Available online at www.sciencedirect.com
SCIENCE DIRECT®
Chemometrics and Intelligent Laboratory Systems 73 (2004) 131–138
Chemometrics and intelligent laboratory systems
www.elsevier.com/locate/chemolab

Multivariate monitoring of a biological wastewater treatment process: a case study at Melbourne Water's Western Treatment Plant

Tarja Miettinen[a,*], Timothy J. Hurse[b], Michael A. Connor[c], Satu-Pia Reinikainen[a], Pentti Minkkinen[a]

[a] *Department of Chemical Technology, Lappeenranta University of Technology, Lappeenranta, Finland*
[b] *Department of Microbiology and Parasitology, University of Queensland, Brisbane, Australia*
[c] *Department of Chemical and Biomolecular Engineering, University of Melbourne, Melbourne, Australia*

Received 24 July 2003; accepted 25 November 2003
Available online 14 March 2004

Abstract

Biological wastewater treatment is a complex, multivariate process, in which a number of physical and biological processes occur simultaneously. In this study, principal component analysis (PCA) and parallel factor analysis (PARAFAC) were used to profile and characterise Lagoon 115E, a multistage biological lagoon treatment system at Melbourne Water's Western Treatment Plant (WTP) in Melbourne, Australia. In this study, the objective was to increase our understanding of the multivariate processes taking place in the lagoon. The data used in the study span a 7-year period during which samples were collected as often as weekly from the ponds of Lagoon 115E and subjected to analysis. The resulting database, involving 19 chemical and physical variables, was studied using the multivariate data analysis methods PCA and PARAFAC. With these methods, alterations in the state of the wastewater due to intrinsic and extrinsic factors could be discerned. The methods were effective in illustrating and visually representing the complex purification stages and cyclic changes occurring along the lagoon system. The two methods proved complementary, with each having its own beneficial features.

Keywords: Wastewater stabilization lagoons; Purification; Sewage; PCA; PARAFAC; Process monitoring

1. Introduction

Melbourne Water is the body responsible for managing Melbourne's water supply catchments, collecting and treating most of Melbourne's sewage, and managing waterways and major drainage systems in and around Melbourne. Melbourne Water's Western Treatment Plant (WTP) is one of the largest sewage treatment plants in the world, covering almost 11,000 ha and processing, on average, 500 million litres of sewage a day. At the WTP, the wastewater is treated using lagoons, land filtration and grass filtration; the treated wastewater is discharged to Port Phillip Bay. Protecting the long-term health of Port Phillip Bay, habitat to more than 1000 species of marine plants and animals, is one of Melbourne Water's key priorities. A 4-year Port Phillip Bay Environmental study found that the Bay was healthy by world standards but recommended reducing nitrogen inflows. Since the WTP makes a substantial contribution to these inflows, Melbourne Water has been taking steps to upgrade WTP treatment processes and thereby to reduce the nitrogen levels in WTP effluents.

The effectiveness of nitrogen removal at the WTP requires an improved understanding of the complex processes occurring in its lagoon systems. An earlier multivariate data analysis carried out on measurements of key pollutants in sewage of purely domestic origin in Melbourne's suburbs [1] had shown the value of this technique in discerning relationships between different parameters. It was therefore decided to apply this technique to data accumulated over a number of years for one of the newer WTP lagoon systems, Lagoon 115E. The treatment process was interpreted and profiled by multivariate methods: principal component analysis (PCA) and

* Corresponding author. Tel.: +358-5-621-2263; fax: +358-5-621-2199.
E-mail address: tarja.miettinen@lut.fi (T. Miettinen).

0169-7439/$ - see front matter © 2003 Elsevier B.V. All rights reserved.
doi:10.1016/j.chemolab.2003.11.012

