrith Weir | Penrith Weir |
| **Yarramundi** | **Yarramundi** | Yarramundi | Yarramundi | Yarramundi | Yarramundi |
| **N. Richmond** | **N. Richmond** | N. Richmond | N. Richmond | N. Richmond | N. Richmond |
| Wilberforce | **Wilberforce** | **Wilberforce** | **Wilberforce** | **Wilberforce** | Wilberforce |
| **Sackville** | **Sackville** | **Sackville** | **Sackville** | **Sackville** | **Sackville** |
| L. Portland | **L. Portland** | **L. Portland** | **L. Portland** | **L. Portland** | **L. Portland** |
| Wisemans Ferry | Wisemans Ferry | Wisemans Ferry | Wisemans Ferry | Wisemans Ferry | Wisemans Ferry |
| Gunderman | Gunderman | Gunderman | Gunderman | Gunderman | |
| Peats Ferry | Peats Ferry | Peats Ferry | Peats Ferry | Peats Ferry | Peats Ferry |

the mid one-third situating in the midst of agricultural farmlands and the lower one-third ending with tidal slopes and alluvial soil pockets. According to Markich and Brown (1998), the HNR catchment has three major geological formations—the Narrabeen Group, the Hawkesbury Sandstone and the Wianamatta Group. Much of the catchment of the Nepean River and its tributaries are underlain by shale yielding high salinity in groundwater and resulting in high conductivity of river water while the headwaters lie mostly on sandstone (Growns et al. 1995). If the catchment geology and different landscape formations had a considerable influence on river waters, we would expect a clustering pattern that reflects these geographical differences. However, the clusters found in this study do not seem to match the above divisions of the catchment based on geological formations.

Alternatively, one might expect a clustering pattern based on tidal influence which extends upstream as far as the Nepean confluence (SPCC 1983). As well, the saline intrusion into the HNR system (due to tidal-inflow) is usually limited to the Colo River junction near Sackville station. If saline water had a profound effect on the six water quality variables considered in this study, the analysis would have grouped the monitoring stations into two clusters each representing upstream and downstream sites. This pattern was also not observed in any cluster dendrograms. In this study, clusters were indicative of pollutant loads entering the river. The middle reach identified by the second cluster (highlighted cells in Table 6) includes Yarramundi, North Richmond and Wisemans Ferry stations. In 1997 and 2000, the polluted reach was indicated by three sites, Wilberforce, Sackville and Lower Portland (Fig. 3). Between these two monitoring stations the HNR receives pollutant loads from three major tributaries (McDonald River, Colo River and Grose River) and four creeks (Eastern Creek, Cattai Creek, Ropes Creek and South Creek) all of which transport substantial amounts of treated effluent from Quakers Hill, Riverstone and St. Mary's sewage treatment facilities. It has been reported that 95 % of dry weather flow at South Creek consists of domestic or treated effluent (Thoms et al. 2000). Therefore, the discharge from STPs has a greater influence on the water quality of major creeks in HNR. The middle stream monitoring stations are also situated in highly peri-urbanised areas of the catchment that drain large quantities of urban runoff during rainfall events. If all inflows are considered, about 90 % of flows through various tributaries are effluent (Thoms et al. 2000). If the hypothesis that most upstream reaches of

**Fig. 3** Dendrograms showing clustering of monitoring stations based on water quality in year 1997 and 2000




a river are relatively less polluted is true then the inclusion of most downstream sites together with most upstream sites in one cluster, as shown through HACA, is indicative of self-purification and an assimilative capacity of the HNR. The river tends to improve its water quality close to the mouth due to tidal influence, less pollutant inflows and water purification effect of dense mangrove vegetation. Thus, the two clusters in this analysis may be considered to represent a 'clean zone' and a 'polluted zone' of the HNR system.

The results of this analysis are consistent with other peri-urban rivers in the world reported by a number of researchers (e.g., Zhou et al. 2007; Pejaman et al. 2009; Simeonov et al. 2003). They report on three clusters representing less polluted areas of the river, areas polluted with sewage effluent from STPs, industrial wastewater and domestic wastewater and areas polluted with agricultural runoff. For example, using 23 variables measured over a 12-month period, Zhou et al. (2007) obtained three clusters of a watercourse in Hong Kong representing a near pristine water quality area, a moderately polluted area and a highly polluted area. Further, Alberto et al. (2001) obtained two clusters belonging to upstream and downstream sampling stations based on 22 parameters monitored over a 2-year period in the Suquia River basin in Argentina. Pejaman et al. (2009) reported three clusters with similar water quality which were related to low, moderate and high polluted areas of the Haraz River basin Iran. Similar to the HNR system, water quality of the Haraz River was affected by pollution originating from agricultural activities, sand mining activities and untreated sewage effluent.

