Multivariate Analytical Insights into Spatial and Temporal Variation in Water Quality of a Major Drinking Water Reservoir

Azadeh Golshan, Craig Evans, Phillip Geary, Abigail Morrow, Zoe Rogers, Marcel Maeder

Abstract—22 physicochemical variables have been determined in water samples collected weekly from January to December in 2013 from three sampling stations located within a major drinking water reservoir. Classical Multivariate Curve Resolution Alternating Least Squares (MCR-ALS) analysis was used to investigate the environmental factors associated with the physico-chemical variability of the water samples at each of the sampling stations. Matrix augmentation MCR-ALS (MA-MCR-ALS) was also applied, and the two sets of results were compared for interpretative clarity. Links between these factors, reservoir inflows and catchment land-uses were investigated and interpreted in relation to chemical composition of the water and their resolved geographical distribution profiles. The results suggested that the major factors affecting reservoir water quality were those associated with agricultural runoff, with evidence of influence on algal photosynthesis within the water column. Water quality variability within the reservoir was also found to be strongly linked to physical parameters such as water temperature and the occurrence of thermal stratification. The two methods applied (MCR-ALS and MA-MCR-ALS) led to similar conclusions; however, MA-MCR-ALS appeared to provide results more amenable to interpretation of temporal and geological variation than those obtained through classical MCR-ALS.

Keywords—Catchment management, drinking water reservoir, multivariate curve resolution alternating least squares, thermal stratification, water quality.

I. INTRODUCTION

ENVIRONMENTAL monitoring of water quality can produce large and complex datasets which are often difficult to interpret, inviting application of multivariate data analysis methods to extract the underlying information. These techniques have been applied in a variety of scenarios for the identification of factors that influence water systems to facilitate more effective pollution control strategies and management of water resources [1], [2]. Among these methods, MCR-ALS [3]-[5] has been shown to be useful for the resolution and interpretation of real environmental sources affecting a particular geographical area over a period of time [6], [7] using existing environmental monitoring data.

Environmental datasets are usually organized in tables or matrices with rows corresponding to geographical and or temporal sampling sites and columns corresponding to measured variables (concentrations of chemical contaminants or other environmental parameters). Classical MCR-ALS assumes a linear model to explain the observed data variance using a reduced number of contamination sources, representing a linear decomposition of the data. More detailed description of this method can be found in previous publications [3], [4], [8]-[10]. This linear decomposition used in the method can be written in matrix form as:

$$\mathbf{D=SL+E} \tag{1}$$

where **D** is the experimental data array arranged in a data matrix. Equation (1) describes the decomposition of matrix **D** on two matrices, the loadings matrix **L** and the scores matrix **S**. The loadings matrix **L**, may identify the main sources of data variance by means of their chemical composition (composition loadings), which eventually may be related with the main patterns and sources of contamination. The scores matrix **S**, provides sample scores for these data variance patterns (i.e. it will indicate the geographical/temporal sample distribution of these patterns). The matrix **E** is residuals (noise) or non-modelled parts of **D**, which has the same dimension as **D**. However, the matrix decomposition in this equation has to be performed under some constraints as MCR-ALS solutions are usually not unique, being potentially scale and rotation ambiguous [3], [4], [11], [12]. One of the main advantages of MCR-ALS is that it uses physically and chemically interpretable constraints [4], [13], [14] such as non-negativity, selectivity, or closure, to give loading and score profiles that are more easily interpreted and more meaningful from an environmental management perspective. Notwithstanding, it is important to recognise that MCR-ALS solutions to (1) under a set of constraints, must be interpreted with caution, as the obtaining of a unique solution is not guaranteed.

MCR-ALS can be easily extended to the simultaneous analysis of several data matrices via column- or row-wise augmented data matrices, as shown in Fig. 1. By application of these methods to simultaneous analysis via column- wise augmented data matrices, the geographical and temporal information of the main sources of contamination can be obtained. This procedure has been applied to the analysis of environmental data, such as water samples from lakes [15] and rivers [16], fish and sediments [16].

C. Evans, P. Geary and M. Maeder are with the School of Environmental and Life Sciences, The University of Newcastle, Callaghan, NSW 2308, Australia

A. Golshan is with the School of Environmental and Life Sciences, The University of Newcastle, Callaghan, NSW 2308, Australia (corresponding author, e-mail: Azadeh.golshan@uon.edu.au).

A. Morrow and Z. Rogers are with the Hunter Water Corporation, Newcastle, NSW 2300, Australia.

Equation (2) describes the decomposition of the column-wise augmented experimental data matrix $\mathbf{D}_{aug}$ on two matrices, the loadings matrix $\mathbf{L}$, the augmented scores matrix $\mathbf{S}_{aug}$ with N factors and the augmented residual matrix $\mathbf{E}_{aug}$.