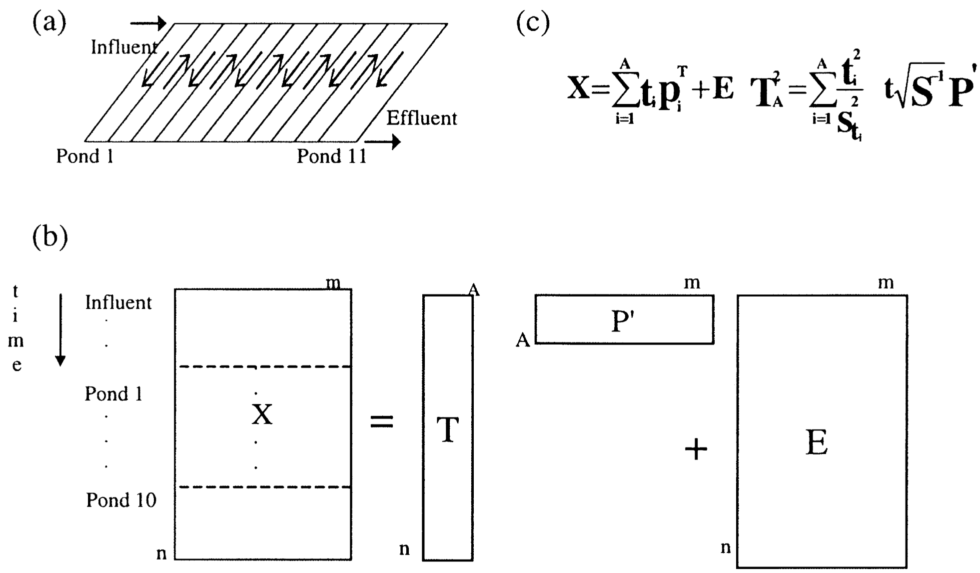


Fig. 1. (a) Schematic of lagoon. (b) Data matrix with objects (samples) as rows and chemical and physical variables as columns. Data decomposition is made into object score vectors **t** and variable loading vectors **p**. (c) Hotelling T^2-values are computed based on the first A PCs, s_{ti}^2 being an estimated variance of t_i. The variable contributions for a T^2 value of a sample are derived by multiplying together a vector of score values **t** and loading matrix **P**. The matrix **S** is a diagonal matrix with diagonal elements equal to eigenvalues and it is used to normalise the score values.

multiway parallel factor analysis (PARAFAC). The primary function of the study was to investigate which are the critical ponds in the process and which factors most affect the levels of purification achieved in the lagoon.

2. Materials and methods

Lagoon 115E comprises 11 ponds in series. A schematic of the lagoon is shown in Fig. 1(a). It takes around 10 weeks

Removal of Organic Carbon

Anaerobic reaction:

anaerobic bacteria

Organics → new cells, organic acids, CH_4, H_2S, H_2O, CO_2, H_2

Aerobic reaction:

aerobic bacteria

Organics + O_2 → new cells, CO_2, H_2O

Removal of Nitrogen

1st stage – Ammonification:

bacteria

Organic-N → NH_3/NH_4^+

2nd stage – Nitrification:

nitrifying bacteria

$NH_4^+ + ½\,O_2$ → $NO_2^- + H_2O + 2H^+$

$NO_2^- + ½\,O_2$ → NO_3^-

3rd stage – Denitrification

denitrifying bacteria

NO_3^- + organics → N_2, CO_2, H_2O, new cells

Fig. 2. Main biochemical reactions relevant to removal of organic carbon and removal of nitrogen from wastewater. Important products and substrates are indicated; some products and substrates are omitted.

Table 1
Chemical and physical variables either routinely measured or calculated from measured variables

	Chemical or physical variable	Unit	
BOD_5	5-day biological oxygen demand	mg/l	*
CA	Chlorophyll *a* concentration	μg/l	
DO	Dissolved oxygen concentration	mg/l	¤ *
H_2S	Hydrogen sulfide concentration	mg/l	
NH_3	Ammonia-N concentration	mg/l	*
NO_2	Nitrite-N concentration	mg/l	*
NO_3	Nitrate-N concentration	mg/l	*
Org-N	Organic-N concentration	mg/l	*
TOC	Total organic carbon concentration	mg/l	*
TSS	Total suspended solids concentration	mg/l	*
Free NH_3	Gaseous dissolved ammonia concentration	mg/l	¤ * #
Free HNO_2	Free nitrous acid concentration	mg/l	¤ * #
TKN	Total Kjeldahl nitrogen concentration (N in form of ammonia-N or organic-N)	mg/l	* #
TON	Total oxidized nitrogen concentration (N in form of nitrate or nitrite)	mg/l	* #
NCABOD	Non-algal biological oxygen demand	mg/l	#
NCATSS	Non-algal suspended solids concentration	mg/l	#
Total-N	Total nitrogen concentration (sum of TKN and TON)	mg/l	* #
pH	pH		
Temp	Temperature	°C	

[#]Variables derived from other measured parameters. *Variables used in PCA for influent. Influent not analysed for ¤ variables, set to zero in PCA.

for the wastewater to flow through the lagoon, whereupon it is discharged to Port Phillip Bay. The ponds are bounded by concrete weirs and they are constructed so that the water flows from one pond to the next under gravity.

In the lagoon, removal of organic carbon and nitrogen is achieved largely by means of biological processes [2,3]. The main reactions occurring are shown in Fig. 2. The wastewater first enters the anaerobic zone, where most of the solids settle out. Both the settled solids and dissolved organic compounds undergo anaerobic decomposition while organic nitrogen is converted to ammonia by ammonification reactions. From the anaerobic zone, wastewater moves on to the aerobic zone, where aerobic bacteria break down more of the organic matter. Removal of nitrogen is achieved by nitrification–denitrification reactions. In the nitrification stage, ammonia is oxidized, first to nitrite and then to nitrate, by nitrifying bacteria. In the final, denitrification stage nitrate is converted under anoxic conditions to nitrogen (N_2) by denitrifying bacteria. This nitrogen escapes to the atmosphere.

