Biosystems Engineering 243 (2024) 42–56

Contents lists available at ScienceDirect

Biosystems Engineering

journal homepage: www.elsevier.com/locate/issn/15375110

Research Paper

Hybrid DE optimised kernel SVR–relied techniques to forecast the outlet turbidity and outlet dissolved oxygen in distinct filtration media and micro-irrigation filters

Paulino José García–Nieto [a,*], Esperanza García–Gonzalo [a], Gerard Arbat [b], Miquel Duran–Ros [b], Toni Pujol [c], 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, 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

Keywords:
Support vector regression (SVR)
Kernel trick
Differential evolution (DE)
Drip irrigation
Clogging

ABSTRACT

In micro-irrigation systems, distinct media filters and filtering materials are employed to remove suspended solids from irrigation water and thereby avoid emitter obstruction. Turbidity is related to suspended solids and dissolved oxygen depends on organic matter load. At this time, no models exist that are trustworthy enough to forecast the dissolved oxygen and turbidity at the outlet when utilising various media configurations and filter types. The objective of this investigation was to construct a model that can identify turbidity and dissolved oxygen at the filter outlet in advance. This study presents an algorithm for meta-heuristic optimisation inspired by populations termed Differential Evolution (DE) in conjunction with Support Vector Regression (SVR) (DE/SVR-relied model). This is an effective machine learning method, with seven kernel types for calculating the output turbidity ($Turb_o$) and the output dissolved oxygen (DO_o) from a dataset comprising 1,016 samples of various reclaimed water-using filter types. The type of media and filter, the height of the filter bed, the cycle duration, and the filtration velocity, as well as the electrical conductivity at the filter inlet, pH, inlet dissolved oxygen, water temperature, and the input turbidity are all tracked and analysed in order to achieve this. The best-fitted DE/SVR-relied model was constructed to predict the $Turb_o$ and DO_o as well as the input variables' relative importance. Determination coefficients for the best-fitted DE/SVR-relied model for the testing dataset were 0.89 and 0.92 for outlet turbidity ($Turb_o$) and outlet dissolved oxygen (DO_o), respectively, showing a good predictive performance which are of great importance for the management of drip irrigation systems.

1. Introduction

In many places, irrigation is crucial to sustaining economic growth and ensuring food production (FAO, 2022). The current rising water demand for non-agricultural uses, declining freshwater sources, and growing irregularity in water availability due to more frequent drought periods pose a challenge to irrigated agriculture (FAO, 2022). The response has been improving on-farm irrigation systems, adopting more efficient irrigation methods, shifting to more effective management practices, and using non-conventional water sources (FAO, 2022; Tarjuelo et al., 2015). Among other irrigation methods, micro-irrigation shows increased water use efficiency as well as decreased energy requirements compared to other types of pressurised systems. However, the complete or partial clogging of emitters is still one of the major problems of micro-irrigation systems (Ayars, Bucks, Lamm, & Nakayama, 2007; Tien, 2012).

Irrigation water quality, which is primarily influenced by the suspended particle load, chemical composition, microbiological population, and their interactions, is directly linked to emitter clogging (Nakayama, Boman, & Pitts, 2007; Tien, 2012). To assess emitter clogging risk, irrigation water quality should be periodically checked to determine if any additional water treatment needs to be conducted. Instead of following time-consuming laboratory procedures for determining quality parameters, the use of specific sensors allows quick and accurate measurements. To this regard, there are sensors developed for analysing pH, electrical conductivity, turbidity (which is related to suspended solids) and dissolved oxygen (which is inversely related to

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

https://doi.org/10.1016/j.biosystemseng.2024.04.020
Received 29 November 2023; Received in revised form 22 March 2024; Accepted 28 April 2024

Nomenclature

Abbreviations

C	Regularization constant
DE	Differential evolution
DO	Dissolved oxygen
PUK	Pearson VII Universal kernel
r	Correlation coefficient
R^2	Coefficient of determination
RBF	SCADA SE Radial basis function Supervisory control and data acquisition Squared-exponential
SVM	Support vector machines
SVR	Support vector regression
Turb	Turbidity, FNU
v	Filtration velocity, m h^{-1}

Symbols

a	Parameter of the polynomial and sigmoid kernels
b	Parameter of the polynomial kernel
DO_i	Dissolved oxygen at filter inlet, mg l^{-1}
DO_o	Dissolved oxygen at filter outlet, mg l^{-1}
EC_i	Electrical conductivity at the inlet, μS mm^{-1}
H	Media bed height (of the filter), m
$k(\mathbf{x}_i, \mathbf{x}_j)$	Kernel function of the SVM model
pH_i	Degree of acidity of solution at the inlet
Tc	Cycle duration of filtration, min
T_i	Water temperature, °C
$Turb_i$	Turbidity at the filter inlet, FNU
$Turb_o$	Turbidity at the filter outlet, FNU
ε	Maximum width of the insensitive tube
ξ^+, ξ^-	Slack variables
σ	Parameter of the RBF and PUK kernels
σ_f	Parameter of the matern32 and Matern52 kernels
σ_l	Parameter of the matern32 and mater52 kernels
ω	Parameter of the PUK kernel

the organic pollution in the water).

A fundamental maintenance procedure for preventing emitter clogging in sand media filters is the removal of suspended inorganic and organic particles (Nakayama et al., 2007). This is especially important when using high particle loaded irrigation water (Capra & Scicolone, 2007; Ravina et al., 1997; Tien, 2012). Since silica sand is readily available and inexpensive, it is the medium most frequently used in micro-irrigation media filters (Nakayama et al., 2007). However, using alternative materials with comparable physical properties, like recycled glass, can improve filter performance and result in more energy-efficient use (Cescon & Jiang, 2020). However, the environmental impact of the media materials depend mainly of the energy used in their manufacture, which is usually higher for recycled glass (Pujol et al., 2022a).

In an effort to lower head loss and energy consumption, a number of authors have proposed redesigning the sand media filters (Tien, 2012; Bové, Arbat, Duran-Ros, et al., 2015; Bové et al., 2017; Mesquita, de Deus, Testezlaf, da Rosa, & Diotto, 2019; Pujol et al., 2020; Pujol et al., 2022b). Differences in underdrain designs of the sand media filters and in their operation conditions, such as filtration velocity and media height, significantly impacted the emitter obstruction (Solé–Torres et al., 2019b), dissolved oxygen, and turbidity removal (Solé–Torres et al., 2019a), which also depends on the filtration material (Duran–Ros et al., 2022).

A proper management of micro-irrigation systems requires an accurate prediction of the efficacy of sand media filters in removing solids and, consequently, lowering the risk of emitter clogging, given the variability of irrigation water characteristics and filter operation conditions.

The prediction of turbidity at the filter outlet ($Turb_o$) and Dissolved oxygen at filter outlet (DO_o) for various types of media filters is a complex, non-linear problem, therefore it is difficult to extrapolate their values in a simple way. A reasonable prediction accuracy of hydraulic parameters, as well as DO_o and $Turb_o$ for media filters, has previously been made feasible by employing advanced methods like neural networks (Hawari & Alnahhal, 2016; Puig–Bargués, Duran–Ros, Arbat, Barragán, & Ramírez de Cartagena, 2012), regression trees with gradient boost and hybrid algorithms (García–Nieto et al., 2017; García–Nieto et al., 2018), gene expression programming (Martí et al., 2013), and Gaussian process regression (García–Nieto et al., 2020a, García–Nieto et al., 2020b). However, further methods that might result in better prediction accuracy should be investigated.

One such potential method is a unique regressive model that relies on the Support Vector Machines (SVMs) approach (Cristianini & Shawe–Taylor, 2000; Hastie, Tibshirani, & Friedman, 2003; Schölkopf, Smola, Williamson, & Bartlett, 2000; Vapnik, 1998). SVMs were initially created to address classification issues. Later they were expanded to address regression issues. In the case of regression, this method is termed Support Vector Regression (SVR) (Bishop, 2006; Deisenroth, 2020; Hansen & Wang, 2005; Hastie et al., 2003; James, Witten, Hastie, & Tibshirani, 2021; Kuhn & Johnson, 2018; Li, Lord, Zhang, & Xie, 2008; Steinwart & Christmann, 2008). The SVR approach is a supervised learning method that can be utilised to address issues with regression because of its extreme robustness and capacity to handle nonlinearities (Cristianini & Shawe–Taylor, 2000; Hastie et al., 2003; Schölkopf et al., 2000; Vapnik, 1998). There are a number of benefits to using the SVM approach over traditional and meta-heuristic regression techniques (Bishop, 2006; Deisenroth, 2020; Hansen & Wang, 2005; Hastie et al., 2003; James et al., 2021; Kuhn & Johnson, 2018; Li et al., 2008; Steinwart & Christmann, 2008): (1) the SVM approach only requires a subset of the training data since only the training data are used in the cost function that constructs the model; (2) support vectors, a subset of the decision function's training points, are used by the SVM model to improve its accuracy and memory efficiency in high-dimensional spaces; (3) SVM technique offers the kernel trick inside the algorithm itself, so that the regression of nonlinear data has become easier and any complex problem can be solved; and (4) SVM approach is robust to outliers when hyper-parameter tuning is done correctly. As it is less influenced by outliers, the precision of the results is improved.

In evolutionary computation, differential evolution (DE) (Storn & Price, 1997) is a method that optimises a problem iteratively, trying to improve a candidate solution with regard to a given measure of quality. Such methods are commonly known as metaheuristics as they make few or no assumptions about the optimised problem and can search very large spaces of candidate solutions. DE is used for multidimensional real-valued functions but does not use the gradient of the problem being optimised, which means DE does not require the optimisation problem to be differentiable, as is required by classic optimisation methods such as gradient descent and quasi-newton methods. In this work DE optimiser is used in combination with SVM technique to obtain its optimal hyperparameters.

Since this type of complex problem has not, to the authors' knowledge, been addressed in previous research, the SVR approximation (Cristianini & Shawe–Taylor, 2000; Hastie et al., 2003; Schölkopf et al., 2000) in conjunction with Differential Evolution (DE) (Chakraborty, 2008; Feoktistov, 2006; Price, Storn, & Lampinen, 2006; Rocca, Oliveri, & Massa, 2011; Storn & Price, 1997; Vinoth Kumar, Oliva, & Suganthan, 2022) for tuning the parameters of the various types of kernels seems like a promising approach to solving the kinds of high-nonlinearities

