Using artificial neural networks to determine sediment concentration at limit of deposition in storm sewers

Determinação do limite de deposição de sedimentos em galerias de drenagem utilizando redes neurais artificiais

Adhemar Romero[1] (ID) & José Junji Ota[1] (ID)

[1]Universidade Federal do Paraná, Curitiba, PR, Brasil
E-mails: adhe.romero@gmail.com (AR), ota.dhs@ufpr.br (JJO)

ABSTRACT

The concept of sediment transport at the limit of deposition in storm sewers represents one operational condition that avoid deposition of sediments maintaining the discharge capacity of the pipes. In this study, this condition was analyzed applying one Artificial Neural Network Multilayer Perceptron (ANN-MLP) model to predict the volumetric concentration at the limit of deposition, using 544 experimental data from literature. It was evaluated different input variables combinations and model configurations, showing the sensitivity of the model with these changes. Through this study, it was demonstrated that the proposed model outperforms the existing equations, leading to more assertive predictions in the determination of volumetric concentrations at the limit of deposition, resulting in values of $R^2 = 0.92$, Mean Absolute Percentage Error (MAPE) = 35.09 % and Mean Average Error (MAE) = 59.84 ppm. With the performed analysis, the study selects one equation to be used for extrapolations when determining the volumetric concentration at the limit of deposition in storm sewers. The selected equation is superior due to its theoretical basis. This work includes one more concept to a better methodology in obtaining the conditions of the flow at the limit of deposition.

Keywords: Limit of deposition; Sediment transport; Artificial neural networks.

RESUMO

O conceito de transporte de sedimentos no limite de deposição em galerias de drenagem representa uma condição operacional que evita deposição de sedimentos, mantendo a capacidade de descarga dos tubos. Neste trabalho, essa condição foi analisada aplicando uma Rede Neural Artificial Perceptron Multi Camadas (RNA-PMC) para a obtenção da concentração volumétrica no limite de deposição, utilizando 544 dados experimentais disponíveis na literatura. Foram avaliadas diferentes combinações de parâmetros de entrada e configurações do modelo, demonstrando sua sensibilidade para tais variações. Através do presente estudo é demonstrado que o modelo proposto supera o desempenho das equações existentes, levando a estimativas mais assertivas na determinação da concentração volumétrica no limite de deposição resultando valores de $R^2 = 0.92$, Erro Percentual Absoluto Médio (MAPE) = 35.09% e Erro Absoluto Médio (MAE) = 59.84ppm. Com base nas análises realizadas, o estudo indica uma equação para uso em casos de extrapolações da concentração volumétrica no limite de deposição em galerias de drenagem. O uso da equação apontada está associado à base teórica utilizada em seu processo de obtenção. Este trabalho inclui, portanto, mais um conceito para uma melhor metodologia de se obter as condições de limite de deposição.

Palavras-chave: Limite de deposição; Transporte de sedimentos; Redes neurais artificiais.

INTRODUCTION

Sediment deposition in storm sewers is one of the most recurrent problems affecting the hydraulic behavior of drainage systems. Commonly, these structures are designed considering criteria based only on the discharge capacity, minimum velocity, and minimum shear stress for the design discharge. Owing to their simplicity, these criteria are easily utilized, but may not be sufficient to ensure that no sediment will deposit at the bottom of the pipe during its operation, which may cause a discharge capacity reduction and efficiency loss of the storm sewer.

To minimize the removal frequency of deposited sediments inside storm sewers, a design concept was developed whereby the flow inside would promote sediment transport without deposition and, consequently, there would be no obstructions and the discharge capacity would not be reduced; this is defined as a self-cleaning design concept.

Commonly the self-cleaning design concept is studied with two assumptions, i. e., with deposition (which assumes a small amount of sediments deposition in the bottom of the pipe), and without deposition (where no deposition is assumed – clean bed). In this research, the limit of deposition was analyzed as clean bed.

Studies, such as May (1982), Mayerle et al. (1991), Ab Ghani (1993), May et al. (1996), Ota (1999), Vongvisessomjai et al. (2010), Ebtehaj et al. (2014), Najafzadeh et al. (2017), Romero (2018), Montes et al. (2020a) and Tafarojnoruz & Sharafati (2020) suggest equations based on laboratory experiments to determine the self-cleaning conditions. Some of these studies utilize Froude number as a function of the volumetric concentration at the limit of deposition to determine this condition. Despite the good correlations between the experimental and calculated Froude number, it is not possible to affirm that using these models the same performance will be achieved when calculating the volumetric concentration at the limit of deposition.