The data set used in this study comprised the analytical results obtained for the 115E Lagoon system over a 7-year period (1986–1994). Over this time, the lagoon influent was routinely analysed for seven chemical and physical parameters, and less frequently for an additional two. Pond waters were routinely analysed for 12 chemical and physical parameters (listed in Table 1); from these, an additional

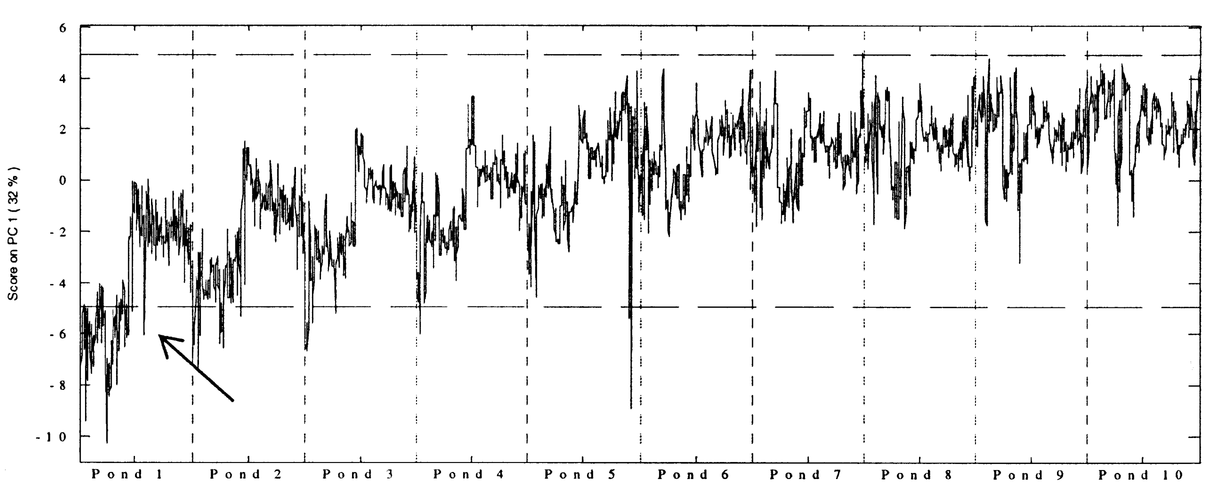


Fig. 3. PCA scores with 95% limits for the first principal component for the analysis of all samples. The arrow indicates the point in time when the process was modified.

seven derived parameters were calculated. The ponds and the influent were sampled contemporaneously at intervals of between 1 week (influent and pond 1) and 3 weeks. Samples were taken at the outlet of the pond concerned so the data for pond 1 relate to the water flowing from pond 1 into pond 2 and so forth. Research at the WTP has shown that individual ponds are well mixed; hence, measurements taken at pond outlets are representative of average pond values. Pond 11 was sampled less frequently, so it was excluded from the data set. During the period from mid-1991 to mid-1993, pond 5 was completely by-passed so in most cases pond 5 data was also excluded. (By-passing was achieved by splitting the effluent from pond 3 between ponds 4 and 6; effluent from both ponds 4 and 6 was directed to pond 7.)

For the purpose of analysis, the influent and bond water data were arranged in a matrix as blocks with equal time periods as shown in Fig. 1(b). On the time axis, the shortest sampling interval (that of pond 1) was used as a reference. In cases where larger sampling intervals gave rise to gaps in the matrix, the corresponding value from the previous sampling round was used to fill the gap. There were two reasons for doing this: firstly it is common practice in process monitoring, and secondly it gives each pond an equal weight in the models. Varying time periods, as well as pond and influent combinations are applied in the following sections.

In PCA, the data was autoscaled, while in PARAFAC, it was scaled within the chemical variables mode and centred across the time (objects) mode. The eigenvalue criterion in PCA and the core consistency diagnostics in PARAFAC were used to estimate the number of significant components [4–7].

3. Results and discussions

3.1. Effects of a process modification detected and assessed by PCA

When the data on all pond water samples collected during the period October 1986–January 1994 were analysed, PCA scores suggested that a major process change had occurred in September–October 1989. This was confirmed by Melbourne Water: it was at this time that the first pond of 115E had been divided into two zones, the first anaerobic and the second aerobic. This was achieved by placing a barrier around 200 m from the start of the pond and by installing a number of mechanical aerators downstream of the barrier. The purpose of the aerators was to maintain a positive dissolved oxygen (DO) concentration in the aerobic zone while the barrier minimized backmixing between the two zones. The scores for the first principal component in Fig. 3 illustrate how the score values change permanently or fluctuate strongly at the point of process enhancement. The influence is seen in up to the fifth principal component.

