Journal of Process Control 130 (2023) 103068

Contents lists available at ScienceDirect

Journal of Process Control

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

A combined experimental-modeling approach for turbidity removal optimization in a coagulation–flocculation unit of a drinking water treatment plant

Agostina Chiavola [a,*], Camilla Di Marcantonio [a], Martina D'Agostini [a], Simone Leoni [b], Marco Lazzazzara [b]

[a] Sapienza University of Rome, Department of Civil, Building and Environmental Engineering (DICEA), Via Eudossiana 18, 00184 Rome, Italy
[b] ACEA ELABORI SpA, Via Vitorchiano 165, Rome, Italy

A R T I C L E I N F O

Article history:
Received 11 April 2023
Received in revised form 1 August 2023
Accepted 27 August 2023
Available online xxxx

Keywords:
Coagulant
Coagulation–flocculation
Drinking water treatment plant
Elbow method
k-means
Turbidity

A B S T R A C T

The drinking water treatment plant plays a key role in providing the consumers with a safe water, in compliance with the United Nations Sustainable Development Goal n. 6 (Clean water and sanitation). The typical drinking water treatment plant includes: coagulation–flocculation, sedimentation, filtration and disinfection. Turbidity removal is among the more common pollutants to remove; this goal is mainly achieved by the coagulation–flocculation process. The present paper shows the results of the application of a combined experimental-modeling approach for turbidity removal optimization in a coagulation–flocculation unit of a full-scale drinking water treatment plant. The applied approach consisted of a laboratory experimental activity aimed at determining the best coagulant type and dosage. Among the different chemicals tested, Poly aluminum chloride (PAC) provided the highest removal at the lowest dosage (90% at 3.5 mg/L PAC, 88% at 18.9 mg/L PACS and 77% at 30 mg/L $FeCl_3$). The addition of polyelectrolytes did not improve the removal at such a level to justify the consequent increase of the costs. Based on the findings of this phase of the study, a data-driven model was implemented using as input variables the historical data of influent and effluent turbidity and influent flow rate. Combining regression models and statistical analysis, it was possible to build up an algorithm for the case-study plant allowing to select the PAC dosage to apply as a function of the influent turbidity, to ensure the constant compliance with the regulation limit on effluent turbidity.

1. Introduction

The new European Drinking Water Directive (DWD), came into force in 2021, aims at further improving the quality of tap water for potable uses, with a better awareness of citizens in order to foster its use in place of the recourse to bottled water, unless strictly needed [1]. These goals fully fit the principles of the Circular Economy and Sustainability, because they allow reduction of packaging as well as the environmental impact of the bottled water industry. The DWD also updates the quality requirements of drinking water based on the more recent findings obtained by ecotoxicological studies. Particularly, the limits for some "old" parameters, like lead, became more stringent, whereas a concern to "new" parameters, like contaminants of emerging concern (CECs), was introduced. The need to ensure access of all to high quality and abundant drinking water, as highlighted by the United Nations Agenda and sustained by the new DWD, boosts the exploitation of different water sources and bears the enhancement of the treatment processes. Surface waters, such as rivers and lakes, can represent an available water source, provided that a proper treatment is applied to make its quality suitable for drinking water purposes. The required treatment depends on the initial characteristics and can span from a simple filtration and disinfection to a more complex layout including different chemical–physical processes [2]. The drinking water treatment plant for surface water is mostly designed to address the following main contaminants: turbidity, microorganisms and organic substance [3]. As a consequence, its typical layout consists of: screening, coagulation–flocculation, sedimentation, filtration and disinfection. In some circumstances, there is also the need to remove metals, metalloids (e.g. arsenic) or biorefractory organic compounds. Therefore, the plant includes additional and more complex stages, like pre-oxidation and/or adsorption [4].

A challenge that must be frequently addressed by the treatment plants for surface waters is represented by the wide fluctuation of the influent quality, due to season changes or uncontrolled pollutant discharges on the river.

* Corresponding author.
E-mail address: agostina.chiavola@uniroma1.it (A. Chiavola).

https://doi.org/10.1016/j.jprocont.2023.103068

The turbidity is among the parameters of main concern because it reduces water limpidity and can be also the vehicle of toxic substances and microorganisms. Turbidity is due to the presence of colloidal particles, which are characterized by a very low size (in the range 1 nm–1 μm): because of that, their removal by free settling or filtration is hindered. Furthermore, they also possess an electrical (mostly negative) charge on the surface which determines a repellent force between similar particles. As a consequence, the colloidal particles often form a stable suspension which cannot be destabilized by simple mixing conditions, and, the particles cannot aggregate into larger settleable/filtrable flocs [5–7]. Coagulation–flocculation is widely present in the water treatment plants with the aim to reduce the turbidity content. By adding positively charged coagulants under rapid mixing and pH-controlled conditions, the neutralization of the negative charges on the colloids is achieved and consequently the electric destabilization of the suspension takes place [6,8]. After the addition of the coagulants under rapid mixing conditions, the particles can bind with the chemicals and each other to form larger particles (flocs); gentle mixing allows aggregation of insoluble flocs and their progressive size increase (flocculation). Finally, settling of flocs to the bottom of the unit leads to the accumulation of a chemical sludge, whereas a clear supernatant, being rid of turbidity, is formed above.

The main parameters affecting the efficiency of turbidity removal in the coagulation–flocculation process are: coagulant type and dosage, pH value, speed of the mixing device, hydraulic retention time.

Aluminum and iron salts are the chemicals commonly used in the coagulation–flocculation process. However, these mineral coagulants can be costly and negatively alter the quality of the drinking water. Therefore, new and more environmentally sustainable organic chemicals have been tested successfully to the scope. For instance, in [9] a biodegradable natural coagulant was obtained from acorn leaves, which are abundantly available in many countries worldwide. Its use in drinking water treatment allowed to achieve a reduction of turbidity in the range 72%–85%, depending on its physical form. In a following study [10], the use of the natural coagulant *Aloe vera* in both powder and liquid forms reduced the water turbidity at natural pH by 28.23% and 87.84%, respectively. Moreover, it was found that this biocoagulant did not influence the main physical characteristics of the drinking water.

The amount of chemicals required to achieve colloid destabilization depends on the initial turbidity content, which can be subjected to frequent variations in the surface water with effects on the efficiency of the process [11]. Other factors which can also influence the dosage are: pH value, alkalinity content, concentrations of other components which might interfere with colloidal particles in the destabilization process. Depending on the turbidity level and type of coagulant used, a different destabilization process can prevail: therefore, the required dosage must be carefully determined and eventually adjusted when some changes in the water quality or other parameters intervene. For instance, if destabilization occurs through the adsorption-charge neutralization mechanism, overdosing the system with absorbable ions can cause a restabilization of the suspension, as a consequence of a reversal of the charge on the colloid particles [12]. Such consequences are of utmost relevance when dealing with drinking water, since the effects of improper treatment can ultimately produce an unsafe water with a possible threat to human health [13].