$$\mathbf{D}_{aug}=\mathbf{S}_{aug}\,\mathbf{L}+\mathbf{E}_{aug} \qquad (2)$$

Geographical and temporal information of main sources of contamination are readily available from scores $\mathbf{S}_{geo}$ and $\mathbf{S}_{temp}$, which are obtained by the refolding and averaging procedure of augmented scores matrix $\mathbf{S}_{aug}$ [7], [17] (Fig. 1).
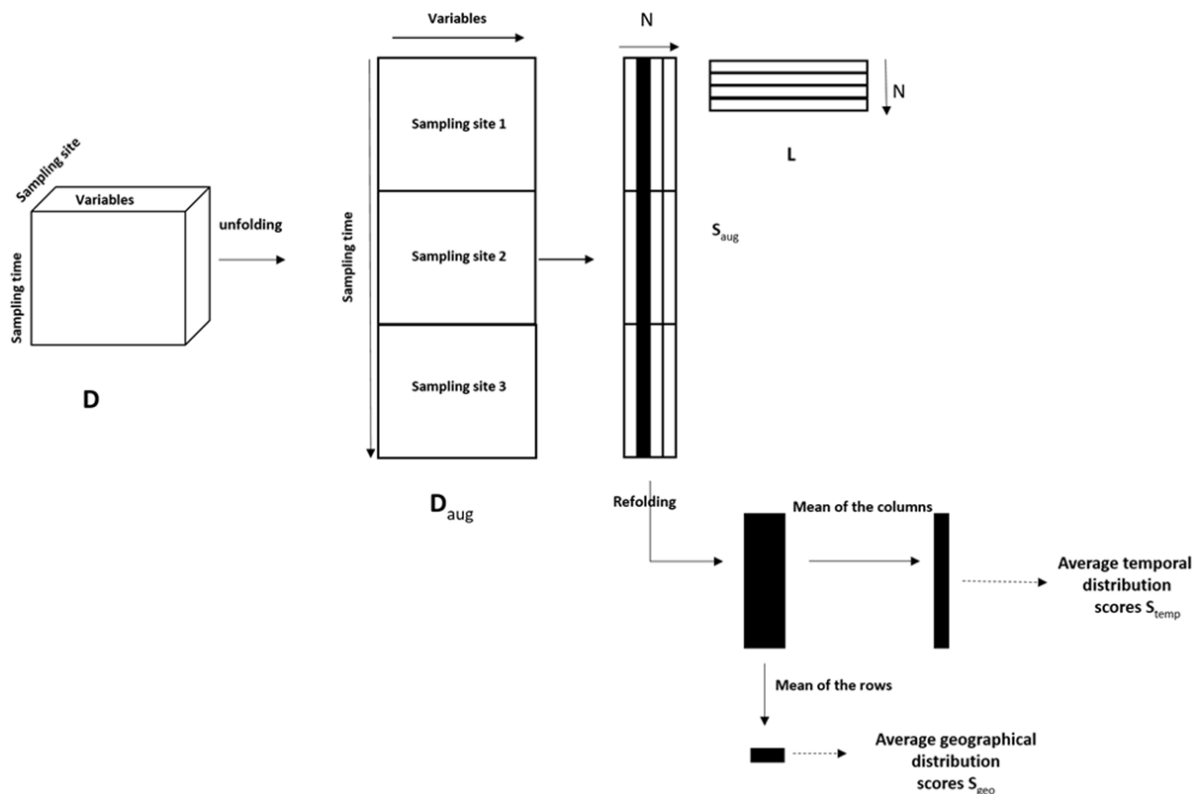

Fig. 1 Using (2), augmented data matrix $\mathbf{D}_{aug}$ of data cube $\mathbf{D}$ can be resolved to the augmented scores, $\mathbf{S}_{aug}$, and loadings, $\mathbf{L}$, matrices considering N factors. For each factor (column filled black), score matrix, $\mathbf{S}$, is constructed by refolding the relevant column of $\mathbf{S}_{aug}$ matrix. Geographical ($\mathbf{s}_{geo}$) and temporal ($\mathbf{s}_{temp}$) profile distributions of each factor are obtained by taking an average of the rows and columns of $\mathbf{S}$ matrix respectively. $\mathbf{S}_{geo}$ and $\mathbf{S}_{temp}$ matrices are obtained by repeating the above procedure for all N factors and combining them. These two matrices give geographical and temporal distributions of these factors

By the extension of MCR-ALS to the simultaneous analysis of the multiple data sets via column- or row-wise augmented data matrices, an additional constraint, so-called trilinearity [18]-[20], can be implemented. By applying this trilinear constraint, temporal profile distributions of each factor in the different geographical sites are forced to have a common shape during each iteration of the ALS optimization and only differ in a scaling factor. This constraint can be applied either to all factors or selectively to some, but with this additional constraint imposed, a unique solution can be achieved. It is therefore strongly recommended to apply the trilinearity constraint whenever there is a trilinear structure in the data. Explanation of the implementation of these additional constraints can be found in previous publications [4], [18], [19], [21]-[24].

In this study, multivariate data analysis including matrix augmentation MCR-ALS was applied in order to better visualize and interpret physicochemical monitoring data collected from three key sites within Grahamstown reservoir, the major drinking water supply for the urban area of Newcastle, New South Wales, Australia. The reservoir is a broad, relatively shallow storage, receiving inputs from a largely unprotected mixed land-use catchment, including urban residential development and a variety of agricultural, industrial and recreational activities. The nature of the catchment lends itself to a variety of potentially adverse water quality outcomes for the reservoir including elevated nutrient levels which, coupled with the physical dimensions of the reservoir, leave it vulnerable to development of algal blooms. In relation to this, stratification of the water column has been considered potentially important in terms of explaining fluctuations in algal productivity within the reservoir. Hence, data relating to temperature gradient within the water column has been included among the physico-chemical parameters

analysed.

The objective of the implementation of an MCR-ALS approach to this data was to identify and assess the spatial and temporal distribution of key water quality variables, and their likely links with the main contamination sources within the Grahamstown catchment. In particular, the aim was to clarify the relative influence of the major inflows from different regions of the catchment along with factors such as stratification, on water quality conditions at the key monitoring points within the reservoir, thereby informing management strategies needed to prevent development of adverse water quality scenarios in the future.