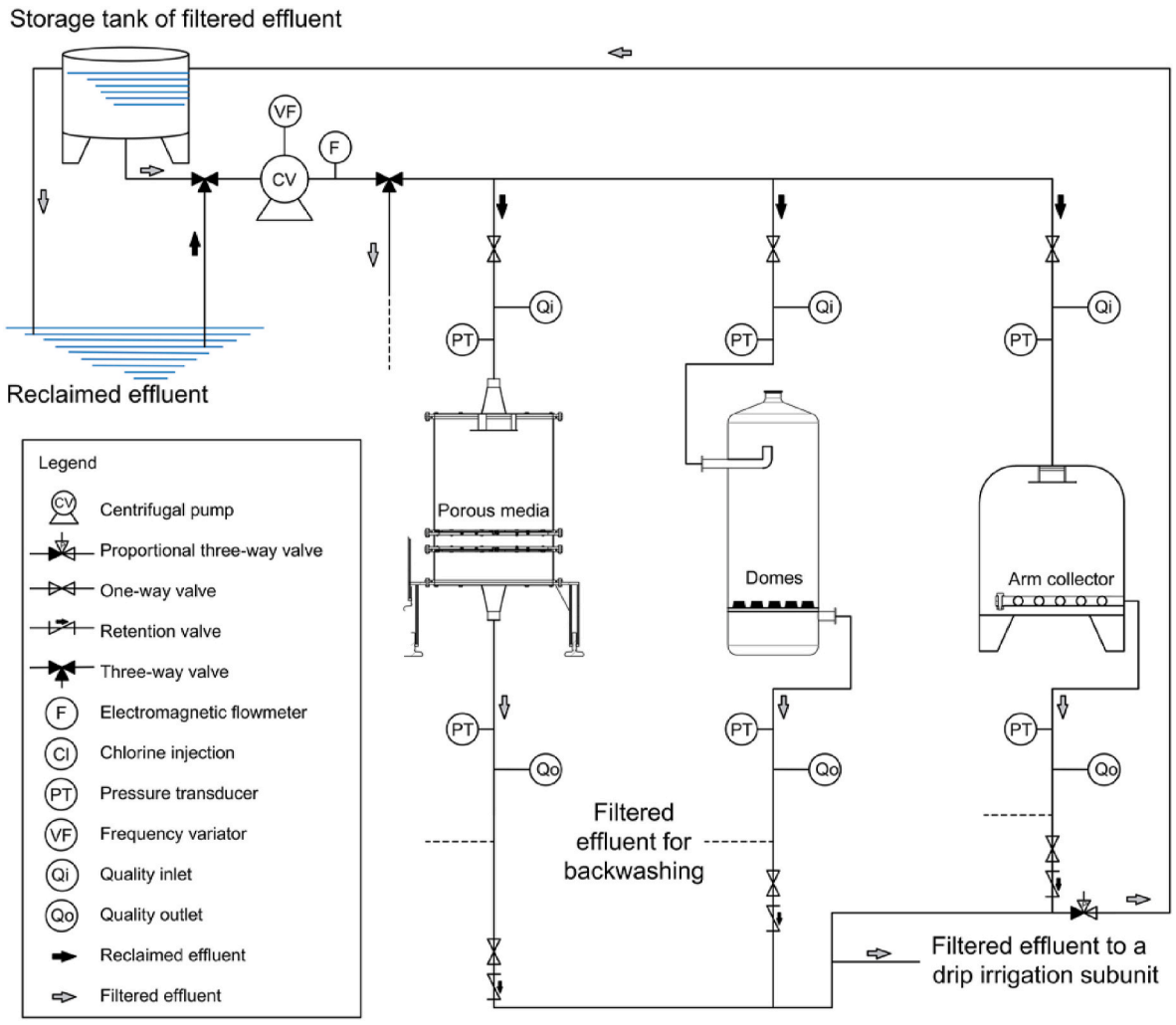

Fig. 1. Setup of the experimental filtration platform.

associated with predicting $Turb_o$ and DO_o. For the purpose of comparison, the same experimental dataset was modified using the DE/SVR–relied technique with seven distinct types of kernels with the purpose of forecasting the $Turb_o$ and DO_o and compare the results obtained with the experimental values observed (Bishop, 2006; Hansen & Wang, 2005; Hastie et al., 2003; Kuhn & Johnson, 2018; Steinwart & Christmann, 2008).

Preceding investigations have shown that SVM is a highly useful tool in many different fields, including hydro-climatic factors (Shrestla & Shukla, 2015), solar energy (Chen, Liu, Li, Zhou, 2013) or photovoltaic power (De Leone, Pietrini, & Giovannelli, 2015), among others, but it has never been applied to media filters. This work, using a DE/SVR-based model, hypotheses that it is an approach that could provide an effective solution in accurately predicting the $Turb_o$ and DO_o in media filters commonly utilised in micro-irrigation systems.

The structure of this paper is as follows: The experimental design is presented in Section 2 along with a breakdown of the factors that went into the investigation. Through the compilation of the DE/SVR results with the experimental values and the order of relevance of the input parameters, Section 3 provides insights obtained from this reliable technique. Finally, Section 4 concludes the study by summarizing its main findings.

2. Materials and methods

2.1. Experimental setup

The wastewater treatment plant (WWTP) of Celrà (Girona, Spain) provided recovered effluents, which were filtered using three different media filters (see Fig. 1) featuring porous media (designed by Bové et al., 2017), domes (model FA-F2-188, Regaber, Parets del Vallès, Spain), and an arm collector (model FA1M, Lama, Gelves, Spain) underdrains.

The recovered effluent was pumped to the filters. Measurements were made of the filter's inlet and outlet pressures as well as its flow. For filter backwashing, water from a storage tank was utilised, and the majority of the filtered effluent was sent to a drip irrigation sub-unit. A membrane pump was used to continuously inject chlorine to reach a concentration of 4 mg l^{-1} in the water used to backwash the filters and 2 mg l^{-1} in the filtered effluent. At the filter inlet, the pH, temperature, and electrical conductivity of the effluent were measured. At the filter's inlet and outlet, Turb and DO were measured, respectively. Further details of the experimental setup can be found in Duran-Ros et al., 2022.

The experimental facility could be operated as well as data on pressure, filter inlet flow, filtration cycle duration, and water quality parameters could be collected and recorded on a minute-by-minute basis because of a system for supervisory control and data acquisition, or SCADA (Duran–Ros, Puig–Bargués, Arbat, Barragán, & Ramírez de Cartagena, 2008). The SCADA system was intended to halt operations when the $Turb_i$ exceeded 50 FTU, thereby preventing operation with

Table 1
The physical operational variables used in this research, average values and standard deviations (STDs).

Input variables	Name of the variable	Mean	STD
Media	Media	–	–
Filter media type	Filter	–	–
Media bed height (of the filter) (m)	H	0.2666	0.0472
Filtration velocity (m/h)	v	52.101	13.219
Cycle duration (of filtration) (min)	Tc	251.90	268.06
Electrical conductivity (μS/cm)	EC_i	1966.6	708.20
Inlet 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**
Output turbidity (FNU)	$Turb_o$	4.7604	1.1381
Output dissolved oxygen (mg/l)	DO_o	3.7104	0.9128

high particle loads.

A CA-07MS silica sand, with an effective diameter of 0.48 mm and a uniformity coefficient of 1.73 and a porosity of 0.39 (Silbeco Minerales, Bilbao, Spain), and NW2 crushed recycled glass Nature Works Tecnologías, with an effective diameter D10 of 0.44 mm, a uniformity coefficient of 1.59, and a porosity of 0.54 (L'Alfàs del Pi, Spain) were the two tested filtration media.

For a total of 250 h, media bed heights of 200 and 300 mm and filtration velocities of 30 and 60 m h^{-1} were checked. Two 4-h sessions per day were dedicated to running the filters for each material and underdrain design. A 3-min automatic backwash was initiated when filter pressure drop reached 50 kPa.

2.2. Materials and model variables

The primary aim of this research is to evaluate the partnership between distinct parameters that were measured experimentally and the inputs that the DE/SVR model required in order to calculate the DO_o and $Turb_o$. The output variable for micro-irrigation systems is the $Turb_o$, which is a measure of the irrigation water's quality and is connected to the probability of physical obstruction of emitters. The operation's input variables are as follows (refer to Table 1):

- Media: every single filtering media (silica sand and recycled glass) mentioned in subsection 2.1. It is a categorical type variable;
- Filter: section 2.1 provided an explanation of the three kinds of sand media filters: those with dome, porous, and arm collector underdrains. This variable is categorical;
- Filter bed height (mm): for sand filters, this is an operational variable. Various filter bed heights (200 and 300 mm) were checked for each filter;
- Filtration velocity (m h^{-1}): this has an impact on filter performance. Two filter test velocities (30 and 60 m h^{-1}) were applied to each filter.
- Cycle duration (s): is the time that the filter operates in filtration mode;
- Electrical conductivity (S mm^{-1}): is a water quality indicator and in particular for water salinity, which can limit the use of the effluent for micro-irrigation (Ravina et al., 1997);
- Input dissolved oxygen (mg l^{-1}): this parameter controls the ability of water to sustain life and is commonly used as a control in wastewater treatment plants because it is inversely correlated with the amount of oxygen that microorganisms consume during the breakdown of organic matter, a process known as biochemical oxygen demand, or BDO;

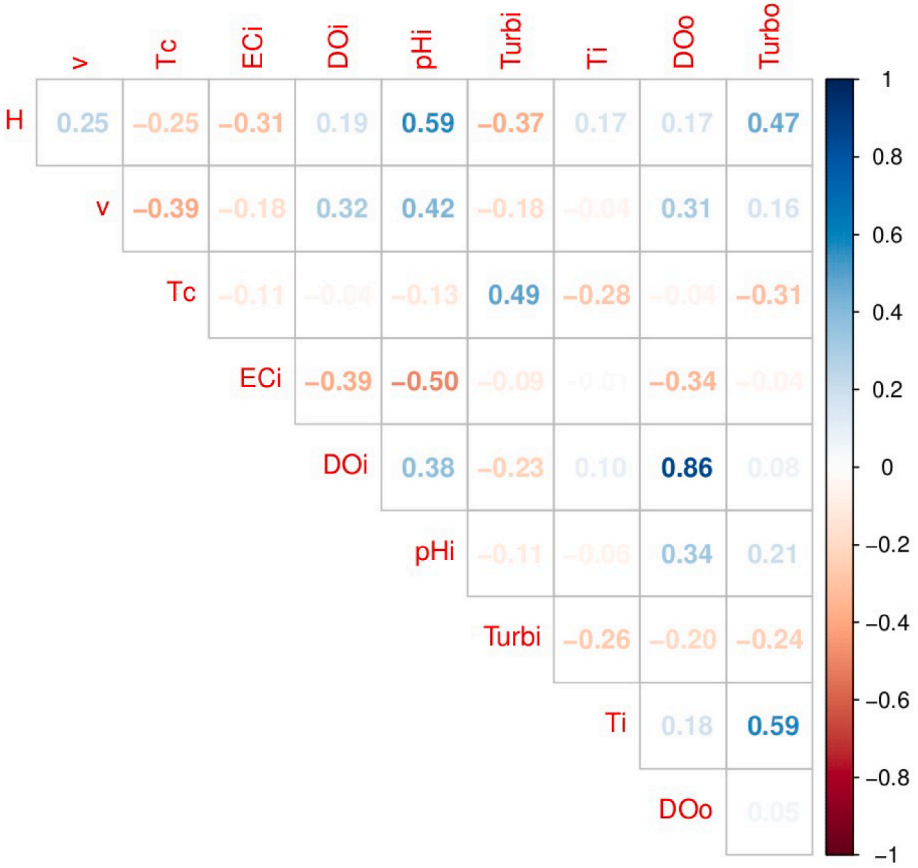

Fig. 2. Correlation matrix of the process variables.

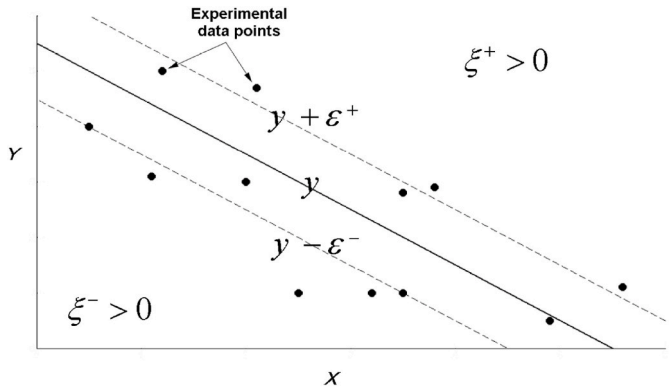

Fig. 3. A single $\varepsilon-$ insensitive tube in a SVM regression's scenario.

- pH: it measures the degree of acidity or alkalinity;
- Water temperature (°C): at the inlet of the filter;
- Input turbidity (measured in formazin nephelometric units or FNU): is an important indicator of the water quality at the inlet of the filter, measures the clarity of the water and is related with the amount of suspended solids.

The operational output variables of the study are as follows:

- Output turbidity (FNU): this factor is used to evaluate the water quality at the filter's outlet. Moreover, measuring the possibility of emitter clogging irrigation water.
- Output dissolved oxygen (mg l^{-1}): it has an impact on how well water can sustain life and if the irrigation water has low DO levels then it may contribute to plant root hypoxia and cause agronomic problems.

As an initial data analysis, a correlation matrix is calculated for all the variables that take part in the process. Fig. 2 shows graphically how close the two variables are to having a linear relationship with each other. Each variable in the Table 1 is correlated with each of the other variables.