## Trend analysis

The MK test revealed a significantly ($p<0.05$) increasing trend for annual median TEMP and significantly decreasing trend for SS over the last two decades on temporal scale (Fig. 4). Trends for both CHL and SIL variables decreased over time and were not significant. Minor increasing trends were observed for DO and NOx on temporal scale. The temperature of river waters plays a considerable role in the energy utilization process, cooling capacity, dissolving capacity and most importantly maintaining healthy biological communities within an aquatic ecosystem. However, water temperature is greatly affected by the atmospheric temperatures and to a lesser extent by discharge rates, although such increases are marginal (Pekárová et al. 2011; Moatar and Gailhard 2006). Pekárová et al. (2011) observed a slight increase (0.12 °C) in water temperature due to an increase of 1.5 °C air temperature over a 50-year period for the river Bela basin in Slovakia (Pekárová et al. 2011). Similarly, for the river Loir (France), the mean annual water temperature increased by 0.8 °C over 95 years (Moatar and Gailhard 2006). In line with such observations, the present study indicates an increasing trend for the water temperature, and this trend may be attributed to the consequences of global warming. As to the variable SS, the significantly decreasing trend observed is an interesting observation from the analysis. In natural waters, presence of SS increases the nitrification process and this leads to an increase in NOx levels (Xia et al. 2004). The SS in the present study indicates the opposite of this relationship on temporal scale (Fig. 4). This may be due to the effects of tide which we have not accounted for in this analysis.

The trends for variation of NOx and CHL are shown in Fig. 5a, b. There are clear highs and lows in both plots. For NOx variation, there are three clear peaks over Sharpes Weir, Wilberforce and Yarramundi in descending order (Fig. 5a). All sites from Lower Portland to the mouth of the river have relatively low levels of NOx for all years. The graph for CHL indicated only two peaks, one at Sharpes Weir and other between Wilberforce and Lower Portland (Fig. 5b). Chlorophyll-a for all other sites remained relatively low due to dominance by attached aquatic plants in those sections. The decreasing spatial trend in NOx is due to the reduced impacts of STP discharge and dilution effect from salt water intrusion towards the mouth of the river system.

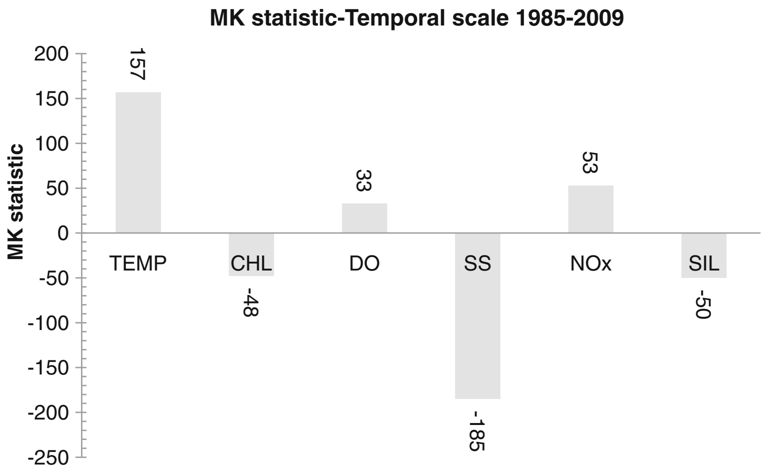



**Fig. 4** Multivariate MK test results for temporal trends. Note that for variables CHL and SS a decreasing trend and for other variables an increasing trend are observed. Also, these trends are significant ($\alpha=0.05$) for variables TEMP and SS

