Contents lists available at ScienceDirect

Journal of Computational and Applied Mathematics

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

Forecast of the outlet turbidity and filtered volume in different microirrigation filters and filtration media by using machine learning techniques

P.J. García-Nieto [a,*], E. García-Gonzalo [a], G. Arbat [b], M. Duran-Ros [b], T. Pujol [c], J. 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, 17003 Girona, Catalonia, Spain*
[c] *Department of Mechanical Engineering and Industrial Construction, University of Girona, C/ Universitat de Girona 4, 17003 Girona, Catalonia, Spain*

ARTICLE INFO

Article history:
Received 25 May 2023
Received in revised form 11 September 2023

Keywords:
Random forest regression (RFR) technique
Differential evolution (DE)
Ridge
Lasso and elastic-net regressions
Microirrigation

ABSTRACT

Different media filters and filtration media are used to eliminate suspended particles in microirrigation and therefore prevent clogging in the emitter. The water volume by filtration cycle is a parameter related to the filter and media capacity to retain particles, while turbidity is a variable related to particles in suspension in the water. Since turbidity can be measured easily and quickly, it is commonly mentioned in recommendations for the reuse of effluents in microirrigation. There are currently no models that are reliable enough to predict the filtered volume in each irrigation filter and outlet turbidity when using different filter types and media configurations. The object of this work was to propose a model that can detect early the filtered volume and the turbidity at the outlet values. This investigation presents an effective machine learning method, the Random Forest regression (RFR) in combination with the population-inspired metaheuristic optimization algorithm, called Differential Evolution (DE), for estimating the output turbidity and the filtered volume from a dataset with 1,016 samples of distinct media filters that use reclaimed effluent. The same experimental dataset was also fitted with Elastic-net, Lasso and Ridge regression machine learning methods also in combination with DE optimizer for comparison. This optimization performs the parameter tuning in the RFR using the training dataset, which considerably improves the accuracy of the regression. To achieve this, the most relevant operation input variables are tracked and analyzed: the kind of medium and filter, filtration velocity (v), height of the filter bed (H), cycle duration and the electrical conductivity for the filter inlet (EC_i), pH_i, dissolved oxygen (DO_i), water temperature (T_i) and the input turbidity ($Turb_i$). There are two kinds of results. Firstly, the importance ranking of the input variables on the outlet turbidity and filtered volume is presented using the DE/RFR model. Secondly, an innovative model for the prediction of the outlet turbidity and filtered volume was built and a regression with optimized parameters was done and coefficients of determination of 0.9331 and 0.8712 for filtered volume and outlet turbidity were obtained with this DE/RFR–based model, respectively. Additionally, the outcomes from the Elastic-net, Lasso and Ridge models are worse than DE/RFR–relied model estimations. The DE/RFR-based model's strong performance was confirmed by the agreement between experimental data and the latter results.

[*] Corresponding author.
E-mail address: lato@orion.ciencias.uniovi.es (P.J. García-Nieto).

https://doi.org/10.1016/j.cam.2023.115606

1. Introduction

The increasing frequency of water scarcity periods due to climatic change has led to the use of more effective irrigation techniques, like microirrigation [1], whose worldwide area has increased by 5 Mha between 2012 and 2021 [2]. Moreover, microirrigation allows a safe use of reclaimed effluents or other marginal water sources, whose use help alleviating freshwater scarcity [3–5]. However, to ensure proper water application uniformity and efficiency for keeping agricultural productivity [6], a filtration treatment is needed due to prevent emitter clogging, which is the main constraint to microirrigation spreading. The most common filter types are screen, media, disc and cyclonic [7], but when using low-quality irrigation waters, the most efficient are the pressurized media filters [8,9]. Simultaneous and different filtration mechanisms like hydrodynamic action, interception, straining, sedimentation, inertia and diffusion, can act across the whole depth of filter media allowing the retention of particles of different sizes and, therefore, increasing filtration efficiency [10]. Recent researches have been focused in improving media filter hydrodynamics [11–14], but the impact of various designs and operation conditions on the quality of filtered water has been few studied [15,16].

Precision irrigation models achieve high water use efficiencies by optimizing irrigation scheduling and selecting proper emitters and operation conditions [17,18]. However, the effect of microirrigation filters has not been included in these models yet. At this regard, the prediction of media filter performance parameters such as filtered volume, head loss and filtered water quality are required. The use of advanced techniques like neural networks [19,20], support vector machines [21], gradient boosted regression trees and hybrid algorithms [22,23], gene expression programming [24] and Gaussian process regression [25,26] allowed reasonable prediction accuracy of pressure loss, filtered volume, and outlet dissolved oxygen and turbidity for media filters. However, alternative approaches with may yield improved prediction accuracy need to be explored.

In this study, the outlet turbidity and filtered volume for several kinds of filters in microirrigation systems have been accurately foretold using a unique regressive model relied on the Random Forest Regression (RFR) approach [27–29]. This method, the RFR approximation [27,30–33] in conjunction with Differential Evolution (DE) [34–39] for tuning the parameters of the main method, could be an attractive methodology to tackle this kind of high-nonlinear problems since, to the authors' knowledge, it has not been covered in earlier research. The same experimental dataset was also adjusted for the Elastic-net, Lasso and Ridge regression models for comparison's sake to foretell the outlet turbidity and filtered volume and contrast the outcomes obtained [29,40–43]. To cope with nonlinearities, including interactions between variables, the RFR approach is a statistical learning procedure that was developed conforming to statistics and mathematical analysis [44,45]. It is a development of linear models that uses complex relationships and nonlinearities between the parameters to model them. Comparing RFR approach to traditional and metaheuristic regression approaches, several advantages are apparent [27,30–33,46,47]: (1) it is one of the most precise learning algorithms obtainable. Indeed, for a large enough dataset, it produces a very accurate regressor; (2) it can operate effectively on huge databases; (3) it is capable of handling hundreds of input variables without excluding any; (4) it provides estimates of the key variables in regression; and (5) it makes possible to model nonlinear interactions between the input operational variables of the process. Also, prior studies have shown that RFR is a highly useful technique for practical applications, such as determining the temperature of the near-surface air in glacier zones [48], the mechanical properties of $\gamma-$ TiAl alloys [49], erodibility of treated unsaturated lateritic soil [50], neighborhood environment's impact on peer-to-peer accommodation [51], etc. However, it has never been previously used in microirrigation media filters.

The organization of this paper starts with the experimental design, then it describes the parameters of in this work, and the DE/RFR, Elastic-net, Lasso and Ridge regression techniques are all presented in Section 2; by compiling the DE/RFR outcomes with the experimental values and the relevance order of the input parameters, Section 3 offers the insights gained with this reliable technique, and Section 4 finishes this study by presenting a summary of the investigation's key findings.

2. Materials and methods

2.1. Experimental setup

The test were carried out in a filtration platform located in the wastewater treatment plant (WWTP) of Celrà (Girona, Spain, 42°02'25.3"N 2°52'19.8"E). Three media filters that contained different designs for the underdrain were tested: porous media (designed by Bové et al. [52]), domes (model FA-F2-188, Regaber, Parets del Vallès, Spain) and arm collector (model FA1M, Lama, Gelves, Spain). In Fig. 1, the experimental setup can be seen.

The reclaimed effluent was pumped from the WWTP outlet chamber using a centrifugal pump (model CR-15-4, Grundfos, Bjerringbro, Denmark), which was governed by a frequency variator (model FRN-4, Fuji Electric, Cerdanyola del Vallès, Spain). The filter inlet flow was measured with an electromagnetic flowmeter (model Isomag MS2500, ISOIL Industria, Cinisello Balsamo, Italy). Two pressure transducers (model TM-01/C, STEP Logística *y* Control, Barcelona, Spain) allowed measuring the pressure at the each filter inlet and outlet. Once filtered, the effluent was conveyed to drip irrigation laterals. Since the filtrated flow was higher than what was needed for the driplines, a proportional electrohydraulic actuator (model SKD32, Siemens, Munich, Germany) operated a three-way valve (model VXG41, Siemens, Munich, Germany) and the excess flow was carried to a 3 m^3 water storage tank, which was used for filter backwashing.

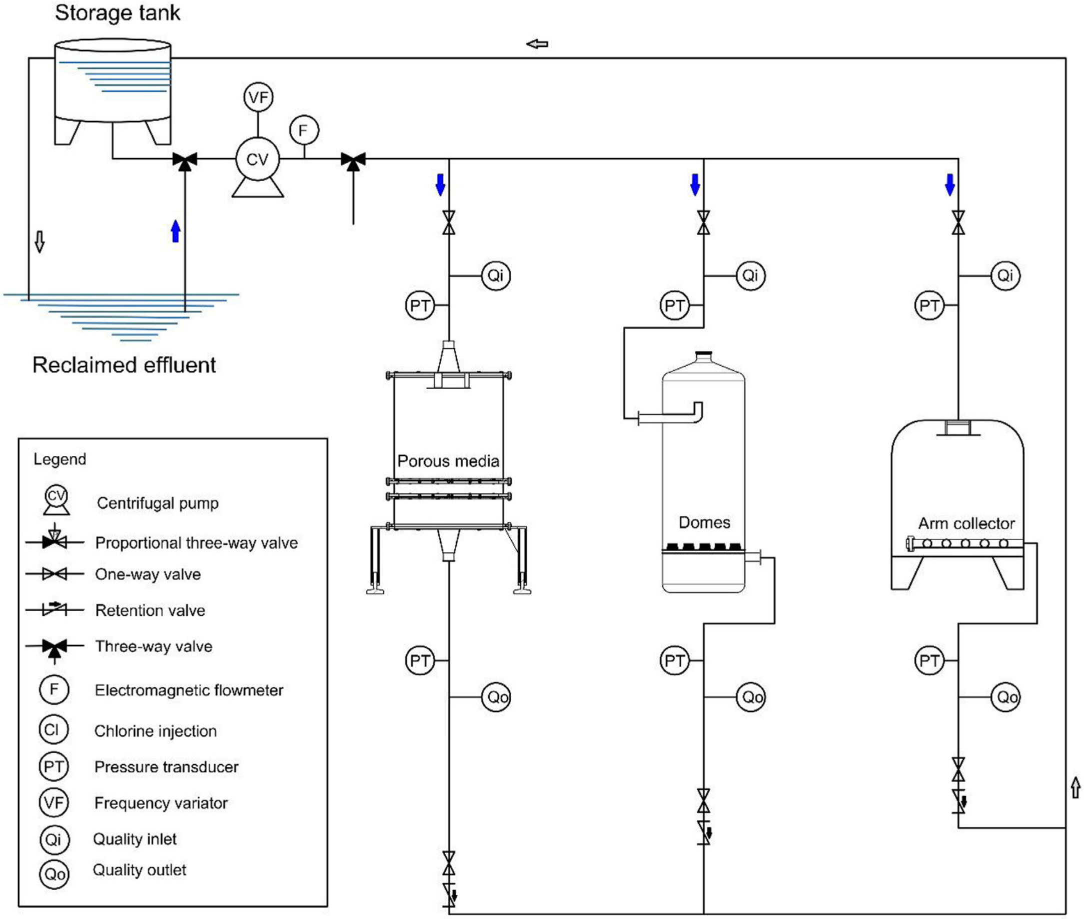


Fig. 1. Experimental filtration platform setup.