Another issue regarding these sediment transport equations comprehends deciding which equation to use or which equation should provide better predictions of the volumetric concentration at the limit of deposition. Moreover, most of those equations were developed using a limited range of experiments. One analysis with a larger dataset can demonstrate the efficiency of the proposed equations in a more general manner.

In order to present a model with a better efficiency to deal with deviations observed between the calculated and the experimental volumetric concentration at the limit of deposition, the proposed research present an application of Artificial Neural Network (ANN) models to predict it. The deviations observed are related to the methodology used to obtain most of the best-fit equations, which are based on a specific range of tests and the type of equation (potential function) is fixed. Several studies analyzed the ANN model in the determination of Froude Number, which showed a good performance, suggesting that the ANN model could outperform the existing equations in determining the volumetric concentrations the limit of deposition. Additionally, the analysis presented in this paper uses a wider range of experiments, aggregated from 544 experimental data sets obtained from research by Mayerle (1988), May (1993), Ab Ghani (1993), Ota (1999), and Montes et al. (2020a). Montes et al. (2020a), differently of the other researches, utilized a larger smooth pipe with 595mm diameter

to develop the experiments. Due this, their data was included in the presented analysis.

Using the abovementioned dataset, this study compares the performance of the equations proposed by Mayerle et al. (1991), Ab Ghani (1993), May et al. (1996), Ota (1999), Vongvisessomjai et al. (2010), Ebtehaj et al. (2014), Najafzadeh et al. (2017), Romero (2018), Montes et al. (2020a) and Tafarojnoruz & Sharafati (2020) in determining the volumetric concentration at the limit of deposition. In addition, there are uncertainties regarding the use of these equations for the extrapolations. Ota (1999) included a more accurate physical consideration on his model, and may be the best solution. This research tries to bring the ANN and a performance analysis with a wider dataset to confirm this fact.

With a more precise estimation of the volumetric concentration at the limit of deposition, a better design approach for storm sewers can be suggested. This approach will lead to the maximum allowable volumetric concentration of sediments that does not reduce the discharge capacity of the structure.

LITERATURE REVIEW

This section presents the most recurrent criteria used to design storm sewers, then it is presented the equations used to determine the limit of deposition. These equations typically use different dimensionless parameters to define such condition. Finally, some definitions related to ANN, and previous studies that have applied artificial intelligence for sediment transport inside storm sewers are presented.

Existing criteria for designing storm sewers

The basic premise in designing storm sewers lies in operating the structure under uniform and free flow. Manning equation is commonly used to determine the diameter and slope of the pipe.

Additional sizing criteria of the storm sewers encompass the determination of the maximum and minimum velocities. Maximum velocities, above 3.0 m/s are avoided to protect the pipe against abrasion. However, depending on the material used, these velocities can be higher (Mays, 1999, 2004). The minimum flow velocities must be in the range between 0.60 m/s and 0.80 m/s, to avoid sediment deposition at the bottom of the pipes (Fendrich et al., 1997; Mays, 1999; Tucci, 2004).

Butler & Davies (2004) emphasized that a simple minimum velocity value, unrelated to volumetric concentration, characteristics of the sediments being transported, and other hydraulic characteristics of the flow, do not represent the capacity of the pipe to transport a given sediment, because higher velocities are necessary to transport it in larger pipes rather than in smaller ones.

Vongvisessomjai et al. (2010) indicated that the structure could be under two operational situations throughout its work-life: low and high flow. Operation during low-flow periods must be such that no deposition occurs inside the pipe, and the structure must be able to transport the design discharge without overflowing during the rainy periods.

According to the American Society of Civil Engineers (2007), during low-flow periods, the flows in the pipes are

commonly lower than the design flows, leading to a lower sediment transport capacity.

Determination of the limit of deposition in storm sewers

Definition of the limit of deposition

The limit of sediment deposition in pipes is studied to determine the flow conditions that do not cause the deposition of solids at the bottom of the pipes. For design purposes, the determination of this limit characterizes the self-cleaning condition of the structure, and is often studied through two types of experiments: allowing a small layer of sediments to deposit at the bottom of the pipe (with deposition) and in clean beds (without deposition).

For both cases, the limit of deposition in storm sewers is usually determined by equations developed from experimental analyses, performed using non-cohesive sands and gravel (May et al., 1996).

