

Article

Data-Driven Surrogate Modelling for Industrial Scale-Up of Packed Bed Columns Using Residual Biomass for Pb(II) Removal

Oscar E. Coronado-Hernández ¹, Angel Villabona-Ortiz ², Mauricio J. Rosso-Pinto ^{3,*},
Candelaria Tejada-Tovar ² and José Quevedo-Cabarcas ²

¹ Instituto de Hidráulica y Saneamiento Ambiental, Universidad de Cartagena, Cartagena de Indias 130001, Colombia; ocoronadoh@unicartagena.edu.co

² Biomass Process Design and Utilization Research Group (IDAB), Department of Chemical Engineering, Universidad de Cartagena, Avenida del Consulado St. 30, Cartagena de Indias 130015, Colombia; avillabonao@unicartagena.edu.co (A.V.-O.); ctejadat@unicartagena.edu.co (C.T.-T.); jquevedoc@unicartagena.edu.co (J.Q.-C.)

³ Department of Environmental Engineering, Faculty of Engineering, Universidad de Córdoba, Carrera 6 No. 76-103, Montería 230002, Colombia

* Correspondence: mauriciorosso@correo.unicordoba.edu.co

Abstract

Pb(II) is a potentially toxic element that can cause problems in the respiratory, digestive, and nervous systems; it can also cause cancer. Most studies have been conducted primarily in the laboratory; therefore, it is necessary to find a way to predict performance on a large scale. Taking this into account, the objective of this study was to simulate an industrial-scale packed adsorption column for the removal of Pb(II) using *Dioscorea rotundata*-based biomass as an adsorbent, employing the Aspen Adsorption software. To this end, a parametric analysis was conducted to evaluate the performance of the bed under different operating conditions, using the Langmuir and Freundlich isotherms in conjunction with Linear Driving Force (LDF) kinetics. Removal efficiencies of up to 92.4% were observed for the Freundlich–LDF model and 92.4% for the Langmuir–LDF model. Additionally, machine learning algorithms were combined with statistical indicators (R^2 , RMSE) to analyze the models' effectiveness. R^2 values of up to 0.999 were obtained in the validation and testing phases. This research demonstrates a novel approach to predicting the potential performance of adsorption columns packed with biomass derived from organic waste in the field of engineering, using computational tools with machine learning algorithms.

Keywords: adsorption; breakthrough curves; isotherms; kinetics; lead; water treatment



Academic Editor: Nelson Belzile

Received: 21 May 2026

Revised: 24 June 2026

Accepted: 26 June 2026

Published: 30 June 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

The presence of pollutants in water bodies represents one of the most critical environmental issues worldwide, as it directly affects the quality and availability of water resources necessary for life and for human activities [1,2]. This situation is primarily caused by the discharge of wastewater from domestic, agricultural, and industrial activities, which introduce various chemicals, organic materials, and emerging contaminants into water sources depending on their source [3,4]. Among the many pollutants that can be found in water, the following stand out: dissolved solids, which increase turbidity and reduce light penetration [5]; organic compounds such as pesticides or hydrocarbons [6,7] and inorganic compounds such as nitrates and phosphates [8], which can cause phenomena such as eutrophication; emerging contaminants such as pharmaceuticals, which are widely

used in the healthcare sector and possess persistent characteristics and long-term effects that are currently under study [9]. However, among all these pollutants, trace metals are a major cause for concern, as they are highly toxic, bioaccumulative, nonbiodegradable, and very difficult to remove [10].

Lead (Pb) is one of the most common metals found worldwide, as it can be found in the Earth's crust, in the air, in food, and in various materials such as paints and tools, among other sources [11]. Its presence is primarily due to various industrial activities, such as mining, metallurgy, the manufacture of paints and batteries, electroplating, and waste disposal [12]. In water bodies, it can be found in dissolved form or as suspended particles, allowing for its easy transport and accumulation in sediments [13]. This substance has strong neurotoxic properties that negatively affect the human body, causing problems in the respiratory, digestive, and nervous systems, as well as the development of cancerous cells and metabolic disorders, among other issues [14].

To remove contaminants from water sources, various techniques have been developed that can be classified into physical methods, such as filtration [15], sedimentation [16], and flotation [17], which enable the separation of suspended particles from water; chemical methods such as coagulation–flocculation and oxidation [18], used to transform or remove dissolved compounds; and biological methods employing microorganisms to degrade organic matter [19]. Among these technologies, adsorption is a versatile and efficient surface technique used to remove contaminants from water, consisting of the accumulation of substances on the surface of the adsorbent, such as lignocellulosic materials and activated carbon, among others [20]. This technique is classified into two types: physisorption, generated by nonspecific Van der Waals forces, and chemisorption, which is generated by attractive forces (ionic or covalent bonds) [21].

Various studies have been conducted in which different types of adsorbents based on agricultural waste have been developed for the removal of trace metals, such as banana [22], coconut [23], oil palm bagasse [24], orange [25], and cocoa [26], among other materials, demonstrating excellent adsorption efficiency. However, most of these studies have been conducted at the laboratory level; therefore, researchers have sought ways to predict the behavior of adsorption processes on a larger scale. To achieve this scaling, technological tools have been developed that enable the modeling and simulation of removal systems on a larger scale, such as Aspen Adsorption, a software product from AspenTech, which implements mathematical models and digital tools to develop adsorption processes at different scales to understand the behavior of these processes in the treatment of contaminated effluents, separation of substances, and purification of systems, among other applications.

On the other hand, Machine Learning (ML) is a technology that uses artificial intelligence to learn patterns from previously obtained information. Its importance lies in its ability to understand and analyze large amounts of data, identify complex relationships, and build highly accurate predictive models, making it essential in scientific and industrial contexts [27]. Its use with results obtained from simulations allows for accelerated analysis, parameter adjustment, and reduced computational costs, facilitating decision-making, experimental designs, and the anticipation of operational system behavior under various operating conditions, making it an excellent computational tool for technological scaling and advanced research [28].

Similarly, the implementation of Simulink serves as a fundamental tool for the integration and digitization of models developed by exporting data obtained from Aspen Adsorption and ML into a dynamic environment. This tool makes it possible to reproduce and analyze the behavior of complex processes in a simpler and more efficient manner [29]. In this context, issues related to digitalization take on greater significance; this approach relies on the integration of computer tools, process simulation, and data analysis to generate

complex systems through virtual environments. The combination of these technologies forms a surrogate modeling methodology, which enables the development of models capable of predicting system performance with greater accuracy and at a low computational cost. Given the above, this study aimed to simulate an adsorption column packed with yam residues on an industrial scale for the removal of Pb(II) using parametric analysis and ML pre-settings. This research examines the potential of computational technologies, in conjunction with machine learning, to predict the performance of adsorption beds in order to obtain highly accurate results, providing quantitative information on a fixed-bed column composed of agricultural waste.

2. Materials and Methods

2.1. Establishing the Conditions for Simulating the Adsorption Column

Aspen Adsorption V11 software was used as a computational tool to simulate the packed adsorption column using *Dioscorea rotundata* for the removal of Pb(II). Therefore, it is necessary to correctly set the configuration and conditions of the adsorption column, first defining the contaminants to be removed from the water bodies; that is, the chemical components are established using the software's database. In this case, Pb(II) was specified as the substance to be removed. Subsequently, the thermodynamic package is set up to calculate the physical properties of the contaminant. In this case, it used the Electrolyte Non-Random Two-Liquid (ELECNRTL) package [30]. Now, various parameters related to the sizing, operation, and modeling of the column are incorporated into the simulation; these are key variables for developing the equipment that will remove Pb(II) from the solution. Consequently, various studies were analyzed to identify the different variables required by the software [31]. Therefore, these parameters were studied, analyzed, and adjusted for this simulation model. Table 1 presents the parameter values for the fixed-bed column configuration.

Table 1. Parameters required for the adsorption column simulation.