Chlorine was continuously injected using a membrane pump (model DosTec AC1/2, ITC, Santa Perpètua de la Mogoda, Spain) from a 0.2 m^3 chlorine deposit to achieve a concentration of 2 mg/l in the effluent after being filtered, and 4 mg/l in the water used for backwashing the filters.

The effluent electrical conductivity, pH and temperature were measured at the filter inlet using sensors (models CLS21-C1E4 A and CPS11D–7BA21, respectively, Endress+Hauser, Gerlingen, Germany), while turbidity and dissolved oxygen were determined at both the filter inlet and outlet with other specific sensors (models CUS31-A2E and COS 61-A1F0, Endress+Hauser, Gerlingen, Germany). Since the experimental facility had only one set of sensors for measuring effluent quality, each filter operated individually.

A supervisory control and data acquisition (SCADA) system previously developed [53] allowed the irrigation system operation and recorded data of pressure, filter inlet flow, filtration cycle duration, and quality parameters every minute. The SCADA system was programmed to stop irrigation when inlet turbidity was above 50 FTU, avoiding operation with high particle loads (see Appendix).

The experiments were carried out using two media materials: CA-07MS silica sand (Silbeco Minerales, Bilbao, Spain) and NW2 crushed recycled glass (Nature Works Tecnologías, L'Alfàs del Pi, Spain). The recycled glass has been investigated as a possible replacement for silica sand due to its similar physical characteristics [10]. The silica sand had a D_{10} diameter (diameter where the sample's 10% weight was finer) of 0.48 mm, a D_{60} diameter (diameter where the sample's 60% weight was finer) of 0.83 mm, a uniformity coefficient (D_{60}/D_{10}) of 1.73, and a porosity (ratio of the media's total volume occupied to the volume of voids) of 0.39. The crushed recycled glass had a D_{10} of 0.44 mm, D_{60} of 0.70 mm, and porosity of 0.54 and uniformity coefficient of 1.59.

Two media bed heights (20 and 30 cm) and two different filtration velocities (30 and 60 m/h), and were tested during 250 h, working in two 4 h daily sessions, for each combination of media material and filter design. All the filters were automatically backwashed by 3 min when the total pressure drop across them reached 50 kPa.

Table 1
The operational variables of physical nature with their means and standard deviations (STDs) used in this study.

Input variables	**Name of the variable**	**Mean**	**STD**
Media	Media	–	–
Filter media type	Filter	–	–
Height of the filter bed (m)	H	0.2666	0.0472
Filtration velocity (m/h)	V	52.101	13.219
Duration of filtration cycle (min)	Tc	251.90	268.06
Electrical conductivity (μS/cm)	EC_i	1966.6	708.20
Dissolved oxygen (mg/l)	DO_i	3.7130	1.1341
pH	pH_i	7.4375	0.2295
Input turbidity (FNU)	$Turb_i$	6.5280	2.8349
Water temperature (°C)	T_i	21.753	3.8305
Output variables	**Name of the variable**	**Mean**	**STD**
Filtered volume (m^3)	V_f	37.098	31.926
Output turbidity (FNU)	$Turb_o$	4.7604	1.1361

2.2. Variables in the model and materials

This study's primary objective was to evaluate the relationship between various experimentally measured parameters and the inputs needed by the DE/RFR model that could be used to calculate the filtrated volume and outlet turbidity. The outlet turbidity, which measures the quality of the filtered effluent and is directly related to the likelihood of physical clogging of emitters, serves as the output variable for microirrigation systems. The following are the operation input variables (see Table 1):

- Media: each one of the two-filtering media (recycled glass and silica sand) described in Section 2.1.
- Filter: the three types of filter—dome, porous and arm collector—are explained in Section 2.1. This is a categorical variable;
- Height of the filter bed (cm): This is an operational variable for sand filters. For each filter, 20 cm and 30 cm of different filter bed heights were tested;
- Filtration velocity (m/h): It affects how filters work. Each filter was tested at two different velocities of filtration (30 and 60 m/h), which fall within the typical range of velocities advised by the manufacturers;
- Duration of filtration cycle (s): it is the time the filter is working at the filtration mode, which is equivalent to the time between two consecutive backwashings;
- Electrical conductivity (μS/cm): it related to salinity, which is an indicator of the quality of the water and a barrier to the use of microirrigation [5];
- Dissolved oxygen (mg/l): It affects how well water can support life and it is a typical control parameter in wastewater treatment facilities;
- pH: it is an indicator of the water alkalinity or acidity;
- Water temperature (°C): at the filter inlet;
- Input turbidity (formazin nephelometric units, FNU): it is a crucial sign of the quality of the water at the filter's entry. It gauges water clarity and is influenced by the number of suspended solids.

The operation output variables of this study are:

- Filtered volume (m^3): it is the amount of water that goes through the filter during each cycle or between two filter backwashings.
- Output turbidity (FNU): it evaluates the water quality at the filter outlet. It is very important to measure the potential clogging of the water used in the microirrigation system.

2.3. Mathematical modeling techniques

2.3.1. Random forest regression (RFR)

A method for lowering an estimated prediction function's variance is bootstrap aggregation or bagging [27,29,30,32]. In particular, trees and other high-variance, low-bias techniques seem to benefit from bagging. Several iterations of the same regression tree are fitted using bootstrap-sampled versions of the training data, and the outcomes are averaged. Bagging has been significantly modified by random forests [27,29–33,46,47,54,55], which averages a sizable collection of de-correlated trees after aggregating them. The training phase of the random forests (RF) ensemble learning technique, which can be used for regression, classification, and other tasks, entails the construction of numerous decision trees. The average estimation of each individual tree is provided when focusing on the regression issue. Random forests perform better overall than decision trees because they compensate for decision trees' propensity to overfit their training dataset.

In fact, the fundamental concept behind the prior technique known as bagging is to lessen variance through the average of many noisy but roughly unbiased models. Because, when developed deeply enough, they have relatively little bias and can catch complicated interaction between the data, trees are excellent candidates for bagging. Trees gain a lot from the averaging because they are known to be noisy. The expectation of any tree is the same as the expectation of an average of B for such trees because each tree formed during bagging is identically distributed (i.d.). Therefore, the bias of the individual trees is the same as that of the bagged trees, and it can only get better the variation. Variance $\frac{1}{B}\sigma^2$ is the variance of B independent random variables with identical distributions (i.i.d.), each one of them has a variance σ^2. The average variance is [27,29–33,46,47,54,55]:

$$\rho\sigma^2 + \frac{1-\rho}{B}\sigma^2 \tag{1}$$

if the variables are simply i.d. (that is, identically distributed, but that does not mean that they are necessarily independent) and with a positive pairwise correlation ρ. Then, the advantages of averaging are constrained by the correlation size between bagged trees pairs since when B rises, while the first term stays, the second one vanishes. By lowering the correlation between the trees, random forests (see algorithm below) aim to enhance bagging variance reduction without substantially raising variance. This is achieved by randomly choosing the input variables during the tree-growing process. In particular, randomly select $m \leq p$ input variables as candidates for splitting when building a tree on a bootstrapped dataset. In most cases, m values are $p/3$ or even 1. The following is the random forest regression predictor after B trees $\{T(x;\Theta_b)\}_1^B$ have grown [27,29–33,46,47,54,55]:

$$\hat{f}_{rf}^{B}(x) = \frac{1}{B}\sum_{b=1}^{B} T(x;\Theta_b) \tag{2}$$

where Θ_b describes cut points at each node, the split variables and terminal-node values of the bth tree of the random forest. It makes sense that decreasing m would decrease the correlation between any two ensemble trees and, as a result, the variance of the average as calculated using Eq. (1) also decreases. The algorithm that corresponds to the RFR is therefore shown below [27,29–33,46,47,54,55]:

Algorithm of Random Forest for regression (RFR)

1. For $b = 1$ to B:
   (a) Take a bootstrap sample Z^* of size N from the training data.
   (b) Recursively repeat the next few steps for each terminal node in the tree T_b until it is reached the minimal node size $n_{\min}$. Extend a random forest tree T_b to the bootstrapped data after that.
      i. Pick m random variables between the p input variables.
      ii. Select the best point for splitting for the variables from the list of m variables.
      iii. Get two child nodes from the node.
2. Result of the set of trees $\{T_b\}_1^B$.

