Contents lists available at ScienceDirect

Process Safety and Environmental Protection

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

Fish canning industry wastewater variability assessment using multivariate statistical methods

Raquel O. Cristóvão[a,*], **Victor M.S. Pinto**[b], **António Gonçalves**[c], **Ramiro J.E. Martins**[a,b], **José M. Loureiro**[a], **Rui A.R. Boaventura**[a]

[a] Laboratory of Separation and Reaction Engineering (LSRE), Associate Laboratory LSRE/LCM, Department of Chemical Engineering, Faculty of Engineering, University of Porto, Rua do Dr. Roberto Frias, 4200-465 Porto, Portugal

[b] Department of Chemical and Biological Technology, Superior School of Technology, Polytechnic Institute of Bragança, Campus de Santa Apolónia, 5301-857 Bragança, Portugal

[c] CRACS/INESC-TEC and Department of Computer Science, Faculty of Sciences, University of Porto, Rua do Campo Alegre 1021/1055, 4169-007 Porto, Portugal

ARTICLE INFO

Article history:
Received 26 October 2015
Received in revised form 18 March 2016
Accepted 24 March 2016
Available online 1 April 2016

Keywords:
Fish canning wastewater
Wastewater variability
Principal component analysis
Cluster analysis
Correlation analysis
Multivariate statistical methods

ABSTRACT

Usually, fish canning industrial wastewaters have a highly variable composition over time. For a good performance of treatment processes it is necessary to limit that variation. However, extended wastewater monitoring, including all relevant analytical parameters, is expensive. This work proposes an efficient approach to minimize the analytical determinations number without compromising the global characterization goal. This way, fish canning industry wastewaters variability was assessed and interpreted through multivariate statistical tools application to analytical data obtained from a monitoring program carried out in a fish canning industry of northern Portugal. 23 physicochemical parameters were determined in 20 samples collected on an 8 months period. The results achieved by correlation analysis, principal component analysis (PCA) and cluster analysis (CA) led to the main water pollution sources identification and to the minimization of physical and chemical parameters number to be analyzed in order to achieve a correct wastewater characterization, at minimum cost. The main pollution sources proved to be the brine and eviscerating step waters. Dissolved organic carbon (DOC), total suspended solids (TSS), conductivity, pH, Ca^{2+}, F^- and one of the parameters $SO_4{}^2$, $NO_3{}^-$ and $PO_4{}^{3-}$ were identified as important parameters that must be monitored in order to obtain an accurate characterization allowing to define the most appropriate wastewater treatment.

1. Introduction

In recent years, there has been a rapid growth of commercial fish markets and industries across the world. Countries with rapid population, income and urbanization growths tend to have the greatest increases of fish products consumption (Delgado et al., 2003). The volume and concentration of wastewater produced by fish canning industries is highly variable, depending on the production season, fish type that is being processed, additives used, processing water source and on the unit processes implemented (Chowdhury et al., 2010). Each plant is unique, so generalizations about water use and wastewater characteristics are difficult.

The treatment of these effluents is complex due to the presence of high content of organic matter, oil and grease and also due to the high NaCl concentration that they normally present (Cristóvão et al., 2015; Gharsallah et al., 2002). Currently most of fish canning industries in Portugal only perform

* Corresponding author. Tel.: +351 22 508 2263; fax: +351 22 508 1674.
E-mail address: raquel.cristovao@fe.up.pt (R.O. Cristóvão).
http://dx.doi.org/10.1016/j.psep.2016.03.016

a pre-treatment of their wastewaters, usually by screening, filtration and/or decanting to remove coarse particles (Cristóvão et al., 2014). However, there is a need to consider the treatment of these wastewaters in order to fulfil the limits imposed by the Portuguese legislation (Decree-Law No. 236/98) for industrial wastewater discharge. The overall treatment efficiency varies according to wastewater characteristics and with the technologies applied. Since the wastewaters from fish canning industrial processes are known to have a high variability, there is a need to know their characteristics in detail in order to decide the best treatment sequence to apply. In fact, for a good performance of certain treatment processes (i.e., to obtain a good efficiency) it is necessary to limit wastewater variation. The usual manner of achieving this is to install a homogenizing tank upstream from the treatment system. For the homogenization tank design it is essential to know the wastewater analytical parameters with higher variability (or that contribute most to the overall variability of the effluent). Knowing these parameters it is possible, for a given level of probability, to calculate the residence time in the homogenization tank (and, therefore, the volume, assuming an approximately constant flow rate) so that the concentrations of the output parameters are within a predefined range. This could be achieved with the design of monitoring programs for collection of different wastewater samples before being launched to the wastewater treatment plant (WWTP) and subsequent characterization of a large number of physicochemical parameters to provide representative and reliable wastewater quality parameters. However, these programs are expensive and produce large data sets which are often difficult to analyze and interpret. In these cases, the use of multivariate statistical analysis methods is appealing.