Parameters	Values
Bed diameter, m	1
Bed porosity, m ³ of void space/m ³ of bed	0.4
Bulk density, kg/m ³	230
Constant mass transfer coefficient, 1/s	2.722×10^{-4}
Total void porosity	0.753
Freundlich parameter : k_F , (mg g ⁻¹ (mg L ⁻¹) ^{1/n})	13.42
Freundlich parameter: 1/n	0.86
Langmuir parameter: b (L/mg)	0.16
Langmuir parameter : q_{max} (mg/g)	98.36

Similarly, the software requires various parameters to be set correctly so that the simulation runs efficiently. Therefore, each of these parameters was analyzed and configured to ensure proper operation. These parameters are presented in Table 2.

Table 2. Criteria specified in the column settings.

Section	Description	Established Criterion
General	Define the discretization technique	Discretization scheme: Upwind Discretization Scheme (UDS1)
Material/Momentum Balance	Define the considerations related to axial dispersion, pressure drop, and flow rate.	No axial dispersion No pressure drops Constant velocity
Kinetic Model	Define the kinetic model	Linear driving force (LDF) model
Isothermal Model	Define the isothermal model	Langmuir model Freundlich model
Energy Balance	Address the established adsorption process.	Isothermal system

2.2. Parametric Evaluation in the Adsorption Column

The operating conditions of a system are a set of fundamental parameters that enable the process to function properly and achieve its established objective. In this study, a parametric analysis was conducted to evaluate the effect of various operating conditions on the performance of the Pb(II) removal system. Therefore, the factors of bed height (h) and inlet flow rate (Q) were varied to assess their impact on adsorption efficiency, breakthrough time (BT), and saturation time (ST) [32–34]. Table 3 presents the parameter ranges used in the parametric study of Pb(II) adsorption. Furthermore, it is assumed that the process occurs at a temperature of 30 °C, an initial contaminant concentration of 1000 mg/L and a pressure of 1 atm.

Table 3. Ranges of study parameters for the adsorption column simulation.

Variables	Unit	Range
Q	m ³ /day	100, 125, 150, 175, 200, 225, 250
h	m	3, 3.25, 3.5, 3.75, 4

2.3. Machine Learning Techniques

Machine learning techniques were used to reproduce and evaluate the simulations developed in Aspen Adsorption. Consequently, a database was constructed based on the results obtained under different operating conditions of the simulated system, which were previously analyzed and organized for proper use. The database consists of a set of 70 input data points obtained from the adsorption column simulations for the removal of Pb(II). This information was entered into an integrated software application in MATLAB R2024b known as Regression Learner, which allows for the identification of the best algorithm to model the evaluated process. Twenty-eight machine learning algorithms integrated into the software's database were applied, as shown in Figure 1.

During the model training phase, the following predictors and responses were selected, as shown in Table 4.

Table 4. Predictors and selected responses for the ML model.

Predictors	Responses
Q h Model type	Breakthrough time (BT) Saturation time (ST) Contaminant removal efficiency ($1 - C_f/C_0$).



Figure 1. Machine learning hyperparameters used in this study.

2.4. Mathematical Foundations

Isothermal models describe the relationship between the concentration of the adsorbate in the liquid phase at a constant temperature and the adsorbent, representing how the system reaches equilibrium [35]. Aspen Adsorption uses various mathematical models to describe the interactions between the amount of solute to be removed and the material used for adsorption. For this study, we used the Langmuir isothermal model, which assumes that the adsorption process occurs on a surface with identical active sites (monolayer) [36], and the Freundlich isothermal model, which assumes adsorption on surfaces with different active sites (multilayer) [37], described by Equations (1) and (2), respectively:

$$q_e = \frac{q_{max} b C_e}{1 + b C_e} \quad (1)$$

$$q_e = k_f C_e^{1/n} \quad (2)$$

On the other hand, kinetic models help us understand how quickly the adsorbate transfers from the fluid to the surface of the adsorbent material. In the software, these models help determine factors such as the process rate and the time required to reach system equilibrium [38]. In this study, the linear driving force (LDF) model was selected as the kinetic model, based on the hypothesis that the system's driving force, when affected by concentration, has an impact on the process's mass transfer [39], as described by Equation (3):

$$\frac{\partial w_k}{\partial t} = MTC_{sk} (w_k^* - w_k) \quad (3)$$

When applying ML techniques, the selection of statistical metrics is key to evaluating model performance and determining the best algorithm for the model, taking into account the analyzed responses. For this study, we used the statistical metrics shown in Table 5.

Table 5. Selected statistical indicators for analyzing model performance.

Metrics	Equation	Source
Coefficient of determination or R-squared (R^2)	$R^2 = 1 - \frac{\sum_{i=1}^N (T_i - P_i)^2}{\sum_{i=1}^N (T_i - \bar{T})^2}$ (4)	[40,41]
Root mean square error (RMSE)	$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (T_i - P_i)^2}$ (5)	

This study used three machine learning hyperparameter sets because they yielded the best results during validation and testing, based on the values of the selected metrics. Furthermore, the results obtained for the response variables BT, ST, and efficiency were best suited to the Gaussian Process Regression (GPR) algorithm, a probabilistic mathematical model that follows a Gaussian distribution [42].

$$P(y | f, X) \sim N\left(y \mid H\beta + f, \sigma^2 I\right) \quad (6)$$

3. Results and Discussion

3.1. Sensitivity Analysis of Inlet Flow Variations

The Q parameter was varied within the range of values presented in Table 2, while keeping the h fixed at 3 m, in order to isolate and evaluate the effect of flow rate on the performance of the Pb(II) removal process. The results obtained following the simulation process revealed that increasing the Q had a positive effect on the efficiency of the contaminant removal process; however, this led to a decrease in the BT and ST times obtained. This phenomenon occurs when a greater flow of liquid enters the column, facilitating the renewal of the boundary layer on the adsorbent surface and promoting mass transfer from the liquid phase to the active sites of the solid [43,44]. This behavior is shown in Table 6.

Table 6. The data were obtained from the variations in the inlet flow rate during the adsorption process.

Contaminant		Lead (Pb(II))		
Mathematical Model	Parameter Evaluated	BT (min)	ST (min)	Efficiency (%)
Freundlich–LDF	250 m ³ /day	143	781	92.4
	225 m ³ /day	160	906	91.6
	200 m ³ /day	181	1127	90.6
	175 m ³ /day	209	1357	89.3
	150 m ³ /day	246	1466	87.7
	125 m ³ /day	297	1833	85.4
	100 m ³ /day	375	2042	82.1
Langmuir–LDF	250 m ³ /day	143	781	92.4
	225 m ³ /day	160	906	91.6
	200 m ³ /day	181	1126	90.6
	175 m ³ /day	209	1356	89.3
	150 m ³ /day	246	1465	87.7
	125 m ³ /day	297	1832	85.4
	100 m ³ /day	375	2041	82.2

This behavior can be observed in Figure 2, which shows the breakthrough profiles constructed from the results obtained from the packed column simulations. These curves describe the change in the outlet concentration of Pb(II) over time. Therefore, a combination of the Freundlich model with LDF (Figure 2a) and the Langmuir model with LDF (Figure 2b) was employed. The curves exhibit a relatively sharp breakpoint, reflecting a rapid change in relative concentration between the phase in which the column effectively retains the contaminant and the point at which it begins to saturate abruptly, indicating that the mass transfer zone is relatively narrow and that the adsorption process through the bed is well defined. Furthermore, the tail effect does not have a significant influence in the saturation region, as the curves reach stability rapidly, indicating that intraparticle diffusion resistance is low; therefore, mass transfer to the active sites occurs efficiently and relatively uniformly under the established conditions. These curves allowed us to identify the fundamental parameters of bed performance: BT, ST, and efficiency, presented in Table 6.