Jar test is a procedure typically used to determine the coagulant dosage by simulating in laboratory the two-stages of the coagulation–flocculation process of the treatment plant [14]. However, jar tests are time-consuming; furthermore, the complexity of the coagulant chemical theory can limit the correct determination of the dosage. Usually in the full-scale water treatment plants, the coagulant dosage selection often relies on the operator's experience, with the risk of wrong actions or improper dosing and consequent increase of the treatment cost or a poor quality of the treated water. A prediction model based on the historical data from the plant operation can be a better tool for optimizing coagulant dosage also under dynamic conditions. Data-driven modeling techniques often combine with the interdisciplinary areas of the fields of statistics, multiple objective analysis, evolutionary algorithms, etc. [15,16]. An example of the application of this type of approach is presented in [17].

The objective of the present study was to optimize the coagulation–flocculation process for turbidity removal in a drinking water treatment plant which receives the surface water from an artificial lake located in the central Italy. The unit is usually able to produce a treated effluent with the quality required in Italy for drinking purposes. However, the high costs of the treatment and the risk of not complying consistently with the limits because of the frequent variations of the influent quality were of concern to the company managing the plant. Therefore, it was highlighted the need to carry out a study to find out the best coagulant type and its dosage, needed to ensure a constant compliance with the regulation limits on drinking water.

A double approach was applied to the purpose: (1) an experimental activity was carried out at the laboratory scale with the aim to determine the best type of coagulant and evaluate the need of a polyelectrolyte to improve the turbidity removal; (2) a data-driven model was implemented based on the historical data of operation of the unit to predicting the dosage of the coagulant selected in (1) as a function of the raw water quality.

This double approach represents the key novel aspect of the present study, since it provides an application of the data-driven modeling technique to optimizing the coagulation–flocculation process of a full-scale drinking water treatment plant. A further novelty is represented by the combination of different tools to build up the predictive model: the cluster analysis to define the main classes of turbidity in the influent and three different regression models. Finally, a chart was generated showing the coagulant dosage to apply as a function of the raw water quality, to achieve a constant effluent turbidity. Therefore, combining a lab-scale activity with a mathematical approach, the study was finally able to a generate a valid tool to apply in the real operation of the plant.

2. Materials and methods

2.1. Drinking water treatment plant

The drinking water treatment plant (DWTP) is located in the central Italy and collects the raw water from a reservoir formed by stopping a surface water with a dam. After the intake structure, the water is subjected to fine screening and pre-aeration and then pumped to the DWTP. The first treatment is made in a water tower where poly-aluminum chloride (PAC) and sodium hypochlorite are dosed together under mixing conditions. Afterwards, the treatment proceeds through two parallel lines, each one having the same lay-out and consisting of the following main processes: coagulation–flocculation–sedimentation, sand filtration, granular activated carbon (GAC) adsorption column, disinfection with sodium hypochlorite Then, the treated water is released into the distribution network. The lay-out of the DWTP is shown in Fig. 1.

The DWTP is provided with a separate line for chemical sludge processing, which consists of a storage tank followed by a static thickener and a filter-press. The liquid centrate produced by the sludge separation is recirculated to the inlet of the DWTP for its treatment.

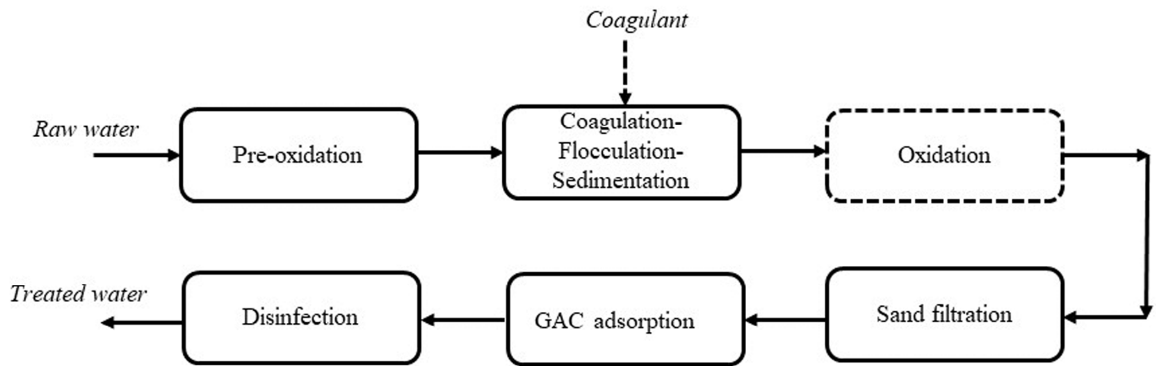

Fig. 1. Layout of the drinking water treatment plant.

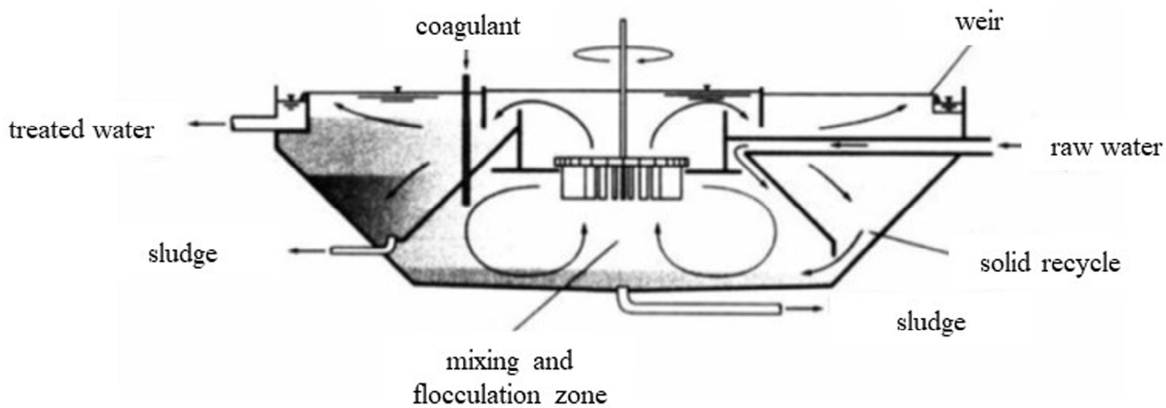

Fig. 2. A schematic view of the Accelerator (Veolia).

Coagulation, flocculation and sedimentation are all carried out in a single-basin unit, the so called "Accelator" (by Veolia) (Fig. 2). It consists of an upper cylindrical part and a lower hopper-shape part. The raw water enters through a pipe in the central part, where the coagulation takes place due to mixing with chemicals. Then, it moves to the flocculation section within the reverse cone (bell-shape) section. This is connected to the settlement space through two openings on the bottom. The treated water is removed by means of small radial channels. The Accelerator employs the principle of internal slurry recirculation to accelerate chemical reactions and dense particle growth. The main advantages the system offers are: different mixing processes for flocculation, coagulation and sedimentation within one single unit, with less footprint; no need for sludge collectors nor electro-pumps for returning sludge from the sedimentation section to the initial coagulating zone; possibility of simultaneous softening and clarification; relatively acceptable yield. Therefore, its implementation in the drinking water treatment plants is increasing [18].

Sizes of the accelerator are listed in Table 1, whereas the values of its main operating parameters are reported in Table 2. They were all set-up by the supplier of the accelerator, based on their previous experiences and preliminary tests carried out to assess the optimal values for the present case-study.