In the case of regression, we will utilize Eq. (2) to obtain a foretelling at a point x: $\hat{f}_{rf}^{B}(x) = \frac{1}{B}\sum_{b=1}^{B} T_b(x)$

2.3.2. Least absolute shrinkage and selection operator (Lasso) regression (LR) and Ridge regression (RR)

The third technique used in this paper for V_f and Turb$_o$ prediction is the Ridge regression (RR) [29,40,41]. Typically, we have n samples that form the training data $(x_1, y_1), \ldots, (x_n, y_n)$, each one with p features (or input variables) and only one output. If y_1 is the output, call $x_i = (x_{i1}, x_{i2}, \ldots, x_{ip})^T$, to the ith sample covariate vector. A regression technique, least squared, is based on minimize a cost function or residual sum of squares (RSS) given by [29,40,41]:

$$RSS(y, y_{pred}) = \sum_{i=1}^{n}\left(y_i - \theta_0 - \sum_{j=1}^{p}\theta_j x_{ij}\right)^2 \tag{3}$$

and $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)^T$ are the coefficients that must be calculated in order to minimize Eq. (3). There is a way to improve the classical linear regression, the regularized linear regression, by including a regularization term in Eq. (3), constituted by squares of the coefficients to compute (excluding the intercept). Then, we choose $\boldsymbol{\theta}$ to minimize [29]:

$$L^{RR}(\boldsymbol{\theta}) = \sum_{i=1}^{n}\left(y_i - \theta_0 - \sum_{j=1}^{p}\theta_j x_{ij}\right)^2 + \lambda\sum_{j=1}^{p}\theta_j^2 = RSS(y, y_{pred}) + \lambda\|\boldsymbol{\theta}\|_2^2 \tag{4}$$

here $\lambda \geq 0$ stands for the complexity parameter or regularization parameter (tuning parameter), that address the overfitting, keeping all the features, but reducing the magnitude/values of parameters θ_j. In fact, Eq. (4), known as Ridge regression (RR), balances two different approaches. The first tries to fit a function that is as closest to the data as possible and it minimizes RSS, while the term, $\lambda \sum_{j=1}^{p} \theta_j^2$, called the shrinkage penalty, aims to get small $\theta_1, \theta_2, \ldots, \theta_p$, and makes the coefficients θ_j as small as possible The λ parameter controls the importance of the two terms in the solution. If $\lambda = 0$, the second term vanishes and, as $\lambda \to \infty$, the shrinkage increases. Therefore, the final solution is dependent on this parameter and different coefficients $\hat{\theta}_{\lambda}^{Ridge}$, will result for different values of λ.

The value of the intercept θ_0 cannot be included in this shrinkage of coefficients because it is the estimated average value when $x_{i1} = x_{i2} = \cdots = x_{ip} = 0$. If we transform linearly the data with $\theta_0 = \bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$, we say that the data is scaled and the average value for all the independent variables is zero.

The main benefit of using Ridge regression instead of traditional linear regression is that it allows for more flexible manipulation of the variance and bias of the regression because as λ decreases, the variance increases but the bias decreases.

This regression controls the overfitting, reducing the magnitude of large θ coefficients but the interpretation of the model is not improved. Ridge regression has the obvious disadvantage that does not drop any predictor in the final model. Unless $\lambda \to \infty$, the term $\lambda \sum_{j=1}^{p} \theta_j^2$ in Eq. (4) will cause the coefficients to contract but not to zero. This model can be effective from a predictive standpoint, but it may be challenging to interpret in situations where there are a lot of input variables.

To address this issue, the *Lasso Regression (LR)* method was developed. In effect, the loss function in the Lasso regression is [29,40,41]:

$$L^{LR}(\boldsymbol{\theta}) = \sum_{i=1}^{n} \left(y_i - \theta_0 - \sum_{j=1}^{p} \theta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^{p} \left| \theta_j \right| = RSS\left(y, y_{pred}\right) + \lambda \left\| \boldsymbol{\theta} \right\|_1^2 \tag{5}$$

Ridge regression and Lasso regression are quite similar, as can be seen by comparing Eqs. (4) and (5), with the exception that the second term in the Ridge regression is different (see Eq. (4)) and has been modified to $\left|\theta_j\right|$ in Lasso regression (see Eq. (5)). It could be said that the penalty term in Lasso regression uses L_1 while Ridge regression uses L_2, because the L_p- norm of a vector $\boldsymbol{\theta}$ is:

$$\left\| \boldsymbol{\theta} \right\|_p = \left(\sum_{j=1}^{p} \left| \theta_j \right|^p \right)^{\frac{1}{p}} \tag{6}$$

Lasso regression also causes the regression function's coefficients to decrease towards zero, but when the penalty L_1 is applied, some of the coefficients are made to disappear. Lasso can thus perform variable selection by dropping variables. As a result, the model becomes simpler to understand. Lasso is said to produce sparse models, or models with a small number of variables.

2.3.3. Elastic-net regression (ENR)

Elastic-net regression (*ENR*) combines the two most widely used Ridge and Lasso linear regression models. Lasso uses an L_1 penalty, while Ridge employs an L_2 penalty. Elastic-net uses both the L_2 and the L_1 penalties, so there is no need to select between these two models. Combining the two penalties now yields the elastic-net loss function [29,42,43]:

$$\begin{aligned} L^{ENR}(\boldsymbol{\theta}) &= \sum_{i=1}^{n} \left(y_i - \theta_0 - \sum_{j=1}^{p} \theta_j x_{ij} \right)^2 + \lambda_1 \sum_{j=1}^{p} \left| \theta_j \right| + \lambda_2 \sum_{j=1}^{p} \theta_j^2 \\ &= RSS\left(y, y_{pred}\right) + \lambda_1 \left\| \theta \right\|_1^2 + \lambda_2 \left\| \theta \right\|_2^2 \end{aligned} \tag{7}$$

We now employ two regularization parameters, one for each penalty, as opposed to regularization parameter λ. The L_1 penalty is managed by λ_1, and the L_2 penalty is managed by λ_2. Now, Elastic-net can be used in the same manner as Ridge or Lasso.

If $\lambda_1 = 0$, then we have ridge regression. If $\lambda_2 = 0$, it is Lasso regression. The L_1-ratio parameter, which establishes the proportion of our L_1 penalty with respect to λ, can also be used in place of the two λ parameters. In this case, our L_1 penalty will be multiplied by 0.5 if L_1-ratio $= 0.5$ and our L_2 penalty by $1 - L_1 - ratio = 0.5$. This value, L_1-ratio $= 0.5$, is employed here in the calculations because it is the most frequently used in most computational codes. Hence, the final equation is as follows [29,42,43]:

$$\begin{aligned} L^{ENR}(\boldsymbol{\theta}) &= RSS\left(y, y_{pred}\right) + \lambda \cdot \left(1 - L_1 ratio\right) \sum_{j=1}^{p} \left| \theta_j \right| + \lambda \cdot L_1 ratio \sum_{j=1}^{p} \theta_j^2 \\ &= RSS\left(y, y_{pred}\right) + \tfrac{\lambda}{2} \cdot \left(\sum_{j=1}^{p} \left| \theta_j \right| + \sum_{j=1}^{p} \theta_j^2 \right) \end{aligned} \tag{8}$$

2.3.4. Differential evolution (DE) optimizer

A problem can be optimized using differential evolution (DE), a metaheuristic method in evolutionary computation, by repeatedly attempting to raise the quality of a potential solution. The optimized function does not have to be differentiable so that the DE optimizer can be employed for multidimensional real-valued data with success. Moreover, DE optimizer can also be used for problems that change over time, are noisy or are not continuous. By using a population of potential solutions, and combining existing ones using a simple formulae, DE optimizes a problem by keeping the a solution that is the fittest for the given problem of optimization [34]. The algorithm represents the optimization problem variables as a vector of real numbers. The population is made up of NP vectors (actual population), and these vectors length n is the number of parameters in the optimization problem.

If p is the index of a vector within the population ($p = 1, \ldots, NP$) and the generation is g, we define the vector as $\mathbf{x}_p^g$. The components of this vector are the variables of the problem $x_{p,m}^g$, where m corresponds to the index of the variable within the individual ($m = 1, \ldots, n$). The variables are contained in intervals limited by the values $\mathbf{x}_m^{\min}$ and $\mathbf{x}_m^{\max}$, respectively, at the minimum and maximum. The DE algorithm has four steps [34,35,37,38]:

- Initialization;
- Mutation;
- Recombination; and
- Selection.

The search starts after initialization. When a stopping criterion (number of generations, length of time, level of solution attained, etc.) is met, the mutation-recombination-selection steps finish.

Initialization

When initializing the population (first generation) at random, each variable's minimum and maximum values are taken into account [37,38]:

$$\mathbf{x}_{p,m}^1 = \mathbf{x}_m^{\min} + rand\,(0,1) \cdot \left(\mathbf{x}_m^{\max} - \mathbf{x}_m^{\min}\right) \ \text{ for } p = 1, \ldots, NP \text{ and } m = 1, \ldots, n \tag{9}$$

so that $rand\,(0,1)$ represents a random number in $[0, 1]$.

Mutation

Three randomly selected individuals, known as the target vectors $\mathbf{x}_a$, $\mathbf{x}_b$ and $\mathbf{x}_c$, are used to create the NP new vectors that make up the mutation. The following is how the new vectors $\mathbf{n}_p^t$ are created [37,38]:

$$\mathbf{n}_p^g = \mathbf{x}_c + F \cdot (\mathbf{x}_a - \mathbf{x}_b) \ \text{ for } p = 1, \ldots, NP \tag{10}$$

with a, b, c and p, that are different. The parameter F, which is in the interval $[0, 2]$, regulates the mutation rate.

Recombination

Following the creation of the NP new vectors, the trial vectors $\mathbf{t}_m^g$ are created by performing a random recombination and comparing the results to the initial vectors $\mathbf{x}_p^g$ [37,38]:

$$t_{p,m}^g = \left\{ \begin{array}{ll} n_{p,m}^g & \text{if } \ rand\,(0,1) < GR \\ x_{p,m}^g & \text{otherwise} \end{array} \right\} \text{ for } p = 1, \ldots, NP \text{ and } m = 1, \ldots, n \tag{11}$$

The recombination rate is controlled by the parameter GR. The test vector will contain both the original vectors and the updated vectors because the comparison is done variable by variable.

Selection

In order to choose the vector of the following generation, which will have the best value given by the fitness function, the test vectors are simply compared to the original vectors [37,38]:

$$\mathbf{x}_p^{g+1} = \left\{ \begin{array}{ll} \mathbf{t}_p^g & \text{if } \ fit\left(\mathbf{t}_p^g\right) > fit\left(\mathbf{x}_p^g\right) \\ \mathbf{x}_p^g & \text{otherwise} \end{array} \right\} \tag{12}$$

2.4. Goodness–of–fit

For the building of the DE/RFR model, ten input variables were used. The dependent estimated variables are the filtered volume (V_f) and outlet turbidity (Turb_o).