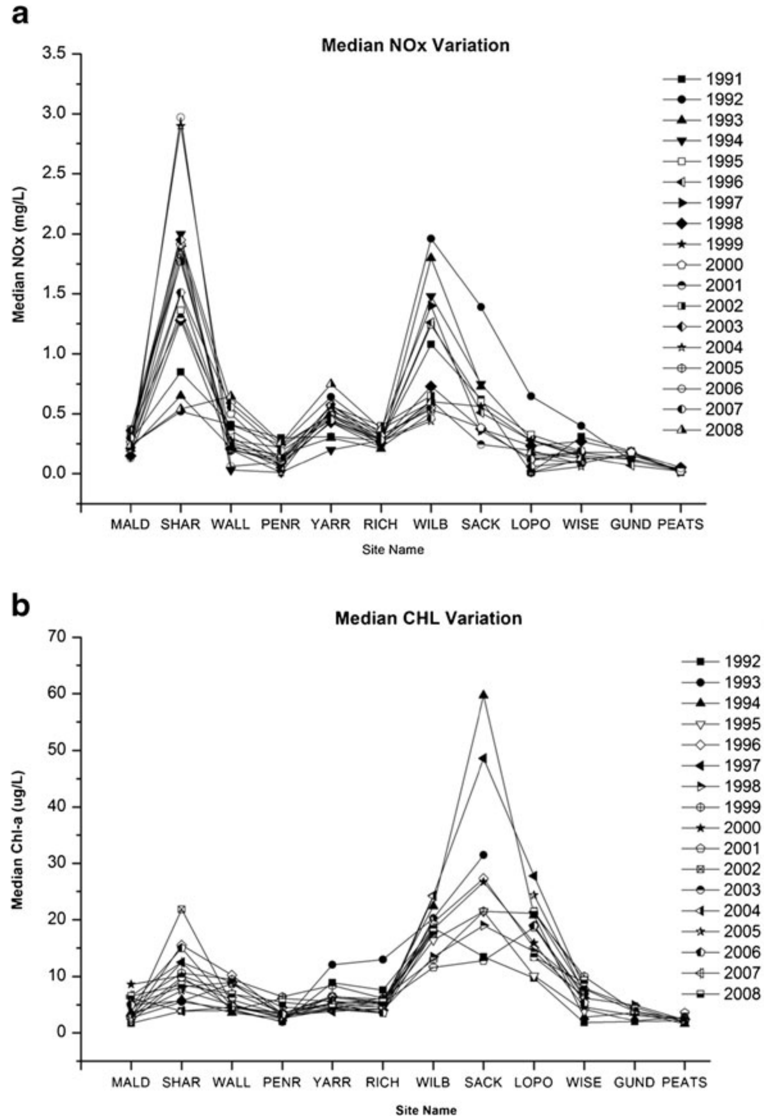



**Fig. 5** **a** Long-term median NOx variation along HNR. **b** Long-term chlorophyll-a variation along HNR. (MALD-Maldon Weir, SHARP-Sharpes Weir, WALL-Wallacia Bridge, PENR-Penrith Weir, YARR-Yarramundi Bridge, RICH-North Richmond, WILB-Wilberforce, SACK-Sackville Ferry, LOPO-Lower Portland, WISE-Wisemans Ferry, GUND-Gunderman, PEATS-Peats Ferry Bridge)

Two steep peaks at Sharpes Weir and Wilberforce were noted. Increased NOx levels observed over consecutive years at Sharpes Weir are most likely due to effluent discharge from adjacent West Camden and Picton STPs. The West Camden STP releases 3112 ML of effluent annually into the environment (Howard 2009). On the other hand, the peak at Wilberforce probably represents nutrients entering the HNR system just upstream through South Creek, one of the most polluted tributaries of the HNR. A detailed report by Krogh et al. (2008) on the analysis of water quality of the HNR system suggested that regardless of minor trends in decreasing nitrogen levels, NOx levels remained above the Australian and New Zealand Environment and Conservation Council and Agricultural and Resource Management Council of Australia and New Zealand (ANZECC and ARMCANZ) guidelines particularly near Sharpes Weir monitoring station. The results obtained from the TA also agree with a modelling exercise by Qin et al. (1995). They showed that the effect of the West Camden STP on total nitrogen levels had dropped considerably by 10 km downstream of the discharge point (Qin et al. 1995). Further, the Wilberforce site exceeded the ANZECC and ARMCANZ guidelines in more than 90 % of the sampling sessions (SCA 2010). In 1990, about 75 % nitrogen loads in Wilberforce reach was related to treated effluent originating from two nearby STPs (James 1997). Therefore, we assume the peak observed at Wilberforce was strongly influenced by McGrath Hill, South Windsor, Riverstone, St Mary's, Quakers Hill and Rouse Hill STPs attached to Cattai and South creeks. However, there was no peak visible at the Penrith monitoring site although it is located near Penrith STP. This is because the effluents from the Penrith STP are discharged via Boundary Creek below Penrith Weir and monitoring is undertaken just above the discharge point. It is assumed that the peaks recorded on consecutive years at Yarramundi are indicative of the effects of effluent discharged by the Penrith STP. Overall, the heights of peaks reduced around the year 2000, due to major upgrades made to the STP.

