Available online at www.sciencedirect.com
ScienceDirect
journal homepage: www.elsevier.com/locate/issn/15375110

Research Paper

# Pressure drop modelling in sand filters in micro-irrigation using gradient boosted regression trees

*Paulino J. García Nieto* [a,*], *Esperanza García–Gonzalo* [a], *Gerard Arbat* [b], *Miquel Duran–Ros* [b], *Francisco Ramírez de Cartagena* [b], *Jaume Puig-Bargués* [b]

[a] *Department of Mathematics, Faculty of Sciences, University of Oviedo, 33007, Oviedo, Spain*
[b] *Department of Chemical and Agricultural Engineering and Technology, University of Girona, 17071, Girona, Catalonia, Spain*

ARTICLE INFO

*Article history:*
Received 7 February 2018
Received in revised form 24 March 2018
Accepted 17 April 2018

*Keywords:*
Regression trees
Gradient boosting
Differential evolution
Drip irrigation
Sand filters

Filters are essential for guaranteeing the good performance of microirrigation systems. Pressure losses across filters should be known for the proper design and management of this irrigation equipment. Pressure losses produced by filtering media in sand filters can be computed using Ergun or Kozeny–Karman equations, which require knowledge, among other parameters, of the sphericity of the filter medium. As this parameter is not easy to determine, it is useful to explore the performance of alternative computing methods that can avoid requiring knowledge of sphericity. In this paper, taking as starting point the nonparametric machine learning approach known as the gradient boosted regression tree (GBRT) approach and hybridising it with the differential evolution (DE) technique, the pressure drop in sand filters used in microirrigation has been modelled. For different filtering materials such as modified glass, crushed glass, silica sand and glass microspheres, experimental data of pressure drop for velocities between 0.004 and 0.025 m $s^{-1}$ was collected and the model built. The results demonstrated that DE–GBRT–based model was able to accurately predict pressure drop. The model also allowed ranking of the importance of the independent variables examined within the model. Taking into account this ranking, and using only the main variables, a simplified method with an improved coefficient of determination was constructed.


## 1. Introduction

Proper irrigation water filtration is essential to ensure the successful continuous long-term operation of microirrigation systems (Clark, Haman, Prochaska, & Yitayew, 2007). By following good maintenance practises, which includes filtration, the longevity of some subsurface microirrigation systems have reached 26.5 years (Lamm & Rogers, 2017). Screen, disc, media and hydro-cyclone filters are common filter types that are used in microirrigation systems. The choice of filter

* *Corresponding author*. Fax: +34 985103354.
E-mail address: lato@orion.ciencias.uniovi.es (P.J. García Nieto).
https://doi.org/10.1016/j.biosystemseng.2018.04.011
1537-5110/

**Nomenclature**

| | |
|---|---|
| $ABC$ | Artificial bee colony |
| $b_{jm}$ | Constant value calculated for the region $R_{jm}$ |
| $C_o$ | Cover of the GBRT algorithm |
| CART | Classification and Regression Trees |
| $D_{eq}$ | Equivalent diameter, m |
| $DE$ | Differential evolution |
| $F_m$ | Weak model that predicts the mean y of the training set |
| $F_q$ | Frequency of the GBRT algorithm |
| $F_0(x)$ | Constant function |
| $\widehat{F}(x)$ | an estimate of the function $F^*(x)$ |
| $G_a$ | Gain of the GBRT algorithm |
| GA | Genetic algorithm |
| GBRT | Gradient boosted regression tree |
| GR | Parameter that controls the recombination rate |
| $H$ | Set of arbitrary differentiable functions |
| $h$ | Weak learner function |
| $h_i(x)$ | Weighted sum of functions |
| $h_m(x)$ | Decision tree |
| $J_m$ | Number of terminal nodes in the tree model |
| NP | Noisy random vectors |
| $L()$ | Loss function |
| $m$ | Weighted medium mass, kg |
| $m_{og}$ | Overall mass of the grains, kg |
| MARS | multivariate adaptive regression splines |
| MCW | Minimum child weight of GBRT algorithm |
| MDS | Minimum delta step of GBRT algorithm |
| $n$ | Number of observed data |
| $N$ | Number of grains |
| $Nrounds$ | Maximum number of iterations of the GBRT algorithm |
| $p$ | Index of the individual in the population |
| PSO | Particle swarm optimization |
| $r_{im}$ | Pseudo-residuals |
| RMSE | Root mean square error |
| $R^2$ | Coefficient of determination |
| SR | Subsample ratio of the GBRT algorithm |
| $SS_{tot}$ | Total sum of squares |
| $SS_{reg}$ | Regression sum of squares |
| $SS_{err}$ | Residual sum of squares |
| $\mathbf{t}_m^g$ | Trial vectors |
| $V_m$ | Medium volume, $m^3$ |
| $V_w$ | Volume of the additional water, $m^3$ |
| $V_f$ | Final volume of the water and medium mixture, $m^3$ |
| $V$ | Mean flow velocity, m $s^{-1}$ |
| $\mathbf{x}_p^g$ | Original vectors |
| $(\Delta p/\Delta L)$ | Pressure drop per unit length |
| $\Delta$ | Step value over each tree's weight estimation |
| $\varepsilon$ | Medium porosity |
| $\eta$ | Learning rate of GBRT algorithm |
| $\rho_b$ | Bulk density of each medium, kg $m^{-3}$ |
| $\rho_r$ | Real density of each medium, kg $m^{-3}$ |
| $\phi$ | Sphericity factor |
| $\gamma$ | Minimum loss reduction of the GBRT algorithm |
| $\Omega$ | Penalty function that controls the model complexity |