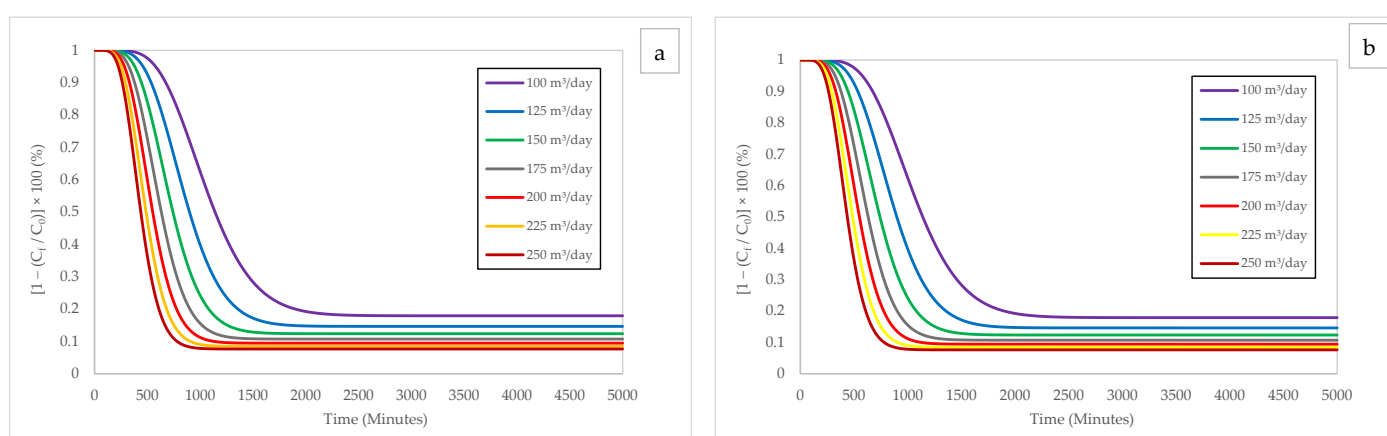


Figure 2. Adsorption profiles of contaminants as a function of Q : (a) Freundlich–LDF model and (b) Langmuir–LDF model.

3.2. Sensitivity Analysis Regarding Variations in Bed Height

The h parameter was varied within the range of values shown in Table 2, while the Q was kept constant at $250 \text{ m}^3/\text{day}$, in order to isolate and evaluate the influence of flow rate on the performance of the Pb(II) removal process. The results obtained after the simulation process showed that increasing the h prolongs the BT and ST times; however, process efficiency is negatively affected, resulting in a reduction in this variable. This phenomenon is due to the presence of a greater volume of adsorbent in the system resulting from the column length, which increases the number of active sites and delays the saturation of the adsorbent material, leading to a longer total process duration [31,45]. This performance can be seen in Table 7.

This behavior can be observed in Figure 3, which shows the breakthrough profiles constructed from the results obtained from the packed column simulations. These curves describe the change in the outlet concentration of Pb(II) over time. Therefore, a combination of the Freundlich model with LDF (Figure 3a) and the Langmuir model with LDF (Figure 3b) was employed. The curves show a consistent trend toward longer break-up and saturation times, while maintaining a well-defined sigmoidal shape and relatively steep break-up fronts, indicating that the mass transfer zone remains concentrated and that the bed is operating efficiently. Furthermore, the tail effect does not exert a significant influence in the saturation region, as the curves reach stability rapidly, indicating that intraparticle diffusion resistance is low; therefore, mass transfer to the active sites occurs efficiently and relatively uniformly under the established conditions. These curves allowed us to

identify the fundamental parameters of bed performance, BT, ST, and efficiency, presented in Table 7.

Table 7. The data were obtained from changes in bed height during the adsorption process.

Contaminant		Lead (Pb(II))		
Mathematical Model	Parameter Evaluated	BT (min)	ST (min)	Efficiency (%)
Freundlich–LDF	3 m	143	781	92.4
	3.25 m	156	873	91.8
	3.5 m	169	981	91.2
	3.75 m	181	1015	90.6
	4 m	194	1137	90.0
Langmuir–LDF	3 m	143	781	92.4
	3.25 m	156	873	91.8
	3.5 m	169	980	91.2
	3.75 m	181	1014	90.6
	4 m	194	1136	90.0

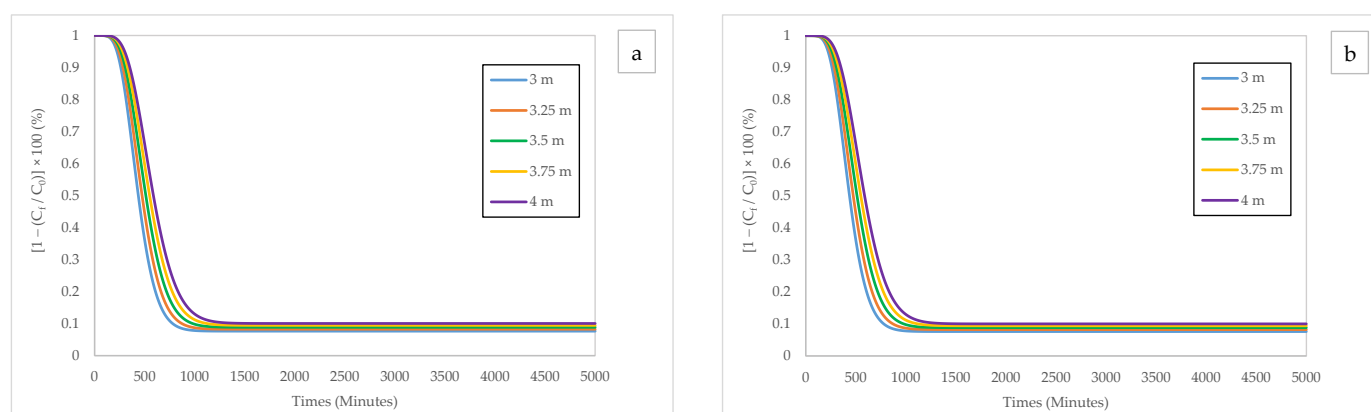


Figure 3. Adsorption profiles of contaminants as a function of h : (a) Freundlich–LDF model and (b) Langmuir–LDF model.

3.3. ML Implementation

ML techniques were used to simulate the responses obtained through the adsorption process developed in Aspen Adsorption. Twenty-eight established machine learning algorithms were applied to a dataset of 70 input data points to select the best statistical indicators for each evaluated response. These algorithms allow for a more accurate estimation of the evaluated response variables, presenting results in a validation phase and a testing phase. The results obtained from applying the set of 70 input data points to the ML implementation can be seen in Table 8. It can be observed that, for the TC, TM, SVMF, and EEL models, R^2 values below 0.8 were obtained for all the response variables studied. On the other hand, the EBO model exhibits low R^2 values for the efficiency response, while the SVMK model exhibits low R^2 values for the BT and ST response variables. This indicates that these models have a limited ability to capture the patterns and relationships within the implemented dataset. Likewise, it is evident that the GPRS model performed the best.

Table 8. Computation of RMSE and R^2 for the 28 ML models.