2.3. Techniques for mathematical modeling

2.3.1. Support vector machines (SVM) method

SVMs were created to deal with issues related to binary classification. Due to the conditions, it was quickly found that the basic concepts could be used to solve a range of issues, such as regression issues (Bishop, 2006; Deisenroth, 2020; Hansen & Wang, 2005; Hastie et al., 2003; James et al., 2021; Kuhn & Johnson, 2018; Li et al., 2008; Steinwart & Christmann, 2008). Next, as an example, for a dataset where the training set consists of the values of the continuous dependent variable $y_i \in \Re, \forall i = 1, 2, ..., m$ and the covariates $\boldsymbol{x}_i \in \Re^p, i = 1,2,...,m$, the SVR technique creates a function $f(\boldsymbol{x}) = \boldsymbol{w}^T\boldsymbol{x} + b$ where $\mathbf{w}$ signifies the hyperplane's perpendicular vector, also known as the director vector, and $b/\ \|\boldsymbol{w}\|$ stands for the distance measured perpendicularly between the hyperplane and the coordinate origin. This approach must accommodate a maximum deviation from the true value y_i of ε while also being as flat as possible for all $\boldsymbol{x}_i$ training cases. The Euclidean norm $\|\boldsymbol{w}\|_2$ is minimised to achieve flatness, and additionally, by penalising the total of deviations that exceed ε. In actuality, the following optimisation problem is what the SVR approach seeks to solve:

$$\min_{\boldsymbol{w},b,\xi^+,\xi^-} \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{m}\left(\xi_i^+ + \xi_i^-\right) \tag{1}$$

liable to

$$\left\{\begin{array}{ll} y_i - \left(\boldsymbol{w}^T\boldsymbol{x}_i + b\right) \geq \varepsilon + \xi_i^+ & i = 1, ..., m \\ \left(\boldsymbol{w}^T\boldsymbol{x}_i + b\right) - y_i \geq \varepsilon + \xi_i^- & i = 1, ..., m \\ \xi_i^+, \xi_i^- \geq 0 & i = 1, ..., m \end{array}\right\} \tag{2}$$

so that $\boldsymbol{\xi}^+, \boldsymbol{\xi}^- \in \Re^m$ are mentioned as slack variables and C is known as the regularization constant. With the purpose of mitigating the overfitting, the factor C in Eq. (1) requires a positive numerical value, which restricts the penalisation imposed on observations outside the interval ε. This value establishes the trade-off between the reduced complexity of the model and the horizontality of the objective function (Cristianini & Shawe–Taylor, 2000; Hastie et al., 2003; Schölkopf et al., 2000; Li et al., 2008; Steinwart & Christmann, 2008; Kuhn & Johnson, 2018; Deisenroth, 2020; James et al., 2021). Every training vector has slack variables, which allow deviations greater than ε but penalises those deviations in the objective function. An $\varepsilon-$ insensitive tube is the region that $y_i \pm \varepsilon, \forall i$ encloses (see Fig. 3).

Predicting DO_o and $Turb_o$ a highly nonlinear problem, so the kernelisation technique can be used. The methodology's foundation is the original dataset's mapping to the feature space, which is space H with a higher dimension. A kernel function $k(\boldsymbol{x}_i, \boldsymbol{x}_j)$ is used to carry out the application, and it finds a scalar product in H. With the purpose of solving the primal optimisation issue as presented by Eq. (1), the problem can be formulated it the following manner, using its dual form. When the Karush-Kuhn-Tucker (KKT) conditions are met, the optimisation problem is dual-formulated (Cristianini & Shawe–Taylor, 2000; Deisenroth, 2020; Hastie et al., 2003; James et al., 2021; Kuhn & Johnson, 2018; Li et al., 2008; Schölkopf et al., 2000; Steinwart & Christmann, 2008) such that:

$$\max_{\boldsymbol{\alpha}^+, \boldsymbol{\alpha}^-} \sum_{i=1}^{m} y_i\left(\alpha_i^+ - \alpha_i^-\right) - \varepsilon\sum_{i=1}^{m}\left(\alpha_i^+ + \alpha_i^-\right) - \frac{1}{2}\sum_{ij=1}^{m}\left(\alpha_i^+ - \alpha_i^-\right)\left(\alpha_j^+ - \alpha_j^-\right)k(\boldsymbol{x}_i, \boldsymbol{x}_j) \tag{3}$$

liable to

$$\left\{\begin{array}{ll} \sum_{i=1}^{m}\left(\alpha_i^+ - \alpha_i^-\right) = 0, & \\ 0 \leq \alpha_i^+ \leq C, & i = 1, ..., m \\ 0 \leq \alpha_i^- \leq C, & i = 1, ..., m \end{array}\right\} \tag{4}$$

The regression forecasting for a modern sample $\mathbf{x}$ can be computed using the function $f(\boldsymbol{x})$ listed below (Eqn. (5)) (Cristianini & Shawe–Taylor, 2000; Deisenroth, 2020; Hastie et al., 2003; James et al., 2021; Kuhn & Johnson, 2018; Li et al., 2008; Schölkopf et al., 2000; Steinwart & Christmann, 2008):

$$f(\boldsymbol{x}) = \sum_{i=1}^{m}\left(\alpha_i^+ - \alpha_i^-\right)k(\boldsymbol{x}, \boldsymbol{x}_i) + b \tag{5}$$

The technical bibliography uses a variety of common functions as kernels (Cristianini & Shawe–Taylor, 2000; Deisenroth, 2020; Hastie et al., 2003; James et al., 2021; Kuhn & Johnson, 2018; Li et al., 2008; Schölkopf et al., 2000; Steinwart & Christmann, 2008). If $q = \left\|\boldsymbol{x}_i - \boldsymbol{x}_j\right\|_2$, then the following kernels types can be expressed as

- Linear kernel:

$$k(\boldsymbol{x}_i, \boldsymbol{x}_j) = \boldsymbol{x}_i \cdot \boldsymbol{x}_j \tag{6}$$

- Polynomial kernel:

$$k(\boldsymbol{x}_i, \boldsymbol{x}_j) = \left(\sigma \boldsymbol{x}_i \cdot \boldsymbol{x}_j + a\right)^b \tag{7}$$

- Sigmoid kernel:

$$k(\boldsymbol{x}_i,\boldsymbol{x}_j)=tanh(\sigma\boldsymbol{x}_i\cdot\boldsymbol{x}_j+a) \tag{8}$$

- RBF kernel (kernel of the Radial Basis Function):

$$k(\boldsymbol{x}_i,\boldsymbol{x}_j)=e^{-\sigma r^2} \tag{9}$$

- Pearson VII Universal kernel (PUK):

$$k(\boldsymbol{x}_i,\boldsymbol{x}_j)=\frac{1}{\left[1+\left(\frac{2r\sqrt{2^{(1/\omega)}-1}}{\sigma}\right)^2\right]^{\omega}} \tag{10}$$

- Matern32:

$$k(\boldsymbol{x}_i,\boldsymbol{x}_j)=\sigma_f\left(1+\frac{\sqrt{3}q}{\sigma_l}\right)e^{-\frac{\sqrt{3}q}{\sigma_l}} \tag{11}$$

- Matern52:

$$k(\boldsymbol{x}_i,\boldsymbol{x}_j)=\sigma_f\left(1+\frac{\sqrt{5}q}{\sigma_l}+\frac{5q^2}{3\sigma_l^2}\right)e^{-\frac{\sqrt{5}q}{\sigma_l}} \tag{12}$$

so that a kernel's typology is determined by the a, b, σ, σ_f, σ_l and ω parameters.

In conclusion, choosing the right kernel type and its ideal parameters is essential with the purpose of mapping data that is nonlinearly separable in a feature space (higher-dimensional space) and using the SVM approach to solve a regression problem.

Moreover, the following succinctly describes representative parameters of the SVR technique (Cristianini & Shawe–Taylor, 2000; Deisenroth, 2020; Hastie et al., 2003; James et al., 2021; Kuhn & Johnson, 2018; Li et al., 2008; Schölkopf et al., 2000; Steinwart & Christmann, 2008):

- ε hyperparameter: This number sets a maximum width for the permitted margin of error. The empirical error is computed using the insensitive loss function, which accounts for errors smaller than ε, which is the second objective function term that was reliant on the ε factor (see Eq. (3)).
- Regularisation constant: The cost function is another name for this constant C. This parameter represents the trade-off between the slack variables and margin. It is one of the SVR hyperparameters that needs to be pre-tuned.
- a, b, σ, σ_f, σ_l and ω: The final model's various kernels' mathematical expressions are defined by these parameters.

Therefore, it is practical to use a mathematical method that accurately determines the aforementioned hyperparameters. The meta-heuristic DE optimiser outlined below was successfully applied in this investigation. Using the python package termed scikit-learn for the SVR (Chang & Lin, 2011), the final hybrid DE/SVR models were built.

2.3.2. Optimisation based on the differential evolution (DE) algorithm

DE is a meta-heuristic approach in evolutionary computation that attempts to improve the quality of a potential solution iteratively with the purpose of optimising a problem. Using the DE optimiser for multidimensional real-valued data can be successful even if the optimised function is not differentiable. Furthermore, problems that are noisy, non-continuous, or change over time can also be solved with a DE optimiser. DE chooses the solution that best fits the specified optimisation problem by utilising a populace of feasible solutions and merging pre-existing ones with simple mathematical formulas (Storn & Price, 1997). The optimisation problem variables are represented by the algorithm as a vector of real numbers. The quantity of parameters in the problem of optimisation, denoted by the length n of the NP vectors that comprise the actual population, makes up the population.

A vector can be defined as $\boldsymbol{x}_p^g$ if p is its index among the populace ($p=1,\ldots,NP$), and g is its generation. The variables in problem $x_{p,m}^g$, where m denotes the variable's index within each individual ($m=1,\ldots,n$), make up the components of this vector. The variables are contained within intervals that are, at minimum and maximum, bound by the values $\boldsymbol{x}_m^{min}$ and $\boldsymbol{x}_m^{max}$, respectively.

The DE algorithm consists of the following four steps (Chakraborty, 2008; Feoktistov, 2006; Price et al., 2006; Rocca et al., 2011; Storn & Price, 1997; Vinoth Kumar et al., 2022): (1) initialisation; (2) mutation; (3) recombination; and (4) selection. The DE optimiser was implemented here using the python package called SciPy (Agresti & Kateri, 2021).

Following initialisation, the search is initiated. The stages of mutation, recombination, and selection come to an end whenever a criterion for stopping is met (number of generations, duration, degree of solution reached, etc …).

2.3.2.1. *Initialisation.*

Every variable is initialised (first generation) at random, taking into consideration the variable's lowest and highest values (Chakraborty, 2008; Feoktistov, 2006; Price et al., 2006; Rocca et al., 2011; Storn & Price, 1997; Vinoth Kumar et al., 2022):

$$\boldsymbol{x}_{p,m}^1=\boldsymbol{x}_m^{min}+rand(0,1)\cdot\left(\boldsymbol{x}_m^{max}-\boldsymbol{x}_m^{min}\right)\quad for\ p=1,\ldots,NP\ \ and\ \ m=1,\ldots,n \tag{13}$$

such that a random number in [0,1] is represented by $rand(0,1)$.

2.3.2.2. *Mutation.*

The three randomly chosen subjects, referred to as the target vectors $\boldsymbol{x}_a$, $\boldsymbol{x}_b$ and $\boldsymbol{x}_c$, are utilised while developing the NP new vectors that comprise the mutation. The new vectors $\boldsymbol{n}_p^t$ are made in the manner described below (Chakraborty, 2008; Feoktistov, 2006; Price et al., 2006; Rocca et al., 2011; Storn & Price, 1997; Vinoth Kumar et al., 2022):

$$\boldsymbol{n}_p^g=\boldsymbol{x}_c+F\cdot(\boldsymbol{x}_a-\boldsymbol{x}_b)\ \ for\ \ p=1,\ldots,NP \tag{14}$$

that differ with a, b, c, and p. The parameter F governs the rate of mutation. Moreover, F lies in the interval [0,2].

2.3.2.3. *Recombination.*

Following the creation of the NP new vectors, a random recombination is performed, and the process of creating the trial vectors $\boldsymbol{t}_m^g$ involves comparing the result to the original vectors $\boldsymbol{x}_p^g$ (Chakraborty, 2008; Feoktistov, 2006; Price et al., 2006; Rocca et al., 2011; Storn & Price, 1997; Vinoth Kumar et al., 2022):

$$t_{p,m}^g=\left\{\begin{array}{ll} n_{p,m}^g & if\quad rand(0,1)<GR \\ x_{p,m}^g & otherwise \end{array}\right\}\ for\ p=1,\ldots,NP\ \ and\ \ m=1,\ldots,n \tag{15}$$

The parameter GR regulates the rate of recombination. The test vector will contain both the original and updated vectors because the comparison is carried out variable by variable.