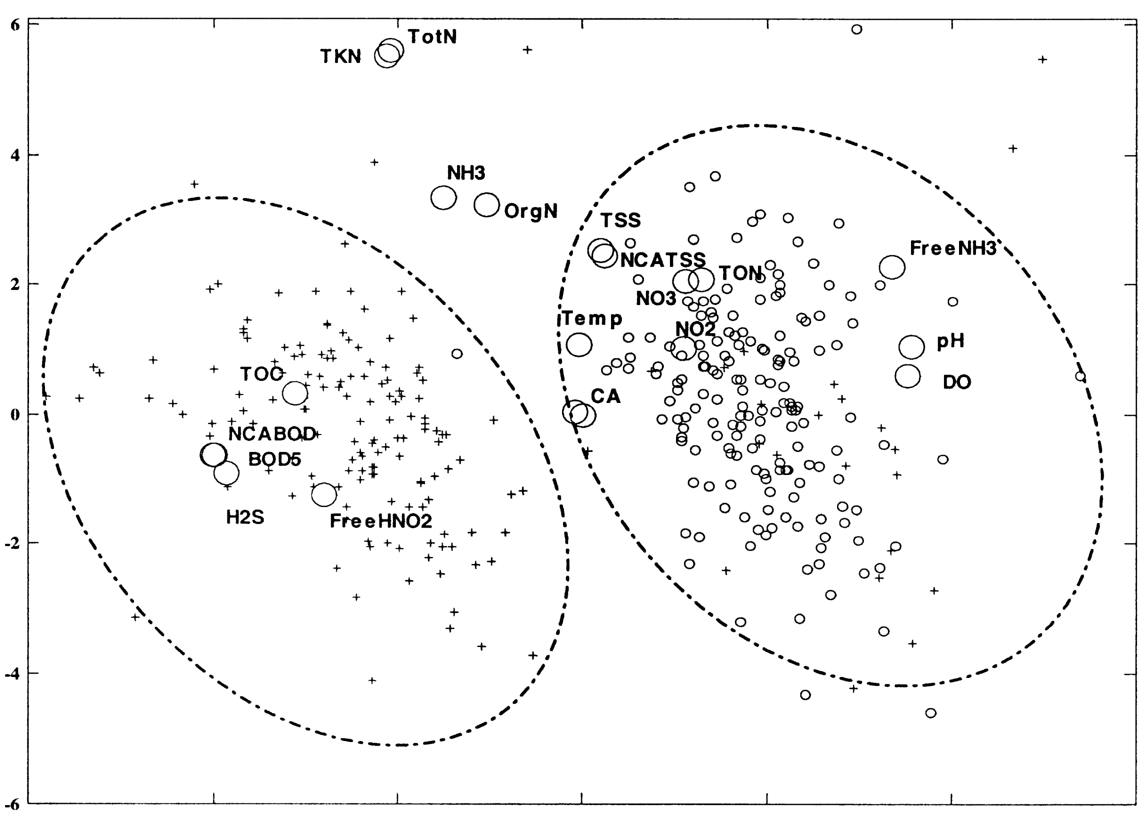


Fig. 4. PCA scores and loadings for the samples of pond 1 collected during a 7-year period between 1986 and 1994. + stands for the samples collected before and ○ for the samples collected after the process enhancements. The ellipses indicate sample groups and have no statistical meaning.

The influence of the process modifications on treatment efficiency was interpreted with PCA for the data on pond 1. PCA scores and loadings for the first two principal components in Fig. 4 show that after the installation of the aerators and the barrier, the DO concentration increased. That in turn increased the removal of biological oxygen demand (BOD_5) and total organic carbon (TOC). Increased concentrations of nitrite (NO_2^-) and nitrate (NO_3^-) in pond 1 effluents suggest that nitrification was now beginning at an earlier stage than before. The pH increased by around half a pH unit, due most likely to more complete degradation of organic acids. The increase in DO level led also to a decrease in the concentration of hydrogen sulfide (H_2S); this is not surprising as H_2S is a product of anaerobic decomposition.

3.2. Treatment stages discerned through use of PCA

PCA of the data for pond water samples collected during the 2 years 1991–1992 was used to describe the treatment stages in the lagoon. The lagoon system was not modified and was rather stable during these years. The score and loading biplot for the first two principal components, shown in Fig. 5, enables the progress of the pollutant removal processes to be visualised. The objects within group 1 are characterised by the variables used to define the content of suspended solids and organic matter in the wastewater. The propagation of ponds 1–4 along arrow 1 is caused by the decreasing amounts of organics. The objects within group 2 are dominated by two factors. One is associated with nitrification and denitrification, the other with seasonal variation, which has quite an influence on biological processes at the WTP. Ponds 6–10 do not propagate clearly in the direction of arrow 2, with the objects tending to subdivide according to the season. Ammonia and TKN levels decrease and the concentrations of intermediates of nitrification–denitrification reactions increase in the direction of arrow 2. As would be expected, DO concentration increases toward the end of the lagoon, as does pH.