II. MATERIALS

A. Location and Description of the Study Site

Grahamstown reservoir is a shallow reservoir located 20 kilometres north of Newcastle, between the townships of Raymond Terrace to the west and Medowie to the east, in the Hunter Region of New South Wales, Australia. It has an average depth of 7 m, maximum depth of 11 m, a total capacity of 193,000 ML and surface area of 28 km^2. Approximately 50% of the inflow volume consists of water pumped from the Williams River, which is transferred to the northern end of the reservoir through the Balickera Canal (Fig. 2). The land use of the Williams River catchment is dominated by agriculture. The remainder of the inflows comprise runoff from the reservoir's own small catchment on the northern and eastern sides, with approximately 75% of total runoff volume coming from the northern part of the catchment (Fig. 2).

The catchment to the north of the reservoir receives a mixture of runoff from Seven and Nine Mile Creeks, the latter draining through Wallaroo State Forest, although there are also some small farm allotments in this part of the catchment. The Seven Mile Creek catchment includes some quarrying and motor sport activity. The catchment to the east of the reservoir also contains some small farm allotments alongside the urban, mainly sewered settlement of Medowie. Runoff waters from this part of the catchment drain to the reservoir along Campvale Canal from which they are ultimately pumped into the reservoir (Fig. 2). These drainage waters must be pumped into Grahamstown reservoir after rainfall to alleviate flooding of low-lying areas as the dam has been constructed over natural drainage lines for that area. Water for treatment and distribution throughout the city of Newcastle is extracted from Grahamstown reservoir via the George Schroder Water Pumping Station which is located at the far southern point of the dam (Fig. 2) [25].

B. Data

For this study, the focus was on physical and chemical analyses at three key water monitoring sites within the reservoir. One site is located on the north side of the reservoir (NTH) closer to Balickera Canal. It is the most protected and shallowest (4-6 m) of the monitoring sites. A second site is located in the central part of the reservoir (MID). This is the most wind exposed and deepest (9-10 m) of the sites, while

the third site, which is the most southern of the sites, is close to the drinking water off take (STH). The reservoir at this site is between 7 and 9 m deep (Fig. 2).
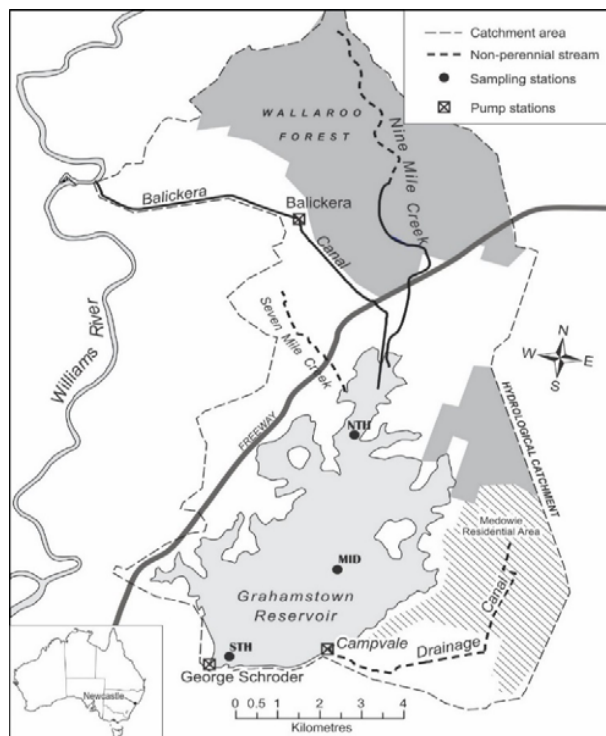

Fig. 2 Location of Grahamstown reservoir and location of the three monitoring sites within the lake. Solid circles mark approximate positions of different monitoring/sampling stations within the reservoir [25]

22 physicochemical parameters (Table I) obtained from each station were used for analysis. Analysis of water quality parameters at the three sampling sites (NTH, MID, STH) was carried out from surface water samples collected weekly from January to December in 2013. An additional variable, the gradient of the change in water temperature with respect to depth (°C/m), was also included in the analyses (described in Table I as 'thermal gradient'). Note that this parameter relates to difference in water temperature between the bottom and the surface, expressed as a gradient, as the limitations of the monitoring undertaken precluded any clear identification of the occurrence of a sharp thermocline and distinct layering within the water column. Discharge data relating to transfers from Balickera and Campvale canals, as well as inflows from Seven and Nine Mile Creeks, were also included as parameters for analysis.

Data sets from the three sampling stations were organised in three data matrices D_{NTH}, D_{MID}, D_{STH} in which the rows of these three matrices were sampling time and the columns were the physico-chemical parameters. These three datasets were analysed individually and simultaneously in a column-wise augmented matrix. The rows in the column-wise augmented matrix represent samples taken on different occasions from the

different geographical sampling sites, and the columns (variables) are the physico-chemical parameters.