2.3.2.4. *Selection.*

We now compare the test vectors to the original vectors with the purpose of determining which vector will be the most valuable in the upcoming generation as indicated by the fitness function (Chakraborty, 2008; Feoktistov, 2006; Price et al., 2006; Rocca et al., 2011; Storn & Price, 1997; Vinoth Kumar et al., 2022), according to:

$$\boldsymbol{x}_p^{g+1}=\left\{\begin{array}{ll} \boldsymbol{t}_p^g & if\quad fit\left(\boldsymbol{t}_p^g\right)>fit\left(\boldsymbol{x}_p^g\right) \\ \boldsymbol{x}_p^g & otherwise \end{array}\right\} \tag{16}$$

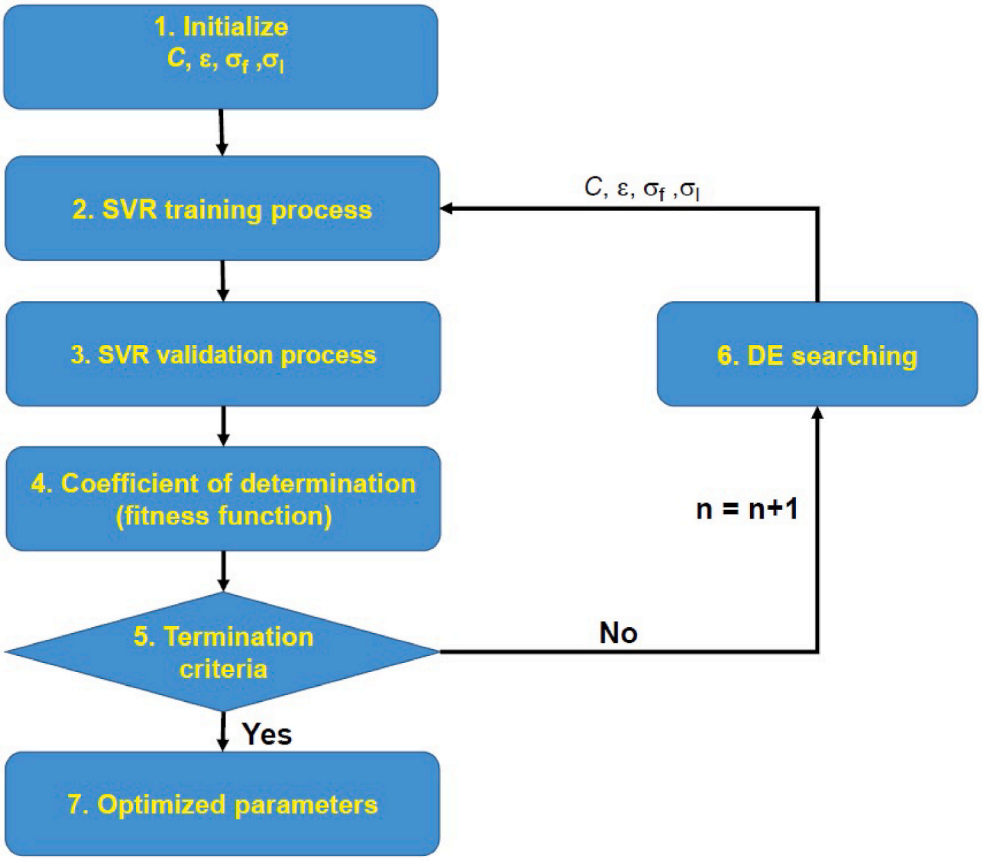

Fig. 4. Process diagram employing the SVR–Matern52 model using DE for parameter optimisation.

Table 2
Intervals of variation for the DE/SVR–based approach's hyperparameters that were fitted for this investigation.

SVR hyperparameters	Lower limit	Upper limit
C	10^{-6}	10^{2}
ε	10^{-6}	10^{2}
$\sigma, \sigma_f, \sigma_l$	10^{-6}	10^{2}
a	10^{-6}	10^{2}
b	0.5	5
ω	1	10

Table 3
The ideal hyperparameters for the $Turb_o$ foretelling in the fitted DE/SVR–based models.

Model–kernel	Values of optimal hyperparameters
DE/SVR–Linear	Regularization factor $C = 6.50 \times 10^{0}$, $\varepsilon = 7.67 \times 10^{-2}$
DE/ SVR–Polynomial	Regularization factor $C = 4.23 \times 10^{-3}$, $\varepsilon = 9.88 \times 10^{-6}$, $\sigma = 2.91 \times 10^{-1} a = 8.97 \times 10^{1}$, $b = 3.91 \times 10^{0}$
DE/SVR–Sigmoid	Regularization factor $C = 4.63 \times 10^{1}$, $\varepsilon = 9.26 \times 10^{-3}$, $\sigma = 2.01 \times 10^{-3} a = 1.20 \times 10^{-5}$
DE/SVR–RBF	Regularization factor $C = 3.56 \times 10^{1}$, $\varepsilon = 1.87 \times 10^{-2}$, $\sigma = 5.79 \times 10^{-1}$
DE/SVR–PUK	Regularization factor $C = 3.53 \times 10^{1}$, $\varepsilon = 6.37 \times 10^{-4}$, $\sigma = 6.42 \times 10^{-1}$, $\omega = 1.11 \times 10^{0}$
DE/ SVR–Mattern32	Regularization factor $C = 1.46 \times 10^{0}$, $\varepsilon = 3.58 \times 10^{-5}$, $\sigma_f = 2.14 \times 10^{0}$, $\sigma_l = 5.01 \times 10^{-1}$
DE/ SVR–Mattern52	Regularization factor $C = 2.06 \times 10^{0}$, $\varepsilon = 7.23 \times 10^{-3}$, $\sigma_f = 1.55 \times 10^{0}$, $\sigma_l = 2.66 \times 10^{-1}$

Table 4
The ideal hyperparameters for the DO_o foretelling in the fitted DE/SVR–based models.

Model–kernel	Values of optimal hyperparameters
DE/SVR–Linear	Regularization factor $C = 5.91 \times 10^{-1}$, $\varepsilon = 6.57 \times 10^{-2}$
DE/ SVR–Polynomial	Regularization factor $C = 3.04 \times 10^{-4}$, $\varepsilon = 1.52 \times 10^{-2}$, $\sigma = 4.85 \times 10^{0} a = 1.49 \times 10^{0}$, $b = 4.65 \times 10^{0}$
DE/SVR–Sigmoid	Regularization factor $C = 2.09 \times 10^{1}$, $\varepsilon = 6.41 \times 10^{-2}$, $\sigma = 7.83 \times 10^{-3} a = 6.52 \times 10^{-3}$
DE/SVR–RBF	Regularization factor $C = 2.42 \times 10^{1}$, $\varepsilon = 1.47 \times 10^{-2}$, $\sigma = 4.20 \times 10^{-1}$
DE/SVR–PUK	Regularization factor $C = 2.15 \times 10^{1}$, $\varepsilon = 1.73 \times 10^{-2}$, $\sigma = 2.79 \times 10^{0}$, $\omega = 1.75 \times 10^{0}$
DE/ SVR–Mattern32	Regularization factor $C = 7.17 \times 10^{-2}$, $\varepsilon = 5.81 \times 10^{-3}$, $\sigma_f = 1.20 \times 10^{1}$, $\sigma_l = 1.64 \times 10^{0}$
DE/ SVR–Mattern52	Regularization factor $C = 8.14 \times 10^{-1}$, $\varepsilon = 3.81 \times 10^{-5}$, $\sigma_f = 2.03 \times 10^{0}$, $\sigma_l = 6.05 \times 10^{-1}$

2.4. Precision of this approach

The DE/SVR model was constructed using ten input variables. The DO_o and $Turb_o$ are the dependent predicted variables. The coefficient of determination (R^2) is the primary goodness–of–fit statistic for the regression problem discussed in this article (Freedman, Pisani, & Purves, 2007; Picard & Cook, 1984). If if t_i and y_i, respectively, represent the observed and predicted values, it considers the subsequent expressions (Freedman et al., 2007; Wasserman, 2003):

- $SS_{tot} = \sum_{i=1}^{n} (t_i - \bar{t})^2$: stands for the sum of squares clarified;

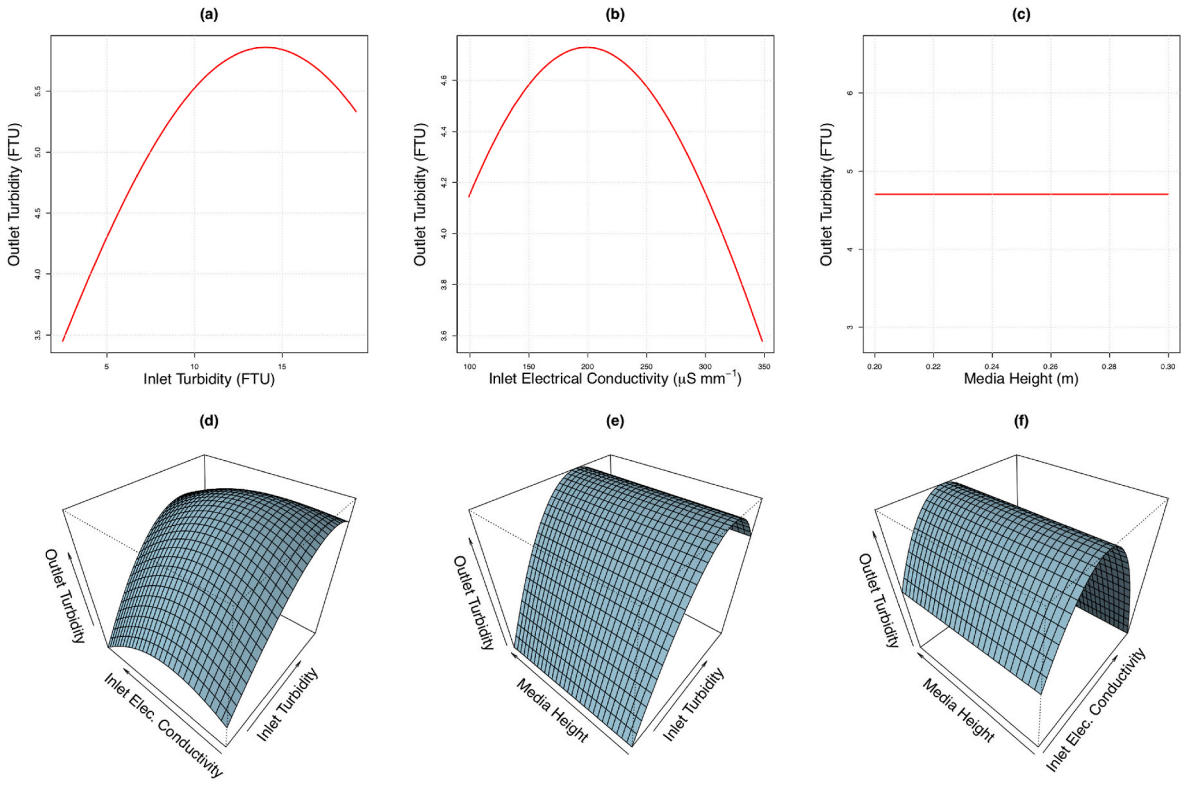

Fig. 5. Terms of first and second order for the DE/SVR–Matern52 model's three key variables that determine how the $Turb_o$ will be predicted.