Model Type	BT (min)				ST (min)				Efficiency			
	RMSE (V)	RMSE (T)	R^2 (V)	R^2 (T)	RMSE (V)	RMSE (T)	R^2 (V)	R^2 (T)	RMSE (V)	RMSE (T)	R^2 (V)	R^2 (T)
LRL	27.441	26.745	0.923	0.852	124.005	110.259	0.917	0.948	0.011	0.008	0.926	0.935
LRI	27.184	28.449	0.925	0.832	134.037	126.221	0.903	0.932	0.011	0.008	0.922	0.932
LRR	28.121	25.929	0.919	0.860	124.165	113.912	0.917	0.945	0.011	0.008	0.931	0.934
SLR	28.163	26.095	0.919	0.859	122.930	110.144	0.919	0.948	0.011	0.008	0.923	0.937
FT	36.119	24.576	0.867	0.875	141.696	131.927	0.892	0.926	0.015	0.010	0.879	0.871
TM	53.546	32.775	0.708	0.777	224.983	215.236	0.727	0.802	0.021	0.019	0.602	0.773
TC	99.045	85.106	-4.44×10^{-16}	-0.503	430.740	486.968	-2.22×10^{-16}	-0.013	0.043	0.032	-0.159	4.44×10^{-16}
SVML	28.342	23.564	0.918	0.885	128.317	124.637	0.911	0.934	0.012	0.006	0.963	0.920
SVMQ	8.243	7.561	0.993	0.988	99.254	93.625	0.947	0.963	0.004	0.004	0.983	0.992
SVMC	8.953	8.376	0.992	0.985	96.874	95.596	0.949	0.961	0.004	0.003	0.988	0.990
SVMF	78.524	58.936	0.371	0.279	349.909	379.128	0.340	0.386	0.036	0.019	0.602	0.310
SVMM	22.255	17.545	0.950	0.936	153.720	171.385	0.873	0.875	0.011	0.006	0.965	0.940
SVMG	36.745	15.098	0.862	0.953	155.348	163.104	0.870	0.886	0.015	0.005	0.973	0.874
ELE	61.144	26.470	0.619	0.855	358.711	460.349	0.306	0.095	0.042	0.027	0.156	0.072
EEL	103.834	62.328	-0.099	0.194	433.766	464.315	-0.014	0.079	0.044	0.030	-0.015	-0.052
EBO	32.339	9.411	0.893	0.982	164.451	111.794	0.854	0.947	0.038	0.038	-0.642	0.242
EBA	40.782	19.735	0.830	0.919	174.018	173.663	0.837	0.871	0.017	0.010	0.893	0.846
GPRS	1.611	0.710	1.000	1.000	93.414	86.945	0.953	0.968	0.001	0.001	1.000	1.000
GPRM	2.127	0.843	0.999	1.000	92.975	87.944	0.953	0.967	0.001	0.000	1.000	1.000
GPPE	12.976	5.671	0.983	0.993	101.991	100.977	0.944	0.956	0.005	0.002	0.995	0.985
GPRR	1.637	0.686	1.000	1.000	93.507	86.945	0.953	0.968	0.001	0.001	1.000	1.000
NNN	7.857	3.005	0.994	0.998	84.519	120.800	0.961	0.938	0.005	0.002	0.994	0.989
NNM	3.209	1.799	0.999	0.999	98.213	79.757	0.948	0.973	0.011	0.008	0.936	0.939
NNW	2.542	2.220	0.999	0.999	91.483	122.692	0.955	0.936	0.017	0.012	0.833	0.838
NNB	2.979	4.850	0.999	0.995	120.966	81.775	0.921	0.971	0.007	0.001	0.999	0.971
NNT	9.179	6.018	0.991	0.992	116.040	110.405	0.927	0.948	0.011	0.005	0.969	0.932
SVMK	96.697	67.684	0.047	0.049	433.570	480.806	-0.013	0.013	0.007	0.005	0.977	0.974
LSRK	34.259	28.512	0.880	0.831	175.933	238.773	0.833	0.757	0.013	0.008	0.933	0.905

Note: T = testing stage, and V = validation stage. The gray cells represent the selected ML model.

It can be observed that, for the GPRS model, R^2 values close to 1 were obtained for all response variables; this indicates that this model provided the best fit for the dataset used. In this context, it can be concluded that the relationship between the input variables and the response variables is highly consistent and reproducible. Figure 4 shows a comparison between the values obtained using Aspen Adsorption (True) and the predicted values (Predicted). The selected ML algorithms showed excellent agreement with the data obtained from the Aspen Adsorption simulation, demonstrating a good correlation between the validation phase and the testing phase, indicating that there was no overfitting of the parameters.

The use of a dataset consisting of 70 data points proved effective for fitting the model, demonstrating that the ML models successfully captured the pattern generated by the data produced by the Aspen Adsorption software. Furthermore, the use of SHAP (Shapley Additive Explanations) plots makes it possible to understand the influence of each predictor variable on the response variable, identifying which parameters have the greatest positive or negative impact on the model's output. As can be seen in Figure 5, the predictor with the greatest effect on the response variables BT (Figure 5a), ST (Figure 5b), and Efficiency (Figure 5c) was the Q parameter, followed by h and finally the model type.

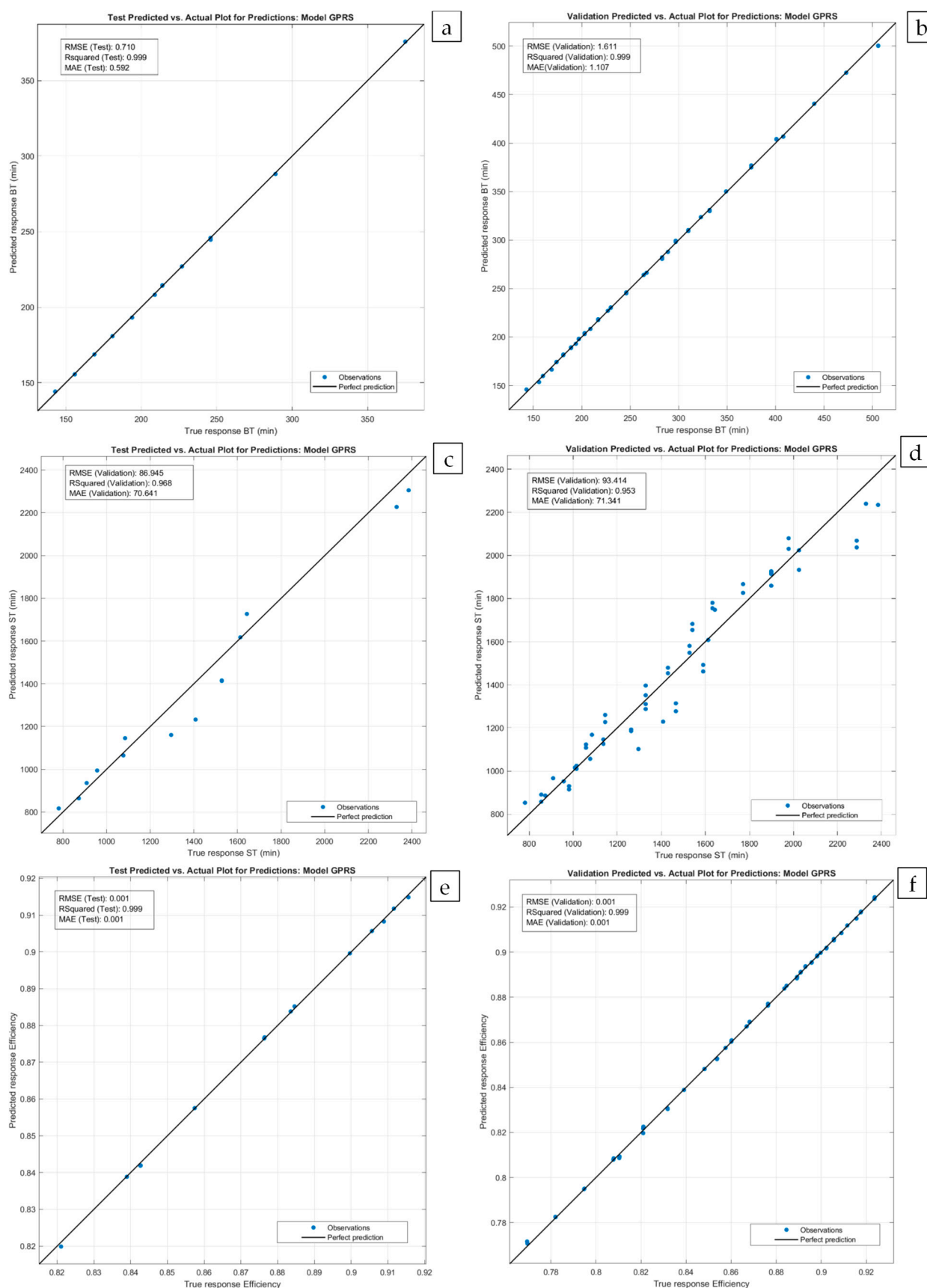


Figure 4. Comparison of actual and predicted values: for BT, (a) validation phase and (b) testing phase; for ST, (c) validation phase and (d) testing phase; for efficiency, (e) validation phase and (f) testing phase.