TABLE I
CHEMICAL AND PHYSICAL WATER QUALITY PARAMETERS IN THE WATER
QUALITY DATA SETS USED IN THE ANALYSES

Water quality/quantity parameter	Abbreviation	Unit
Water temperature (at surface)	Temp	°C
Thermal stratification	TS	°C/m
pH	pH	-
Alkalinity at pH 4.5	ALK	mg CaCO$_3$/L
Dissolved oxygen	DO	mg/L
Secchi disk depth	Secchi	m
Electrical conductivity	EC	µS/cm
Suspended solids	SS	mg/L
Turbidity	Turbidity	NTU
Total organic carbon	TOC	mg/L
Total ammonia	NH3	mg N/L
Total Kjeldahl nitrogen	TKN	mg N/L
Nitrate	NO_3^-	mg N/L
Nitrite	NO_2^-	mg N/L
Total phosphorus	TP	mg P/L
Soluble reactive phosphorus	SRP	mg P/L
Sulfate	SO_4^{2-}	mg/L
Silica	SiO_2	mg/L
Chlorophyll α	Chl-α	µg/L
Balickera transfers	-	ML
Campvale transfers	-	ML
Seven and Nine Mile Creeks	Creeks	ML

Constraints used in this work during the MCR-ALS were non-negativity, normalisation of loadings to equal height and, the trilinearity constraint in the case of simultaneous analysis of the multiway data sets via column-wise matrix augmentation. To apply non-negativity constraints during the alternating least-squares optimisation, the experimental data matrices D_{NTH}, D_{MID}, D_{STH} and D_{aug} were normalized (Min-Max normalisation) by divided difference between each element and the minimum in a column to the difference between maximum and minimum of the same column [26], [27].

III. RESULTS AND DISCUSSION

Decomposition of the three individual scaled data matrices and scaled augmented experimental data matrices D_{aug}, was performed using the MCR-ALS method, from which four principle components were identified.

A. MCR-ALS Results of Individual Data Matrices

MCR-ALS results for three sites that were analysed individually are summarized in Fig. 3, which shows the loading plots characterizing the main factors in each analysed dataset. In these diagrams, the magnitude of the loading (i.e. height of the column) indicates the relative significance of the variable in relation to the identified factor

The first resolved composition profile at Fig. 3 exhibited the highest contributions of dissolved inorganic nitrogen (NH$_3$, NO$_3^-$ and NO$_2^-$), SRP, Chl-α, and TOC in comparison with the other factors. It was most likely that thermal stratification (or at least the existence of a significant temperature gradient) was responsible for this observation. Since thermal stratification may be frequently interrupted by mixing of the entire water column, this may increase nutrient availability in the euphotic zone, which may in turn promote phytoplankton growth [28], as suggested by the high Chl-α value in this factor.

The second resolved composition profile was largely characterised by variations in water temperature with the three monitoring sites exhibiting a relatively similar response in relation to this parameter. The third resolved composition profile was characterised by large contributions from Chl-α, TP, NH$_3$, turbidity, and SiO$_2$ in the northern (NTH) and central (MID) zones of the reservoir. Transfers from Balickera Canal and inflows from Seven and Nine Mile Creeks appeared more prominent in this profile than transfers from Campavle Canal, suggesting that pumping from the Williams River and other inflows at the northern end, are the main vehicles for increases in these parameters within the reservoir. The prominence of turbidity, SiO$_2$ and TP in this factor for the NTH and MID sampling sites would appear to be consistent with the dominance of agricultural land use in the Williams River catchment and to a lesser extent the activities within the Seven Mile Creek catchment. Turbidity and TP were clearly less significant parameters in the profile of the STH sampling site, probably reflecting its spatial separation from the northern inflows. Conversely, SS was the strongest component of the STH and MID site profiles, but not so for the NTH, suggesting that this parameter may be more closely linked with inflows from Campvale Canal.

In other hand, the relative significance of N species (NH$_3$, NO$_3^-$ and NO$_2^-$) in the northern and southern region of the reservoir (NTH, STH) could be attributed to contributions from all areas of the catchment. This result likely reflects the influence of agricultural run-off, especially from the Williams River catchment via Balickera canal in the NTH, but with regard to the eastern (Campvale Canal) part of the catchment area, may be more indicative of sewage leaks or overflows.