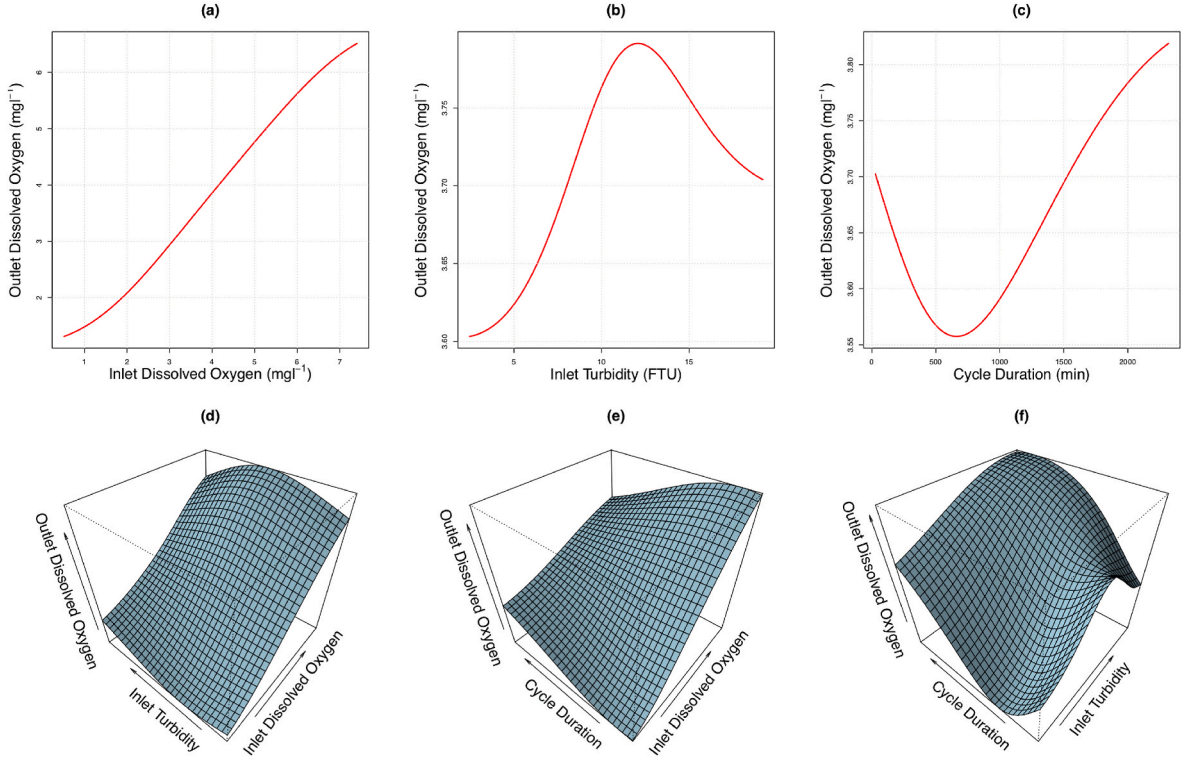

Fig. 6. Terms of first and second order for the DE/SVR–Matern52 model's three key variables that determine how the DO_o will be predicted.

- $SS_{reg} = \sum_{i=1}^{n} (y_i - \bar{t})^2$: this summation has a direct correlation with the sample variance;
- $SS_{err} = \sum_{i=1}^{n} (t_i - y_i)^2$: is referred to as the residual squared sum.

being $\bar{t}$ the mean value of the experimental data provided by:

$$\bar{t} = \frac{1}{n} \sum_{i=1}^{n} t_i \tag{17}$$

Table 5
Correlation coefficients (r), coefficients of determination (R^2), root mean squared errors (RMSE) and mean absolute errors (MAE) for the DE/SVM–based method fitted with distinct kinds of kernels for the $Turb_o$ in this study. In bold, the values corresponding to the best DE/SVM kernel.

Technique–kernel	R^2	r	RMSE	MAE
DE/SVM–Linear	0.57	0.77	0.70	0.52
DE/SVM–Polynomial	0.79	0.89	0.50	0.35
DE/SVM–Sigmoid	0.56	0.76	0.71	0.54
DE/SVM–RBF	0.83	0.91	0.44	0.27
DE/SVM–PUK	0.88	0.94	0.38	0.24
DE/SVM–Matern32	0.89	0.95	0.35	0.23
DE/SVM–Matern52	**0.89**	**0.95**	**0.35**	**0.23**

Table 6
Correlation coefficients (r), coefficients of determination (R^2), root mean squared errors (RMSE) and mean absolute errors (MAE) for the DE/SVM–based method fitted with distinct kinds of kernels for the DO_o in this study. In bold, the values corresponding to the best DE/SVM kernel.

Technique–kernel	R^2	r	RMSE	MAE
DE/SVM–Linear	0.77	0.86	0.48	0.33
DE/SVM–Polynomial	0.88	0.94	0.34	0.20
DE/SVM–Sigmoid	0.771	0.88	0.48	0.33
DE/SVM–RBF	0.89	0.94	0.33	0.19
DE/SVM–PUK	0.90	0.95	0.31	0.19
DE/SVM–Matern32	0.91	0.96	0.29	0.17
DE/SVM–Matern52	**0.92**	**0.96**	**0.29**	**0.18**

Table 7
The ranking of significance of the input factors in the most excellent DE/SVR–relied method for the $Turb_o$ prediction, based on the matching weights, in absolute descending order.

Input variable	Weight
Inlet turbidity ($Turb_i$)	0.4311
Electrical Conductivity (EC_i)	0.1843
Media bed height (H)	0.0994
Media	0.0907
pH	0.0778
Inlet dissolved oxygen (DO_i)	0.0752
Filtration velocity (v)	0.0520
Filter	0.0455
Cycle duration (Tc)	0.0327
Water temperature (T_i)	0.0061

Hence, the determination coefficient is given by (Marsland, 2014; Wasserman, 2003):

$$R^2 \equiv 1 - \frac{SS_{err}}{SS_{tot}} \tag{18}$$

The difference between the experimental and predicted data is less the closer the R^2 statistic is to 1.0.

Two additional criteria considered in this study were the root mean square error (RMSE) and mean absolute error (MAE) (Hastie et al., 2003; Wasserman, 2003). These statistics are also used frequently to evaluate the forecasting capability of a mathematical model. The root mean square error (RMSE) and mean absolute error (MAE) are given by the

Table 8
The ranking of significance of the input factors in the most excellent DE/SVR–relied method for the DO_o prediction, based on the matching weights, in absolute descending order.

Input variable	Weight
Inlet dissolved oxygen (DO_i)	0.8422
Inlet turbidity ($Turb_i$)	0.1073
Cycle duration (Tc)	0.0364
pH	0.0207
Filter	0.0180
Water temperature (T_i)	0.0173
Electrical conductivity (EC_i)	0.0122
Media bed height (H)	0.0115
Media	0.0026
Filtration velocity (v)	0.0026

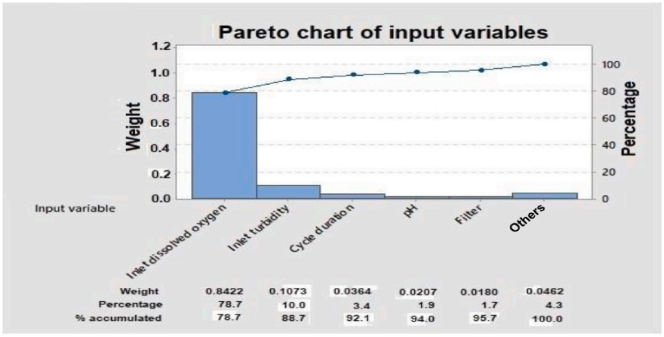

Fig. 8. Pareto chart of input variables: order of relevance for the input variables employed in the best–fitted DE/SVR–relied approach for the prediction of the DO_o.

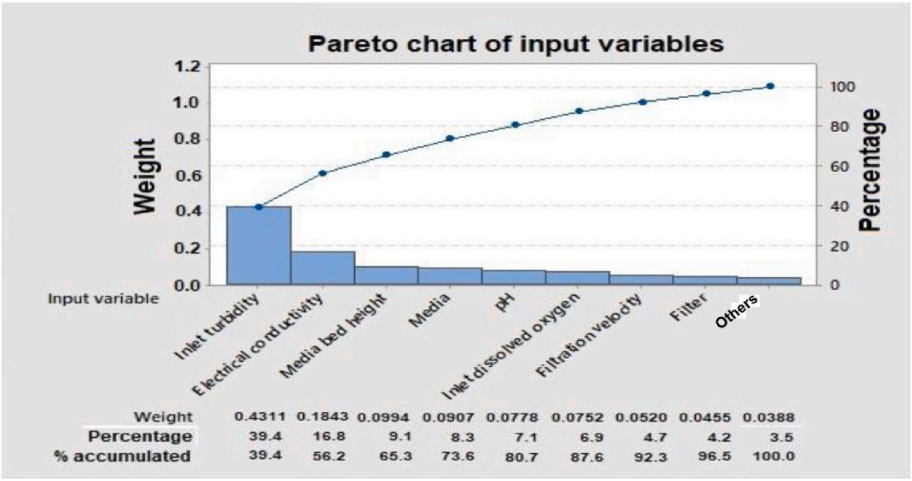

Fig. 7. Pareto chart of input variables: order of relevance for the input variables employed in the best–fitted DE/SVR–relied approach for forecasting the $Turb_o$.

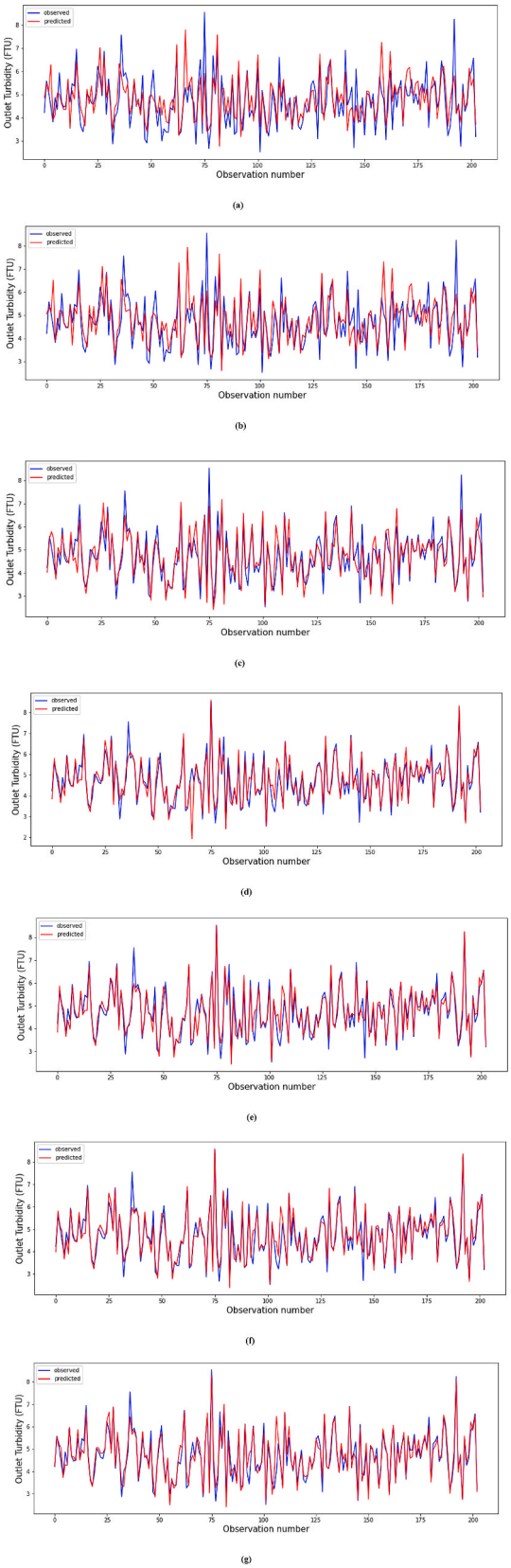

Fig. 9. Observed and predicted $Turb_o$ values regarding the test dataset utilising two distinct filter media (silica sand and crushed recycled glass): (a) DE/SVR–Sigmoid model (R^2 = 0.5589); (b) DE/SVR–Linear model (R^2 = 0.5722); (c) DE/SVR–Polynomial model (R^2 = 0.7855); (d) DE/SVR–RBF model (R^2 = 0.8287); (e) DE/SVR–PUK model (R^2 = 0.8771); (f) DE/SVR–Matern32 model (R^2 = 0.8924); and (g) DE/SVR–Matern52 model (R^2 = 0.8944).