A more specific investigation of the purification stages was carried out by studying variable contributions for Hotelling T^2 values of objects (Fig. 1(c)). Such contributions describe how far sample values are from a model centre and quantify the contribution of the variables concerned. Differences in variable contributions between the pond blocks were compared using a change in contribution levels as an indicator of a process transition. As the influent differs greatly from the pond waters, it dominates PCA models. For that reason, two PCA models were used: (a) the model used earlier in this section and (b) a model for the data on pond waters and influent.

The variable contributions of PCA model b in Fig 6(a) show that most of the BOD_5, TSS and TOC are removed in

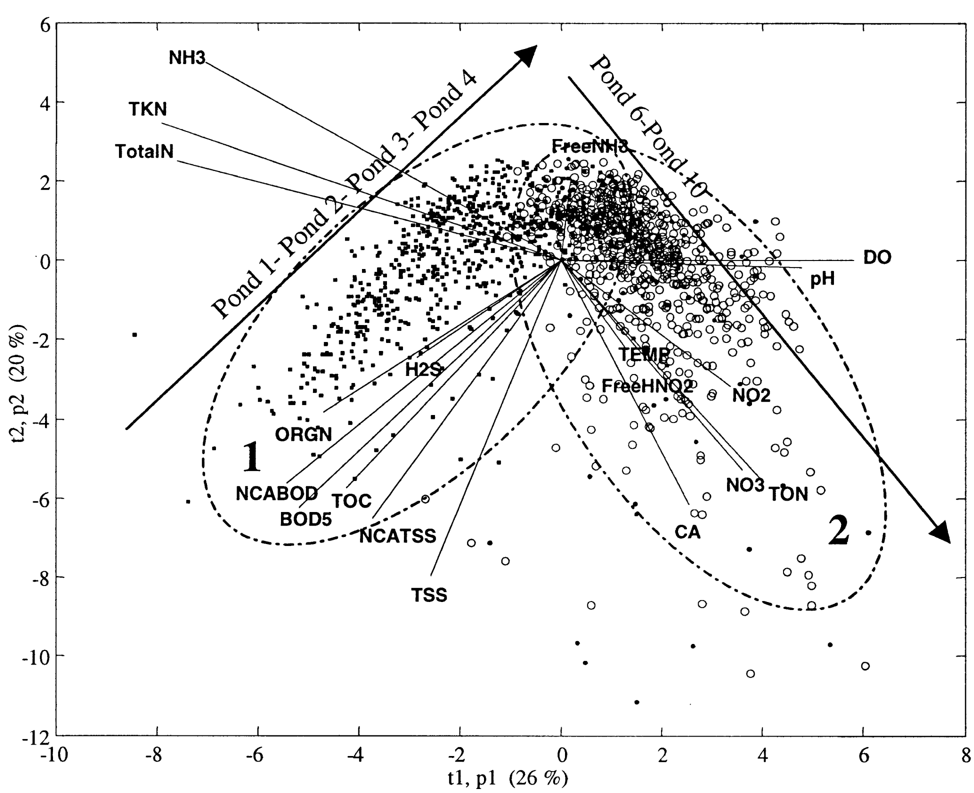


Fig. 5. Score and loading biplot for PCA on the pond samples collected during 2 years describes advancing purification. Group 1 comprises ponds 1–4. The ponds propagate in the direction of arrow 1 describing the removal of organics. Group 2 comprises ponds 6–10. The ponds propagate slightly in the direction of arrow 2 describing removal of nitrogen. The samples within group 2 tend to subdivide according to the season.