The variation of CHL was relatively high at Sharpes Weir and between Wilberforce & Wisemans Ferry compared to other sites on the river (Fig. 5b). Rahman and Salbe (1995) found that South Creek was a major carrier of diffused source pollution into the main river. The two creeks carry runoff, stormwater and effluent from sub-catchments that are highly urbanised but include agricultural areas. Wilberforce and Sackville monitoring stations are situated downstream of Cattai Creek and South Creek confluences. Thus, effects of nutrient loads from these creeks have probably influenced the increased CHL levels. Change in river morphology is another possible factor for increased CHL levels. Previous river depth reports indicate that the depth of the HNR reach between Cattai Creek and the Colo River confluences ranges from 4 and 6.5 m and is deeper than the depths in both Cattai Creek and the Colo River (SPCC 1984). Sackville station is located in this deeper reach of the river. Slow moving water resulting from increased depth and high nutrient loads from point and non-point sources favour the native and exotic floating macrophyte species in the HNR (Thiebaud and Williams 2007). Similarly, Krogh et al. (2008) reported that increased nutrients (Nitrogen and Phosphorus) were related to STP discharges in the HNR.

The DO variable exhibited a slight decrease towards the mouth of the river but no obvious corresponding pattern of SIL or TEMP was observed for the years considered in this analysis. However, the median temperature was consistent along the HNR with slight fluctuations due to local weather conditions.

Predominantly the peaks observed at some monitoring stations were directly associated with treated effluent discharged by STPs and pollutant loads released by major creeks. South Creek and Cattai Creeks are the two most influential pollutant carriers into the HNR. However, in recent years, agricultural land contributed more (nitrogen 64 %) to the total nutrients in the river than STPs (nitrogen 27 %) (EPA 2002). The patterns of highs and lows for a particular variable were similar when they occurred for both annual median values and overall mean values calculated using all raw data (see Table 2 and Fig. 5a, b). This further confirms the trends we observed for each variable. The peaks were clearly noted for the monitoring site located downstream of the pollutant source, and they were almost invisible for corresponding years for the monitoring site located upstream of the pollutant source. Nevertheless, these effects do not appear to present along the river at the same level. Our results further confirm the zonation of river obtained by the HACA to some degree as the peaks obtained for NOx at Yarramundi and Wilberforce sites and peaks obtained for CHL between Wilberforce and Lower Portland were within the polluted zone identified by the HACA.

Our study serves as a window to an understanding of the major impacts and challenges to peri-urban rivers in the present time. The findings based on long-term historic trends are particularly applicable to all major rivers that flow through rural–urban fringes of the terrain and impacted by agricultural activities and effluent discharges. While the usual suspects of pollutant sources for urban rivers are common for peri-urban rivers, we found that the trends in water quality variables are more complex and affected by seasonal factors such as runoff from agricultural sources.

## Conclusions

The long-term monitoring of river water quality is likely to suffer from data gap due to funding cuts and other reasons; nevertheless, we need to assess river health based on the available information. In this study, we demonstrated how the FA, HACA and TA techniques can be applied to evaluate long-term historic data sets. The FA was useful in grouping variables based on some meaningful interactions within the river system. In particular, it was useful in accounting for the interactions of wastewater inputs, nutrient inputs, effluent discharge and geomorphologic effects and helped to group the variables based on the factor loadings. Using the HNR system as a case study, the application of FA indicated that there are three main factors, viz., bio-geographical, natural and nutrient pollutant, affecting the overall condition of river system in a peri-urban context. The three components extracted with the application of PCA explained more than 70 % of the total variance of the data. The bio-geographical and nutrient pollution factors more likely related to the direct influence of changes and activities of peri-urban natures and accounted for about 50 % of variability in water quality.