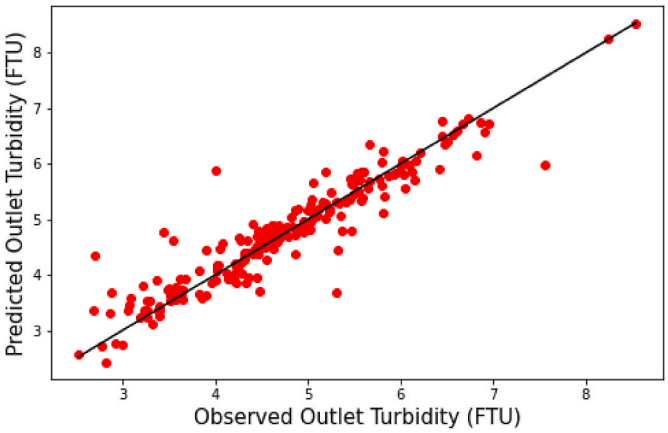

Fig. 10. Observed vs. foretold outlet turbidity ($Turb_o$) scatterplot from the testing dataset for the best-fitted model, DE/SVR–Matern52 model.

expressions (Freedman et al., 2007; Wasserman, 2003):

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(t_i - y_i)^2}{n}} \tag{19}$$

$$MAE = \frac{\sum_{i=1}^{n}|t_i - y_i|}{n} \tag{20}$$

If the RMSE has a value of zero, there is no difference between the predicted and observed data. Mean Absolute Error (MAE) is the average vertical distance between each point and the identity line. MAE is also the average horizontal distance between each point and the identity line. MAE has a clear interpretation as the average absolute difference between t_i and y_i.

Currently, a model has been developed using the $Turb_o$ and DO_o as dependent variables (specifically, in this investigation, the novel DE/SVR– model), taken from the ten input factors in granular filters (Bové et al., 2015a, 2015b), examining their impact with the purpose of enhancing its computation by looking at the coefficient of determination R^2.

Prior research has generally employed grid search, also known as a parameter sweep, which is essentially a thorough search over a subset of the hyperparameter space's values, to tune the hyperparameters. Because the DE optimiser works well with the purpose of resolving associated optimisation issues, it was used in this study to conduct a more economical search and successfully find these optimal parameters (Chakraborty, 2008; Feoktistov, 2006; Price et al., 2006; Rocca et al., 2011; Storn & Price, 1997; Vinoth Kumar et al., 2022). The DE optimiser is an evolutionary computing technique that solves optimisation problems by iteratively trying to improve a candidate solution's quality. Because these methods can search very large spaces for possible solutions with little to no assumptions about the objective function, they are called meta-heuristics (Das, Mullick, & Suganthan, 2016; Onwubolu & Babu, 2004).

As a result, by assessing the effect of ten operation input variables and efficiently utilising the R^2 to optimise the computation, with the help of this innovative hybrid DE/SVR–based approach, the $Turb_o$ and DO_o (output factors) have been successfully predicted. For instance, the process diagram for the DE/SVR–based model with Matern52 kernel constructed in this investigation is described in Fig. 4.

Furthermore, a common method for determining the true model accuracy is by calculating the R^2 from cross-validation (Agresti & Kateri, 2021; Picard & Cook, 1984). The best hyperparameters for the DE/SVR model for the seven different types of kernels can be chosen using 10-fold cross-validation (Chakraborty, 2008; Cristianini &

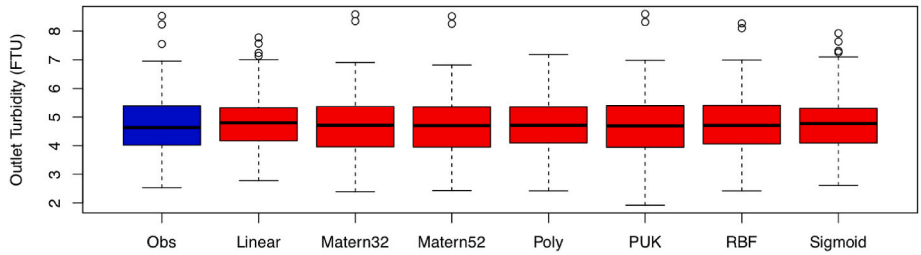

Fig. 11. Observed and foretold outlet turbidity ($Turb_o$) boxplots from the testing dataset for the seven fitted models.

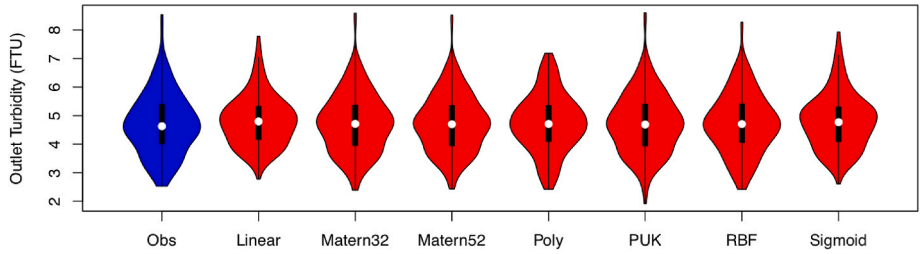

Fig. 12. Observed and foretold outlet turbidity ($Turb_o$) violinplots from the testing dataset for the seven fitted models.

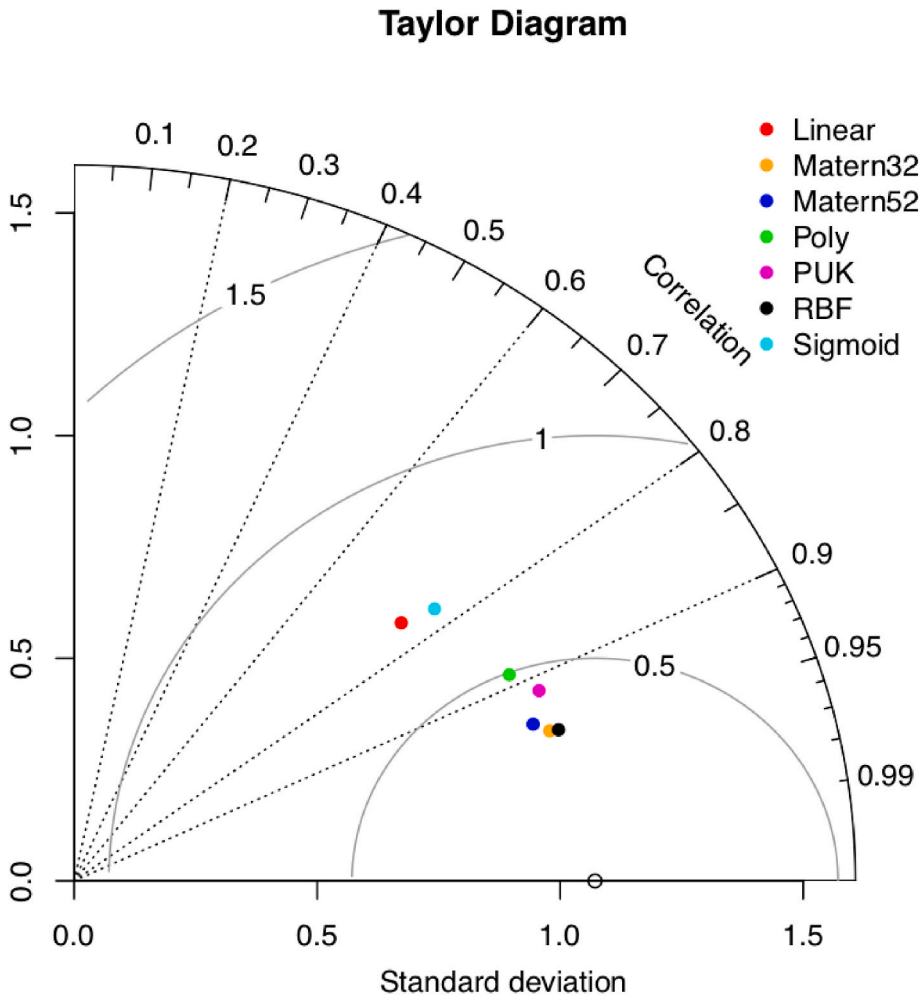

Fig. 13. Taylor diagram for the outlet turbidity ($Turb_o$) using the testing dataset for the seven fitted models.

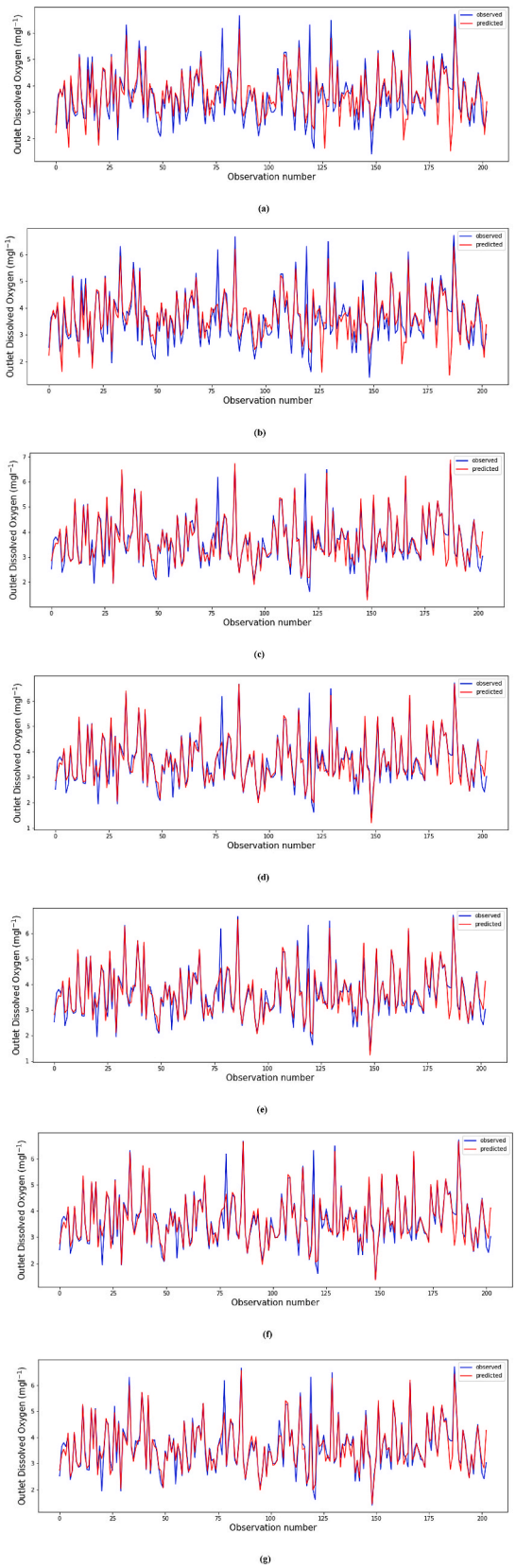

Fig. 14. Observed and predicted DO_o values regarding the test dataset utilising two distinct filter media (silica sand and crushed recycled glass): (a) DE/SVR–Sigmoid model (R^2 = 0.7661); (b) DE/SVR–Linear model (R^2 = 0.7658); (c) DE/SVR–Polynomial model (R^2 = 0.8788); (d) DE/SVR–RBF model (R^2 = 0.8873); (e) DE/SVR–PUK model (R^2 = 0.8994); (f) DE/SVR–Matern32 model (R^2 = 0.9120); and (g) DE/SVR–Matern52 model (R^2 = 0.9156).

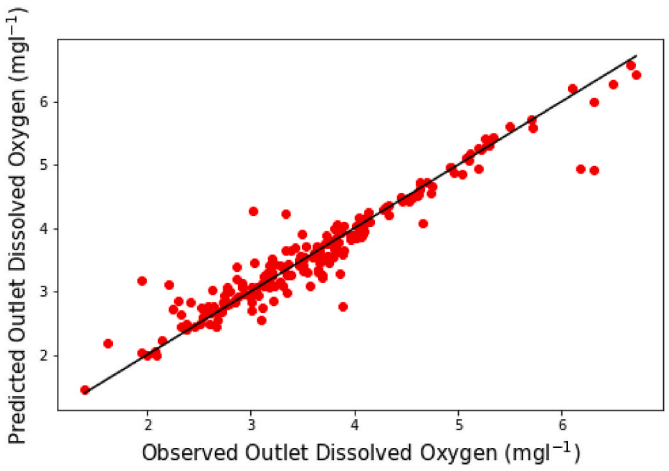

Fig. 15. Observed vs. foretold outlet dissolved oxygen (DO_o) scatterplot from the testing dataset for the best-fitted model, DE/SVR–Matern52 model.