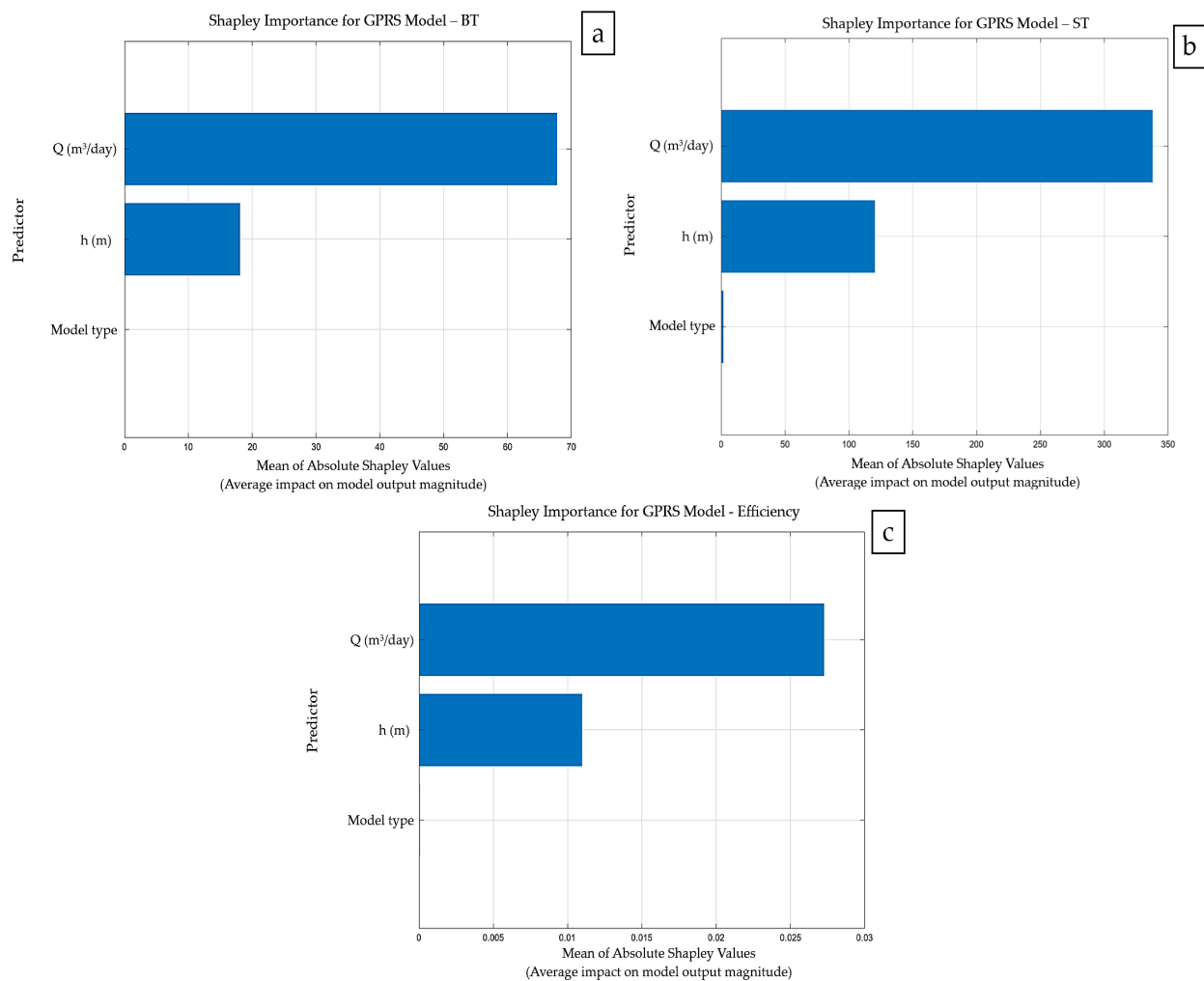


Figure 5. Shapley evaluates the use of a GPRS model to describe the impact of predictors on response variables. (a) BT, (b) ST, (c) Efficiency.

Other important factors for evaluating the performance of ML models are prediction speed and training time, which help determine the models' computational efficiency and feasibility of implementation. Training time represents the model's ability to learn from the dataset within a given period.

Figure 6 shows that the range of training time values for the models varies between 0 and 20 s, indicating that, for the response variable BT, most ML models had training times greater than 10 s. Meanwhile, for the response variables ST and efficiency, the values ranged from 0 to 10 s, with the results obtained for efficiency showing the lowest values in most models.

On the other hand, prediction speed refers to how quickly the model generates results after training. Figure 7 shows the prediction speed ranges for the ML models used. It can be observed that the prediction speed results ranged from 0 to 5100 obs/sec. It is evident that, for the BT response variable, the lowest average prediction values were obtained, indicating that, for this parameter, the models generate results more quickly and efficiently, requiring less computational time to process new data and generate responses. Meanwhile, the efficiency response variable obtained the highest prediction speed values, indicating that, for this parameter, the models require more computational time to produce a response.

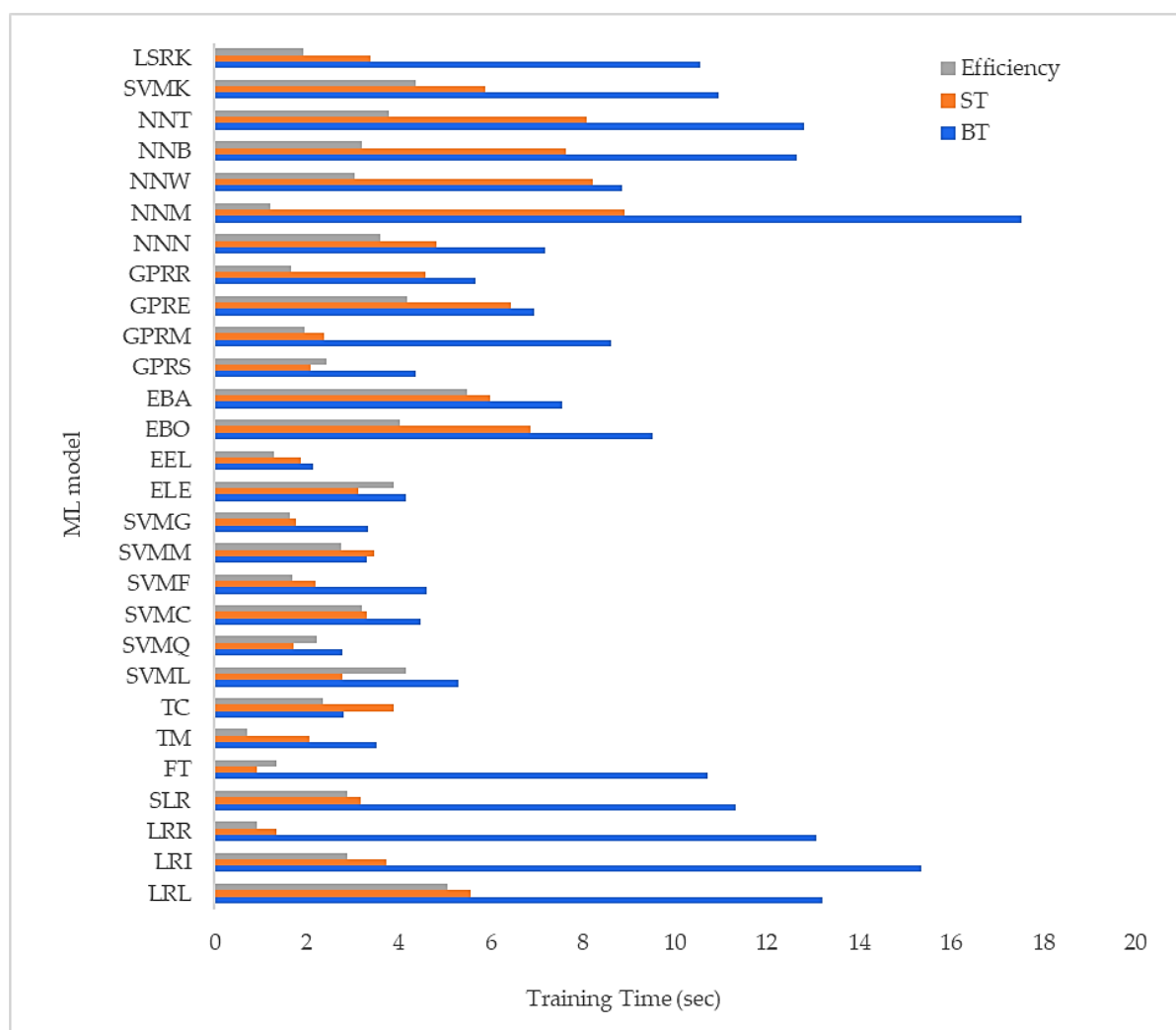


Figure 6. Training Time of ML models.