The model that approaches best the experimental data in order to predict both of the two operation variables from the other ten input operation parameters is searched. In this sense, the coefficient of determination R^2 [56,57] was the criterion taken into consideration to assess the goodness-of-fit. Each value of the dataset t_i, has a corresponding value estimated by the model y_i. The first are known as the observed values, whereas the second are frequently called the predicted values. The following sums of squares are used to measure the dataset variability [57,58]:

- $SS_{tot} = \sum_{i=1}^{n} \left(t_i - \bar{t}\right)^2$: it is called the total sum of squares and it is proportional to the sample variance;

Table 2
Initial intervals for the hyperparameters values.

RFR parameters	Lower limit	Upper limit
node_size	5	30
ntree	30	1000
n_per	1	10
mtry	1	8

- $SS_{reg} = \sum_{i=1}^{n} \left(y_i - \overline{t}\right)^2$: known as the regression sum of squares, also termed the explained sum of squares;
- $SS_{err} = \sum_{i=1}^{n} (t_i - y_i)^2$: it is the residual sum of squares.

Note that, $\overline{t}$ is the average of the n data in the experimental set:

$$\overline{t} = \frac{1}{n}\sum_{i=1}^{n} t_i \tag{13}$$

The coefficient of determination can be defined [58]:

$$R^2 \equiv 1 - \frac{SS_{err}}{SS_{tot}} \tag{14}$$

If the coefficient of determination is 1.0 the fitting is perfect and without error.

The RFR hyperparameters are [27,29–33]:

- Node size (node_size): Minimum size of terminal nodes. Smaller trees grow when this parameter is increased.
- Number of regression trees (ntree): amount of trees to be grown. Model construction will cost more to compute the larger the tree. The 500 trees setting is the default value.
- Number of permutations (n_perm): Number of times that the out of bag (OOB). To determine the relevance of each variable, data are permuted per tree.
- Number of input variables per node (mtry): it deals with the amount of variables we should choose randomly during a node split. One-third of the full set of input variables, p, is taken as the default value. To prevent overfitting, we must always make an effort to avoid using small values of mtry.

Currently, model has been built (particularly in this work, the innovative DE/RFR- model) using the filtered volume (V_f) and outlet turbidity (Turb_o) as dependent variables, taking from the other ten input variables in granular filters [59,60], investigating their influence to successfully improve its computation with the examination of the coefficient of determination R^2.

It is crucial to keep in mind that the RFR method heavily depends on finding the two aforementioned optimal hyperparameters: number of regression trees (ntree); node size (n_size); number of permutations (n_perm) and number of input variables per node (mtry).

Parameter sweep or *grid search*, which is just an exhaustive search through a subset of values in the hyperparameter space, has been the common way of tuning the hyperparameters until now in previous investigations. The *Differential Evolution (DE)* method was applied in this study to perform a more economic search due to its effectiveness in resolving related optimization issues [38] here with success to find these optimal parameters. The DE optimizer is a technique used in evolutionary computation to solve optimization problems by repeatedly attempting to raise the quality of a candidate solution. These techniques are referred to as metaheuristics because they can search very large spaces of potential solutions and make little to no assumptions about the objective function [61,62].

Hence, the filtered volumen and output turbidity (output variables) have therefore been successfully predicted using this novel hybrid DE/RFR–based method by evaluating the influence of 10 operation input variables and successfully optimizing the computation using the coefficient of determination R^2. The flowchart for the DE/RFR–relied model created in this study is in Fig. 2.

The cross-validation methodology also utilized the optimum coefficient of determination (R^2) [63]. The best hyperparameters for the DE/RFR model can be chosen using ten-fold cross-validation [28]. The dataset is randomly divided into ten subsets of comparable size, and then a set of parameters is chosen. A model is built using nine subsets and then tested with the final one to determine its goodness-of-fit [64]. Ten times this procedure is performed and a different subset is used each time as the testing set. The final value for the chosen set of parameters is the average of the goodness-of-fit.

The DE algorithm directs the choice of these parameter sets to test based on their fitness until it chooses a specific set of ideal hyperparameters [28,64].

The initial intervals of the hyperparameters for DE/RFR models are shown in Table 2.

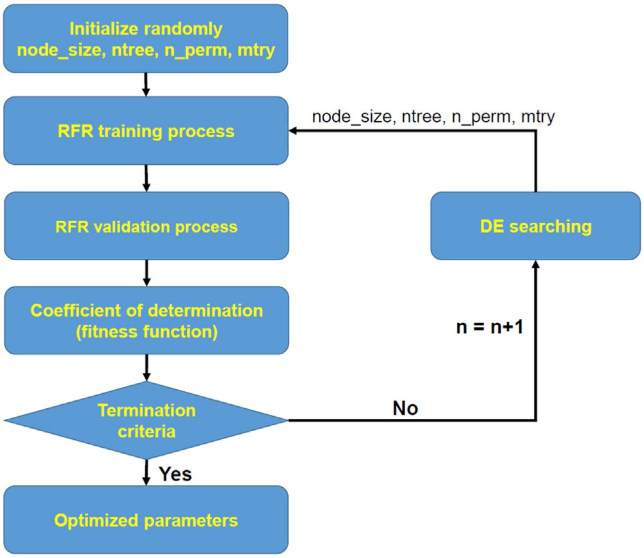


Fig. 2. Process flow diagram for the DE/RFR model's parameter optimization.

Table 3
Optimized parameters for the DE/RFR models.

RFR optimal parameters	Filtered volume	Outlet turbidity
node_size	5	5
ntree	137	626
n_perm	5	4
mtry	8	4

Table 4
Optimal parameter λ for the Elastic-net, Lasso and Ridge regression models with the DE optimizer.

Optimal λ	Filtered volume	Outlet turbidity
Ridge	2.9111	0.068155
Lasso	0.052069	0.00084024
Elastic-net	0.078777	0.0012712

3. Results and discussion

Ten different operation variables were used as input variables in the newly created predictive model. In Table 1, they were all previously presented. We used data from 1,014 filtration cycles. The samples with missing data have been removed from the total 1,016 samples that were measured experimentally.

To tackle this study, the dataset was split into two sets: a set that contains 80% of the data for training set and a testing set with the remainder data. With the help of the training data, a model was built, optimized, and then put to the test using the test data set.

The proposed DE/RFR-based model uses the outlet turbidity and filtered volume as output dependent variables. As previously mentioned, the prediction made using the independent variables [59] was successful. The objective function value and choice of RFR hyperparameters (node size, number of permutations, number of regression trees and number of input variables per node) affect the DE/RFR technique.

Table 3 displays the RFR hyperparameters that were obtained after the RFR was tuned using the metaheuristic DE optimizer.

For comparison purposes, the Elastic-net, Lasso and Ridge regression were used in this work. Their accuracy depends on the parameter λ [29,40–43]. In this way, the optimal value of parameter λ for the Elastic-net, Lasso and Ridge regression models determined using the DE optimizer is indicated in Table 4.

Using the testing set with the optimized model, the R^2 value was calculated. The final DE/RFR models were built using R project [65] jointly with R package DEoptim (DE module) [66].

Fig. 3 shows the first and second order terms for the three variables more relevant in the DE/RFR model for the prediction of the filtered volume.

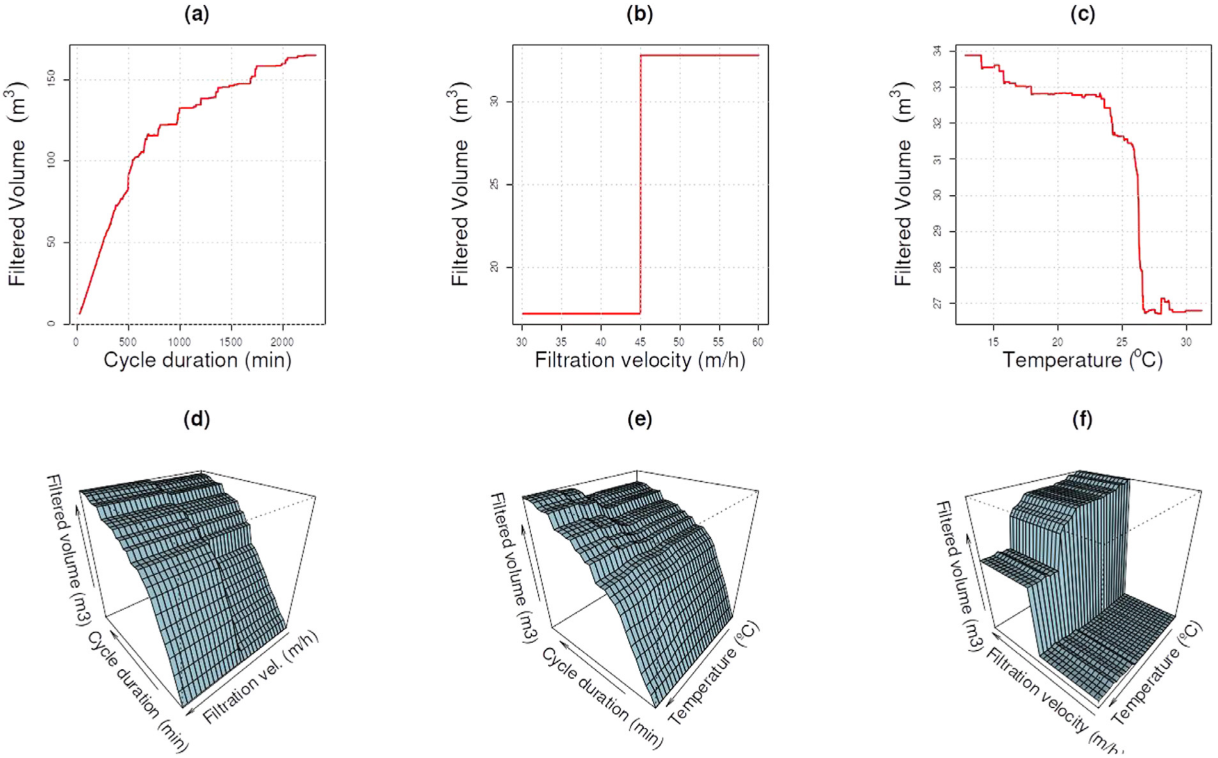


Fig. 3. First and second order terms for the three most relevant variables in the DE/RFR model for the prediction of the filtered volume (V_f).