Shawe–Taylor, 2000; Feoktistov, 2006; Hansen & Wang, 2005; Hastie et al., 2003; Picard & Cook, 1984; Price et al., 2006; Rocca et al., 2011; Storn & Price, 1997; Vinoth Kumar et al., 2022). Ten subsets of similar size are randomly selected from the dataset, and after that, a set of parameters is chosen. To assess a model's goodness-of-fit, nine subsets are used in its construction, and the last subset is used in the test (Chen, Liu, Li, & Zhou, 2022; Picard & Cook, 1984; Wasserman, 2003). This procedure is done ten times, with a different subset serving as the testing set each time. The ultimate value for the chosen collection of parameters is represented by the goodness-of-fit average.

Until it selects a particular set of optimal hyperparameters, based on their fitness, the parameter sets to be checked are chosen according to the DE algorithm (Chakraborty, 2008; Chen et al., 2022; Feoktistov, 2006; Price et al., 2006; Rocca et al., 2011; Storn & Price, 1997; Vinoth Kumar et al., 2022).

Table 2 displays the hyperparameters' intervals of variation for the DE/SVR models.

3. Results and discussion

3.1. SVR simulation results and optimisation process

This new predictive model had ten distinct operation variables as input variables. Each of them had previously been displayed in Table 1. 1,014 filtration cycles' worth of data were employed. The samples with missing data have been removed from the 1,016 samples that underwent experimental measurement (see Appendix A).

The dataset was divided into two sets with the purpose of tackling this study: 80% of the data were used in the training set, and the remaining 20% were used in the testing set. With the help of the training data, a model was created, refined, and tested on the test dataset.

The suggested DE/SVR–based model uses the $Turb_o$ and DO_o as output dependent variables. Furthermore, as was already mentioned, the cost constant (C), the size of the permitted error margin (ε parameter), and, in the end, the six parameters a, b, σ, σ_f, σ_l and ω which specify how the different kernels are expressed mathematically, are the SVM hyperparameters that have the greatest influence on the DE/SVM approximation.

Tables 3 and 4 display the optimised hyperparameters that were obtained after the SVR models with seven different kernels that were tuned by employing the meta-heuristic DE optimiser the DO_o and $Turb_o$, respectively.

The R^2 value was computed utilising the optimised model on the testing set.

The three variables' first and second order terms that were more

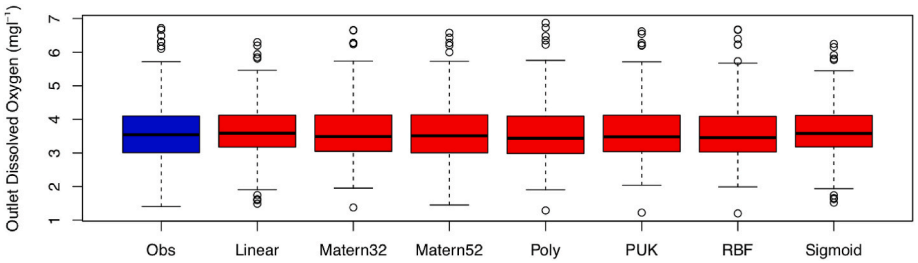

Fig. 16. Observed and foretold outlet dissolved oxygen (DO_o) boxplots from the testing dataset for the fitted seven models.

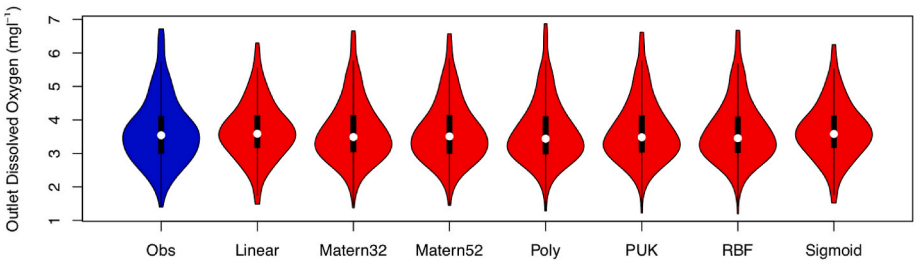

Fig. 17. Observed and foretold outlet dissolved oxygen (DO_o) violinplots from the testing dataset for the fitted seven models.

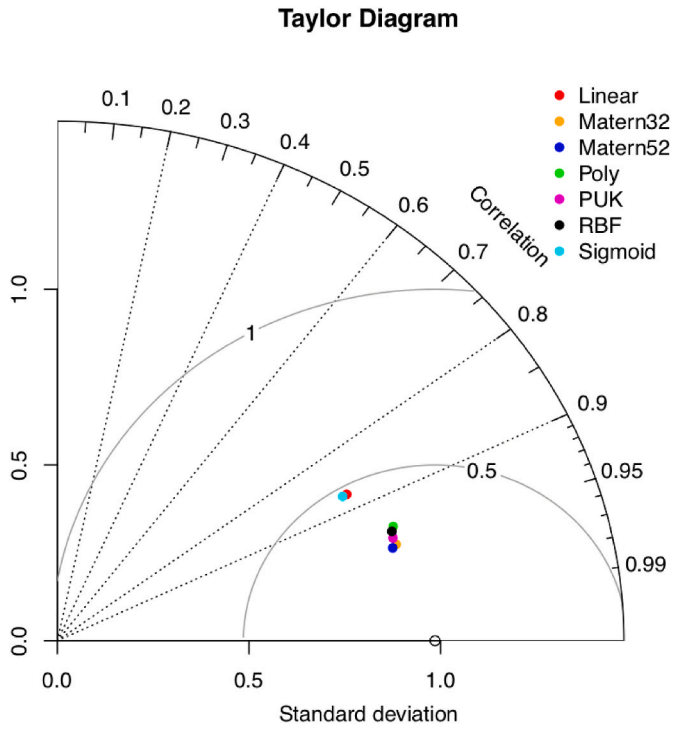

Fig. 18. Taylor diagram for the outlet dissolved oxygen (DO_o) using the testing dataset for the seven fitted models.

significant in the best-fitted DE/SVR model (using a Mater52 kernel) for outlet turbidity prediction ($Turb_o$) are displayed in Fig. 5.

Furthermore, the first- and second-order terms for the three crucial parameters in the best-fitted DE/SVR model for the DO_o prediction are displayed in Fig. 6.

Based on the results, the DE optimiser and SVR technique can be used jointly to create highly effective models for the evaluation of the $Turb_o$ and DO_o in micro-irrigation media filters. In fact, Table 5 shows the R^2 and correlation (r-value) with the test set for the SVR approach using distinct kinds of kernels for the response $Turb_o$.

TaV this study. In bold, the values corresponding to the best DE/SVM kernel.

In a similar vein, Table 6 presents the R^2 and r values from employing the test dataset for the SVR approach with different types of kernels referred to the DO_o.

A computer with a CPU Intel Core i7-4770 @ 3.40 GHz with eight cores and 15.5 GB RAM memory was used, taking 2,031 s (~34 min) to optimise the parameters of the model and < 1 s to obtain the final output dissolved oxygen model (DO_o). For outlet turbidity ($Turb_o$) model, the respective time values were 580 s (~6mins) and, again, < 1 s.

3.2. Relevance of the input variables

Table 7 and Fig. 7 display the rankings of relevance for the independent variables that were utilised in this study to predict the dependent variable, $Turb_o$. Hence, the most significant component is the input variable, $Turb_i$, which is followed by EC_i in the best-fitted DE/SVR model's outlet turbidity prediction, followed by H, media (type of medium: sand or recycled glass), pH, DOi, v, filter type, Tc and T_i.

Since the $Turb_i$ values have a significant impact on the filtering accomplished by the media filters (Solé–Torres et al., 2019a), the $Turb_i$ was the most important variable to consider when forecasting the $Turb_o$. The EC_i was the second variable in relative importance for explaining outlet turbidity. The EC_i depends on dissolved solids, which may affect some of the mechanisms, such as diffusion (Cescon & Jiang, 2020), that allow particle retention across the media bed and a reduction in turbidity. The other water quality parameters (pH and DO_i) play a less relevant role, probably because they do not have a direct link with dissolved solids. The third and four variables were H and media material, respectively. While the impact of various media materials on this parameter was evident, prior research (Solé–Torres et al., 2019a) did not demonstrate a definite pattern of H on turbidity removal (Duran–Ros et al., 2022). On the other hand, v and filter type were relevant variables in explaining turbidity reduction (Solé–Torres et al., 2019a) but their effect was minor in the present analysis. Further research with a wider range of filtration velocities and filter types should be conducted to verify these results.

Similarly, Table 8 and Fig. 8 show the relative importance of the independent variables in the DE/SVR model for the DO_o prediction. The SVR model in this case indicated that the most significant factor in

forecasting the DO_o was the input variable DO_i. This was followed by the $Turb_i$, Tc, pH, filter type, T_i, EC_i, H, media (category of medium), and v.

In the present experiment there was a moderate DO increase at the filter outlet due to the removal of organic pollutants across the media bed (Duran–Ros et al., 2022; Solé–Torres et al., 2019a) and the reduction of microorganisms caused by backwashing with chlorinated water, which meant that less oxygen was consumed in the filters. The high dependence of DO_o on its initial value, i.e. DO_i, is obvious in processes that depend on water quality parameters and was already observed in previous studies using different prediction techniques (Martí et al., 2013; García–Nieto et al., 2020a). The second variable, with much less importance, was inlet turbidity, which is also related to organic particles whose oxidation by microorganisms consume oxygen and therefore reduces the DO. All the other variables had a weight smaller than 0.04, in concordance with the results of similar works (García–Nieto et al., 2020a). Given that the DO is a parameter that depends on temperature, the lowest weight assigned to the temperature variable was explained by the higher average irrigation water temperatures during the present experiment that caused fewer variations in the DO values.

In this investigation, as shown in Fig. 9, the ten independent operation input variables have been used to predict the $Turb_o$ in micro-irrigation systems, making use of the contrast between the predicted and observed $Turb_o$ values, using the seven previously defined models with the DE/SVR–Matern52 model (Fig. 9(g)) being the best-fitted model. Additionally, Figs. 10–13 indicate the comparison between the experimental and predicted $Turb_o$ values employing the seven kinds of models for the test dataset. Consequently, to solve this nonlinear regression problem, it was fundamental to bring together the SVR–Matern52 procedure with the DE optimiser to produce an original hybrid strategy that was significantly more reliable and efficient than the other six regression methods.

The comparison of the predicted and observed DO_o values is displayed in Fig. 14, for the seven proposed models against the testing dataset. Specifically, Figs. 15–18 indicate the comparison between the experimental and predicted DO_o values using the seven types of models for the test dataset. Consequently, to solve this nonlinear regression problem, it is essential to bring together the SVR–Matern52 procedure with the DE optimiser to produce an original hybrid strategy that was significantly more reliable and efficient than the other six regression methods.

4. Conclusions

A substitute diagnosis method using the new differential evolution and support vector regression (DE/SVR) method with optimised hyperparameters allowed reasonable predictions of outlet turbidity $Turb_o$ ($R^2 = 0.93$) and outlet dissolved oxygen DO_o ($R^2 = 0.89$) in sand media filters which are commonly used in micro-irrigation systems. The relative importance of the input variables was determined for each parameter. Thus, the parameter with the biggest impact on the calculation of the $Turb_o$ was the inlet turbidity $Turb_i$, and after it, inlet electrical conductivity EC_i, media height H, type of media, pH, inlet dissolved oxygen DOi, filtration velocity v, type of sand filter design, duration of filtration cycle Tc, and water inlet temperature T_i. Additionally, the input variable DO_i has the biggest impact on the DO_o estimate prior to the $Turb_i$, Tc, pH, T_i, filter design type, EC_i, H, type of media and v. The approach developed in this work may be effectively employed for various filtration mechanisms utilising the same or distinct kinds of filter media if the particular characteristics of the different filters and experiments are considered.

Data availability

Dataset is made available on request.