The volumetric flowrate as well as the turbidity in the influent to the DWTP are characterized by a wide fluctuation over time, as shown by the Q values reported in Table 2. These variations are due to the torrential nature of the river barred by the dam; as a consequence, the turbidity values change with the season (from 5–7 NTU during dry weather, up to even 1700 NTU in the winter time). During the extreme rain events, a high quantity of solids is released to the lake.

Table 1
Sizes of the coagulation–flocculation–sedimentation unit.

Parameter	Value	Unit
Diameter, D	17	m
Horizontal surface, A	227	m^2
Coagulation–Flocculation tank volume, V_{CF}	200	m^3
Sedimentation tank volume, V_S	800	m^3
Total volume, V_{tot}	1000	m^3

During the period of jar tests, samples of influent showed turbidity between 15 and 17 as average, with a peak to 24 and a minimum of 1.7. About the other quality parameters of the influent to the DWT, the range values were as follows (with the average in the brackets): 7.6–8.1 (7.8) pH; 390–1238 (1036) conductivity (μS/cm); 185–544 (331) total aluminum (mg/L); 1.3–209 (108) filtered aluminum (mg/L); 0.99–3.64 (1.9) TOC (mg/L); 0.74–3 (1.4) filtered COD (mg/L); 0.5–14 (4.1) TSS (mg/L).

The treated water from the DWTP complies with the regulation limits on drinking water (D. Lgs 18/2023, which received the European directive 2020/2184) [19,20].

2.2. Lab-scale investigation

The laboratory activity was conducted through a series of jar tests using 1 L volume glass beakers, to determine the best type of coagulant, the optimal dosage under controlled conditions and finally whether there was the need of adding also a coagulant aid. Table 3 shows the list of chemicals used in the study, whereas Tables 4 and 5 highlight the conditions and concentrations applied to each test. For PAC, all the dosages were expressed as concentration of Al_2O_3, making the data reliable and comparable aside from the commercial chemical used.

Table 2
Main operating parameters of the coagulation–flocculation–sedimentation unit.

Parameter	Value			Unit
	Min	Max	Av	
Influent volumetric flow rate, Q	113.5	227	197.5	m^3/h
Surface loading rate, C_{is}	0.5	1.0	0.9	$m^3/(m^2 \times h)$
Coagulation–Flocculation Hydraulic Retention Time, $\Theta_{H,CF}$	1.7	0.9	1.0	h
Sedimentation Hydraulic Retention Time, $\Theta_{H,S}$	7.0	3.5	4.0	h
Total Hydraulic Retention Time, $\Theta_{H,tot}$	8.7	4.4	5.0	h

Legend: Min = at the minimum influent volumetric flow rate; Max = at the maximum influent volumetric flow rate; Av = at the average influent volumetric flow rate.

Table 3
Types and characteristics of the chemicals used in the study.

Chemical formula/ Commercial name	Chemical name and description	Characteristics (concentration, density)
$Al_2(OH)_5Cl$, PAC	Poly aluminum chloride	9.5% Al_2O_3, 1.2 mg/L
$FeCl_3$	Ferric chloride	41%, 1.48 mg/L
PACS	Sulfate poly aluminum chloride	19% Al_2O_3, 1.3 mg/L
NaOCl	Sodium hypochlorite	14%, 1.2 mg/L
PLT 35	Cationic polyelectrolyte	40%, 1.1 mg/L
PLT 37	Cationic polyelectrolyte	40%, 1.1 mg/L
DFC 104	On-average branched out, low-medium molecular weight	50%, 1.1 mg/L
DFC 105	Highly branched out, high molecular weight	50%, 1.1 mg/L
DFC 106	Highly branched out, high molecular weight	50%, 1.1 mg/L
DFC 535 C	Cationic flocculant, average molecular weight	50%, 1.1 mg/L
DFC 201 N	Non-ionic flocculant, average molecular weight	50%, 1.1 mg/L
DFC 203 A	Anionic flocculant, average molecular weight	50%, 1.1 mg/L
CP	Carbon powder	40%

Table 4
Conditions of the tests on coagulant type and dosage selection.

Coagulant	Dosage (mg/L)
Al_2O_3/NaOCl	0, 2, 3.5, 5, 6.5, 8/2
$FeCl_3$/NaOCl	0, 3, 6, 9, 12, 15/2
$FeCl_3$/NaOCl	0, 20, 26, 28, 44, 52/2
Al_2O_3/$FeCl_3$/NaOCl	3.5/0, 5, 10, 15, 20, 30/2
PACS (1)/NaOCl	0, 1.89, 5.66, 9.43, 18.87, 28.3/2
PACS (2)/NaOCl	0, 0.24, 0.72, 1.2, 2.4, 3.6/2
CP/Al_2O_3/NaOCl	0, 5, 10, 20, 40/3.5/2

Table 5
Conditions of the tests on coagulant aid selection.

Coagulant aid	Dosage (mg/L)
Al_2O_3/PLT 35	3.5/0, 0.5, 1, 2, 5, 10
Al_2O_3/PLT 37	3.5/0,0.5, 1, 2, 5, 10
Al_2O_3/DFC 104	3.5/0, 0.1, 0.2, 0.5, 1, 2
Al_2O_3/DFC 105	3.5/0, 0.1, 0.2, 0.5, 1, 2
Al_2O_3/DFC 106	3.5/0, 0.1, 0.2, 0.5, 1, 2
Al_2O_3/DFC 535 C	3.5/0, 0.1, 0.2, 0.5, 1, 2
Al_2O_3/DFC 201 N	3.5/0, 0.1, 0.2, 0.5, 1, 2
Al_2O_3/DFC 203 A	3.5/0, 0.1, 0.2, 0.5, 1, 2

The ranges of concentrations used in the experiments were chosen based on the values usually applied at the full-scale to address similar turbidity levels, as suggested by the operators of the case-study plant. Ferric chloride and PAC were firstly tested at a similar dosage and separately. Then, ferric chloride concentration only was increased to evaluate if it was possible to boost the formation of more dense and heavy flocs with a better settling capacity. Further tests were carried by mixing together PAC and $FeCl_3$. Sodium hypochlorite was always added to the tests at 2 mg/L dosage, as it is done in the real scale plant.

A series of tests was also carried out using carbon powder (CP), due to its capability of removing the natural organic matter (NOM) which might be responsible of the toxic disinfection by-product formation.

Since PAC provided the highest removal, then the second phase was carried out using only this chemical at its optimal dosage as determined in the first phase. This second series of experiments aimed at determining if there was a coagulant aid capable of further enhancing the turbidity removal obtained by using PAC. To this purpose, several polyelectrolytes were tested, such as cationic, anionic and non-ionic, differing for the degree of branching and the molecular weight (as reported in Table 3).

Each test was conducted according to the following schedule: (1) 4 min under rapid mixing at 60 rpm, (2) 35 min under gentle mixing at 20 rpm, (3) sedimentation without mixing for 20 min. These conditions were selected based on preliminary tests carried out in the past on the same influent.