The sediment transport in the sewers is well described by Ota (1999) as when the shear stress produced by the flow at the bottom of the pipe is sufficiently high, a sediment with a certain size may be transported in the pipes without deposition. Nevertheless, for the same flow condition, if the sediment concentration increases, there is a limit of volumetric concentration in the flow at which the solids begin to deposit. This is the limit of deposition, as presented by Pedroli (1963) and Novak & Nalluri (1975), and the condition analyzed in this paper (without deposition – clean beds). As the volumetric concentration increases, the sediments tend to deposit in small amounts but continue to move forming separate dunes, as cited by May (1982). By further increasing the volumetric concentration, continuous deposition can occur, diminishing considerably the discharge capacity of the pipe due to reductions in the cross-sectional area (Ota, 1999).

Several empirical equations were developed to determine the limit of deposition in clean bed state, i. e., the self-cleaning in rigid boundary conditions. These equations are used to predict the minimum velocity and volumetric concentration to satisfy the nondeposition conditions (Safari et al. 2018). The most recurrent equations found in the literature, the theoretical discussions and the experimental dataset used to obtain them are presented in the following section.

Developed equations to determine the limit of deposition in clean beds

The following studies present models used to determine the limit of deposition in storm sewers, which were developed through experiments with clean beds.

Mayerle et al. (1991) investigated the sediment transport inside pipes without depositions using graded sand with diameters ranging from 0.5 mm to 8.74 mm, with a density of 2,550 kg/m^3 in one 152 mm diameter and 20.50 m long pipe. The study affirms that the velocity needed to guarantee flow with no depositions depends on variables such as density, hydraulic radius, sediment concentration and grain size. Based on the results of their experiments, Mayerle et al. (1991) presented Equation 1:

$$\frac{V_c}{\sqrt{g\, d\, (S_s - 1)}} = 4.32\, C_v^{\,0.23} \left(\frac{d}{R_h}\right)^{-0.68} \tag{1}$$

where V_c is the critical velocity for non-deposition in m/s, g is the acceleration of gravity in m/s^2, d is the diameter of the particle in m, S_s is the relative density of the solid, C_v is the volumetric concentration in ppm and the R_h is the hydraulic radius in m. Mayerle et al. (1991) showed that the critical velocity needed to meet the self-cleaning criteria decreased with the particle size and increased with the hydraulic radius and volumetric concentration.

Ab Ghani (1993) developed experiments with three different pipes. Two with diameters of 154mm and 305mm with a length of 20.50 m and one with a diameter of 450 mm and length of 21.00 m. The roughness of the 305 mm diameter pipe was settled using sand with diameters ranging from 0.5 mm to 1 mm. In the 154 mm and 305 mm diameter pipes, the diameter of the sediments ranged from 0.5 mm to 10 mm of non-cohesive uniformly distributed material with densities ranging from 2,530 kg/m^3 to 2,593 kg/m^3. For the analysis in the 450 mm diameter pipe, the diameter of the sediment was equivalent to 0.72 mm with a density of 2,620 kg/m^3. In addition, using a larger set of experimental observations with data from Mayerle et al. (1991), May (1993), and Loveless (1991), Ab Ghani (1993) determined an empirical equation with the volumetric concentration as the dependent variable, to represent the limit of deposition in pipes, defined by Equation 2.

$$C_v = 0.03 \left(\frac{V_c}{\sqrt{g\, d_{50}\, (S_s - 1)}}\right)^{3.54} D_{gr}^{\,-0.11} \left(\frac{d_{50}}{R_h}\right)^{2.47} \lambda_s^{\,0.57} \tag{2}$$

$$\lambda_s = 1.13\, \lambda_0^{\,0.98}\, C_v^{\,0.02}\, D_{gr}^{\,0.01} \tag{3}$$

$$D_{gr} = d_{50} \left[\frac{(S_s - 1)g}{v^2}\right]^{1/3} \tag{4}$$

where d_{50} represents the median diameter of particles in a mixture in m, λ_0 is the clean water friction factor, and D_{gr} is the dimensionless sediment diameter, defined by Equation 4. Ab Ghani (1993) showed that the discrepancies between the observed and predicted concentrations using Equation 2 increase with decreasing sediment concentration.