The application of HACA helped in identifying similarities and dissimilarities in monitoring stations based on the six water quality parameters, viz., TEMP, CHL, DO, NOx, SS and SIL. This analysis indicated that the monitoring stations of the HNR could be divided into two distinct zones, 'clean' and 'polluted'. Further, the present study demonstrated that HACA can be employed for objective classification of river waters for future monitoring programs by government agencies, particularly to design cost-effective monitoring programs. The TA indicated considerable temporal and spatial variations in NOx and CHL in places where the HNR system is impacted by point-source discharges from sewage treatment plants. The TEMP and SS variables showed significantly increasing and decreasing trends in spatio-temporal scales for the last 26 years. Direct impact by the diffused source of pollution is not very clear from the results, although TA supported the findings of HACA.

**Acknowledgments** The authors acknowledge the Sydney Catchment Authority for permission to use the data in this publication. We also sincerely thank Ms. Tracey Schultz for her great support in sending us the data sets promptly, Ms. Joycelyn Applebee for the assistance in the preparation of this manuscript and two anonymous referees who provided some helpful comments.

## References

Alberto, W., Diaz, M. P., Ame, M. V., Pesce, S. B., Hued, A. C., & Bistoni, M. A. (2001). Pattern recognition techniques for the evaluation of spatial and temporal variations in water quality. A Case Study: Suqu a River Basin (Córdoba–Argentina). *Water Research, 35*, 2881–2894.

Anders, G. (2010). *Multitest Version 5.2*. Sweden: Linkopings University.

Askey-Doran, M., Hart, B., Ladson, T., & Read, M. (2009). *Tasmanian river condition index*. Tasmania: South Horbart.

Baginska, B., Pritchard, T., & Krogh, M. (2003). Roles of land use resolution and unit-area load rates in assessment of diffuse nutrient emissions. *Journal of Environmental Management, 69*, 39–46.

Burt, T., Howden, N., Worrall, F., & Whelan, M. (2010). Long-term monitoring of river water nitrate: how much data do we need? *Journal of Environmental Monitoring, 12*, 71–79.

Buxton, M., Tieman, G., Bekessy, S., Budge, T., Mercer, D., Coote, M., & Morcombe, J. (2006). *Change and continuity in peri-urban Australia—State of the Peri-urban Regions: a review of the literature*. Melbourne: RMIT University.

Chapman, P. (1992). Ecosystem health synthesis: can we get there from here? *Journal of Aquatic Ecosystem Stress and Recovery, 1*, 69–79.

Clarke, K., & Ainsworth, M. (1993). A method of linking multivariate community structure to environmental variables. *Marine Ecology Progress Series, 92*, 205–205.

Deepesh, M., & Madan, K. J. (2012). *Hydrologic time series analysis: theory and practice*. New York: Springer. 280 p.

Diamond, R. (2004). *Water and Sydney's future—balancing the value of our rivers and economy [online]*. Sydney: Department of Infrastructure, Planning and Natural Resources [Accessed 11 June 2012].

Dillon, P., & Kirchner, W. (1975). The effects of geology and land use on the export of phosphorus from watersheds. *Water Research, 9*, 135–148.

Environmental Protection Agency (EPA) (2002). *Green offset for sustainable development. Concept paper.* [Online]. Available: http://www.environment.nsw.gov.au/resources/greenoffsets/greenoffsets.pdf [Accessed 01 December 2010].

Fields, A. (2009). *Discovering statistics using SPSS* (p. 856). London: SAGE Publications.

Filik Iscen, C., Emiroglu, O., Ilhan, S., Arslan, N., Yilmaz, V., & Ahiska, S. (2008). Application of multivariate statistical techniques in the assessment of surface water quality in Uluabat Lake, Turkey. *Environmental Monitoring and Assessment, 144*, 269–276.

Ford, T. (1999). Understanding population growth in the peri-urban region. *International Journal of Population Geography, 5*, 297–311.

Gavin, B., Nicole, S., & Pieter, S. (1998). The environmental status of Hawkesbury river sediments. *Australian Geographical Studies, 36*, 37–57.