Table 5
Test set coefficient of determination (R^2), correlation coefficient (r), root mean square error (RMSE) and mean absolute error (MAE) for the RFR, Ridge, Lasso, and Elastic-net regression models fitted for the prediction of the filtered volume (V_f) using the DE optimizer.

Technique	R^2	r	*RMSE*	*MAE*
RFR	0.9331	0.9661	8.4383	2.4474
Ridge	0.8577	0.9375	12.3057	7.0434
Lasso	0.8823	0.9426	11.1912	6.5243
Elastic-net	0.8821	0.9426	11.1988	6.5282

Table 6
Test set coefficient of determination (R^2), correlation coefficient (r), root mean square error (RMSE) and mean absolute error (MAE) for the RFR, Ridge, Lasso, and Elastic-net regression models fitted for the prediction of the outlet turbidity ($Turb_o$) using the DE optimizer.

Technique	R^2	r	*RMSE*	*MAE*
RFR	0.8712	0.9366	0.3931	0.2719
Ridge	0.5533	0.7444	0.7322	0.5777
Lasso	0.5509	0.7448	0.7342	0.5747
Elastic-net	0.5510	0.7449	0.7341	0.5746

Additionally, Fig. 4 shows the terms of first and second order for the three most relevant parameters in the DE/RFR model for the prediction of the outlet turbidity.

In light of the outcomes, the DE optimizer and RFR technique can be used to create models for the estimation of the outlet turbidity and filtered volume in microirrigation media filters that have high performance. Indeed, the coefficient of correlation the coefficient of determination (r and R^2) with the test set for the RFR, Elastic-net, Lasso and Ridge regression approaches for the response variable filtered volume (V_f) (see Table 5).

Similarly, Table 6 illustrates the coefficient of correlation and coefficient of determination (r and R^2) for the test set with the RFR, Elastic-net, Lasso and Ridge regression approaches for the response variable outlet turbidity ($Turb_o$).

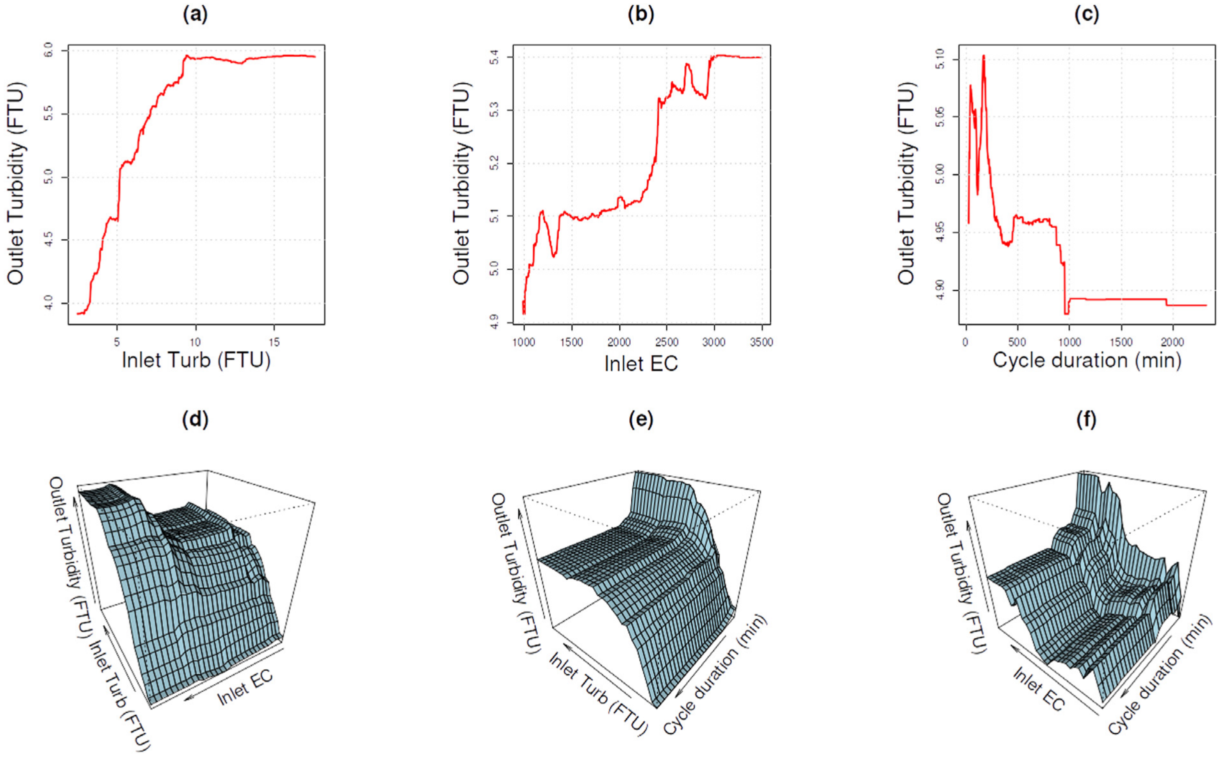


Fig. 4. First and second order terms for the three most relevant parameters in the DE/RFR model for the prediction of the outlet turbidity.

Table 7
Relative importance of the variables within the DE/RFR model for the filtered volume prediction as stated in the associated weights in absolute decreasing order.

Input variable	Weight
Cycle duration	80.431524
Filtration velocity	45.571252
T_i	7.630397
$\mathrm{EC_i}$	5.687074
Media	5.119279
pH	4.875634
$\mathrm{DO_i}$	4.359500
Filter	3.748693
$\mathrm{Turb_i}$	2.651496
Media bed height	1.894156

Additionally, an iMac with a CPU Intel Core i5 @ 3.2 GHz with four cores and 8 GB RAM memory was used, taking 16,846 s (approximately 4.7 h) to obtain the optimal model for the filtered volume (Vf) and 15,635 s (approximately 4.3 h) for the outlet turbidity ($\mathrm{Turb_o}$).

3.1. Importance of the variables

Additionally, as a result of these calculations, Table 7 and Fig. 5 display the relevance rankings for the independent variables used to predict the filtered volume (dependent variable) in this study. Thus, the cycle duration, followed by filtration velocity, is the most important factor in the RFR model's filtered volume prediction, followed by water temperature, electrical conductivity, type of medium, pH, inlet dissolved oxygen, filter, inlet turbidity, and media bed height.

The filtered volume was higher with longer filtration cycles and greater filtration velocities since more water was filtered. The filtered volume in media filter is a complex variable since depends on the double interactions between filtration velocity, filter design and media bed height [15], as well as media material and filter type [16]. In our case, filtration velocity and, to a lesser extent, media material have more relevance than filter design and media bed height. The

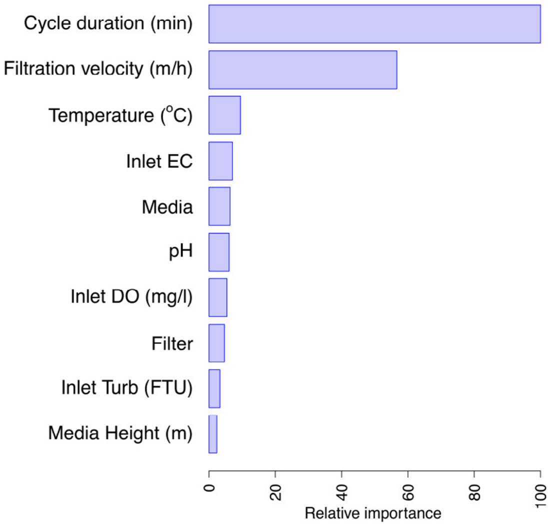


Fig. 5. Relative relevance of the independent variables in the DE/RFR model for the prediction of the filtered volume (V_f) calculated with respect to the variable with the maximum weight.

Table 8
Relative importance of the variables within the DE/RFR model for the outlet turbidity prediction as stated in the associated weights in absolute decreasing order.

Input variable	Weight
$\mathrm{Turb_i}$	121.30774
$\mathrm{EC_i}$	50.00736
Cycle duration	40.52834
T_i	39.83786
Media bed height	33.83612
$\mathrm{DO_i}$	28.86789
pH	25.31847
Filter	23.81075
Media	20.18713
Filtration velocity	18.36561

filtered volumes observed using reclaimed effluents are usually highly variable [15] due to effluent changing composition with time. The temperature, which may affect biological growth, and electrical conductivity, related to dissolved solids, are the two effluent characteristics that have more effect on the filtered volume. Since both temperature and EC are easy and inexpensive to measure, our results suggest that both parameters should be monitored for assessing media filter performance.

Similar to this, Table 8 and Fig. 6 illustrate the relative significance of the independent variables in the DE/RFR model for the prediction of outlet turbidity ($\mathrm{Tub_o}$). In this instance, the inlet turbidity ($\mathrm{Turb_i}$) is determined by the RFR model to be the most important factor in outlet turbidity prediction, followed by the electrical conductivity, cycle duration, water temperature, media height (height of the filter bed), dissolved oxygen, pH, filter, type of medium, and filtration velocity.

The fact that the inlet turbidity appears to be the most relevant variable for explaining the outlet turbidity was expected since the turbidity removal achieved by media filters are highly dependent on turbidity inlet values [15]. During the filtration cycles, reduced turbidity results from the retention of particles throughout the media bed. Those variables that allow higher residence time of the reclaimed effluent in the filter contribute to improve filtered water quality [7]. This explains why filtration cycle duration and media bed height show relevance in explaining the outlet turbidity. On the other hand, reclaimed effluent have several compounds and organisms that interact between them with chemical reactions and biological processes, which affect water quality parameters. According to our results, both electrical conductivity and temperature play a role on turbidity, but their acting mechanism needs to be further explored.

In this investigation, the filtered volume (V_f) has been foretold from the ten independent operation input variables in microirrigation systems as shown in Fig. 7 utilizing the comparison among the observed and foretold V_f examples using the Ridge (Fig. 7(a)), Lasso (Fig. 7(b)), Elastic-net (Fig. 7(c)), and DE/RFR (Fig. 7(d)) regressions. The fourth model is the best.