The experiments were carried out using the raw water entering the DWTP, and therefore with initial turbidity changing in the range 2–24 NTU (average 6 NTU) to apply the real conditions. The following parameters were measured on all the samples from the tests: turbidity, pH, electrical conductivity, total and dissolved aluminum, iron, total organic carbon (TOC), dissolved organic carbon (DOC) and total suspended solids (TSS).

The efficiency of the process was evaluated based on the turbidity removal rate by applying the following equation:

$$Removal\ [\%] = \frac{(NTU_{in} - NTU_{out})}{NTU_{in}} \times 100 \quad (1)$$

where: NTU_{in}, NTU_{out} = turbidity in the influent and effluent of the Accelerator, respectively [NTU] However, the effects on the other measured parameters were also considered.

2.3. Analytical methods

All the experiments were conducted according to the Italian Metodi Analitici per le Acque by APAT IRSA-CNR (Agenzia per la Protezione dell'Ambiente e per i Servizi Tecnici, Istituto di Ricerca sulle Acque, Consiglio Nazionale delle Ricerche) and to Standard Methods for Water and Wastewater Examination [21, 22]. Particularly, the following procedures were applied for the determinations of the different parameters: UNI (Ente Nazionale Italiano di Unificazione) EN (Comité Européen de Normalisation) ISO (International Organization for Standardization) 7027-1:2016/1 for turbidity; APAT CNR IRSA 2060 Man 29 2003/1 for

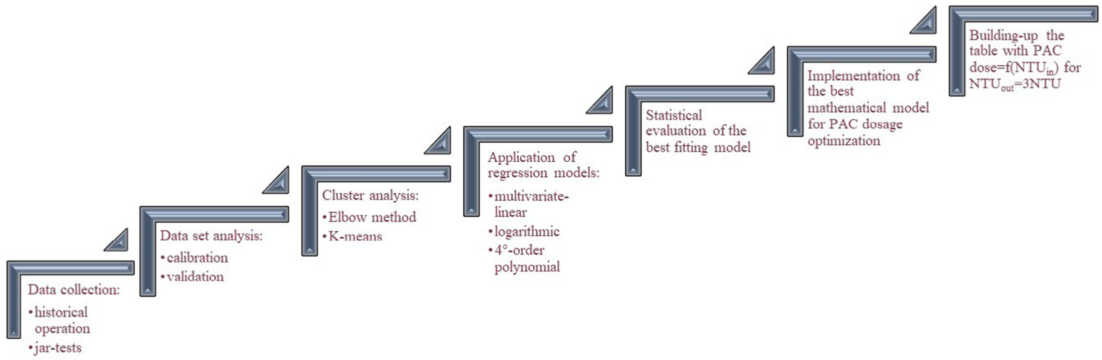

Fig. 3. Flow-chart of the main steps of the experimental-modeling study.

pH; UNI EN 27888:1995/1 for electrical conductivity; APHA SM 3030 A + APHA SM 3125/1 for total and dissolved Al and Fe; UNI EN 1484:1999/1 for filtered dissolved organic carbon (DOC) (filtrated at 0.2 μm) and total organic carbon (TOC); APAT CNR IRSA 2090 B Man 29 2003/1 for total suspended solids (TSS).

The results obtained were analyzed by means of Excel and R software.

2.4. Data-driven modeling

The aim of this part of the study was to implement a mathematical model based on the historical data of operation of the DWTP (a data-driven model) to optimizing the dosage of the coagulant selected in the experimental part of the study, i.e. PAC, as a function of the influent turbidity. For instance, in the laboratory tests, the best coagulant type was selected and the dosage was optimized based on a narrow range of variation of the influent turbidity. The proposed model would allow to consider a much wider range of variation of the influent turbidity, as it occurs in the real operation of the DWTP. Therefore, the model can represent a tool for the automatic control of PAC addition in the full-scale treatment.

The model was implemented following the approach proposed by Wang et al. (2021) to optimizing coagulant dosage in a wastewater treatment process for Cu removal [17].

The input data were collected from October 2020 till November 2021. These data regarded the following variables: turbidity and volumetric flow rate in the influent to the accelerator, turbidity in the effluent of the accelerator, coagulant dosage in the accelerator. The pH value was not included in the data set, since it is not subjected to adjustment being always within the range required by the legislation in force for drinking water, i.e. 6.5–9.5.

The model was implemented throughout the following main steps: (1) selection and collection of data for calibration and validation; (2) assessment of influent turbidity operating classes by means of the cluster analysis, applying the k-means algorithm and the Elbow method to identify the optimal number of clusters, thorough the R packages "Stats" and "Factorextra" [23,24]; (3) application of different regression models (i.e. multivariate-linear, logarithmic and 4th-order polynomial); (4) identification of the best fitting model through the statistical indices (e.g. determination coefficient, R^2, correlation coefficients); (5) building up a chart with coagulant dosage values as a function of the influent turbidity to obtain a prefixed turbidity in the effluent of the Accelerator.

About point (3), three different relationships were tested according to the independent variables selected for the definition of the dose/dosage of the coagulant agent, as represented by the following equations:

$$Dosage = f\ (NTU_{in}, NTU_{out}, Q) \quad (2)$$

$$Dosage = f\ (NTU_{in} \times Q, NTU_{out}) \quad (3)$$

$$Dose = f\ (NTU_{in}, NTU_{out}) \quad (4)$$

where:

Dosage = volumetric flowrate of coagulant [L/h];

Dose = concentration of coagulant [mg/L];

NTU_{in}, NTU_{out} = turbidity in the influent and effluent of the Accelerator, respectively [NTU];

Q = influent volumetric flow rate [L/s]

The multivariate-linear model did not provide any good fitting. Therefore, in the Results section only the data obtained by the logarithmic and 4th-order polynomial models will be presented.

Fig. 3 shows the sequence of the different steps of the experimental and modeling study.

3. Results

3.1. Lab-scale investigation

3.1.1. Coagulant selection

Fig. 4 shows the turbidity content in the effluent of the jar tests versus the dosage of the different coagulants tested in the study. The removal in a decreasing order for each coagulant dosed alone was as follows, along with the applied dosage: 90% at 3.5 mg/L Al_2O_3, 88% at 18.9 mg/L PACS and 77% at 30 mg/L $FeCl_3$. These values highlight that the highest removal was achieved by PAC (expressed as Al_2O_3) which also required the lowest dosage. A high removal was also obtained when PAC was dosed in combination with $FeCl_3$, although the value was lower than that achieved by PAC alone and required a higher dosage. Indeed, it is referred that PAC is a partially hydrolyzed aluminum chloride incorporating a small amount of sulfate. The results obtained in previous experiments with PAC alone as coagulant were equivalent to using aluminum sulfate in conjunction with a polyelectrolyte [8].

Regarding the behavior of the other parameters, a similar removal of TSS, TOC and DOC was observed using PAC and $FeCl_3$, whereas the effect on aluminum was different. Indeed, Al concentration did not change after the addition of $FeCl_3$, likely due to the fact that Al is not present in this coagulant molecule and therefore the two elements, Al and Fe, cannot combine with each other. When PAC was dosed, the behavior of Al concentration changed depending on the dosage: it increased at a low coagulant dosage, whereas it decreased at a higher dosage. This different behavior was likely due to fact that at a higher dosage, also aluminum was dragged in the coagulation process and therefore removed from the liquid phase.