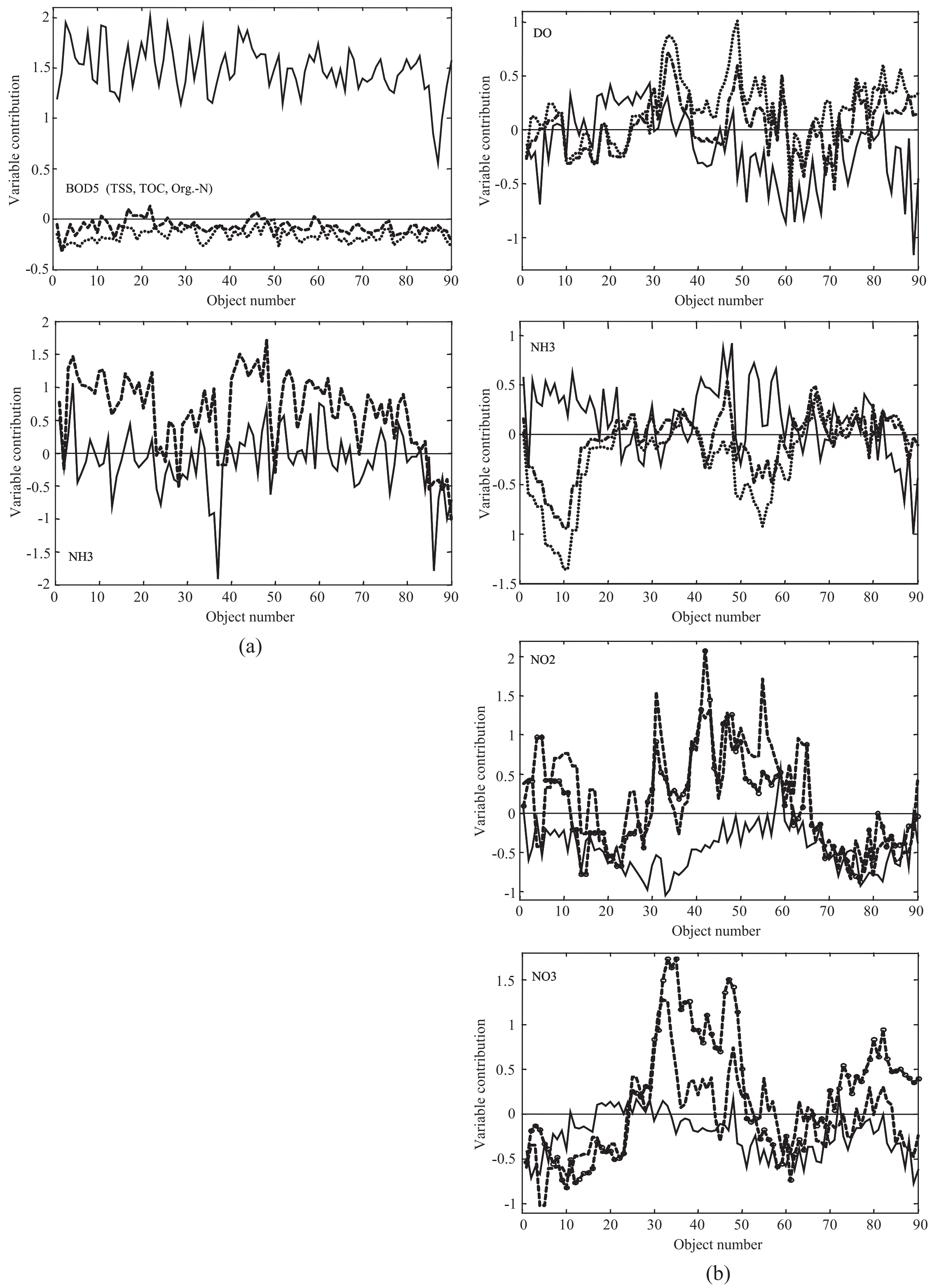
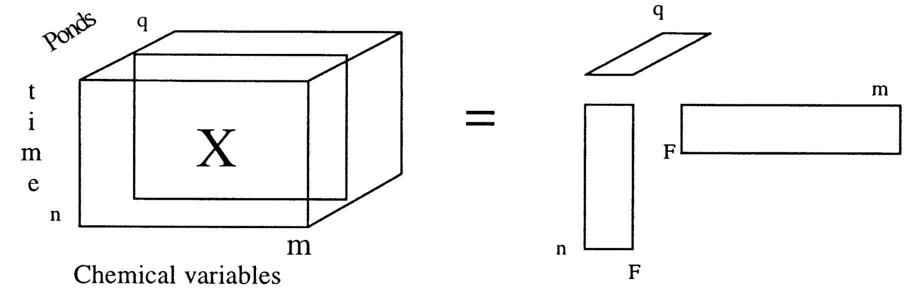


Fig. 6. Variable contributions for T^2-values: (a) derived from PCA model for the data on influent and pond samples: — influent; - - pond 1; · · · pond 2. Only BOD5 is presented, TSS, TOC and organic-N behave identically. (b) Derived from PCA model for the data on pond samples: — pond 1; - - pond 4; --- pond 7; · · · pond 8; -●- pond 9.

pond 1. Additionally, organic-N is converted to ammonia by ammonification reactions. The removal of organics continues to a smaller extent in pond 2. Once BOD_5 drops to a level where the bacterial demand for oxygen falls below the amount of oxygen diffusing into the pond from the air, the DO level starts to rise. There are periods in data a where the



$$X_{ijk} = \sum_{f=1}^{F} a_{if} b_{jf} c_{kf} + e_{ijk}$$

Fig. 7. Three-way data array and data decomposition in PARAFAC. A decomposition of the data is made into trilinear components or factors A, B, and C with elements a_{if}, b_{jf}, and c_{kf}, respectively. The trilinear model is found to minimize the sum of squares of the residuals, e_{ijk} in the model. A, B and C represent loading matrices of row-, column- and layer-items of X, respectively.