It is shown in Fig. 8 the comparison between the outlet turbidity ($\mathrm{Turb_o}$) predicted and observed values with Ridge model (Fig. 8(a)), Lasso model (Fig. 8(b)), Elastic-net model (Fig. 8(c)), and RFR–based model (Fig. 8(d)) using the testing

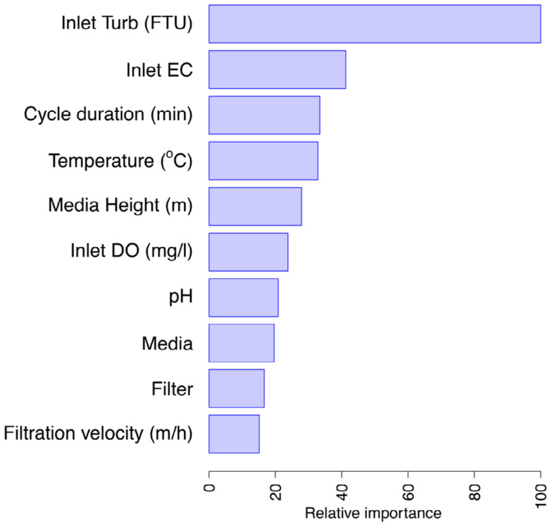


Fig. 6. Relative significance of the independent variables in the DE/RFR model for the prediction of the outlet turbidity (Turb_o) calculated with respect to the variable with the maximum weight.

Table 9
Test set coefficient of determination (R^2), correlation coefficient (r), root mean square error (RMSE) and mean absolute error (MAE) for the RFR reduced models fitted for the prediction of the filtered volume (V_f) using the DE optimizer with decreasing number of variables.

Number of variables	R^2	r	*RMSE*	*MAE*
10	0.9331	0.9661	8.4383	2.4474
8	0.9367	0.9678	8.2086	2.1744
6	0.9370	0.9680	8.1993	2.0819
4	0.9348	0.9669	8.3375	2.1351
2	0.9313	0.9654	8.5537	2.6559

Table 10
Test set coefficient of determination (R^2), correlation coefficient (r), root mean square error (RMSE) and mean absolute error (MAE) for the RFR reduced models fitted for the prediction of the outlet turbidity (Turb_o) using the DE optimizer with decreasing number of variables.

Number of variables	R^2	r	*RMSE*	*MAE*
10	0.8712	0.9366	0.3931	0.2719
8	0.8556	0.9284	0.4163	0.2857
6	0.8415	0.9200	0.4362	0.2991
4	0.8163	0.9050	0.4685	0.3134
2	0.6902	0.8348	0.6084	0.4622

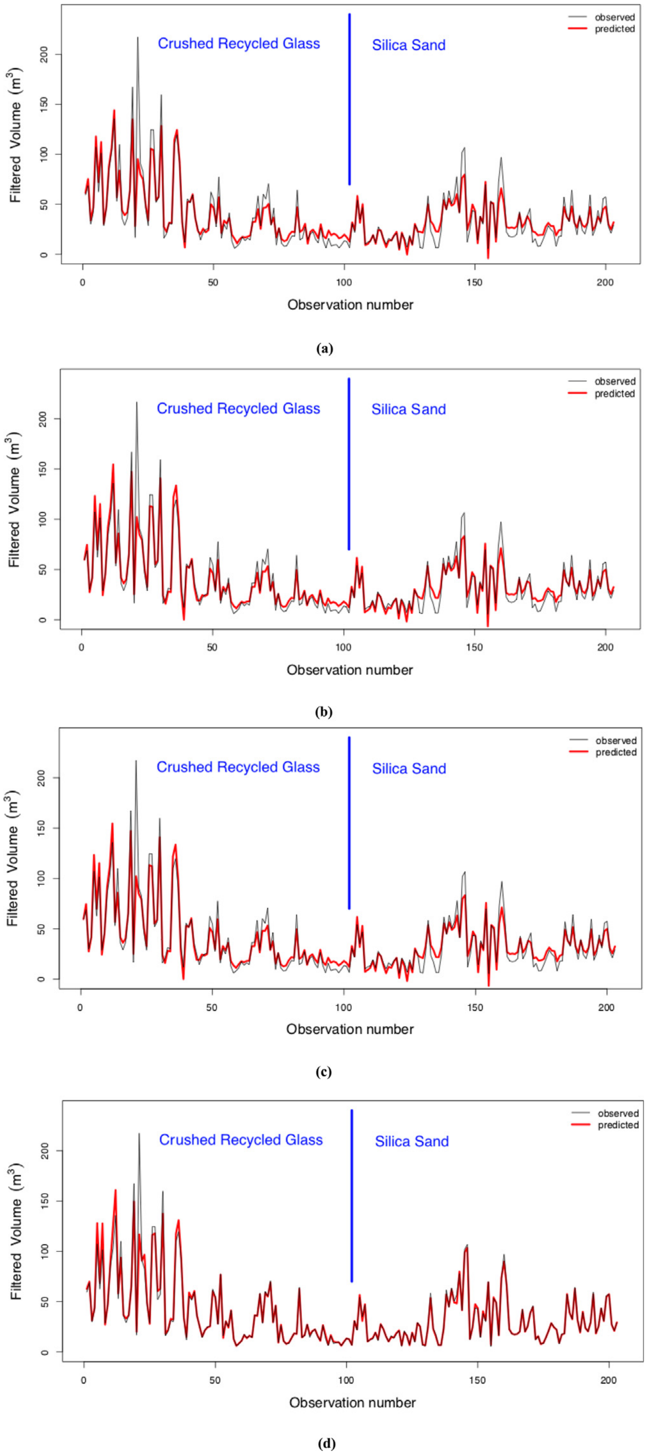
dataset. To achieve the most efficient solution to this regression problem, a RFR model with a DE optimizer is therefore required.

Additionally, we have obtained and calculate the goodness-of-fit for reduced models with only the 2, 4, 6 and 8 first ranked variables (the full models use 10 independent variables) and the results differ. In the case of the estimation of the filtered volume (see Table 9), two variables, cycle duration and filtration velocity are much more important than the remaining 8. Here a very good model can be constructed using only these two variables with a coefficient of determination of 0.9313 very similar to the full model, 0.9331. Also, the addition of the other variables do not add significantly to the accuracy of the model and even, from a certain point (6 variables) it only adds noise.

On the other hand, in the estimation of the outlet turbidity (see Table 10), even though there is a very important variable, the inlet turbidity, it is not able to explain enough the outlet turbidity and the model improves steadily as we add the other variables from a coefficient of determination of 0.6902 to 0.812 for the full model.

4. Conclusions

The main conclusions of this study can be summed up as follows after accounting for the experimental and numerical results:



(a)

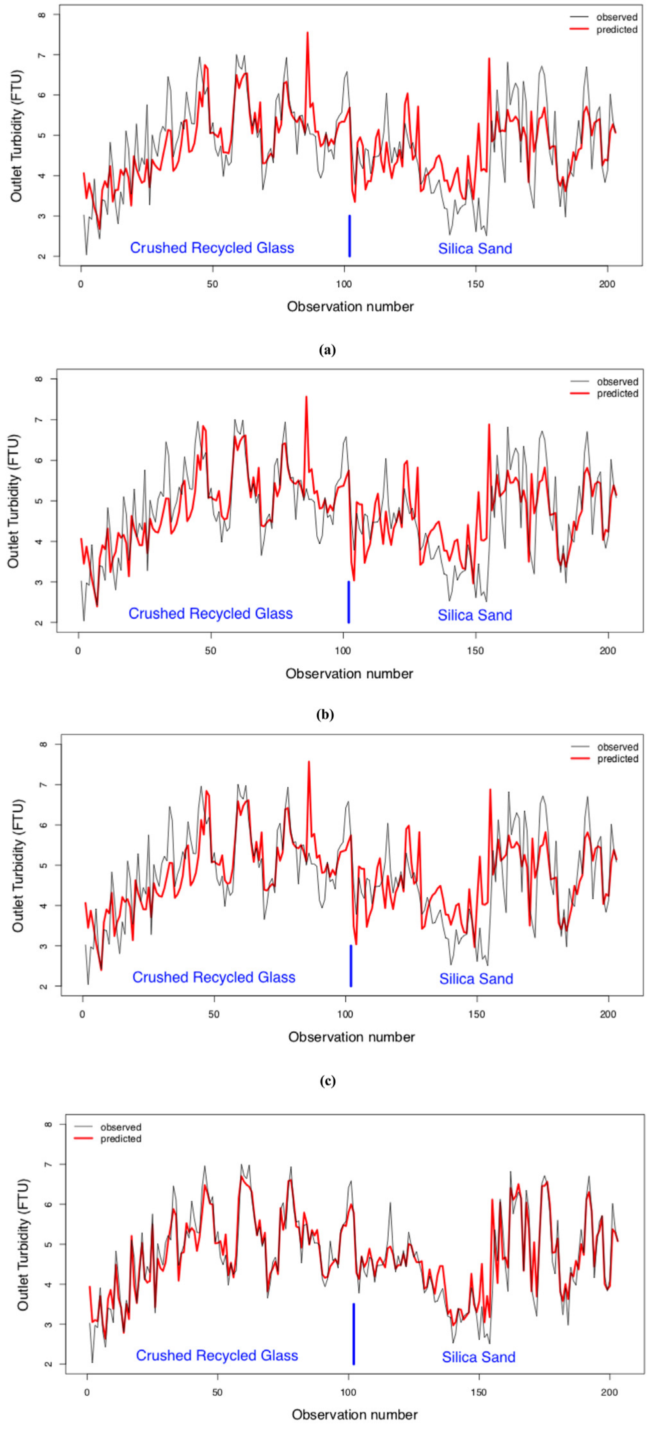
(b)

(c)

(d)

Fig. 7. Predicted and observed V_f for the test dataset with two different filter media (crushed recycled glass and silica sand) using: (a) Ridge model ($R^2 = 0.8577$); (b) Lasso model ($R^2 = 0.8823$); (c) Elastic-net model ($R^2 = 0.8821$); and (d) RFR model ($R^2 = 0.9331$).