type will basically depend on the quality of water source, the flow rate of the irrigation system and the desired filtered water quality for avoiding emitter clogging (Clark et al., 2007).

Irrigation engineers require knowledge of the pressure drop across the filter to properly design and manage this important system component which is related to water and energy consumption as well as pollutant removal efficiency (Duran-Ros, Puig-Bargués, Arbat, Barragán, & Ramírez de Cartagena, 2009). Mathematical models have been developed using dimensional analysis for describing pressure drops across screens (Wu, Chen, Liu, Yin, & Niu, 2014b; Zong, Zheng, Liu & Li, 2015), disc (Wu et al., 2014a; Yurdem, Demir, & Degirmencioglu, 2008), hydrocyclone (Yurdem et al., 2008) and in sand media filters (Elbana, Ramírez de Cartagena, & Puig-Bargués, 2013). These models did not consider the specific effect of the different filter components (filtration zone and auxiliary elements) on pressure loss. In sand media filters, pressure loss clearly vary across the filter media, the underdrain and diffuser platter, and the backflushing valve (Bové et al., 2015b; Burt, 2010; Mesquita, Testezlaf, & Ramirez, 2012).

Bové et al. (2015a) experimentally analysed the pressure drop across different sand and recycled glass media in a microirrigation sand filter. Although the Ergun equation showed the best prediction accuracy for predicting the pressure drop, multi linear regression equations had better performance than the Kozeny–Carman equation, which is a simplification of the Ergun equation. However, these equations require parameters defining the filter media such as equivalent diameter and sphericity which are difficult to obtain.

García-Nieto et al. (2017) used a hybrid model artificial bee colony (ABC)-multivariate adaptive regression splines (MARS) which satisfactorily computed pressure loss across filtration beds without the need for sphericity. This work suggests that other alternative methods, specifically a hybrid methodology that combines the gradient boosted regression tree (GBRT) approach with the differential evolution (DE) optimisation algorithm (Feoktistov, 2006; Price, Storn, & Lampinen, 2005; Rocca, Oliveri, & Massa, 2011; Storn & Price, 1997), could also be used to predict pressure drops in the granular filters used in microirrigation systems.

GBRT models are supervised machine learning procedures that can be used for multivariate classification and regression (Bühlmann & Hothorn, 2007; Friedman, 2002; Hastie, Tibshirani, & Friedman, 2003; Schapire, 2003; Vapnik, 1998). GBRT models build competitive, highly robust procedures that are particularly appropriate for treating not very clean data (Hastie et al., 2003). They are very flexible models that can be easily be customised for any data-driven task. They are straightforward to implement and have been very successful in data-mining and machine-learning challenges (Natekin & Knoll, 2013). One of the reasons for their success could be that tree boosting takes the bias-variance trade-off into consideration while fitting the models (Nielsen, 2016). For example, GBRT models have been