Subsequently, the ML models were exported to Simulink in order to build a digital system based on the trained model. This conversion of the machine learning model to Simulink facilitates the model's interaction with various input and output variables in real time, providing a highly practical way to use simulations under more realistic conditions and enabling more precise control of process conditions. This makes it an excellent tool applicable to various systems, generating more reliable predictions based on real-time operational data. Figure 8 presents a diagram of the system exported to Simulink, showing that the results obtained with this computational tool are similar to those obtained from the simulation process in Aspen Adsorption.

The combined use of these computational tools enables the implementation of a technique known as surrogate modeling, a method for creating mathematical models or simplified computations that provides a simpler approximation of the behavior of complex systems and their costs. This allows the data obtained from the simulation in Aspen Adsorption to be transformed into predictive models using ML for subsequent export to Simulink to build a digital system, making it a convenient tool for data prediction and calibration to ensure the fit, accuracy, and robustness of estimates in various processes.

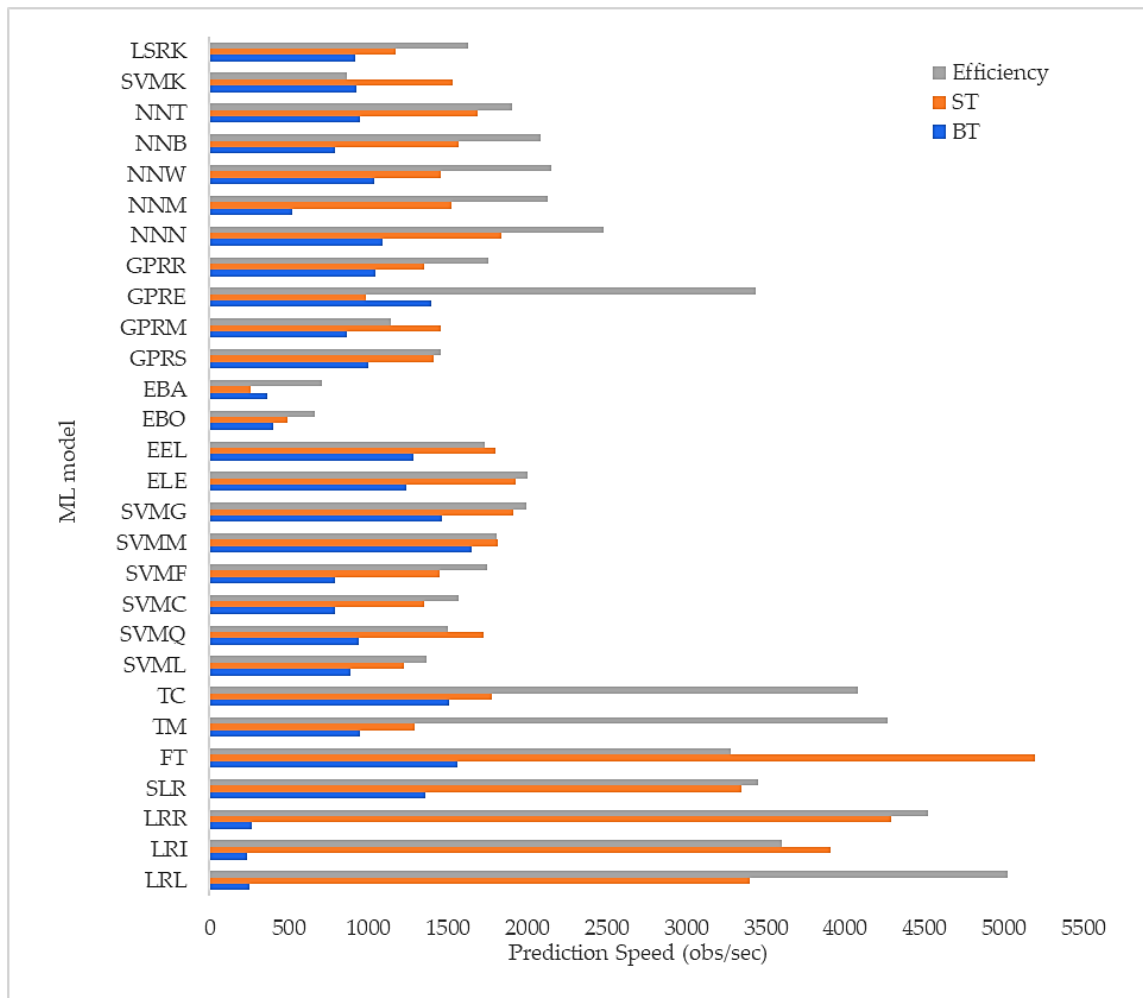


Figure 7. Prediction speed of ML models.

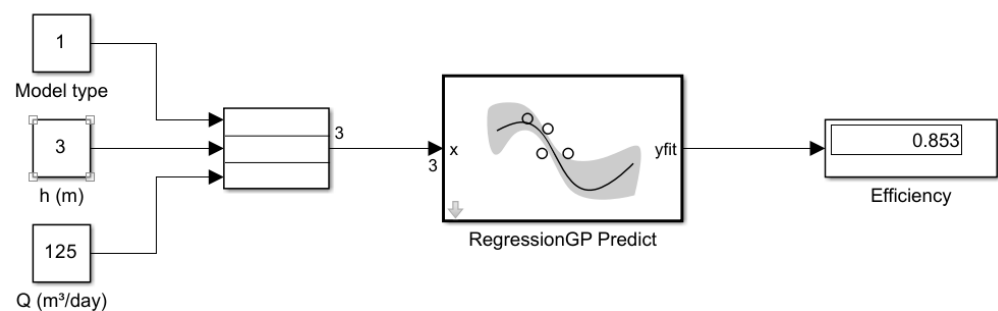


Figure 8. System diagram created by exporting the results of the trained model to Simulink.

4. Conclusions

This study is presented as exploratory research, demonstrating the potential performance of a packed bed containing *Dioscorea rotundata* as an adsorbent for the removal of Pb(II) from water sources. The data obtained from the simulations provide valuable information for trend analysis and the formulation of hypotheses regarding the system's performance under full-scale conditions.

The implementation of a parametric analysis highlighted the impact of changes in flow rate and bed height on system performance, revealing that a taller column prolongs the time required for system saturation to occur, while a higher flow rate enhances process performance, demonstrating that the appropriate selection of these variables is fundamental for system modeling and operation.

In addition, the development of validation and testing phases using machine learning algorithms made it possible to efficiently evaluate the model's performance, achieving R^2 values of up to 0.99, which lays the groundwork for the feasibility of implementing machine learning techniques to replicate the analyzed process.

The combination of Aspen Adsorption with ML algorithms represents an excellent tool for the design and optimization of the adsorption process, allowing for the integration of system performance prediction through simulation with ML, thereby improving scenario evaluation by adjusting fundamental variables to enhance the model's predictive capability. Similarly, exporting the results to Simulink made it possible to build a digital system based on the trained model. This implementation of surrogate modeling allowed for the creation of a simpler model that approximates the performance of a more complex system.

In the future, a technical-economic assessment will be conducted to determine the feasibility and sustainability of the process, with the aim of expanding the current approach by implementing experimental validation. This would allow us to move from the exploratory simulation stage to a phase with more established, well-founded, and experimentally supported characteristics, enabling the design and optimization of adsorption systems.

Author Contributions: Conceptualization, C.T.-T., A.V.-O. and O.E.C.-H.; methodology, C.T.-T., A.V.-O. and J.Q.-C.; software, A.V.-O., O.E.C.-H. and J.Q.-C.; validation, A.V.-O., O.E.C.-H. and M.J.R.-P.; formal analysis, A.V.-O., O.E.C.-H. and J.Q.-C.; investigation, C.T.-T. and A.V.-O.; data curation, O.E.C.-H. and M.J.R.-P.; writing—original draft preparation, C.T.-T., O.E.C.-H. and M.J.R.-P.; writing—review and editing, A.V.-O., O.E.C.-H. and M.J.R.-P.; visualization, O.E.C.-H. and M.J.R.-P.; supervision, O.E.C.-H. and M.J.R.-P.; project administration, C.T.-T. and A.V.-O. All authors have read and agreed to the published version of the manuscript.