May et al. (1996) used a set of 332 sediment transport experiments to develop another equation for defining the volumetric concentration of the sediments at the limit of deposition. In those experiments, the diameter of the pipe used ranged from 77 mm to 450 mm, the median diameter was between 160 µm and 8,300 µm and the volumetric concentration ranged from 2.3 to 2,110 ppm. Equations 5 and 6 were obtained through best-fit model.

$$C_v = 0.0303 \left(\frac{D^2}{A} \right) \left(\frac{d_{50}}{D} \right)^{0.6} \left(1 - \frac{V_t}{V} \right)^4 \left(\frac{V^2}{g D (S_s - 1)} \right)^{1.5} \tag{5}$$

$$V_t = 0.125 \left[g \, d_{50} (S_s - 1) \right]^{0.5} \left(\frac{y}{d_{50}} \right)^{0.47} \tag{6}$$

where V_t is the threshold velocity, y is the water depth in the pipe in m, V is the mean velocity of the flow in m/s, D is the pipe diameter in m, and A is the cross-sectional area of the flow in m².

Ota (1999), in contrast to the other researchers, looked for an approach with a better theoretical basis, not using only best-fit models with the experimental data. Among the observations presented by Ota (1999), there is a variation in shear stress as a function of depth according to Equation 7.

$$\xi = \frac{\tau}{\tau_0} = \left(\frac{u^*}{V^*} \right)^2 = f \left(\frac{y}{D} \right) \tag{7}$$

where ξ represents the relationship between the shear stress at the bottom of the pipe (τ) and the mean value ( ), u^* is the shear velocity at the bottom of the pipe in m/s and, V^* is the mean shear velocity. Figure 1 shows the variation of ξ with the relative depth inside the drainage pipe, as presented by Ota (1999).

Another contribution of Ota (1999) to determine the limit of deposition in pipes involves the relation between particle diameter and pipe roughness (d_{50}/k), representing the internal angle of friction. This relation results in the definition of the compensated dimensionless shear stress, according to Equation 8, to determine the limit of deposition.

$$\psi \left(\frac{d_{50}}{k} \right)^{\frac{2}{3}} = \frac{\tau_0}{\rho g d_{50} (S_s - 1)} \left(\frac{d_{50}}{k} \right)^{\frac{2}{3}} \tag{8}$$
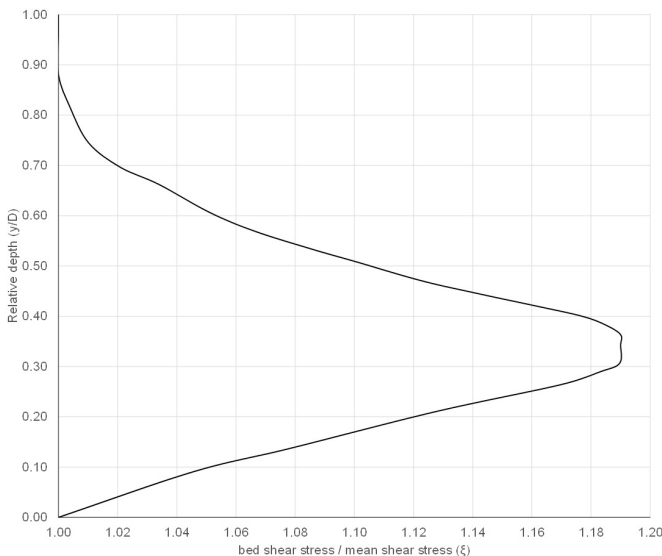

Figure 1. Variation of the ξ with the relative depth inside the pipe.
Source: Adapted from Ota (1999).

where ψ is the dimensionless shear stress, k is the pipe roughness in m, and ρ is the density of fluid in kg/m³.

Ota (1999) determined the limit of sediment deposition experimentally but included the results of other studies for calibration, utilizing data presented by Ab Ghani (1993), Mayerle et al. (1991) and May (1993), resulting in 508 experimental data points.