The fourth resolved composition profile, was characterised largely by variation in DO, pH and Secchi depth for the three monitoring sites, suggesting the possible influence of elevated algal abundance and activity. A marked increase in pH in association with algal blooms is not uncommon, as a result of super saturation of DO during the day due to elevated algal photosynthesis, with rapid concurrent consumption of carbon dioxide [29].
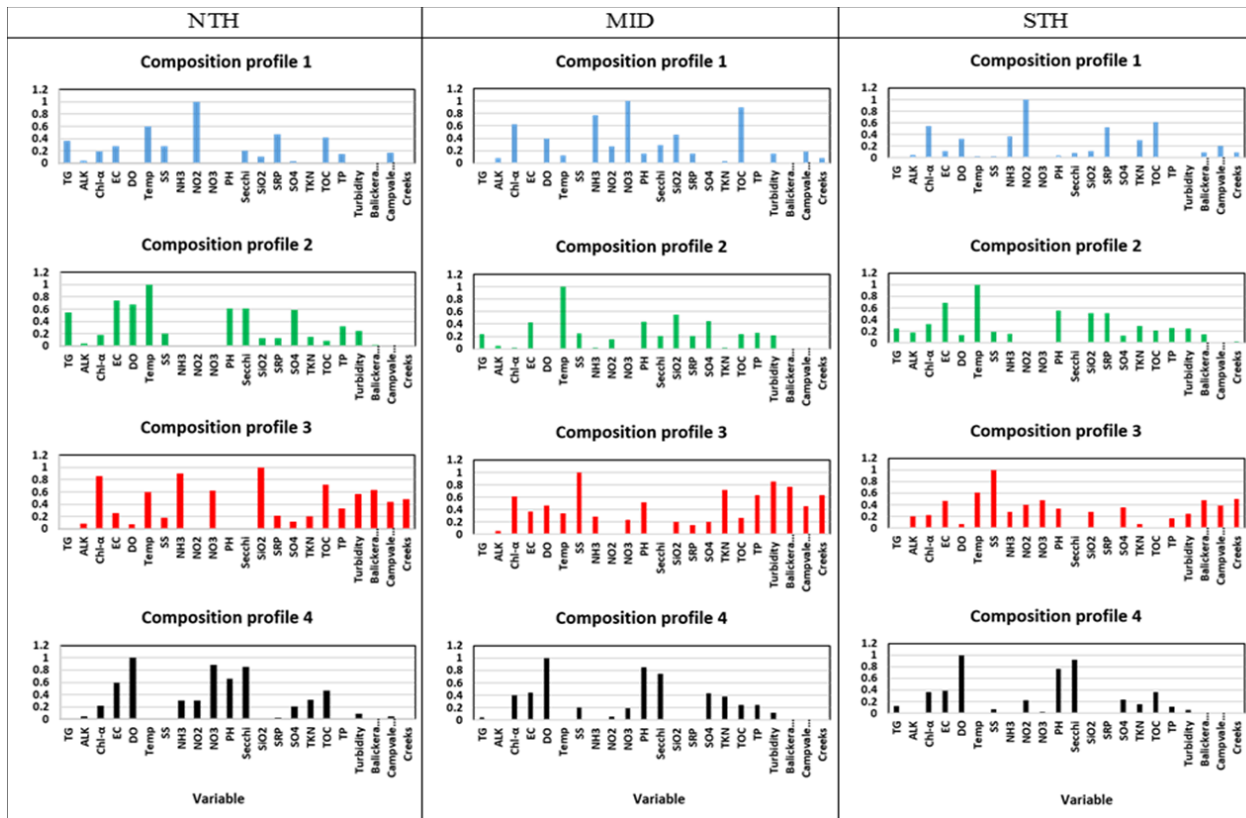
Fig. 3 Four factors were resolved for the NTH, MID and STH sites within the reservoir. Variables are displayed on the x-axis while the y-axis indicates their contribution to the identified factors

Although the composition profiles resolved by classic MCR-ALS provided for meaningful interpretation, further interpretation of the geographical and temporal distribution of some factors (first and third) resolved by this method revealed some inconsistencies. It is likely that these inconsistencies, and the consequent difficulty with interpretation of the geographical and temporal profiles, were likely due to the effects of rotational ambiguities inherent to the classic MCR-ALS method. In order to eliminate these ambiguities, MCR-ALS analysis was subsequently applied to the column-wise augmented matrix in the presence of a trilinearity constraint, with results described below.

B. MCR-ALS Results of Columnwise Augmented Data Matrix

By applying matrix augmentation MCR-ALS methods, it was possible to compare water quality at the three sites simultaneously, providing an easier interpretation of correlation between variables and their geographical and temporal distributions within the water quality variables. The four composition profiles extracted from the MCR-ALS application to D_{aug} are displayed in Fig. 4.

The first resolved composition profile exhibited the highest contributions of NH_3, NO_2^-, Chl-α, SRP, TOC holding some similarities with the first factor obtained of individual data matrix studies. These results suggested that thermal stratification or other processes such as internal water column recycling [30] (bacterial loop or zooplankton excretion) may play an important role in relation to the nutrient released from the sediments. Since TOC values are high in this factor, this result may reflect high dissolved oxygen consumption due to microbial degradation of organic matter in bottom waters, resulting in anoxia and potential increase in nutrient release. This scenario can be facilitated by periods of stratification leading to accumulation of dissolved inorganic nutrients in bottom waters, which can subsequently be released during unstable weather conditions.

The second resolved composition profile had the highest contributions of temperature, aligning it most closely with the second factor obtained by separate analysis of individual sites. The third resolved composition profile had large contributions of TP, turbidity, Chl-α and SiO_2, similar to the third factor obtained in the individual data matrix studies. This likely reflected the quality of the water from Williams River (via Balickera Canal) where the land use of the river's catchment is dominated by agriculture. Compared to the other identified factors, the relative significance of NO_3^- value in this factor, is perhaps also reflective of the contributions of run-off from agricultural activity. While Balickera transfers stand out as the dominant factor in this profile, the relative significance of inflows from Seven and Nine Mile Creeks as well as transfers from Campvale Canal, suggests that the distribution of

nitrogenous nutrient levels within the reservoir is a function of contributions from all areas of the catchment. Similar to the final factor obtained by individual site studies, the fourth factor was characterised largely by variation in DO, pH and