Funding: The APC of this research was covered by the Universidad de Córdoba–Colombia.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding authors.

Acknowledgments: The authors would like to thank the Universidad de Cartagena for providing them with equipment, reagents, and technical, personnel, and computational support.

Conflicts of Interest: The authors declare that they have no conflicts of interest.

Nomenclature

The following abbreviations are used in this manuscript:

b	Langmuir constant (L/mg);
C_e	Concentration of contaminant in solution at equilibrium (mg/L);
C_f	Output concentration at the saturation point (mg/L);
f	Latent variable introduced in each observation;
H	Matrix of basic functions;
I	Identify matrix;
k_F	Freundlich constant indicating adsorption capacity [(mg/g)·(mg/L) ⁿ];
MTC_{sk}	Mass transfer parameter that controls the rate at which the contaminant approaches equilibrium within the adsorbent material (1/s);
N	Total number of observations;
P_i	Predicted value of a response;
q_e	Adsorption capacity of the contaminant (mg/g);
q_{max}	Maximum loading capacity (mg/g);
$\bar{T}$	Average value of observations;
T_i	True value of a response;
w_k	Concentration adsorbed in the solid for component k (mg/g);
w_k^*	Amount that should be adsorbed if the system were in instantaneous equilibrium

	with the fluid phase (mg/g);
$(w_k^* - w_k)$	Expression represents the driving force of mass transfer (mg/g);
$\frac{\partial w_k}{\partial t}$	Expression represents the rate of change in the adsorbed amount with respect to time (mg/g $\times$ s)
X	Predictors;
y	Response variable;
β	Coefficient computed from the dataset;
σ^2	Error variance.
$1/n$	Effect of initial concentration on adsorption capacity;
$(1 - C_f/C_0)$	Indicates the efficiency of the adsorption process (%);

References

- Lin, L.; Yang, H.; Xu, X. Effects of Water Pollution on Human Health and Disease Heterogeneity: A Review. *Front. Environ. Sci.* **2022**, *10*, 880246. [\[CrossRef\]](#)
- Babuji, P.; Thirumalaisamy, S.; Duraisamy, K.; Periyasamy, G. Human Health Risks due to Exposure to Water Pollution: A Review. *Water* **2023**, *15*, 2532. [\[CrossRef\]](#)
- Shekhar, C.; Khosya, R.; Sharma, A.K.; Thakur, K.; Mahajan, D.; Kumar, R.; Kumar, S.; Sharma, A.K. A systematic review on health risks of water pollutants: Classification, effects and innovative solutions for conservation. *Toxicol. Res.* **2024**, *14*, tfaf014. [\[CrossRef\]](#) [\[PubMed\]](#)
- Xu, H.; Gao, Q.; Yuan, B. Analysis and identification of pollution sources of comprehensive river water quality: Evidence from two river basins in China. *Ecol. Indic.* **2022**, *135*, 108561. [\[CrossRef\]](#)
- Adjovu, G.E.; Stephen, H.; James, D.; Ahmad, S. Measurement of Total Dissolved Solids and Total Suspended Solids in Water Systems: A Review of the Issues, Conventional, and Remote Sensing Techniques. *Remote Sens.* **2023**, *15*, 3534. [\[CrossRef\]](#)
- Beljin, J.; Đukanović, N.; Anojčić, J.; Simetić, T.; Apostolović, T.; Mutić, S.; Maletić, S. Biochar in the Remediation of Organic Pollutants in Water: A Review of Polycyclic Aromatic Hydrocarbon and Pesticide Removal. *Nanomaterials* **2025**, *15*, 26. [\[CrossRef\]](#) [\[PubMed\]](#)
- Adam, A.B.; Mu'azu, J.B.; Titilayo, O.A.; Ba'aku, A.E.; Gani, J.; Abubakar, M.Y.A. The Role of Organic Pollutants in Water Pollution A Review. *J. Chem. Technol.* **2025**, *1*, 142–159. [\[CrossRef\]](#)
- Kordbacheh, F.; Heidari, G. Water Pollutants and Approaches for their Removal. *Mater. Chem. Horiz.* **2023**, *2*, 139–153. [\[CrossRef\]](#)
- Egaña, I.; Akhrimenko, V.; Ayerbe, M.; Abasolo, E.; Lertxundi, U.; Orive, G. Promoting sustainable pharmacy for tackling environmental pharmaceutical pollution. *Drug Discov. Today* **2026**, *31*, 104559. [\[CrossRef\]](#) [\[PubMed\]](#)
- Kadim, M.K.; Risjani, Y. Biomarker for monitoring heavy metal pollution in aquatic environment: An overview toward molecular perspectives. *Emerg. Contam.* **2022**, *8*, 195–205. [\[CrossRef\]](#)
- Collin, S.; Baskar, A.; Geevarghese, D.M.; Ali, M.N.V.S.; Bahubali, P.; Choudhary, R.; Lvov, V.; Tovar, G.I.; Senatov, F.; Koppala, S.; et al. Bioaccumulation of lead (Pb) and its effects in plants: A review. *J. Hazard. Mater. Lett.* **2022**, *3*, 100064. [\[CrossRef\]](#)
- Luo, H.; Su, Z.; Liu, Y.; Yuan, D.F.; Wang, R.; Ning, Y.H.; Zhang, D.J.; Chen, X.K.; Wang, Z.B.; Gao, X.Y.; et al. Effective removal of Pb from industrial wastewater: A new approach to remove Pb from wastewater based on engineered yeast. *J. Hazard. Mater.* **2025**, *481*, 136516. [\[CrossRef\]](#) [\[PubMed\]](#)
- Acuña, E.; Castillo, B.; Queupuan, M.; Casanova, M.; Tapia, Y. A review of phytoremediation of environmental lead (pb) contamination. *Chemosphere* **2024**, *362*, 142691. [\[CrossRef\]](#) [\[PubMed\]](#)
- Zhao, H.; Ouyang, X.K.; Yang, L.Y. Adsorption of lead ions from aqueous solutions by porous cellulose nanofiber–sodium alginate hydrogel beads. *J. Mol. Liq.* **2021**, *324*, 115122. [\[CrossRef\]](#)
- Abdiyev, K.; Azat, S.; Kuldeyev, E.; Ybyraiymkul, D.; Kabdrakhmanova, S.; Berndtsson, R.; Khalkhabai, B.; Kabdrakhmanova, A.; Sultakhan, A.S. Review of Slow Sand Filtration for Raw Water Treatment with Potential Application in Less-Developed Coun-tries. *Water* **2023**, *15*, 2007. [\[CrossRef\]](#)
- Wu, Y.; Li, S.; Yu, K.; Hu, J.; Chen, Q.; Sun, W. Wastewater treatment plant effluents exert different impacts on antibiotic resistome in water and sediment of the receiving river: Metagenomic analysis and risk assessment. *J. Hazard. Mater.* **2023**, *460*, 132528. [\[CrossRef\]](#) [\[PubMed\]](#)
- Mao, J.; Hu, G.; Deng, W.; Zhao, M.; Li, J. Industrial wastewater treatment using floating wetlands: A review. *Environ. Sci. Pollut. Res.* **2023**, *31*, 5043–5070. [\[CrossRef\]](#)
- Lucas, M.S.; Teixeira, A.R.; Jorge, N.; Peres, J.A. Industrial Wastewater Treatment by Coagulation–Flocculation and Advanced Oxidation Processes: A Review. *Water* **2025**, *17*, 1934. [\[CrossRef\]](#)