- Firstly, the creation of alternative diagnostic techniques is crucial because currently there are not analytical equations that can accurately estimate the filtered volume (V_f) and outlet turbidity (Turb_o) based solely on experimental values. In this regard, choosing to evaluate the filtered volume (V_f) and outlet turbidity (Turb_o) in microirrigation systems with media filters using the new DE/RFR-based method is a good choice;



(a)

(b)

(c)

(d)

Fig. 8. Predicted and observed values of Turb$_o$ for the test dataset with two different filter media (crushed recycled glass and silica sand) using: (a) Ridge model ($R^2 = 0.5533$); (b) Lasso model ($R^2 = 0.5509$); (c) Elastic-net model ($R^2 = 0.5510$); and (d) RFR model ($R^2 = 0.8712$).

- Secondly, it was confirmed that this DE/RFR model in granular filters can model the outlet turbidity (Turb$_o$) and filtered volume (V_f) diagnosis;
- Thirdly, when this RFR model was used with the dataset that contains the variables outlet turbidity (Turb$_o$) and the filtered volume (V_f) respectively, reasonable coefficients of determination, 0.9331 and 0.8712 are found.

- Fourthly, to estimate the outlet turbidity ($\mathrm{Turb_o}$) and filtered volume (V_f) in media filters, the ranking in importance of the input variables used in this process was established. In particular, input variable cycle duration (tc) may be considered as the parameter that has the greatest influence on the estimation of the filtered volume (V_f), and after it, filtration velocity (v), water temperature (T_i), electrical conductivity ($\mathrm{EC_i}$), type of medium, pH, dissolved oxygen ($\mathrm{DO_i}$), filter, inlet turbidity ($\mathrm{Turb_i}$), and media height (filter bed height) (H). Also, the variable inlet turbidity ($\mathrm{Turb_i}$) might have the greatest bearing on the estimation of the outlet turbidity ($\mathrm{Turb_o}$) before the electrical conductivity ($\mathrm{EC_i}$), cycle duration (tc), water temperature (T_i), media height (filter bed height) (H), dissolved oxygen ($\mathrm{DO_i}$), pH, filter, type of medium and filtration velocity (v);
- Finally, it was established how the RFR approach's hyperparameter settings affected the output turbidity ($\mathrm{Turb_o}$) and filtered volume (V_f) regression performances.

In conclusion, this method could be successfully applied to different filtration mechanism involving the same or different types of filter media, but it must be taken into account the unique properties of the different filters and experiments. As a result, a good solution to the issue of determining the filtered volume (V_f) and outlet turbidity ($\mathrm{Turb_o}$) in media filters widely used in microirrigation systems is a DE/RFR-based model.

CRediT authorship contribution statement

P.J. García-Nieto: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Writing – original draft, Visualization, Supervision. **E. García-Gonzalo:** Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Writing – original draft, Visualization. **G. Arbat:** Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Writing – original draft, Visualization. **M. Duran-Ros:** Conceptualization, Methodology, Formal analysis, Investigation, Data curation, Writing – original draft. **T. Pujol:** Methodology, Formal analysis, Investigation, Data curation, Writing – original draft, Visualization. **J. Puig-Bargués:** Methodology, Formal analysis, Investigation, Data curation, Writing – original draft, Visualization.

Data availability

Data will be made available on request.

Acknowledgments

The Mathematics Department at the University of Oviedo provided computational assistance, and grants AGL2015-63750-R and RTI2018-094798-B-100 were provided by the Spanish Research Agency and the European Regional Development Fund respectively. The Municipality of Celrà assisted the authors in conducting this experiment. Nature Works Tecnologas donated recycled glass media. Finally, Anthony Ashworth edited the manuscript spelling and grammar in English.

Appendix. Supplementary data

Supplementary data to this article can be found online at https://docs.google.com/spreadsheets/d/1c1HArNkMm2zMkqwJ4PjRWf5_KaAwipxy/edit?usp=drive_link&ouid=106109749140390362054&rtpof=true&sd=true.

References

[1] J.M. Tarjuelo, J.A. Rodriguez-Diaz, R. Abadía, E. Camacho, C. Rocamora, M.A. Moreno, Efficient water and energy use in irrigation modernization: Lessons from spanish case studies, Agric. Water Manag. 162 (2015) 67–77, http://dx.doi.org/10.1016/j.agwat.2015.08.009.

[2] Agricultural water management for sustainable rural development, annual report 2021–22, in: International Commission on Irrigation and Drainage, (ICID), 2022, https://icid-ciid.org/icid_data_web/ar_2021.pdf, (accessed 30 2023).

[3] T.P. Trooien, D.J. Hills, Application of biological effluent, in: F.R. Lamm, J.E. Ayars F.S. Nakayama (Eds.), Microirrigation for Crop Production: Design, Operation and Management, Elsevier, Amsterdam, 2007, pp. 329–356.

[4] Y. Shen, J. Puig-Bargués, M. Li, Y. Xiao, Q. Li, Y. Li, Physical, chemical and biological emitter clogging behaviors in drip irrigation systems using high-sediment loaded water, Agric. Water Manag. 270 (2022) 107738, http://dx.doi.org/10.1016/j.agwat.2022.107738.

[5] A. Tal, Rethinking the sustainability of Israel's irrigation practices in the drylands, Water Res. 90 (2016) 387–394, http://dx.doi.org/10.1016/j.watres.2015.12.016.

[6] C. Solé–Torres, F.R. Lamm, M. Duran–Ros, G. Arbat, F. Ramírez de Cartagena, J. Puig–Bargués, Assessment of microirrigation field distribution uniformity procedures for pressure-compensating emitters under potential clogging conditions, Trans. ASABE 64 (3) (2021) 1063–1071, http://dx.doi.org/10.13031/trans.14486.

[7] F.S. Nakayama, B.J. Boman, D.J. Pitts, Maintenance, in: F.R. Lamm, J.E. Ayars, F.S. Nakayama (Eds.), Microirrigation for Crop Production, Elsevier, Amsterdam, 2007, pp. 389–430.

[8] I. Ravina, E. Paz, Z. Sofer, A. Marm, A. Schischa, G. Sagi, Z. Yechialy, Y. Lev, Control of clogging in drip irrigation with stored treated municipal sewage effluent, Agric. Water Manage. 33 (2–3) (1997) 127–137, http://dx.doi.org/10.1016/S0378-3774(96)01286-3.

[9] A. Capra, B. Scicolone, Recycling of poor quality urban wastewater by drip irrigation systems, J. Clean. Prod. 15 (16) (2007) 1529–1534, http://dx.doi.org/10.1016/j.jclepro.2006.07.032.

[10] A. Cescon, J.-Q. Jiang, Filtration process and alternative filter media material in water treatment, Water 12 (2020) 3377, http://dx.doi.org/10.3390/w12123377.

[11] M. Mesquita, F.P. de Deus, R. Testezlaf, L.M. da Rosa, A.V. Diotto, Design and hydrodynamic performance testing of a new pressure sand filter diffuser plate using numerical simulation, Biosyst. Eng. 183 (2019) 58–69, http://dx.doi.org/10.1016/j.biosystemseng.2019.04.015.

[12] F.P. de Deus, M. Mesquita, J.C.S. Ramirez, R. Testezlaf, R.C. de Almeida, Hydraulic characterisation of the backwash process in sand filters used in micro irrigation, Biosyst. Eng. 192 (2020) 188–198, http://dx.doi.org/10.1016/j.biosystemseng.2020.01.019.

[13] J. Graciano–Uribe, T. Pujol, D. Hincapie–Zuluaga, J. Puig–Bargués, M. Duran–Ros, G. Arbat, F. Ramírez de Cartagena, Bed expansion at backwashing in pressurised porous media filters for drip irrigation: Numerical simulations and analytical equations, Biosyst. Eng. 223 Part A (2022) 277–294, http://dx.doi.org/10.1016/j.biosystemseng.2022.09.008.

[14] T. Pujol, J. Puig-Bargués, G. Arbat, M. Chaves, M. Duran-Ros, J. Pujol, F. Ramírez de Cartagena, Numerical study of the hydraulic effect of modifying the outlet pipe and diffuser plate in pressurized sand filters with wand type underdrains, J. ASABE 65 (3) (2022) 609–619, http://dx.doi.org/10.13031/ja.14710.

[15] C. Solé–Torres, J. Puig–Bargués, M. Duran–Ros, G. Arbat, J. Pujol, F. Ramírez de Cartagena, Effect of underdrain design, media height and filtration velocity on the performance of microirrigation sand filters using reclaimed effluents, Biosyst. Eng. 187 (2019) 292–304, http://dx.doi.org/10.1016/j.biosystemseng.2019.09.012.

[16] M. Duran–Ros, J. Puig–Bargués, S. Cufí, C. Solé–Torres, G. Arbat, J. Pujol, F. Ramírez de Cartagena, Effect of different filter media on emitter clogging using reclaimed effluents, Agricult. Water Manag. 258 (2022) 107591, http://dx.doi.org/10.1016/j.agwat.2022.107591.

[17] J. García Morillo, M. Martín, E. Camacho, J.A. Rodríguez Díaz, P. Montesinos, Toward precision irrigation for intensive strawberry cultivation, Agri. Water Manag. 151 (2015) 43–51, http://dx.doi.org/10.1016/j.agwat.2014.09.02.

[18] C. Alcaide Zaragoza, R. González Perea, I. Fernández García, E. Camacho Poyato, J.A. Rodríguez Díaz, Open source application for optimum irrigation and fertilization using reclaimed water in olive orchards, Comp. Electron. Agric. 173 (2020) 105407, http://dx.doi.org/10.1016/j.compag.2020.105407.

[19] J. Puig–Bargués, M. Duran–Ros, G. Arbat, J. Barragán, F. Ramírez de Cartagena, Prediction by neural networks of filtered volume and outlet parameters in micro-irrigation sand filters using effluents, Biosyst. Eng. 111 (1) (2012) 126–132, http://dx.doi.org/10.1016/j.biosystemseng.2011.11.005.

[20] A.H. Hawari, W. Alnahhal, Predicting the performance of multi-media filters using artificial neural networks, Water Sci. Tech. 74 (9) (2016) 2225–2233, http://dx.doi.org/10.2166/wst.2016.380.

[21] P.J. García–Nieto, E. García–Gonzalo, G. Arbat, M. Duran–Ros, F. Ramírez de Cartagena, J. Puig–Bargués, A new predictive model for the filtered volume and outlet parameters in micro-irrigation sand filters fed with effluents using the hybrid PSO–SVM–based approach, Comput. Electron. Agric. 125 (2016) 74–80, http://dx.doi.org/10.1016/j.compag.2016.04.031.