Growns, I. G., & Growns, J. E. (2001). Ecological effects of flow regulation on macroinvertebrate and periphytic diatom assemblages in the Hawkesbury–Nepean River, Australia. *Regulated Rivers: Research & Management, 17*, 275–293.

Growns, J. E., Chessman, C., McEvoy, P. K., & Wright, I. A. (1995). Rapid assessment of rivers using macroinvertebrates: case studies in the Nepean River and Blue Mountains, NSW. *Austral Ecology, 20*, 130–141.

Growns, I., Gehrke, P. C., Astles, K. L., & Pollard, D. A. (2003). A comparison of fish assemblages associated with different riparian vegetation types in the Hawkesbury–Nepean River system. *Fisheries Management and Ecology, 10*, 209–220.

Healthy Rivers Commission. (1998). *Independent inquiry into the Hawkesbury Nepean River system: final report August*

*1998/Healthy Rivers Commission of New South Wales*. Sydney: Healthy Rivers Commission of New South Wales.
Heathwaite, L. (1994). Eutrophication. *Geographical Review, 7*, 31–37.
Howard, M. (2009). *Aquatic Ecosystems productivity 'IS' reliant on water managers and sustainable cities.* Proceedings of 12th International River Symposium. Available: http://riversymposium.com/media/proceedings/2009-proceedings/#Papers [Accessed 10 December 2011].
James, D. (1997). *Environmental incentives: Australian experience with economic instruments for environmental management, Environmental Economics Research Paper 5, Environment Australia*. Canberra, Environment Australia: Commonwealth of Australia.
Krogh, M., Wright, A., & Miller, J. (2008). *Hawkesbury Nepean River environmental monitoring program: final technical report* [Online]. Sydney: Department of Environment and Climate Change NSW and Sydney Catchment Authority. Available: http://www.environment.nsw.gov.au/resources/water/09112hnrempfintechrpt.pdf [Accessed 16 February 2012].
Lettenmaier, D. P., Hooper, E. R., Wagoner, C., & Faris, K. B. (1991). Trends in stream quality in the continental United States, 1978–1987. *Water Resources Research, 27*, 327–339.
Liu, C., Lin, K., & Kuo, Y. (2003). Application of factor analysis in the assessment of groundwater quality in a blackfoot disease area in Taiwan. *Science of the Total Environment, 313*, 77–89.
Mann, H. B. (1945). *Nonparametric tests against trend* (pp. 245–259). Econometrica: Journal of the Econometric Society.
Markich, S. J., & Brown, P. L. (1998). Relative importance of natural and anthropogenic influences on the fresh surface water chemistry of the Hawkesbury–Nepean River, south-eastern Australia. *Science of the Total Environment, 217*, 201–230.
Mazlum, N., Ozer, A., & Mazlum, S. (1999). Interpretation of water quality data by principal components analysis. *Turkish Journal of Engineering and Environmental Sciences, 23*, 19–26.
Moatar, F., & Gailhard, J. (2006). Water temperature behaviour in the River Loire since 1976 and 1881. *Comptes Rendus Geosciences, 338*, 319–328.
Nnane, D. E., Ebdon, J. E., & Taylor, H. D. (2011). Integrated analysis of water quality parameters for cost-effective faecal pollution management in river catchments. *Water Research, 45*, 2235–2246.
Ouyang, Y. (2005). Evaluation of river water quality monitoring stations by principal component analysis. *Water Research, 39*, 2621–2635.
Ouyang, Y., Nkedi-Kizza, P., Wu, Q., Shinde, D., & Huang, C. (2006). Assessment of seasonal variations in surface water quality. *Water Research, 40*, 3800–3810.
Painuly, A. S., Shrestha, S., Hackney, P. & Kabbes, K. C. (2007) Distribution of metals and speciation of sediment grabs in Lake Burragorang in Sydney, Australia. Proceedings of the *World Environmental and Water Resources Congress* (pp.1–9). Available: http://ascelibrary.org/doi/pdf/10.1061/40927(243)65 [Accessed 20 September 2012].
Pejaman, A. H., Bidhendi, N. G. R., Karbassi, A. R., Mehrdadi, N., & Bidhendi, E. M. (2009). Evaluation of spatial and seasonal variation in surface water quality using multivariate statistical techniques. *International of Environmental Science and Technology, 6*, 467–476.