Secchi depth, suggesting the possible influence of elevated algal photosynthesis as it might increase DO levels with rapid uptake of carbon dioxide, which usually produces a marked increase in pH.
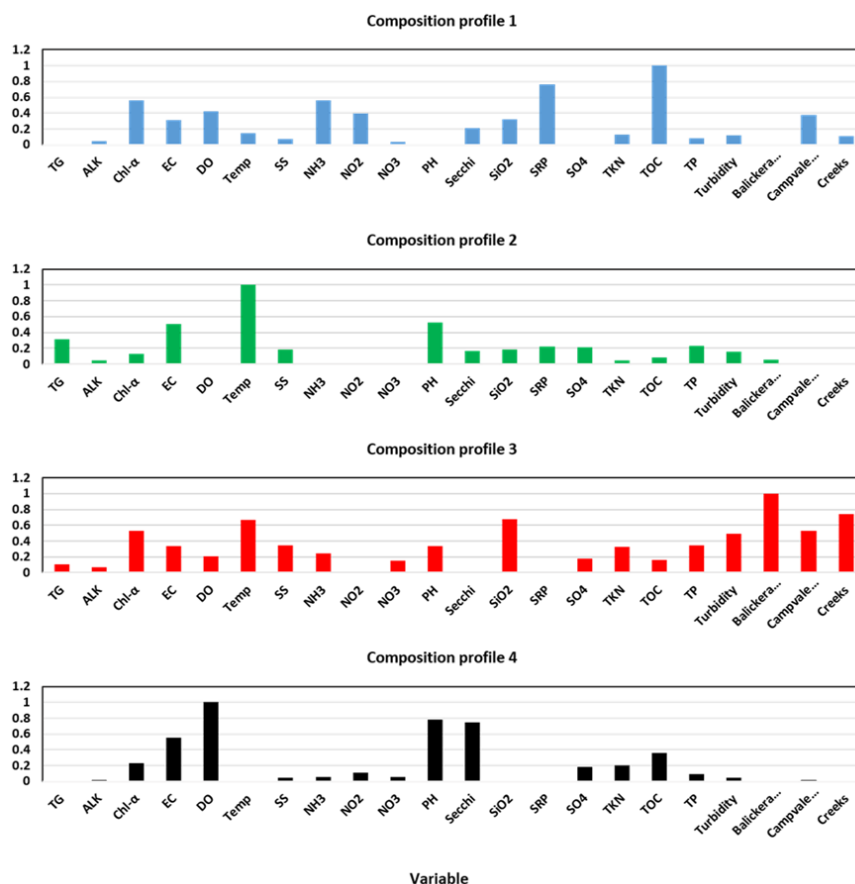

Fig. 4 MCR-ALS composition profiles for the column-wise augmented data matrix. Variables are displayed on the x-axis and the y-axis indicates the contribution of each of the variables for each identified factor

Fig. 5 illustrates the averaging MCR-ALS unfolded resolved scores, which were plotted in two different ways, the temporal and geographical distributions for four environmental factors. As can be observed in Fig. 5 (a), in terms of the temporal distributions the first factor was closely linked to variation in TOC; the second factor appeared to be closely aligned with seasonal temperature variation; the pattern of fluctuations in the Balickera transfers profile resembled that of the third factor; and the pattern of fluctuations in the dissolved oxygen profile resembled the temporal distributions of the fourth factor.

The temporal distributions (S_{temp}) resolved by MCR-ALS clearly identify differences between factors across the sampling period, while the contribution of sampling stations was relatively similar for all factors due to highly overlapped patterns of factors (see the pattern of dotted lines in Fig. 5). It can be observed from Fig. 5 (b), that NTH and STH showed the same contribution of the first factor for which an obvious interpretation was not evident, as these sites are most likely to be influenced by inflows from different regions of the catchment. Further investigation of sediment characteristics and the complexity of interplay between physical, chemical and biological processes in these parts of the reservoir, may be required before firm conclusions can be drawn. Since the second factor was driven largely by broad climatic conditions, as characterised by variations in water temperature, the three monitoring sites exhibited a similar response. The third factor was found to be closely aligned with transfers from Balickera Canal, and as such the contribution was clearly greatest at NTH, the most proximal site to the Balickera inlet. The final factor, characterised largely by variation in DO, was slightly more prominent in the central part of the reservoir, possibly reflecting the extended fetch and therefore greater potential for wind-induced mixing at the surface in this region of the reservoir.
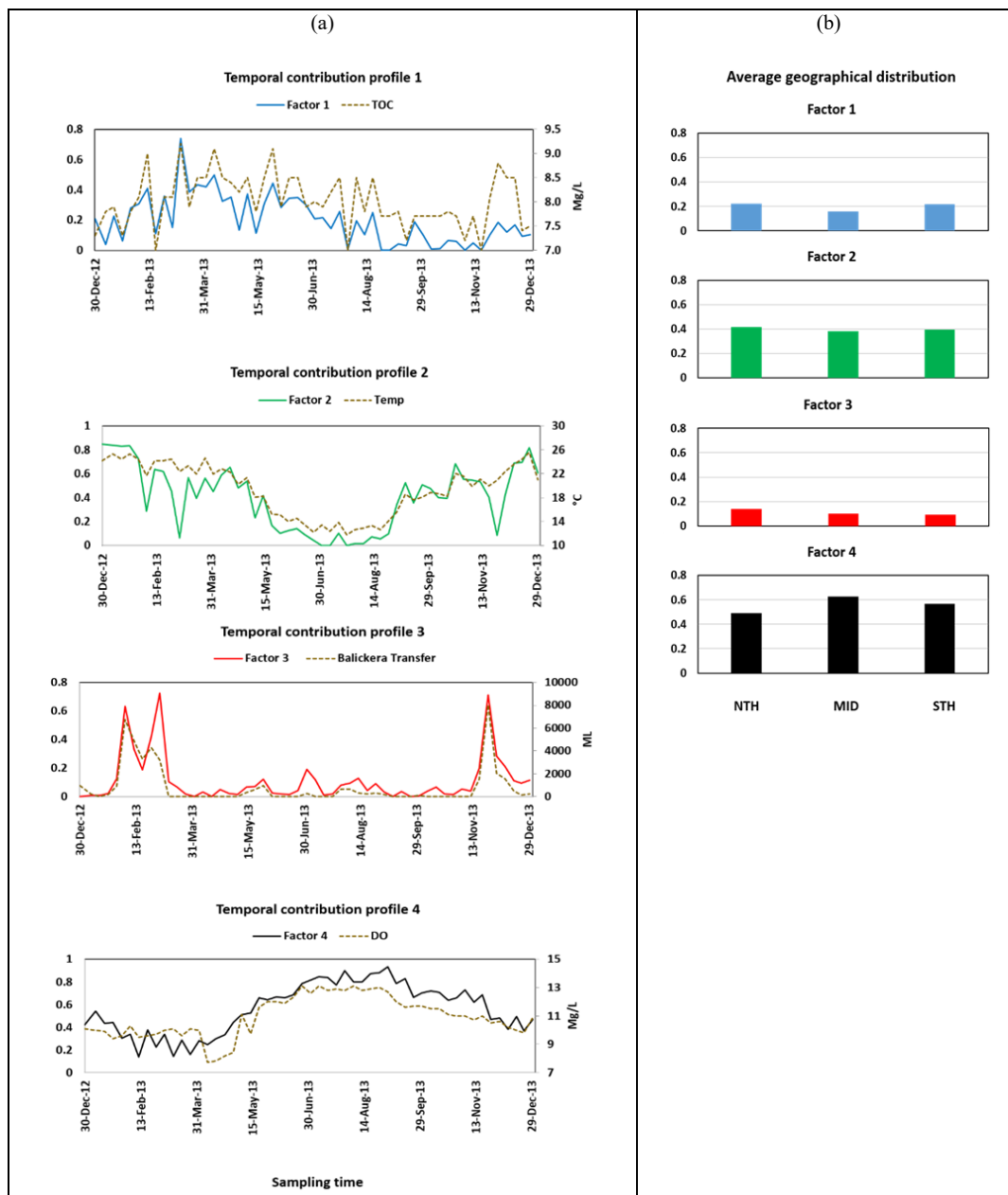
Fig. 5 Temporal (a) and geographical (b) distribution of MCR-ALS resolved scores for scaled augmented data matrix after refolding arrangement described in Fig. 1. Sampling time in the score plots are displayed on the x-axis for the three consecutively years. The y-axis indicates the contribution of the identified factors. Temporal distributions of TOC, Temp, Balickera transfers, and DO are presented as dotted lines, which follow the pattern of the resolved factors most closely