PACS produced a negligible removal of TSS, TOC and DOC; the behavior of Al was similar to that observed with PAC.

The addition of carbon powder (CP) along with Al_2O_3 produced a decrease of turbidity although by a lower value than

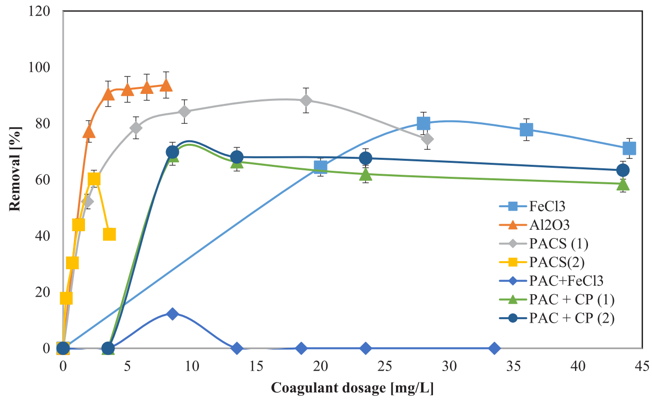

Fig. 4. Turbidity removal efficiency versus coagulant dosage. PAC + CP (1): NTU_{in} = 23.5 NTU; PAC + CP (2): NTU_{in} = 11 NTU.

that achieved by using PAC alone. Furthermore, above 5 mg/L CP dosage, the turbidity started growing and a gray color forming in the water. This behavior was particularly evident when the initial turbidity was higher (test n. 1). A different effect was observed on TOC and DOC: they both decreased after CP addition.

3.1.2. Coagulant aid evaluation

The following series of experiments was performed with the aim to evaluate the effects of coagulant aids on turbidity removal. Based on the previous experiments, PAC at 3.5 mg/L was chosen as coagulant and dosage. Different polyelectrolytes and different concentration were tested.

Fig. 5 shows the turbidity removal efficiency versus polyelectrolytes dosage. The addition of PLT 35 and PLT 37 determined an increase of the removal rate proportional to the dosage.

A different effect was determined by DFC polyelectrolytes. Indeed, a minimum turbidity was observed corresponding to the minimum dosage; afterwards, a continuous rise of the turbidity was registered as the dosage increased. Therefore, DFCs was unable to produce the same high level of removal as PLT 37.

The highest removal observed with the different polyelectrolytes and the corresponding dosages were in the following order: 23% at 0.1 mg/L DFC 535 C, 25% at 0.4 mg/L DFC 105 and 59% at 1.0 mg/L PLT 37. These values show that, among the different polyelectrolytes tested, PLT 37 can be considered the best coagulant aid for the present case. The better performance was likely due to its higher branching level which boosted the floc formation.

Although the improvement of the turbidity removal provided by PLT 37 addition, the high dosage needed to achieve this goal prevented the choice to add it routinely at the full-scale because it would increase too significantly the cost of the process. Therefore, it was decided to recommend the operators of the plant to use PAC only, without any coagulant aid.

3.2. Data-driven modeling

3.2.1. Selection and collection of data

The first step of the model implementation was the selection of the input data to use for calibration and validation. The dataset from the monitoring activity of the Accelerator unit from October 2020 to June 2021 was used for calibration. It included 5822 observations of the following 4 variables:

- influent flowrate to the plant in the range 30 ÷ 138 L/s, with an average value of 114 L/d;
- influent turbidity to the plant in the range 4 ÷ 1800 NTU, with an average value of 43 NTU;

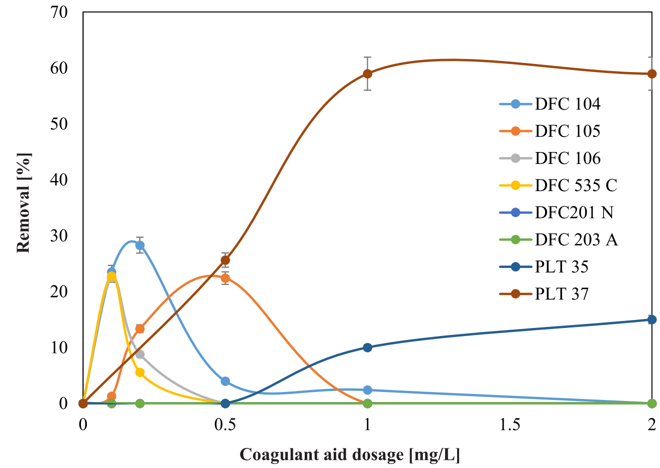

Fig. 5. Turbidity removal efficiency versus coagulant aid dosage. PAC = 3.5 mg Al_2O_3/L.

Table 6
Classification in clusters with corresponding centroids and data number.

Cluster	NTU_{in}	NTU_{out}	N° of data
1	14	2.81	4892
2	176	4.24	366
3	498	4.25	115
4	1039	4.85	48

- effluent turbidity from the accelerator in the range 2 ÷ 18 NTU, with an average value of 3 NTU;
- PAC dosage in the range 1–51 L/h, with an average value of 10.75 L/h.

The validation dataset consisted of the monitoring data from July 2021 to November 2021 and included 143 observations of the following 4 variables:

- influent flowrate to the plant in the range 68 ÷ 130 L/s, with an average value of 109 L/d;
- influent turbidity to the plant in the range 3 ÷ 145 NTU, with an average value of 11 NTU;
- effluent turbidity from the accelerator in the range 1.8 ÷ 4.4 NTU, with an average value of 3 NTU;
- PAC dosage in the range 3.3–26 L/h, with an average value of 6.5 L/h.

3.2.2. Assessment of influent turbidity operating classes

The second step aimed at selecting the process variables which resulted to more influence the dosage. To this aim, the influent turbidity values were divided into classes by means of the cluster analysis. The k-means algorithm and the Elbow method allowed to define the optimal number of clusters. For instance, the explained variance of an increasing number of clusters (i.e. squared sum of the distances from the centroid and the values) was plotted as a function of the number of clusters (see Fig. 6) and the optimal number was selected in correspondence of the curve bends (the Elbow of the curve).

For the calibration dataset, the optimal number of clusters resulted to be equal to 4. The following table shows the clusters and corresponding centroids for each one, as resulting from the application of the k-means algorithm with R software. The graphical representation in the plane [NTU_{in}, NTU_{out}] is depicted in Fig. 7.