Pekárová, P., Miklánek, P., Halmová, D., Onderka, M., Pekár, J., Kucarova, K., Liova, S., & Skoda, P. (2011). Long-term trend and multiannual variability of water temperature in the pristine Bela River basin (Slovakia). *Journal of Hydrology, 400*, 333–340.
Qin, D., Fisher, I., & Maheswaran, S. (1995). Modelling Nepean River water quality due to proposed effluent and dam releases. *Environment International, 21*, 591–596.
Rahman, M., & Salbe, I. (1995). Modelling impacts of diffuse and point source nutrients on the water quality of South Creek catchment. *Environment International, 21*, 597–603.
SCA (2010). *Water quality at Hawkesbury–Nepean River sites* [Online]. Sydney Catchment Authority. Available: http://www.sca.nsw.gov.au/publications/awqmr08/stream/hnriver [Accessed 7 May 2012].
Semhi, K., Amiotte Suchet, P., Clauer, N., & Probst, J. (2000). Dissolved silica in the Garonne River waters: changes in the weathering dynamics. *Environmental Geology, 40*, 19–26.
Shin, P., & Lam, W. (2001). Development of a marine sediment pollution index. *Environmental Pollution, 113*, 281–291.
Shrestha, S., & Kazama, F. (2007). Assessment of surface water quality using multivariate statistical techniques: a case study of the Fuji river basin, Japan. *Environmental Modelling and Software, 22*, 464–475.
Simeonov, V., Stratis, J., Samara, C., Zachariadis, G., Voutsa, D., Anthemidis, A., Sofoniou, M., & Kouimtzis, T. (2003). Assessment of the surface water quality in northern Greece. *Water Research, 37*, 4119–4124.
Simonovski, J., Owens, C., & Birch, G. (2003). Heavy metals in sediments of the Upper Hawkesbury–Nepean River. *Australian Geographical Studies, 41*, 196–207.
SPCC. (1983). *Water quality in the Hawkesbury–Nepean River—a study recommendations* (p. 207). Sydney: State Pollution Control Commission.
SPCC. (1984). *Sand and gravel extraction in the Upper Hawkesbury River* (p. 18). Sydney: State Pollution Control Commission.
St-Hilaire, A., Brun, G., Courtenay, S., Ouarda, T., Boghen, A., & Bobée, B. (2004). Multivariate analysis of water quality in the Richibucto Drainage basin (New Brunswick, Canada). *Journal of the American Water Resources Association, 40*, 691–703.
Thiebaud, I., & Williams, R. (2007). *Distribution of freshwater macrophytes in the Hawkesbury Nepean River from Warragamba Dam to Wisemans Ferry.* Nelson Bay: NSW Department of Primary Industries Port Stephens Fisheries Centre.
Thoms, M., Parker, C., & Simons, M. (2000). The dispersal and storage of trace metals in the Hawkesbury River valley. In S. Brizga & B. Finlayson (Eds.), *River management: the Australasian experience* (pp. 197–219). Sydney: John Wiley and Sons.
Turner, L., & Erskine, W. (2005). Variability in the development, persistence and breakdown of thermal, oxygen and salt stratification on regulated rivers of south-eastern Australia. *River Research and Applications, 21*, 151–168.
Vega, M., Pardo, R., Barrado, E., & Debán, L. (1998). Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis. *Water Research, 32*, 3581–3592.

Waters, T. F. (1995). Sediment in streams. Sources, biological effects, and control. *American Fisheries Society Monograph, 7*, 251.

Winter, T., Mallory, S., Allen, T., & Rosenberry, D. (2000). The use of principal component analysis for interpreting ground water hydrographs. *Ground Water, 38*, 234–246.

Xia, X., Yang, Z., Huang, G., Zhang, X., Yu, H., & Rong, X. (2004). Nitrification in natural waters with high suspended-solid content: a study for the Yellow River. *Chemosphere, 57*, 1017–1029.

Yue, S., Pilon, P., & Cavadias, G. (2002). Power of the Mann–Kendall and Spearman's rho tests for detecting monotonic trends in hydrological series. *Journal of Hydrology, 259*, 254–271.

Zhou, F., Liu, Y., & Guo, H. (2007). Application of multivariate statistical methods to water quality assessment of the watercourses in Northwestern new territories, Hong Kong. *Environmental Monitoring and Assessment, 132*, 1–13.