This study has illustrated the usefulness of MCR-ALS technique for analysing and interpreting environmental data studies. Based on the results of this method, it is concluded that while similar patterns for four factors were found for both individual and augmented data sets, variations between the two analyses were noted. Similar temporal patterns for the second and fourth factors were separately identified by both classical MCR-ALS and matrix augmentation MCR-ALS, with the remaining two patterns appearing to be more clearly resolved by matrix augmentation MCR-ALS. It is likely that

these differences reflect the effect of rotational ambiguities in classical MCR-ALS. This suggests that matrix augmentation MCR-ALS may be more flexible and adaptable to environmental studies where a trilinearity constraint can be used.

IV. CONCLUSIONS

The use of MCR-ALS to provide an insight into water quality in Grahamstown reservoir showed the usefulness of

such multivariate analysis for analysing and interpreting environmental data sets.

In terms of water quality assessment, this work has assisted in the identification pollution sources/factors and an understanding of the temporal/geographical variations in water quality of Grahamstown reservoir which may be used to assist with improved water quality management. The result of applying MCR-ALS has confirmed that the influence of inputs from the reservoir's own catchment area, via Campvale Drainage Canal and Seven and Nine Mile Creeks, was different from that exerted by inputs from the Williams River catchment via Balickera Canal. While all inflows were implicated in elevating levels of nitrogenous nutrients (mostly in the form of NO_3^-) within the reservoir, SiO_2 and TP levels appeared to be influenced more by transfers from the Williams River catchment (via Balickera Canal) at the northern end of the reservoir. SRP, NO_2^- levels appeared to be influenced by thermal gradient within the water column, most likely linked to release from the sediment as a result of either periods of anoxia or microbiologically mediated processes. However, clarification of the likely occurrence and significance of stratification to water quality at the northern and southern ends of the reservoir, as indicated by the existence of a significant thermal gradient within the water column at these locations, requires further investigation.

The results derived from the MCR-ALS of the individual data sets were in good agreement with those from MA-MCR-ALS. However, use of matrix augmentation MCR-ALS appeared to yield an interpretation of geographical and temporal information which is more flexible and adaptable to environmental studies than classic MCR-ALS.

REFERENCES

[1] Reghunath, R., T. S. Murthy, and B. Raghavan, *The utility of multivariate statistical techniques in hydrogeochemical studies: an example from Karnataka, India.* Water Research, 2002. 36(10): p. 2437-2442.

[2] Simeonov, V., et al., *Lake water monitoring data assessment by multivariate statistics.* Journal of Water Resource and Protection, 2010. 2(04): p. 353.

[3] Tauler, R., *Multivariate curve resolution applied to second order data.* Chemometrics and Intelligent Laboratory Systems, 1995. 30(1): p. 133-146.

[4] Tauler, R., A. Smilde, and B. Kowalski, *Selectivity, local rank, three-way data analysis and ambiguity in multivariate curve resolution.* Journal of Chemometrics, 1995. 9(1): p. 31-58.

