Chemosphere 155 (2016) 152–162

Contents lists available at ScienceDirect

Chemosphere

journal homepage: www.elsevier.com/locate/chemosphere

Principal factor and hierarchical cluster analyses for the performance assessment of an urban wastewater treatment plant in the Southeast of Spain

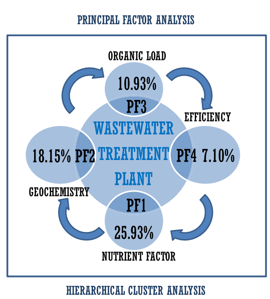
Javier Bayo [a, *], Joaquín López-Castellanos [b]

[a] Depart. of Chemical and Environmental Engineering, Technical University of Cartagena, Paseo Alfonso XIII, 44 E-30203 Cartagena, Spain
[b] Hidrogea, Capitanes Ripoll, 8 E-30203 Cartagena, Spain

HIGHLIGHTS

- The inherent structure of a water quality dataset is proposed.
- The geochemical background of water proved to positively score with the decrease of anthropogenic activity.
- The organic loading factor proved to be well-represented during mid-week.
- Structural improvements in the plant were observed through the efficiency principal factor.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 29 July 2015
Received in revised form
7 March 2016
Accepted 10 April 2016
Available online 22 April 2016

Handling Editor: Jörg E. Drewes

Keywords:
Principal factor analysis
Hierarchical cluster analysis
Wastewater
Chemometrics

ABSTRACT

Process performance and operation of wastewater treatment plants (WWTP) are carried out to ensure their compliance with legislative requirements imposed by European Union. Because a high amount of variables are daily measured, a coherent and structured approach of such a system is required to understand its inherent behavior and performance efficiency. In this sense, both principal factor analysis (PFA) and hierarchical cluster analysis (HCA) are multivariate techniques that have been widely applied to extract and structure information for different purposes. In this paper, both statistical tools are applied in an urban WWTP situated in the Southeast of Spain, a zone with special characteristics related to the geochemical background composition of water and an important use of fertilizers. Four main factors were extracted in association with nutrients, the ionic component, the organic load to the WWTP, and the efficiency of the whole process. HCA allowed distinguish between influent and effluent parameters, although a deeper examination resulted in a dendrogram with groupings similar to those previously reported for PFA.

1. Introduction

A continuous monitoring of the effluent of wastewater treatment plants (WWTP) is always required, in order to ensure smooth operation, adjustment to prescribed legislative requirements and safety of the receiving water bodies (Tchobanoglous et al., 2003). The efficiency of the sewage treatment plant and the quality of the effluent are essential to fulfill the stringent effluent discharge standards (Directive 91/271/CEE), especially for nitrogen and phosphorus, both identified as the two

* Corresponding author.
E-mail address: javier.bayo@upct.es (J. Bayo).

http://dx.doi.org/10.1016/j.chemosphere.2016.04.038
0045-6535/

main nutrients involved in aquatic eutrophication (Zuthi et al., 2013). For this purpose, physicochemical parameters are widely used as a reliable assessment procedure which allows us to check the quality of the water influent and effluent to the wastewater treatment plant. However, due to the large amount of data yearly available, the great variability of processes occurring in the plant and temporal variations due to seasonal changes, it is difficult to assess the overall performance with simple univariate statistical methods such as analysis of variance (ANOVA). The special features of monitoring water quality parameters, including spatial and temporal variations, make them difficult to interpret in a holistic manner, avoiding reach important conclusions that help to understand the process as a whole.

In recent years, multivariate statistical methods known as chemometric techniques, including principal factor analysis (PFA) and hierarchical cluster analysis (HCA), have been used in the water field, to investigate the complex structure that underlies many analytical data, reducing the datasets to a few significant variables, factors or components, without losing significant information (Olawoyin et al., 2014). In particular, PFA is a powerful data reduction technique that was conducted to transform a huge amount of water quality variables and data into a manageable set of uncorrelated factors that explain most of the variance observed, allowing us to generate hypothesis about the identified principal components and their significance. HCA nicely complements PFA, grouping and discovering latent intrinsic structures or underlying behavior of data (Vialle et al., 2011).

Multivariate techniques have been widely applied to grouping water quality variables for different purposes; i.e., the composition of sewage of domestic origin (Singh et al., 2005), surface water (Hussain et al., 2008; Zhang et al., 2010), water quality from roof runoff (Vialle et al., 2011), river watersheds (Olsen et al., 2012), distribution of organic pollutants (Saba and Su, 2013; Zhang et al., 2013), heavy metal distribution in waters (Bingöl et al., 2013), new policy instruments in domestic wastewater treatment systems (Naughton and Hynds, 2014), mineral waters (Astel et al., 2014), and wastewater treatment plants (Aguado and Rosen, 2008; Ouali et al., 2009; Platikanov et al., 2014), among others.