19. Sharma, M.; Agarwal, S.; Malik, R.A.; Kumar, G.; Pal, D.B.; Mandal, M.; Sarkar, A.; Bantun, F.; Haque, S.; Singh, P.; et al. Recent advances in microbial engineering approaches for wastewater treatment: A review. *Bioengineered* **2023**, *14*, 2184518. [[CrossRef](#)] [[PubMed](#)]
20. Raji, Z.; Karim, A.; Karam, A.; Khalloufi, S. Adsorption of Heavy Metals: Mechanisms, Kinetics, and Applications of Various Adsorbents in Wastewater Remediation—A Review. *Waste* **2023**, *1*, 775–805. [[CrossRef](#)]
21. Tran, H.N. Adsorption Technology for Water and Wastewater Treatments. *Water* **2023**, *15*, 2857. [[CrossRef](#)]
22. Zhao, D.; Zhang, S.; Deng, H.; Hu, L.; Li, A. Eco-Friendly Adsorbent: Insights Into the Performance and Adsorption Mechanisms of Banana Fruit Shaft Biochar for the Removal of Mn(II), Cd(II), Pb(II), and Cu(II). *Appl. Organomet. Chem.* **2025**, *39*, e70036. [[CrossRef](#)]
23. Yang, X.; Chen, L.; He, G.; Yang, M.; Geng, S.; Sun, Y.; Feng, X.; Ma, C.; Wei, Q.; Zhao, H.; et al. Chitosan-loaded coconut fiber for highly-enhanced heavy metal adsorption from wastewater. *Sep. Purif. Technol.* **2025**, *368*, 133105. [[CrossRef](#)]
24. Villabona-Ortíz, A.; Tejada-Tovar, C.; González-Delgado, Á.D. Elimination of Chromium (VI) and Nickel (II) Ions in a Packed Column Using Oil Palm Bagasse and Yam Peels. *Water* **2022**, *14*, 1240. [[CrossRef](#)]
25. Tejada-Tovar, C.; González-Delgado, Á.; Villabona-Ortíz, Á. Adsorption Kinetics of Orange Peel Biosorbents for Cr (VI) Uptake from Water. *Contemp. Eng. Sci.* **2018**, *11*, 1185. [[CrossRef](#)]
26. Tejada-Tovar, C.; Villabona-Ortiz, A.; Alvarez-Bajaire, G.; Jattin-Torres, L.; Acevedo, D. Evaluation of *Theobroma cacao* Waste Performance in Nickel Removal (II) in Continuous System. *Int. J. Chemtech Res.* **2018**, *11*, 186–194. [[CrossRef](#)]
27. Caron, C.; Lauret, P.; Bastide, A. Machine Learning to speed up Computational Fluid Dynamics engineering simulations for built environments: A review. *Build. Environ.* **2025**, *267*, 112229. [[CrossRef](#)]
28. Tejada-Tovar, C.; Villabona-Ortiz, Á.; Coronado-Hernández, O.E.; Haeckermann-Ruiz, E.; Méndez-Anillo, R.D. Assessing Banana-Based Activated Carbon as a Biomaterial for the Adsorption of Drug Metabolites in Wastewater: Simulation of an Industrial-Scale Packed Column. *Processes* **2026**, *14*, 129. [[CrossRef](#)]
29. Makinde, K.A.; Bakare, M.S.; Akinloye, B.O.; Amole, A.O.; Adewuyi, O.B.; Zubair, U.O.; Owonikoko, W.O. Simulation based testing and performance investigation of induction motor drives using matlab simulink. *SN Appl. Sci.* **2023**, *5*, 73. [[CrossRef](#)]
30. Agarwal, A.; Upadhyay, U.; Sreedhar, I.; Anitha, K.L. Simulation studies of Cu(II) removal from aqueous solution using olive stone. *Clean. Mater.* **2022**, *5*, 100128. [[CrossRef](#)]
31. Ortega-Toro, R.; Tejada-Tovar, C.N.; Villabona-Ortiz, Á.; González-Delgado, Á.D.; Vergara-Villadiego, J.C. Computational prediction of packed-bed reactor performance for hexavalent chromium removal from aqueous solution. *J. Water Land Dev.* **2025**, *66*, 13–18. [[CrossRef](#)]
32. Hameed, A.; Hameed, B.H.; Almomani, F.A.; Usman, M.; Ba-Abbad, M.M.; Khraisheh, M. Dynamic simulation of lead(II) metal adsorption from water on activated carbons in a packed-bed column. *Biomass Convers. Biorefin.* **2024**, *14*, 8283–8292. [[CrossRef](#)]
33. Grznár, P.; Gregor, M.; Mozol, Š.; Mozolová, L.; Krump, H.; Mizerák, M.; Trojan, J. A Comprehensive Analysis of Sensitivity in Simulation Models for Enhanced System Understanding and Optimisation. *Processes* **2024**, *12*, 716. [[CrossRef](#)]
34. Upadhyay, U.; Gupta, S.; Agarwal, A.; Sreedhar, I.; Anitha, K.L. Process Optimization at an Industrial Scale in the adsorptive removal of Cd²⁺ ions using Dolochar via Response Surface Methodology. *Environ. Sci. Pollut. Res.* **2021**, *30*, 8403–8415. [[CrossRef](#)] [[PubMed](#)]
35. Rahman, M.A.; Lamb, D.; Kunhikrishnan, A.; Rahman, M.M. Kinetics, isotherms and adsorption–Desorption behavior of phosphorus from aqueous solution using zirconium–iron and iron modified biosolid biochars. *Water* **2021**, *13*, 3320. [[CrossRef](#)]
36. Ragadhita, R.; Bayu, A.; Nandiyanto, D. Curcumin Adsorption on Zinc Imidazole Framework-8 Particles: Isotherm Adsorption Using Langmuir, Freundlich, Temkin, And Dubinin-Radushkevich Models. *J. Eng. Sci. Technol.* **2022**, *17*, 1078–1089.
37. Debnath, S.; Das, R. Strong adsorption of CV dye by Ni ferrite nanoparticles for waste water purification: Fits well the pseudo second order kinetic and Freundlich isotherm model. *Ceram. Int.* **2023**, *49*, 16199–16215. [[CrossRef](#)]
38. Cojocar, V.C.; Cristea, I.N.; Paris, I.A.; Ionescu, I.A.; Chiriac, F.L. Exploring Sustainable Solutions: Dynamic Adsorption, Isotherm Models, and Kinetics of Organic Contaminants on Polystyrene Microplastics. *Sustainability* **2024**, *16*, 7743. [[CrossRef](#)]
39. Nikam, S.; Mandal, D.; Dabhade, P. LDF based parametric optimization to model fluidized bed adsorption of trichloroethylene on activated carbon particles. *Particuology* **2022**, *65*, 72–92. [[CrossRef](#)]
40. Chicco, D.; Warrens, M.J.; Jurman, G. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Comput. Sci.* **2021**, *7*, e623. [[CrossRef](#)]
41. Hodson, T.O. Root-mean-square error (RMSE) or mean absolute error (MAE): When to use them or not. *Geosci. Model Dev.* **2022**, *15*, 5481–5487. [[CrossRef](#)]
42. Singh, V.; Nimalipuri, P.; Vitankar, V.; Das, H.C. Multifluid computational fluid dynamics simulation for sawdust gasification inside an industrial scale fluidized bed gasifier. *Mater. Today Proc.* **2023**; in press. [[CrossRef](#)]
43. Tejada-Tovar, C.; Villabona-Ortíz, Á.; Gonzalez-Delgado, Á.; Ortega-Toro, R.; Ortega-Puente, S. Simulation of Pb(II) and Ni(II) Adsorption in a Packed Column: Effects of Bed Height, Flow Rate, and Initial Concentration on Performance Metrics. *Processes* **2025**, *13*, 2141. [[CrossRef](#)]

44. Yousef, R.; Qiblawey, H.; El-Naas, M.H. Evaluation of activated carbon fiber packed-bed for the treatment of gas-to-liquid wastewater: Experimental, modeling and ASPEN Adsorption simulation. *Emergent Mater.* **2024**, *8*, 1591–1603. [[CrossRef](#)]
45. Noorbakhsh, Y.; Bozorgghomi, S.; Ghaemi, A. Simulation and optimization of copper, nickel, cadmium, and zinc removal from industrial wastewater using Aspen adsorption and RSM. *Sci. Rep.* **2025**, *15*, 40606. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.