[5] Tauler, R., B. Kowalski, and S. Fleming, *Multivariate curve resolution applied to spectral data from multiple runs of an industrial process.* Analytical Chemistry, 1993. 65(15): p. 2040-2047.

[6] Salau, J. S. I., et al., *Input characterization of sedimentary organic contaminants and molecular markers in the Northwestern Mediterranean Sea by exploratory data analysis.* Environmental Science & Technology, 1997. 31(12): p. 3482-3490.

[7] Tauler, R., D. Barcelo, and E. M. Thurman, *Multivariate correlation between concentrations of selected herbicides and derivatives in outflows from selected US Midwestern reservoirs.* Environmental

Science & Technology, 2000. 34(16): p. 3307-3314.

[8] Jolliffe, I. T., *Principal Component Analysis and Factor Analysis*, in *Principal component analysis*. 1986, Springer. p. 115-128.

[9] Massart, D., et al., *Straight line regression and calibration.* Handbook of Chemometrics and Qualimetrics. Part A, 1998: p. 171-230.

[10] Wold, S., K. Esbensen, and P. Geladi, *Principal component analysis.* Chemometrics and Intelligent Laboratory Systems, 1987. 2(1-3): p. 37-52.

[11] Lawton, W. H. and E. A. Sylvestre, *Self modeling curve resolution.* Technometrics, 1971. 13(3): p. 617-633.

[12] Paatero, P., et al., *Understanding and controlling rotations in factor analytic models.* Chemometrics and Intelligent Laboratory Systems, 2002. 60(1): p. 253-264.

[13] Ruckebusch, C. and L. Blanchet, *Multivariate curve resolution: A review of advanced and tailored applications and challenges.* Analytica Chimica Acta, 2013. 765(Supplement C): p. 28-36.

[14] Mouton, N., et al., *Hybrid hard- and soft-modeling approach for the resolution of convoluted femtosecond spectrokinetic data.* Chemometrics and Intelligent Laboratory Systems, 2011. 105(1): p. 74-82.

[15] Giussani, B., et al., *A chemometric approach to the investigation of major and minor ion chemistry in lake Como (Lombardia, Northern Italy).* Annali Di Chimica, 2006. 96(5-6): p. 339-346.

[16] Terrado, M., D. Barceló, and R. Tauler, *Identification and distribution of contamination sources in the Ebro river basin by chemometrics modelling coupled to geographical information systems.* Talanta, 2006. 70(4): p. 691-704.

[17] Tauler, R., et al., *Chemometric modeling of main contamination sources in surface waters of Portugal.* Environmental Toxicology and Chemistry, 2004. 23(3): p. 565-575.

[18] Tauler, R., I. Marqués, and E. Casassas, *Multivariate curve resolution applied to three-way trilinear data: Study of a spectrofluorimetric acid–base titration of salicylic acid at three excitation wavelengths.* Journal of Chemometrics, 1998. 12(1): p. 55-75.

[19] Alier, M., et al., *Trilinearity and component interaction constraints in the multivariate curve resolution investigation of NO and O_3 pollution in Barcelona.* Analytical and Bioanalytical Chemistry, 2011. 399(6): p. 2015-2029.

[20] Smilde, A., R. Bro, and P. Geladi, *Multi-way analysis: applications in the chemical sciences*. 2005: John Wiley & Sons.

[21] Bosco, M., M. Callao, and M. Larrechi, *Resolution of phenol, and its di-hydroxyderivative mixtures by excitation–emission fluorescence using MCR-ALS: Application to the quantitative monitoring of phenol photodegradation.* Talanta, 2007. 72(2): p. 800-807.

[22] De Juan, A. and R. Tauler, *Comparison of three-way resolution methods for non-trilinear chemical data sets.* Journal of Chemometrics, 2001. 15(10): p. 749-771.

[23] Galera, M. M., M. G. García, and H. Goicoechea, *The application to wastewaters of chemometric approaches to handling problems of highly complex matrices.* TrAC Trends in Analytical Chemistry, 2007. 26(11): p. 1032-1042.

[24] Jaumot, J., et al., *A graphical user-friendly interface for MCR-ALS: a new tool for multivariate curve resolution in MATLAB.* Chemometrics and Intelligent Laboratory Systems, 2005. 76(1): p. 101-110.

[25] Cole, B. and B. Williams, *Grahamstown DamWater Quality & Aquatic Ecological Functioning, A compilation of Scientific and Adaptive Management Studies.* 2011 (unpublished report).

[26] van den Berg, R.A., et al., *Centering, scaling, and transformations: improving the biological information content of metabolomics data.* BMC Genomics, 2006. 7(1): p. 142.

[27] Smilde, A. K., et al., *Fusion of Mass Spectrometry-Based Metabolomics Data.* Analytical Chemistry, 2005. 77(20): p. 6729-6736.

[28] Wilhelm, S. and R. Adrian, *Impact of summer warming on the thermal characteristics of a polymictic lake and consequences for oxygen, nutrients and phytoplankton.* Freshwater Biology, 2008. 53(2): p. 226-237.

[29] Verspagen, J. M., et al., *Rising CO_2 levels will intensify phytoplankton blooms in eutrophic and hypertrophic lakes.* PLoS One, 2014. 9(8): p. e104325.

[30] Müller, S., S. M. Mitrovic, and D. S. Baldwin, *Oxygen and dissolved organic carbon control release of N, P and Fe from the sediments of a shallow, polymictic lake.* Journal of Soils and Sediments, 2016. 16(3): p. 1109-1120.