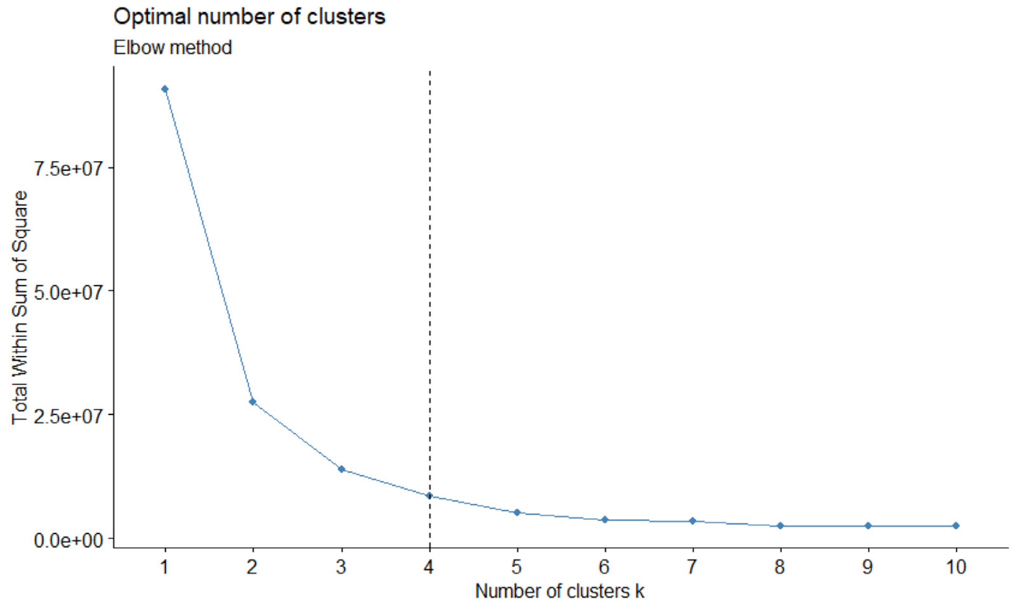

Fig. 6. Elbow method curve: explained variance vs number of clusters.

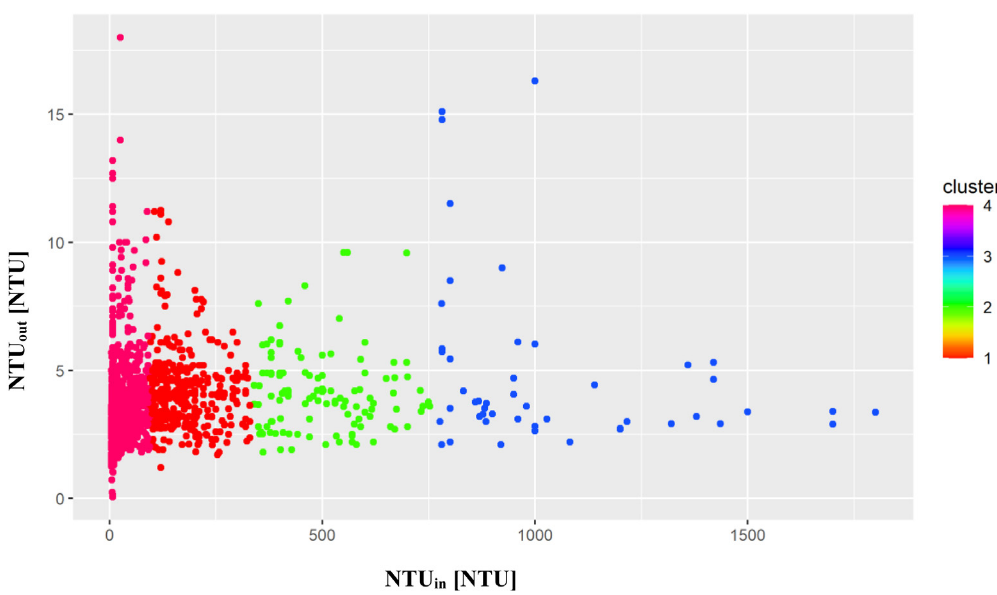

Fig. 7. Experimental points of each centroid.

To better represent the real operation, the 4 clusters were further divided into operating subclasses, by applying again the k-means algorithm with R software. The results obtained are shown in Table 7.

Table 7
Sub-classification of clusters and operating classes of turbidity.

Cluster	Operating classes of turbidity		
	a	0	13
1	b	14	31
	c	32	58
	d	59	94
	e	95	140
2	f	141	190
	g	191	250
	h	251	330
3	i	331	510
	l	510	752
4	m	753	1140
	n	1141	1800

3.2.3. Application of the regression models

In the third step, the selected regression models were applied to each of Eqs. (2), (3) and (4) of Section 2.4. Each regression model was fitted with the whole dataset and also with the different classes as above determined. Correspondingly, a single equation was obtained in the former case, whereas 4 different equations were determined in the latter one.

3.2.4. Identification of the best fitting model

The regression coefficients were used to identify the best fitting model. Table 8a and 8b list the values of the determination coefficient, R^2, for each case, the whole dataset and the different clusters, respectively, in the calibration and validation phases.

Table 8
Determination coefficients of the regression models. (In bold and underlined, the highest values of R^2.)

(a) Whole dataset

		R^2					R^2		
		Eq. (1)	Eq. (2)	Eq. (3)			Eq. (1)	Eq. (2)	Eq. (3)
Logarithmic	Calibration	0.93	0.93	0.91	*4th-order polynomial*	Calibration	0.88	0.89	**<u>0.88</u>**
	Validation	0.84	0.84	0.84		Validation	0.9	0.9	**<u>0.91</u>**

(b) Cluster dataset

Cluster			R^2					R^2		
			Eq. (1)	Eq. (2)	Eq. (3)			Eq. (1)	Eq. (2)	Eq. (3)
1	*Logarithmic*	Calibration	0.85	0.84	0.81	*4th-order polynomial*	Calibration	0.86	0.87	**<u>0.82</u>**
		Validation	0.75	0.74	0.76		Validation	0.81	0.82	**<u>0.84</u>**
2		Calibration	0.37	0.38	0.36		Calibration	0.37	0.4	0.37
3		Calibration	0.53	0.49	0.31		Calibration	0.55	0.54	0.35
4		Calibration	0.32	0.16	0.09		Calibration	0.46	0.36	0.3

Table 9
Correlation between dosage and other variables.

Correlation	Value
Dosage – Flowrate	−0.27
Dosage – NTU_{in}	0.74
Dosage – NTU_{out}	0.47

It can be noted that, only for cluster 1, both calibration and validation were possible, since the range of values of the validation dataset overlapped the range of cluster 1 of the calibration dataset.

Comparing the results, it is highlighted that the determination coefficients were mostly higher for the 4th-polynomial regression model than for the other models. However, the R^2 values did not differ very much for the different equations. Therefore, it was decided to consider also the correlation between the controlled variable, i.e. the coagulant dosage, and the independent variables, i.e. the influent and effluent turbidity and the influent flowrate. The coefficients so obtained, based on the historical data of operation of the Accelerator unit, are listed in Table 9. These coefficients show a low correlation between the flowrate and the dosage; therefore, the flowrate does not seem to influence the required dosage.

The results obtained so far indicate that for the present case-study, Eq. (4) of Section 2.4, which correlates the dosage with the turbidity in the influent and effluent, and the 4th-polynomial regression model provide the best fitting of the real data.