[22] P.J. García–Nieto, E. García–Gonzalo, J. Bové, G. Arbat, M. Duran–Ros, J. Puig–Bargués, Modeling pressure drop produced by different filtering media in microirrigation sand filters using the hybrid ABC–MARS–based approach, MLP neural network and M5 model tree, Comput. Electron. Agric. 139 (2017) 65–74, http://dx.doi.org/10.1016/j.compag.2017.05.008.

[23] P.J. García–Nieto, E. García–Gonzalo, G. Arbat, M. Duran–Ros, F. Ramírez de Cartagena, J. Puig–Bargués, Pressure drop modelling in sand filters in micro-irrigation using gradient boosted regression trees, Biosyst. Eng. 171 (2018) 41–51, http://dx.doi.org/10.1016/j.biosystemseng.2018.04.011.

[24] P. Martí, J. Shiri, M. Duran–Ros, G. Arbat, F. Ramírez de Cartagena, J. Puig–Bargués, Artificial neural networks vs. Gene expression programming for estimating outlet dissolved oxygen in micro-irrigation sand filters fed with effluents, Comput. Electron. Agric. 99 (2013) 176–185, http://dx.doi.org/10.1016/j.compag.2013.08.016.

[25] P.J. García–Nieto, E. García–Gonzalo, J. Puig–Bargués, M. Duran–Ros, F. Ramírez de Cartagena, G. Arbat, Prediction of outlet dissolved oxygen in micro-irrigation sand media filters using a Gaussian process regression, Biosyst. Eng. 195 (2020) 198–207, http://dx.doi.org/10.1016/j.biosystemseng.2020.05.009.

[26] P.J. García–Nieto, E. García–Gonzalo, J. Puig–Bargués, C. Solé–Torres, M. Duran–Ros, G. Arbat, A new predictive model for the outlet turbidity in micro-irrigation sand filters fed with effluents using Gaussian process regression, Comput. Electron. Agric. 170 (2020) 105292, http://dx.doi.org/10.1016/j.compag.2020.105292.

[27] T. Hastie, R. Tibshirani, J.H. Friedman, The Elements of Statistical Learning, Springer-Verlag, New York, USA, 2003.

[28] M. Kuhn, K. Johnson, Applied Predictive Modeling, Springer, New York, 2018.

[29] G. James, D. Witten, T. Hastie, R. Tibshirani, An Introduction to Statistical Learning: With Applications in R, Springer, New York, 2021.

[30] C.M. Bishop, Pattern Recognition and Machine Learning, Springer, Cambridge, UK, 2006.

[31] C. Smith, M. Koning, Decision Trees and Random Forests: A Visual Introduction for Beginners, Blue Windmill Media, British Columbia, Canada, 2017.

[32] M.P. Deisenroth, Mathematics for Machine Learning, Cambridge University Press, New York, USA, 2020.

[33] R. Genuer, J.-M. Poggi, Random Forests with R, Springer, San Francisco, USA, 2020.

[34] R. Storn, K. Price, Differential evolution - A simple and efficient heuristic for global optimization over continuous spaces, J. Global Optim. 11 (1997) 341–359, http://dx.doi.org/10.1023/A:1008202821328.

[35] V. Feoktistov, Differential Evolution: In Search of Solutions, Springer, New York, USA, 2006.

[36] K. Price, R.M. Storn, J.A. Lampinen, Differential Evolution: A Practical Approach to Global Optimization, Springer, New York, USA, 2006.

[37] U.K. Chakraborty, Advances in Differential Evolution, Springer, Berlin, 2008.

[38] P. Rocca, G. Oliveri, A. Massa, Differential evolution as applied to electromagnetics, IEEE Antennas Propag. 53 (1) (2011) 38–49, http://dx.doi.org/10.1109/MAP.2011.5773566.

[39] B. Vinoth Kumar, D. Oliva, P.N. Suganthan, Differential Evolution: From Theory to Practice, Springer, Singapore, 2022.

[40] A.J. Izenman, Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning, Springer, Berlin, 2013.

[41] T. Hastie, R. Tibshirani, M. Wainwright, Statistical Learning with Sparsity: The Lasso and Generalizations, CRC Press, Boca Raton, FL, USA, 2016.

[42] L. Zhang, A. Tedde, P. Ho, C. Grelet, F. Dehareng, E. Froidmont, N. Gengler, Y. Brostaux, D. Hailemariam, J. Pryce, H. Soyeurt, Mining data from milk mid-infrared spectroscopy and animal characteristics to improve the prediction of dairy cow's liveweight using feature selection algorithms based on partial least squares and elastic net regressions, Comput. Electron. Agric. 184 (2021) 106106, http://dx.doi.org/10.1016/j.compag.2021.106106.

[43] A.J. Keeney, C.L. Beseler, S.S. Ingold, County-level analysis on occupation and ecological determinants of child abuse and neglect rates employing elastic net regression, Child Abuse Negl. 137 (2023) 106029, http://dx.doi.org/10.1016/j.chiabu.2023.106029.

[44] V. Vapnik, Statistical Learning Theory, Wiley-Interscience, New York, 1998.

[45] S. Rogers, M. Girolami, A First Course in Machine Learning, Chapman and Hall/CRC, Boca Raton, FL, USA, 2016.

[46] K.P. Murphy, Machine Learning: A Probabilistic Perspective, The MIT Press, Cambridge, MA, USA, 2012.

[47] A.L. Mather, R.L. Johnson, Event-based prediction of stream turbidity using a combined cluster analysis and classification tree approach, J. Hydrol. 530 (2015) 751–761, http://dx.doi.org/10.1016/j.jhydrol.2015.10.032.

[48] Y. He, C. Chen, B. Li, Z. Zhang, Prediction of near-surface air temperature in glacier regions using ERA5 data and the random forest regression method, Remote Sens. Appl.: Soc. Environ. 28 (2022) 100824, http://dx.doi.org/10.1016/j.rsase.2022.100824.

[49] S. Kwak, J. Kim, H. Ding, X. Xu, R. Chen, J. Guo, H. Fu, Machine learning prediction of the mechanical properties of γ-TiAl alloys produced using random forest regression model, J. Mater. Res. Technol. 18 (2022) 520–530, http://dx.doi.org/10.1016/j.jmrt.2022.02.108.

[50] K.C. Onyelowe, T. Gnananandarao, A.M. Ebid, Estimation of the erodibility of treated unsaturated lateritic soil using support vector machine-polynomial and -radial basis function and random forest regression techniques, Clean. Mater. 3 (2022) 100039, http://dx.doi.org/10.1016/j.clema.2021.100039.

[51] H. Jiang, L. Mei, Y. Wei, R. Zheng, Y. Guo, The influence of the neighbourhood environment on peer-to-peer accommodations: A random forest regression analysis, J. Hosp. Tour. Manage. 51 (2022) 105–118, http://dx.doi.org/10.1016/j.jhtm.2022.02.028.

[52] J. Bové, J. Puig–Bargués, G. Arbat, M. Duran–Ros, T. Pujol, J. Pujol, F. Ramírez de Cartagena, Development of a new underdrain for improving the efficiency of microirrigation sand media filters, Agric. Water Manage. 179 (2017) 296–305, http://dx.doi.org/10.1016/j.agwat.2016.06.031.

[53] M. Duran–Ros, J. Puig–Bargués, G. Arbat, J. Barragán, R. Ramírez de Cartagena, Definition of a SCADA system for a microirrigation network with effluents, Comp. Electron. Agric. 64 (2) (2008) 338–342, http://dx.doi.org/10.1016/j.compag.2008.05.023.

[54] L. Breiman, Random forests, Mach. Learn. 45 (2001) 5–32, http://dx.doi.org/10.1023/A:1010933404324.

[55] S. Marsland, Machine Learning: An Algorithmic Perspective, Chapman and Hall/CRC Press, Boca Raton, FL, USA, 2014.

[56] R. Picard, D. Cook, Cross-validation of regression models, J. Amer. Statist. Assoc. 79 (387) (1984) 575–583, http://dx.doi.org/10.2307/2288403.

[57] D. Freedman, R. Pisani, R. Purves, Statistics, W.W. Norton & Company, New York, 2007.

[58] L. Wasserman, All of Statistics: A Concise Course in Statistical Inference, Springer, New York, 2003.

[59] C. Tien, Principles of Filtration, Elsevier, Kidlington, Oxford, UK, 2012.

[60] J. Bové, G. Arbat, M. Duran–Ros, T. Pujol, J. Velayos, J. Puig–Bargués F. Ramírez de Cartagena, Pressure drop across sand and recycled glass media used in micro irrigation filters, Biosyst. Eng. 137 (2015) 55–63, http://dx.doi.org/10.1016/j.biosystemseng.2015.07.009.

[61] G.C. Onwubolu, B.V. Babu, New Optimization Techniques in Engineering, Springer, Berlin, 2004, https://link.springer.com/book/10.1007/978-3-540-39930-8.

[62] S. Das, S.S. Mullick, P.N. Suganthan, Recent advances in differential evolution – An updated survey, Swarm Evol. Comput. 27 (2016) 1–30, http://dx.doi.org/10.1016/j.swevo.2016.01.004.

[63] A. Agresti, M. Kateri, Foundations of Statistics for Data Scientists: With R and Python, Chapman and Hall/CRC Press, Boca Raton, FL, USA, 2021.

[64] Y. Chen, R. Liu, Y. Li, X. Zhou, Research and application of cross validation of fault diagnosis for measurement channels, Prog. Nucl. Energy 150 (2022) 104324, http://dx.doi.org/10.1016/j.pnucene.2022.104324.

[65] A. Liaw, M. Wiener, Classification and regression by randomforest, R News 2 (3) (2002) 18–22, https://cogns.northwestern.edu/cbmg/LiawAndWiener2002.pdf.

[66] D. Ardia, K. Boudt, P. Carl, K.M. Mullen, B.G. Peterson, Differential evolution with deoptim: an application to non-convex portfolio optimization, R J. 3 (1) (2011) 27–34, https://journal.r-project.org/archive/2011-1/RJournal_2011-1_Ardia~et~al.pdf.