CRediT authorship contribution statement

Paulino José García–Nieto: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Writing – original draft, Supervision, Visualization. **Esperanza García–Gonzalo:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Supervision, Visualization, Writing – original draft. **Gerard Arbat:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Supervision, Visualization, Writing – original draft. **Miquel Duran–Ros:** Data curation, Investigation, Writing – original draft. **Toni Pujol:** Data curation, Investigation, Writing – original draft. **Jaume Puig–Bargués:** Data curation, Investigation, Writing – original draft.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The Department of Mathematics of the University of Oviedo only provided us with the computational support (no monetary support), that is, the use of powerful computers to carry out the numerical calculations of this research. The other two organizations (specifically, the Spanish Research Agency and European Regional Development Fund (Belgium)) provided the financial funds necessary to carry out the experimental setup as well as data collection.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.biosystemseng.2024.04.020.

References

Agresti, A., & Kateri, M. (2021). *Foundations of statistics for data scientists: with R and Python*. Boca Raton, FL, USA: Chapman and Hall/CRC Press.

Ayars, J. E., Bucks, D. A., Lamm, F. R., & Nakayama, F. S. (2007). Introduction. In F. R. Lamm, F. R., J. E. Ayars, & F. S. Nakayama (Eds.), *Microirrigation for crop production: Design, operation and management* (pp. 1–26). Amsterdam: Elsevier.

Bishop, C. M. (2006). *Pattern recognition and machine learning*. Cambridge, UK: Springer.

Bové, J., Arbat, G., Duran-Ros, M., Ramírez de Cartagena, F., Velayos, J., & Puig-Bargués, J. (2015). Reducing energy requirements for sand filtration in microirrigation: Improving the underdrain and packing. *Biosystems Engineering, 140*, 67–78. https://doi.org/10.1016/j.biosystemseng.2015.09.00

Bové, J., Arbat, G., Duran–Ros, M., Pujol, T., Velayos, J., Ramírez de Cartagena, F., et al. (2015). Pressure drop across sand and recycled glass media used in micro irrigation filters. *Biosystems Engineering, 137*, 55–63. https://doi.org/10.1016/j.biosystemseng.2015.07.009

Bové, J., Puig–Bargués, J., Arbat, G., Duran–Ros, M., Pujol, T., Pujol, J., et al. (2017). Development of a new underdrain for improving the efficiency of microirrigation sand media filters. *Agricultural Water Management, 179*, 296–305. https://doi.org/10.1016/j.agwat.2016.06.031

Capra, A., & Scicolone, B. (2007). Recycling of poor quality urban wastewater by drip irrigation systems. *Journal of Cleaner Production, 15*(16), 1529–1534. https://doi.org/10.1016/j.jclepro.2006.07.032

Cescon, A., & Jiang, J.-Q. (2020). Filtration process and alternative filter media material in water treatment. *Water, 12*, 3377. https://doi.org/10.3390/w12123377

Chakraborty, U. K. (2008). *Advances in differential evolution*. Berlin, Germany: Springer.

Chang, C.-C., & Lin, C.-J. (2011). Libsvm: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology, 2*(3), 1–27. https://doi.org/10.1145/1961189.1961199

Chen, J.-L., Li, G.-S., & Wu, S.-J. (2013). Assessing the potential of support vector machine for estimating daily solar radiation using sunshine duration. *Energy Conversion and Management, 75*, 311–318. https://doi.org/10.1016/j.enconman.2013.06.034

Chen, Y., Liu, R., Li, Y., & Zhou, X. (2022). Research and application of cross validation of fault diagnosis for measurement channels. *Progress in Nuclear Energy, 150*, Article 104324. https://doi.org/10.1016/j.pnucene.2022.104324

Cristianini, N., & Shawe–Taylor, J. (2000). *An introduction to support vector machines and other kernel–based learning methods*. New York: Cambridge University Press.

Das, S., Mullick, S. S., & Suganthan, P. N. (2016). Recent advances in differential evolution – an updated survey. *Swarm and Evolutionary Computation, 27*, 1–30. https://doi.org/10.1016/j.swevo.2016.01.004

De Leone, R., Pietrini, M., & Giovannelli, A. (2015). Photovoltaic energy production forecast using support vector regression. *Neural Computing & Applications, 26*, 1955–1962. https://doi.org/10.1007/s00521-015-1842-y

Deisenroth, M. P. (2020). *Mathematics for machine learning*. New York, USA: Cambridge University Press.

Duran–Ros, M., Puig–Bargués, J., Arbat, G., Barragán, J., & Ramírez de Cartagena, R. (2008). Definition of a SCADA system for a microirrigation network with effluents. *Computers and Electronics in Agriculture, 64*(2), 338–342. https://doi.org/10.1016/j.compag.2008.05.023

Duran–Ros, M., Puig–Bargués, J., Cufí, S., Solé–Torres, C., Arbat, G., Pujol, J., et al. (2022). Effect of different filter media on emitter clogging using reclaimed effluents. *Agricultural Water Management, 258*, Article 107591. https://doi.org/10.1016/j.agwat.2022.107591

FAO. (2022). *The state of the world's land and water resources for food and agriculture – systems at breaking point. Main report*. Rome: FAO. https://doi.org/10.4060/cb9910en

Feoktistov, V. (2006). *Differential evolution: In search of solutions*. New York, USA: Springer.

Freedman, D., Pisani, R., & Purves, R. (2007). *Statistics*. New York, USA: W.W. Norton & Company.

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

García–Nieto, P. J., García–Gonzalo, E., Bové, J., Arbat, G., Duran–Ros, M., & Puig–Bargués, J. (2017). 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. *Computers and Electronics in Agriculture, 139*, 65–74. https://doi.org/10.1016/j.compag.2017.05.008

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

García–Nieto, P. J., García–Gonzalo, E., Puig–Bargués, J., Solé–Torres, C., Duran–Ros, M., & Arbat, G. (2020b). A new predictive model for the outlet turbidity in micro-irrigation sand filters fed with effluents using Gaussian process regression. *Computers and Electronics in Agriculture, 170*, Article 105292. https://doi.org/10.1016/j.compag.2020.105292

Hansen, T., & Wang, C. J. (2005). Support vector based battery state of charge estimator. *Journal of Power Sources, 141*, 351–358. https://doi.org/10.1016/j.jpowsour.2004.09.020

Hastie, T., Tibshirani, R., & Friedman, J. H. (2003). *The elements of statistical learning*. New York, USA: Springer-Verlag.

Hawari, A. H., & Alnahhal, W. (2016). Predicting the performance of multi-media filters using artificial neural networks. *Water Science and Technology, 74*(9), 2225–2233. https://doi.org/10.2166/wst.2016.380

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An introduction to statistical learning: With applications in R*. New York, USA: Springer.

Kuhn, M., & Johnson, K. (2018). *Applied predictive modeling*. New York, USA: Springer.

Li, X., Lord, D., Zhang, Y., & Xie, Y. (2008). Predicting motor vehicle crashes using Support Vector Machine models. *Accident Analysis & Prevention, 40*, 1611–1618. https://doi.org/10.1016/j.aap.2008.04.01

Marsland, S. (2014). *Machine learning: An algorithmic perspective*. Boca Raton, FL, USA: Chapman and Hall/CRC Press.

Martí, P., Shiri, J., Duran–Ros, M., Arbat, G., Ramírez de Cartagena, F., & Puig–Bargués, J. (2013). Artificial neural networks vs. Gene Expression Programming for estimating outlet dissolved oxygen in micro-irrigation sand filters fed with effluents. *Computers and Electronics in Agriculture, 99*, 176–185. https://doi.org/10.1016/j.compag.2013.08.016

Mesquita, M., de Deus, F. P., Testezlaf, R., da Rosa, L. M., & Diotto, A. V. (2019). Design and hydrodynamic performance testing of a new pressure sand filter diffuser plate using numerical simulation. *Biosystems Engineering, 183*, 58–69. https://doi.org/10.1016/j.biosystemseng.2019.04.015

Nakayama, F. S., Boman, B. J., & Pitts, D. J. (2007). Maintenance. In F. R. Lamm, J. E. Ayars, & F. S. Nakayama (Eds.), *Microirrigation for crop production: Design, operation and management* (pp. 389–430). Amsterdam: Elsevier.

Onwubolu, G. C., & Babu, B. V. (2004). *New optimisation techniques in engineering*. Berlin: Springer. https://link.springer.com/book/10.1007/978-3-540-39930-8.

Picard, R., & Cook, D. (1984). Cross-validation of regression models. *Journal of the American Statistical Association, 79*(387), 575–583. https://doi.org/10.2307/2288403

Price, K., Storn, R. M., & Lampinen, J. A. (2006). *Differential evolution: A practical approach to global optimisation*. New York, USA: Springer.

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

Pujol, J., Espinach, F. X., Duran-Ros, M., Arbat, G., Pujol, T., Ramírez de Cartagena, F., & Puig-Bargués, J. (2022a). Environmental assessment of underdrain designs for granular media filters in drip irrigation systems. *Agriculture, 12*(6), 810. https://doi.org/10.3390/agriculture12060810.

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

Pujol, T., Puig-Bargués, J., Arbat, G., Vegas, A., Duran-Ros, M., Pujol, J., et al. (2020). Numerical study of the effects of pod, wand and spike type underdrain systems in pressurised sand filter. *Biosystems Engineering, 200*, 338–352. https://doi.org/10.1016/j.biosystemseng.2020.10.018

Ravina, I., Paz, E., Sofer, Z., Marm, A., Schischa, A., Sagi, G., et al. (1997). Control of clogging in drip irrigation with stored treated municipal sewage effluent. *Agricultural Water Management, 33*(2–3), 127–137. https://doi.org/10.1016/S0378-3774(96)01286-3

Rocca, P., Oliveri, G., & Massa, A. (2011). Differential evolution as applied to electromagnetics. *IEEE Transactions on Antennas and Propagation, 53*(1), 38–49. https://doi.org/10.1109/MAP.2011.5773566

Schölkopf, B., Smola, A. J., Williamson, R., & Bartlett, P. (2000). New support vector algorithms. *Neural Computation, 12*(5), 1207–1245. https://doi.org/10.1162/089976600300015565

Shrestla, N. K., & Shukla, S. (2015). Support vector machine based modeling of evapotranspiration using hydro–climatic variables in a sub–tropical environment. *Agricultural and Forest Meteorology, 200*, 172–184. https://doi.org/10.1016/j.agrformet.2014.09.025

Solé–Torres, C., Puig–Bargués, J., Duran–Ros, M., Arbat, G., Pujol, J., & Ramírez de Cartagena, F. (2019a). Effect of underdrain design, media height and filtration velocity on the performance of microirrigation sand filters using reclaimed effluents. *Biosystems Engineering, 187*, 292–304. https://doi.org/10.1016/j.biosystemseng.2019.09.012

Solé–Torres, C., Puig–Bargués, J., Duran–Ros, M., Arbat, G., Pujol, J., & Ramírez de Cartagena, F. (2019b). Effect of different sand filter underdrain designs on emitter clogging using reclaimed effluents. *Agricultural Water Management, 223*, Article 105683. https://doi.org/10.1016/j.agwat.2019.105683

Steinwart, I., & Christmann, A. (2008). *Support vector machines*. New York, USA: Springer.

Storn, R., & Price, K. (1997). Differential evolution - a simple and efficient heuristic for global optimisation over continuous spaces. *Journal of Global Optimisation, 11*, 341–359. https://doi.org/10.1023/A:1008202821328

Tarjuelo, J. M., Rodriguez–Diaz, J. A., Abadía, R., Camacho, E., Rocamora, C., & Moreno, M. A. (2015). Efficient water and energy use in irrigation modernization: Lessons from Spanish case studies. *Agricultural Water Management, 162*, 67–77. https://doi.org/10.1016/j.agwat.2015.08.009

Tien, C. (2012). *Principles of filtration*. Kidlington, Oxford, UK: Elsevier.

Vapnik, V. (1998). *Statistical learning theory*. New York, USA: Wiley–Interscience.

Vinoth Kumar, B., Oliva, D., & Suganthan, P. N. (2022). *Differential evolution: From theory to practice*. Singapore: Springer.

Wasserman, L. (2003). *All of statistics: A concise course in statistical inference*. New York, USA: Springer.