Biochemical Engineering Journal 111 (2016) 43–53

Contents lists available at ScienceDirect

Biochemical Engineering Journal

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

Regular article

Managing the touristic pressure: performances prediction of an advanced biological system by means of regression trees

E. Barca, G. Del Moro*, M. De Sanctis, G. Mascolo, G. Passarella, C. Di Iaconi

Istituto di Ricerca Sulle Acque, Consiglio Nazionale delle Ricerche, Viale F. De Blasio 5, 70132 Bari, Italy

ARTICLE INFO

Article history:
Received 12 November 2015
Received in revised form 12 February 2016
Accepted 12 March 2016
Available online 15 March 2016

Keywords:
Fixed-bed bioreactors
Modelling
Waste-water treatment
Control
Regression trees
Touristic pressure

ABSTRACT

A regression tree model has been used to make predictions of six gross parameters (COD, COD_{sol}, $N\text{–}NH_4^+$, TN, P_{tot} and TSS) of an innovative SBBGR reactor. R^2 values ranging from 0.94 to 0.97 were found for ammonia and total phosphorus, respectively. This application showed its usefulness as a decision support system for wastewater treatment plants in tourist areas which typically operate under high stress conditions due to the sharp fluctuations of wastewater flow and composition. A forecast of the bioreactor's performance would help the plant manager to put in place the required practices and procedures. The regression tree model could be part of the automation and control system of the SBBGR plant, allowing the change of operating conditions to be carried out automatically and in an effective way to face the touristic stress issue.

1. Introduction

Tourism is the largest and fastest-growing economic sector in the world. Globally, tourism and related economic activities generate 11% of Global Domestic Product, employ 200 million people, and transport nearly 700 million international travellers per year. These numbers are expected to double by 2020, especially in some of the world's least developed countries. Massive influxes of tourists, often to a relatively small area, have a huge impact. They add to the pollution, waste, and water needs of the local population, putting local infrastructure and habitats under enormous pressure [1]. Monitoring has shown that wastewater is one of the main sources of pollution of seas and lakes surrounding tourist attractions, damaging the flora and fauna. Sewage pollution can threaten human and animal health [2].

Because of the above, tourism requires a high standard of water supply and wastewater services. In fact, the development of tourism facilities such as accommodation, water supplies, restaurants and recreation facilities can cause an increase in the extent of wastewater pollutant load and flow. Sound infrastructure engineering can contribute significantly to improved and more sustainable water and wastewater services and more efficient use of resources [3]. Currently, the main managerial solution for the wastewater treatment in tourist areas is entrusted to the decentralization of treatment plants [4]. Usually, the wastewater treatment plant employed is based on extended aeration activated sludge processes. Unfortunately, these systems suffer the shock loads, sludge bulking, and absence of regular supervision and maintenance which are typical of tourist areas. As a result, the effect is a poor effluent quality [5]. To sidestep the above-mentioned problems, more reliable and robust wastewater treatment technologies with better operational flexibility should be provided in tourist areas. The Water Research Institute (IRSA) of the National Research Council of Italy (CNR) recently developed a new system whose acronym is SBBGR (Sequencing Batch Bio-filter Granular Reactor). The SBBGR system combines the advantages of attached biomass systems (i.e., greater robustness and compactness) with those of periodic systems (i.e., greater flexibility and stability). It is a unique system in virtue of the particular type of biomass growing in it (a mixture of biofilm and granules packed in a filling material) which allows a greater retention of the biomass in the reactor to be obtained [6,7]. Consequently, the SBBGR reactor is the technical solution, at the state of the art, that is most appropriate for addressing tourist area wastewater treatment [8]. However, even these SBBGR reactors are usually designed on average flow rates and wastewater composition, often disregarding peak conditions [9]. As an example, during the longer-term shock (lasting days to weeks) induced by touristic pressure, the biomass requires a longer time to achieve a new steady state. This instability can result in an effluent quality which exceeds the discharge limits. Furthermore, due to the

* Corresponding author.
E-mail address: guido.delmoro@ba.irsa.cnr.it (G.D. Moro).