Multivariate statistical methods are useful for the interpretation of large and complex water and wastewater quality data sets, evaluating redundant measurements in the environment, allowing, this way, the classification and the grouping of pollutants according to their sources, achieving a small number of underlying factors without losing too much information (Singh et al., 2005). Correlation analysis is a well-known statistical method to assess the relationships between parameters. The resulting value, the correlation coefficient, can range between ±1 and shows if the variation of one variable is correlated to the variation of other variable. The closer the correlation is to +1 or −1, the closer it is to a perfect relationship (Babu et al., 2014). Although useful, sometimes correlation analysis leads to a large number of variables that are difficult to examine and the correlations between the variables can be better observed and qualitatively visualized using cluster analysis. CA is a classification method used to split a data set into a number of groups of observations that share observed properties and are distinct from each other in terms of variables values (Paoletti et al., 2002). There are different clustering techniques, but hierarchical agglomerative clustering is the most important and widely used. In clustering, the objects are grouped such that similar objects fall into the same class (cluster). The hierarchical agglomerative clustering is based on distances between clusters. Given an initial cluster, first the two clusters that are nearest are merged to form a new cluster. This is repeated each time merging the two closest clusters, until just one cluster, of all the data points, exists. The levels of similarity at which observations are merged are used to construct a dendrogram (Hand et al., 2001; Magyar et al., 2013). A third data analysis method, principal components analysis, can also be used to explore the relationships among several samples, being at the same time, a variable reduction procedure. PCA is a statistical technique that transforms the original set of inter-correlated variables into a new set with a small number of independent uncorrelated variables or principal components (PCs) that are linear combinations of the original variables and account for most of the variance in the observed variables. This way, principal components do not present multicollinearity probably present in original variables (Hatcher and Stepansku, 1994; Song et al., 2006). The aim of this technique is to capture the intrinsic variability in the data and to identify groups of variables based on the loadings (the weight by which each standardized original variable should be multiplied to get the component score), i.e., correlations between the variables and the principal components (Boruvka et al., 2005). Loadings show how well a variable is taken into account by the model components and can be used to understand how much each variable contributes to the meaningful variation in the data and to interpret variable relationships. Loadings are also useful for interpreting the meaning of each component. This is a useful way of reducing the dimensionality of a data set, either to ease interpretation or as a way to avoid overfitting and to prepare for subsequent analysis (Hand et al., 2001).

Using correlation analysis with CA and PCA provides more information than using each method alone. These multivariate statistical analyses can be efficient tools for evaluating water/wastewater quality and also for identifying latent sources that influence their characteristics, reducing the dimensionality of a data set and offering a valuable tool for reliable management of water resources, as well as effective solutions to pollution problems (Magyar et al., 2013; Lee et al., 2008; Ouali et al., 2009; Wan et al., 2011; Wang et al., 2013; Yoo et al., 2003; Zhao et al., 2012).

In this work, correlation analysis, PCA and CA were performed to analyze data from twenty different wastewater samples resulting from a sampling program carried out in a fish canning industry of northern Portugal, from November 2013 to June 2014, aiming at assessing linear relationships between wastewater characterization parameters and to evaluate main wastewater pollution sources in order to optimize and reduce the number of monitoring parameters (redundant and correlated information), preserving the maximum of information whereas minimizing the analysis costs.

2. Materials and methods

2.1. *Fish canning industry process*

In a traditional fish canning production process, the generated wastewaters come essentially from (i) the fish thawing stage (when exists), (ii) the brine stage, (iii) eviscerating, baking, cooling, canning and washing of fish and cans and (iv) the clean-up of washing areas. In Fig. 1 is presented, as an example, the production process flowchart of the cannery under study (from northern Portugal), where it is shown the steps where water is used and the main sources of wastewater generated, together with the most important contaminants, to help understand the variability of these wastewaters and the results obtained from multivariate statistical analysis. The detailed fish canning production process, as well as the corresponding flowchart presented in Fig. 1 are described in a previous published work of our group (Cristóvão et al., 2012).