The following equation was finally obtained by the application of the 4th-polynomial regression model to the whole dataset, with the coefficient values determined by means of R software:

$$\begin{aligned} Dose = {} & 1.278 + 5.902 \times 10^{-2} \times NTU_{in} - 1.31 \times 10^{-4} \times NTU_{in}^2 \\ & + 1,18 \times 10^{-7} \times NTU_{in}^3 - 3.476 \times 10^{-11} \times NTU_{in}^4 \\ & + 7.045 \times 10^{-2} \times NTU_{\text{out}} \end{aligned} \quad (5)$$

Figs. 8a and 8b show the results of the calibration and validation phase with the whole dataset, respectively. It can be noted that the model provided a pretty good fitting of the real data up to 1000 NTU influent turbidity. Above this value, the agreement goodness decayed, likely due to the reduced availability of turbidity data. Indeed, such a high turbidity in the influent is pretty rare, since related to short term extreme rain events which occur at a low frequency. This is clearly highlighted by the number of available data for each cluster, as reported in Table 6. It is also noteworthy that when turbidity rises, the content of suspended solids (with a larger size than colloidal particles) increases significantly and this might interfere in the analytical quantification of turbidity by means of the nephelometric method (as that used in the present study).

The model was also applied to each cluster. Figs. 8c and 8b show the results of the calibration and validation phase for cluster 1 only, respectively.

From cluster 1 to 4, it was not possible to carry out the model validation due to the limited number of real data.

Based on the results so far obtained, it can be concluded that the model best represents the real data for turbidity values in the range 0–94 NTU, corresponding to cluster 1. The extension of the model to the whole dataset allows to consider a wider range of possible turbidities in the influent, up to 1000 NTU; however, under the circumstances of such high values, the raw water should not be sent to the plant, because of the possible failure of the treatment units.

3.2.5. Building up a chart for coagulant dosing

The results so obtained were used to build-up a chart showing the recommended PAC dosage as a function of the influent turbidity with the aim to maintain the effluent turbidity at the prefixed value of 3 NTU (Table 10). The coagulant dosage is expressed in terms of volumetric flowrate (L/h). Particularly, starting from the dose calculated by means of Eq. (5) in terms of coagulant concentration (mg/L), the corresponding dosage flowrate (L/h) could be determined based on the density and concentration of the PAC actually used in the full-scale plant (i.e. 1.2 mg/L and 9.5% Al_2O_3, respectively).

The target of 3 NTU in the effluent from the treatment was established based on the experience of the operators of the full-scale plant. Indeed, it was observed that at higher levels of turbidity in the effluent, the following sand filtration unit gets easily clogged because of the high content of solids; by the other hand, effluent turbidity values below 3 NTU lead to unnecessary coagulant overdosing with high costs of the treatment.

The values of coagulant dosage, influent turbidity and effluent turbidity listed in the chart can be considered as the optimal conditions of operation of the Accelerator unit.

As highlighted previously, the high levels of turbidity cannot be considered of interest to practical applications of the data of the chart. Coagulant dosage [L/h] as a function of influent turbidity and flow rate.

4. Conclusions

The combined experimental-modeling approach allowed to optimize the turbidity removal in the coagulation–flocculation process of a full-scale drinking water treatment plant

Firstly, the experimental activity using the raw water of the plant showed PAC to be the preferable coagulant, with an

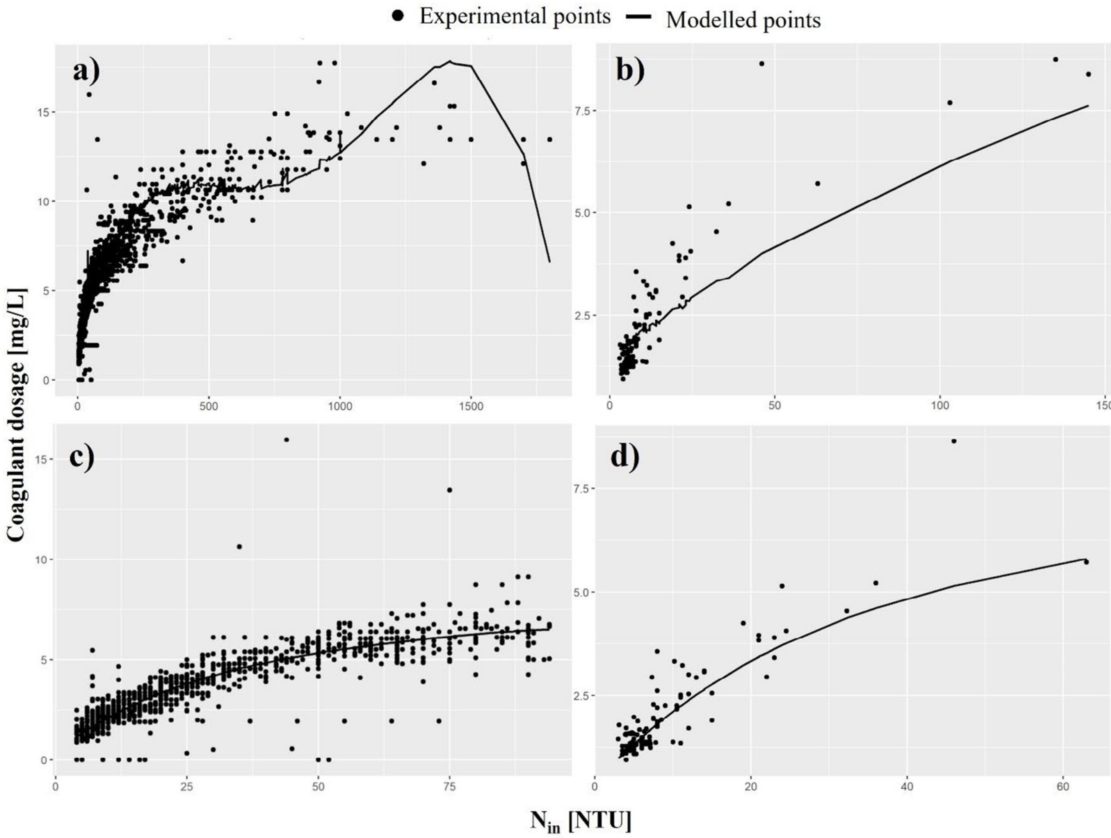

Fig. 8. Application of Eq. (5): (a) calibration with the whole dataset; (b) validation with the whole dataset; (c) calibration with cluster 1 dataset; (d) validation with cluster 1 dataset.

Table 10
Coagulant dosage [L/h] as a function of influent turbidity [NTU] and flow rate (Q, [L/s]).