This paper presents the results of both techniques, PFA and HCA, applied to water quality parameters from an urban wastewater treatment plant situated in the Region of Murcia (Southeast of Spain), to better characterize the associations present in the entire dataset in a three-year period (2010–2012). The final objective was to investigate a data reduction that allows grouping water quality variables into selected factors with common features, in order to describe the behavior of the plant in terms of pollution sources and treatment efficiency with a few uncorrelated components.

2. Materials and methods

2.1. Sample collection and dataset

A maximum of 1085 and a minimum of 155 wastewater samples were measured for different parameters during 2010, 2011, and 2012, from the influent (after primary treatment), the activated sludge reactor, the recirculation, and the effluent of the urban WWTP “Cabezo Beaza” situated in Cartagena (Spain). This plant consists of a conventional system with a primary treatment and two activated sludge reactors, serving about 140000 equivalent inhabitants. Raw wastewater arrives into the plant through a security spillway of 6 m length, undergoing screening and sieving for the removal of coarse and fine solids. After that, there are two grit and grease chambers of 250 m^3 for each unit with two compressed-air powered pumps of 650 $Nm^3\ h^{-1}$ each one. The primary settling consists of two circular units of 30 × 3 m, with two 40 $m^3\ h^{-1}$ submersible primary sludge pumps. The biological reactor consists of an anaerobic tank for pre-denitrification followed by a facultative tank, both completely mixed and with a volume of 2025 m^3 each, ahead of the plug-flow aerobic reactor with a total net volume of 12,150 m^3 and 1008 fine bubble membrane diffusers. Biological treatment ends in two circular secondary clarifiers of 40 × 3.6 m each unit. Before discharging the effluent, water is disinfected with chlorine in a 384 m^3 chlorination labyrinth.

The following water quality parameters were included in the dataset: water flow [WF] ($m^3\ d^{-1}$), pH (pH units), water temperature [WT] (°C), electrical conductivity [EC] (mS cm^{-1}), suspended solids [SS] (mg l^{-1}), mixed liquor suspended solids [MLSS] (mg l^{-1}), mixed liquor volatile suspended solids [MLVSS] (mg l^{-1}), volume of the settled sludge after 30 minutes sedimentation [V30] (ml l^{-1}), sludge volume index [SVI] (ml g^{-1}), chemical oxygen demand [COD] (mg l^{-1}), biochemical oxygen demand [BOD] (mg l^{-1}), ammonium nitrogen [NH_4–N] (mg l^{-1}), nitrate nitrogen [NO_3–N] (mg l^{-1}) and phosphate phosphorus [PO_4–P] (mg l^{-1}). Acronyms were used with sub-indexes to properly indicate influent to the biological process (I), biological reactor (BR), recirculation (R) or effluent (E) from the WWTP. EC was determined with a Crison GLP 32 conductimeter (Barcelona, Spain), weekly calibrated with standard solutions of 1413 ± 12 μS cm^{-1} and 12.88 ± 0.11 mS cm^{-1}, and pH with a Crison GLP 22 pH meter (Barcelona, Spain), calibrated by means of standard solutions of pH 4.01 ± 0.002, pH 7.00 ± 0.002, and pH 9.25 ± 0.002. Values of both parameters are verified within a control chart according to specific calibration instructions. COD was determined using the Spectroquant NOVA 30 from Merck (Darmstadt, Germany), calibrated on a quarterly basis with a 100 mg l^{-1} COD solution, stable for one month. Oxytop respirometers (WTW, Weiheim, Germany) were used for BOD_5 analysis, using a standard solution of glucose and glutamic acid, with a BOD_5 of 198 ± 30.5 mg l^{-1}, measured by inter-laboratory tests and considered a limit control for individual laboratories. All the other parameters were calculated according to the Standards Methods for the Examination of Water and Wastewater (APHA, 2012). During 2010 and 2011, the laboratory participated in two laboratory and sampling intercomparison studies for all the previous parameters.

2.2. Statistical analysis of experimental data

Statistical treatment of data was carried out with the SPSS (Statistical Package for Social Sciences) 22.0 software. Pearson's correlation coefficient (r) was computed between different parameters in order to determine the extent to which values of both parameters were correlated. Tukey-Kramer method was the multiple comparisons procedure used for the simultaneous estimation of pairwise differences of means in one-way ANOVA, as recommended by Storline (1981). Besides, a general linear model for repeated measurements with estimated marginal means test and the fitting performance of analysis of variance by means of *F-Senedecor* value were used. Table 1 depicts minimum, maximum, mean, median and standard deviation corresponding to the measured parameters in the wastewater treatment plant, avoiding values under the quantification limit for each analytical technique.

2.3. Multivariate data analysis

Principal factor analysis (PFA) was carried out to quantify the significance of different variables in the dataset (Zhang et al., 2010). It is a multivariate statistical technique widely used in the environmetrics to reduce correlated data into a few uncorrelated components or factors explaining the maximum variance (Vialle et al., 2011). Principal factors or components (PFs) were calculated based on a correlation matrix; in our case, varimax orthogonal