http://dx.doi.org/10.1016/j.bej.2016.03.005
1369-703X/

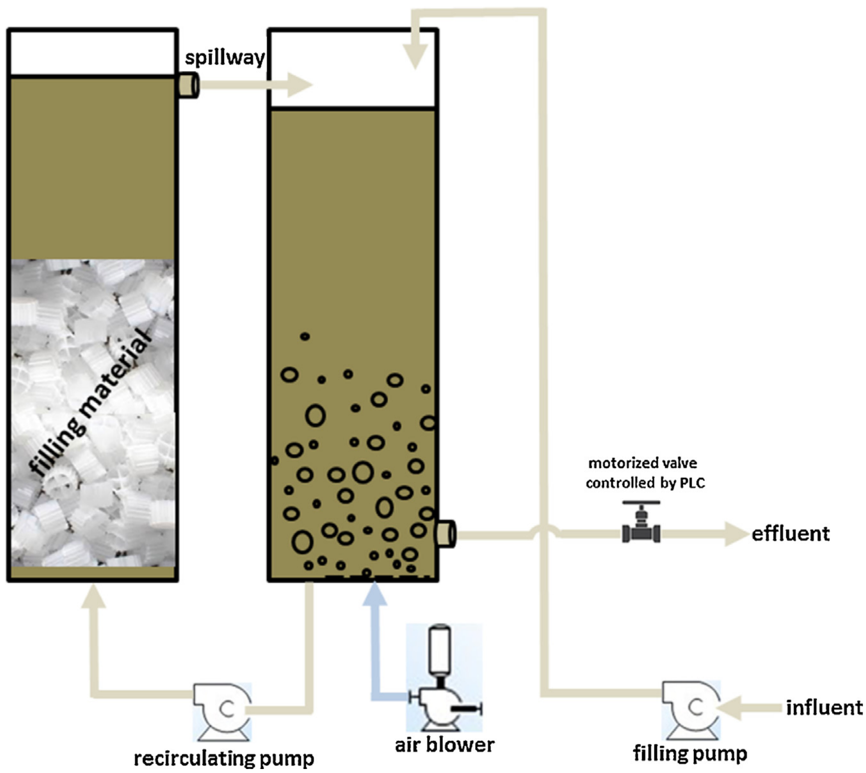

Fig. 1. Sketch of the SBBGR (Sequencing Batch Biofilter Granular Reactor) pilot plant used in the investigation.

deficiencies of the sewerage system, which makes the hydrolysis process of particulate matter less extensive, considerable variations in dissolved organic matter content can be observed during wastewater flow variations. This can greatly affect the performance of the nitrification and denitrification processes (in addition to the carbon removal), with the consequent possible failure to achieve discharge limits. Therefore, it would be of great value to be able to predict, with good reliability, the system performance during these shocks, and thus to be able to make the necessary operating changes.

The problem of predicting sewage treatment plant behaviour, even if characterized by different technologies from the object of the present study, has been addressed by different authors, each proposing a different approach. Some approach the problem through physically based models [10–16] while others approach it using data-driven grey and black box models [17–24]. From the scientific literature it can be concluded that the relationship between the gross parameters before and after treatment is of a non-linear kind [19,20,22,24,25]. A statistical model that can potentially take into account the non-linearity of the phenomenon is the regression tree. The term Classification and Regression Tree (CART) analysis is a broad term first introduced by Breiman et al. [26]. When the data has many features which interact in complicated, non-linear ways, assembling a single global model can be very difficult and quite confusing when it succeeds. A regression tree sub-divides or partitions the intervals into smaller parts, where the interactions between the involved parameters are more manageable. The partition is repeated again to obtain sub-divisions (this is called recursive partitioning) until finally the data intervals are arranged in a way which allows one to fit simple models to them [27]. The algorithm just described is susceptible to a graphical representation as a tree, hence the name of the technique. The initial dataset corresponds to the tree root and the leaves are sub-intervals resulting from the subdivisions made. Each of the terminal nodes, or leaves, of the tree represents a cell of the partition, and has attached to it a simple model which applies in that cell only [28].

Despite its application in various scientific fields, there are, so far, few attempts to apply the CART technique to predicting the behaviour of gross parameters for wastewater treatment plants. To the best of the authors' knowledge, there are just three works in the literature about applying a regression tree to model a wastewater treatment plant. However, these papers refer to large municipal plants that are significantly different from the studied reactor. Only the work of Atanasova and Kompare [29] reports an attempt to model some of the gross parameters considered in the present paper. On the other hand, Kusiak et al. [30] deal with a single

Table 1
Characterization of studied parameters related to the training phase. All parameters are expressed in mg/L except for kurtosis and skewness which are dimensionless.

	Influent						effluent					
	COD	COD_{sol}	$N–NH_4^+$	TN	P_{tot}	TSS	COD	COD_{sol}	$N-NH_4^+$	TN	P_{tot}	TSS
Mean	651.6	183.1	58.9	80.1	6.4	324.1	42.6	29.5	2.9	26.3	5.1	10.2
Std error	66.1	15.9	3	4.7	0.1	42.2	2.4	1.8	0.6	2.1	0.2	0.7
Std dev	540.4	133.3	26.1	34.2	1.5	298.4	20	15.2	5.2	15.3	1.7	5.3
Kurtosis	3.5	3.5	0.4	−0.8	10.8	2.6	1	1.8	16.2	2.5	5.5	0.7
Skewness	1.9	1.8	0.9	0.4	1.6	1.7	1.2	1.1	3.5	1.1	2.6	0.9
Min	134	19	15.4	26.4	1.3	57	17	0	0	2.8	1.5	2.5
Max	2640	651	131.6	152.8	13	1302	101	78	32.6	84	19.4	27