Turbidity [NTU]			Q [L/s]												
Min	**Max**	**Centroid**	**70**	**75**	**80**	**85**	**90**	**95**	**100**	**105**	**110**	**115**	**120**	**125**	**130**
0	**13**	**7.24**	3.93	4.22	4.50	4.78	5.06	5.34	5.62	5.90	6.18	6.46	6.75	7.03	7.31
14	**31**	**19.93**	6.39	6.84	7.30	7.75	8.21	8.67	9.12	9.58	10.03	10.49	10.95	11.40	11.86
32	**58**	**43.36**	9.02	9.66	10.31	10.95	11.59	12.24	12.88	13.53	14.17	14.81	15.46	16.10	16.75
59	**94**	**73.74**	11.76	12.60	13.45	14.29	15.13	15.97	16.81	17.65	18.49	19.33	20.17	21.01	21.85
95	**140**	**116.48**	14.93	15.99	17.06	18.12	19.19	20.26	21.32	22.39	23.45	24.52	25.59	26.65	27.72
141	**190**	**166.38**	17.80	19.07	20.34	21.61	22.88	24.15	25.42	26.69	27.97	29.24	30.51	31.78	33.05
191	**250**	**217.25**	19.98	21.40	22.83	24.26	25.68	27.11	28.54	29.96	31.39	32.82	34.24	35.67	37.10
251	**330**	**289.59**	22.04	23.61	25.19	26.76	28.34	29.91	31.49	33.06	34.64	36.21	37.78	39.36	40.93
331	**510**	**415.82**	23.66	25.35	27.04	28.73	30.42	32.11	33.80	35.49	37.18	38.87	40.56	42.25	43.94
511	**752**	**613.27**	24.10	25.82	27.54	29.26	30.98	32.71	34.43	36.15	37.87	39.59	41.31	43.03	44.76
753	**1140**	**891.83**	26.34	28.22	30.10	31.98	33.86	35.74	37.62	39.50	41.38	43.27	45.15	47.03	48.91
1141	**1800**	**1434.77**	37.41	40.08	42.75	45.42	48.09	50.77	53.44	56.11	58.78	61.45	64.13	66.80	69.47

optimum dosage equal to 3.5 mg Al_2O_3/L under the laboratory conditions. The addition of polyelectrolytes resulted not to be worthy.

A data-driven model was then implemented to predict the optimal PAC dosage under the conditions of variable influent turbidity typical of the real plant. The input to the model were the historical data of operation of the plant. The model was optimized by means of the regression and statistical analyses. After implementation, it was used to build-up a chart indicating the dosage of PAC to apply as a function of influent turbidity for a prefixed turbidity.

The approach herewith proposed allows to constantly comply with the required drinking water quality, despite the variations of the influent characteristics. It can be recommended for the optimization of the coagulation–flocculation unit of other plants and also of different types of chemical–physical processes.

CRediT authorship contribution statement

Agostina Chiavola: Writing – original draft, Writing – review & editing, Supervision. **Camilla Di Marcantonio:** Formal analysis, Data curation. **Martina D'Agostini:** Formal analysis, Investigation, Data curation. **Simone Leoni:** Conceptualization, Methodology, Resources. **Marco Lazzazzara:** Conceptualization, Methodology, Resources.

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.

Data availability

Data will be made available on request.

References

[1] The European parliament and the council of the European union, in: DIRECTIVE (EU) 2020/2184 of the European Parliament and of the Council of 16 December 2020 on the Quality of Water Intended for Human Consumption, 2020, https://eur-lex.europa.eu/eli/dir/2020/2184/oj.

[2] R.L. Droste, Theory and Practice of Water and Wastewater Treatment, John Wiley & Sons, Inc., NJ (USA), 2018.

[3] D.S. Sarai, Water Treatment Made Simple for Operators, John Wiley & Sons, Inc., NJ (USA), 2006.

[4] American Water Works Association, J. Edzwald, Water Quality & Treatment: A Handbook on Drinking Water, sixth ed., McGraw-Hill, Inc., NY (USA), 2011.

[5] Metcalf & Eddy, Wastewater Engineering Treatment and Resource Recovery, fifth ed., 2015, http://dx.doi.org/10.1007/978-1-349-06927-9_4.

[6] L.D. Benefield, J.F. Judkins, B.L. Weand, Process Chemistry for Water and Wastewater, Prentice-Hall, Inc., Englewood Cliffs, NJ (USA), 1982.

[7] World Health Organization, Guidelines for Drinking-Water Quality: Fourth Edition Incorporating the First and Second Addenda, Licence: C, Geneva, 2022.

[8] Bratby, Coagulation and Flocculation with an Emphasis on Water and Wastewater Treatment, Uplands press LTD, 1980.

[9] A. Benalia, K. Derbal, A. Panico, F. Pirozzi, Use of acorn leaves as a natural coagulant in a drinking water treatment plant, Water 11 (2019) 57, http://dx.doi.org/10.3390/w11010057.

[10] A. Benalia, K. Derbal, A. Khalfaoui, R. Bouchareb, A. Panico, C. Gisonni, G. Crispino, F. Pirozzi, A. Pizzi, Use of aloe vera as an organic coagulant for improving drinking water quality, Water 13 (2021) 2024, http://dx.doi.org/10.3390/w13152024.

[11] S. Tolouei, J.-B. Burnet, L. Autixier, M. Taghipour, J. Bonsteel, S.V. Duy, S. Sauvé, M. Prévost, S. Dorner, Temporal variability of parasites, bacterial indicators, and wastewater micropollutants in a water resource recovery facility under various weather conditions, Water Res. 148 (2019) 446–458, http://dx.doi.org/10.1016/j.watres.2018.10.068.

[12] D.P.W. Sheng, B. Muhammad Roil, S. Norazanita, Assessment and optimization of coagulation process in water treatment plant: A review, ASEAN J. Sci. Eng. 3 (2023) 79–100, http://dx.doi.org/10.17509/ajse.v3i1.45035.

[13] C.R. O'Melia, Coagulation in wastewater treatment, in: Sci. Basis Flocculation, Springer Netherlands, Dordrecht, 1978, pp. 219–268, http://dx.doi.org/10.1007/978-94-009-9938-1_10.

[14] S. Haghiri, A. Daghighi, S. Moharramzadeh, Optimum coagulant forecasting by modeling jar test experiments using ANNs, Drink. Water Eng. Sci. 11 (2018) 1–8, http://dx.doi.org/10.5194/dwes-11-1-2018.

[15] Z. Pandilov, M. Stojkov, Application of intelligent optimization tools in determination and control of dosing of flocculent in water treatment, Ann. Fac. Eng. Hunedoara 17 (2019) 109–116.

[16] F.C.R. dos Santos, A.F.H. Librantz, C.G. Dias, S.G. Rodrigues, Intelligent system for improving dosage control, Acta Sci. Acta Sci. Technol. 39 (2017) 33–38.

[17] K.-J. Wang, P.-S. Wang, H.-P. Nguyen, A data-driven optimization model for coagulant dosage decision in industrial wastewater treatment, Comput. Chem. Eng. 152 (2021) 107383, http://dx.doi.org/10.1016/j.compchemeng.2021.107383.

[18] M.H. Cholicheh, B. Neamati1, H. Farrokhzadeh, M. Hajian, N. Khanjani, Evaluating the performance of accelator clarifier in eliminating turbidity by using chemical coagulants, Environ. Sci. 13 (2013) 565–569, http://dx.doi.org/10.5829/idosi.aejaes.2013.13.04.741.

[19] DECRETO LEGISLATIVO 23 febbraio 2023, n. 18.

[20] DIRECTIVE (EU) 2020/2184 of the European Parliament and of the Council of 16 December 2020 on the quality of water intended for human consumption.

[21] APAT IRSA-CNR, Analytical methods for water 29/2003, 2003.

[22] APHA, Standard methods for the examination of water and wastewater, 2017.

[23] A. Kassambara, F. Mundt, Factoextra: Extract and visualize the results of multivariate data analyses, R Package Version 1.0.7, 2020. https://cran.r-project.org/package=factoextra.

[24] R Core Team, R: a language and environment for statistical computing, 2021, http://www.r-project.org/.