DO level starts to increase in pond 4 but as can be seen in Fig. 6(b), the major increase in DO concentration occurs in pond 7 (which at this time received the outflow from pond 4).

As nitrite is only an intermediate product in nitrification, it tends not to accumulate in systems, except when conditions are marginal for nitrification and the more resilient bacteria that oxidise ammonia to nitrite outperform the bacteria that oxidise nitrite to nitrate. The small rise in nitrite concentrations in pond 2 suggests that conditions here are just tolerable for ammonia-oxidisers but not for nitrite-oxidisers, while the small rise in nitrate concentrations in pond 3 suggests that conditions here are more acceptable to the nitrite-oxidisers. The conditions in these ponds contrast with those in the seventh and subsequent ponds. Here the

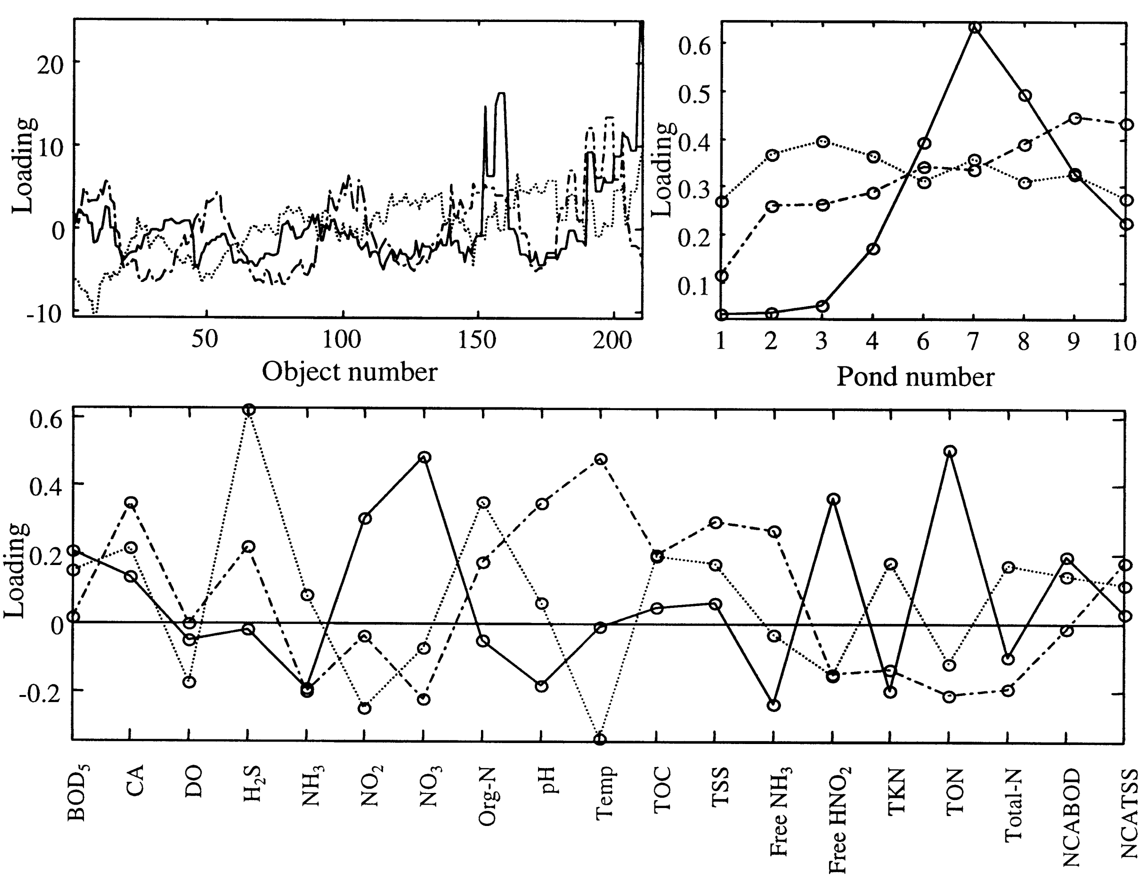


Fig. 8. PARAFAC loadings for the data on samples taken during a period of 4 1/2 years between 1990 and 1994. The first factor (solid line) describes nitrification, the second (dashed line) seasonal variation and the third (dotted line) a trend over time due to weakening process conditions during the examined period of time.

concentration of DO increases and the high ammonia removal rates indicate that conditions are highly favourable for ammonia-oxidisers. They are also favourable for nitrite-oxidisers, with the result that nitrite concentrations start to fall, while nitrate concentrations continue to build through ponds 8–10. Nitrate is itself a substrate for denitrifying bacteria but the accumulation of nitrate in these later ponds suggests that denitrification rates are lower than nitrification rates. This is not unexpected as in these highly aerobic ponds regions sufficiently anoxic for denitrification to occur are found only in the pond sediments.

3.3. Process profiling by three-way modelling

Parallel factor analysis (PARAFAC) is a decomposition method, which conceptually can be compared to bilinear PCA, or rather should be regarded as one generalization of bilinear PCA [6]. In the three-way case, the data are arranged in a cube instead of a matrix as in standard multivariate data sets. A decomposition of the data is made into trilinear components or factors, but instead of one score vector and one loading vector as in bilinear PCA, each component consists of three loading vectors (Fig. 7). The three-way PARAFAC technique was applied to the samples collected after the installation of the aerators, during a 4 1/2 year period between 1990 and 1994. The first mode in the three-way array (Fig. 7) is time (objects), the second is the 19 chemical and physical variables, and the third is the ponds. The appropriate number of components or factors was found to be three.

PARAFAC loadings are presented in Fig. 8. The first factor describes nitrification and denitrification. The largest loadings over the chemical variables are for NO_2^- and NO_3^-, which correlate negatively with ammonia and positively with BOD_5. Nitrate has a slightly larger loading than nitrite presumably due to its tendency to accumulate in the system before it is converted to nitrogen gas by denitrification reactions (see Section 3.2). The loadings over the ponds suggest that nitrification starts in pond 2, then accelerates and reaches a maximum in pond 7, which ties in well with the observations made in the previous section.

The loadings over time for the second factor are characterised by a cycle of around 50 weeks, referrable to seasonal variation. The loadings over chemical variables reveal that the levels of TSS, TOC and organic-N are highest in summer, which ties in well with the observations made in reference [2] on pollutant levels in sewage. The level of nitrate is lowest in summer, due presumably to more efficient denitrification in the summer season. The loadings over ponds show how the influence of temperature/seasonal variation becomes more dominant as purification proceeds and the system becomes less complex (see Section 3.2).

The loadings over time for the third factor reveal a trend over time. The loadings over chemical and physical variables indicate that the trend is a result of decreasing levels of DO and increasing levels of BOD_5, TOC, TSS and H_2S, especially in ponds 2 and 3. The third factor also embodies seasonal variation, and the loadings over chemical variables reveal, not surprisingly, that dissolved oxygen deficiency occurs particularly in winter.

4. Conclusions

Multivariate data analysis methods, PCA and PARAFAC, proved to be efficient tools for profiling and visualising the multivariate, complex biological wastewater treatment system. The successive stages in the purification process could be identified and the ponds where particular reactions mostly occurred could be fairly well located. The effect of changes in process conditions could be discerned. Dissolved oxygen concentration was found to be a key process parameter with a marked influence on treatment efficiency.

The results obtained using PCA and PARAFAC were consistent and the two methods complemented each other well, with both having certain advantageous features for illustrating and visualising the lagoon process. As these multivariate methods proved to be applicable and sensitive tools for analysing wastewater treatment processes, a natural extension of the study was to apply multivariate statistical process control (MSPC) to the data. Experiments using MPSC have already been undertaken, with interesting results (to be published later).

Acknowledgements

The authors wish to thank Melbourne Water for making available the data used in this study and Mr. B. McLean for his help in providing access to the data.

References

[1] M. Pantsar-Kallio, S.-P. Mujunen, G. Hatzimihalis, P. Koutoufides, P. Minkkinen, P.J. Wilkie, M.A. Connor, Anal. Chim. Acta 393 (1999) 181–191.
[2] D. Orhon, N. Artan, Modelling of Activated Sludge Systems, Technomic Publishing, USA, 1994.
[3] J. Wanner, Activated Sludge Bulking and Foaming Control, Technomic Publishing, USA, 1994.
[4] D.L. Massart, B.G.M. Vandeginste, L.M.C. Buydens, S. De Jong, P.J. Lewi, J. Smeyers-Verbeke, Handbook of Chemometrics and Qualimetrics, Elsevier, Netherlands, 1997.
[5] J.E. Jackson, A User's guide to Principal Components, Wiley, New York, 1991, pp. 20–23, 36–38, 342–344.
[6] R. Bro, PARAFAC: tutorial and applications, Chemom. Intell. Lab. Syst. 38 (1997) 149–171.
[7] C.A. Andersson, R. Bro, The N-way toolbox for MATLAB, Chemom. Intell. Lab. Syst. 52 (1) (2000) 1–4.