With these considerations, Ota (1999) correlated the transport parameter and the compensated dimensionless shear stress, obtaining Equation 9:

$$\phi = 2.9 \left[\psi \left(d_{50}/k \right)^{\frac{2}{3}} \right]^{2.7} \tag{9}$$

where ϕ is the transport parameter, described in Equation 10.

$$\phi = \frac{C_v R_h V}{d_{50}^{3/2} \sqrt{g (S_S - 1)}} \tag{10}$$

To facilitate the use of Equation 9, Ota (1999) rewrote his equation in a simplified form, presented in Equation 11, valid for the limit of deposition:

$$C_v = 2.9 \xi^{2.7} \left(\frac{R_h}{d_{50}} \right)^{1.2} \left(\frac{S}{(S_s - 1)} \right)^{2.2} \sqrt{\frac{\lambda_0}{8}} \left(\frac{d_{50}}{k} \right)^{1.8} \tag{11}$$

The adoption of $\psi \left(d_{50}/k \right)^{\frac{2}{3}}$ parameter in the model $\phi - \psi \left(d_{50}/k \right)^{\frac{2}{3}}$ appears to be more reliable for extrapolating the limit of deposition for pipe diameters larger than 500 mm (Ota & Nalluri, 2003).

Rushforth et al. (2003) presented experiments in a full-scale model to evaluate the performance of two models developed to determine the self-cleaning condition. The experiments were carried out in a pipe with 60 m long and 800 mm diameter. Although this comparison was performed with a deposited bed model, they noticed that the best-fit equations such as May (1993) underestimate the flow capacity to transport sediments, confirming Ota & Nalluri's (2003) conclusion.

Vongvisessomjai et al. (2010) performed 28 new experiments using polyvinyl chloride (PVC) pipes. The laboratory analysis conducted was developed considering the sediment transport in pipes with partial-full-flow without deposition, using two pipes with dimensions of 100 mm and 150 mm, both 22 m long. Three uniform grain sizes were used in their study: 0.20 mm, 0.30 mm, and 0.43 mm. Vongvisessomjai et al. (2010) determined a dimensionless parameter defined as the densimetric Froude number (Fr^*) to determine the sediment transport at the limit of deposition using the experiments conducted and obtained Equation 12.

$$Fr^* = \frac{V_c}{\sqrt{g \, d_{50} (S_s - 1)}} = 4.31 \, C_v^{0.226} \left(\frac{d_{50}}{R_h} \right)^{-0.616} \tag{12}$$

Vongvisessomjai et al. (2010) mentioned that the calculated values of those equations are very similar to those calculated by Equations 5 and 6, and when compared with the experimental results, Equations 12 presented slightly better results than those equations. Vongvisessomjai et al. (2010) also concluded that the relationship between the grain size and the hydraulic radius is the most important parameter for sediment deposition in storm sewers.

Ebtehaj et al. (2014), using the experimental datasets obtained by Ab Ghani (1993) and Vongvisessomjai et al. (2010), developed another sediment transport equation to determine the limit of deposition of sediments in storm sewers. To validate the proposed equation, Ebtehaj et al. (2014) used a dataset from Ota & Nalluri (1999). Ota & Nalluri (1999) data were aggregated from 24 experiments with six different sediment diameters (ranging from 0.71 mm to 5.61 mm) and by 20 more experiments with a 2 mm sediment diameter. Through multiple regression using the root mean square error and the mean absolute relative error, as error functions, Ebtehaj et al. (2014) developed an equation to define the densimetric Froude number. Akaike Information Criterion (AIC) was used to measure the accuracy of the model. The resulting equation presented by Ebtehaj et al. (2014) is presented by Equation 13.

$$\frac{V_c}{\sqrt{g\,d_{50}\left(S_s-1\right)}} = 4.49\,C_v^{\,0.21}\left(\frac{d_{50}}{R_h}\right)^{-0.54} \tag{13}$$

Ebtehaj et al. (2014) concluded that the new equation had a simpler form, and showed better accuracy than Vongvisessomjai et al. (2010) and May et al. (1996) equations.

Najafzadeh et al. (2017) used the Ab Ghani's (1993) dataset to evaluate and develop another equation through EPR-MOGA-XL (Evolutionary Polynomial Regression – Multi-objective optimization strategy) model. The EPR-MOGA-XL model was compared with an MT (Model Tree) model and with the equations of Novak & Nalluri (1975), Mayerle et al. (1991), Ab Ghani (1993), Vongvisessomjai et al. (2010) and Azamathulla et al. (2012). Najafzadeh et al. (2017) concluded that the developed models surpassed all empirical equations analyzed, which over (or under) predicted the densimentric Froude number. The Equation 14 was obtained through EPR-MOGA-XL model.

$$\frac{V_c}{\sqrt{g\,d_{50}\left(S_s-1\right)}} = 0.404\left(\frac{R_h}{d_{50}}\right)^{0.5} + 23.25\left(\frac{R_h}{d_{50}}\right)^{0.5}C_v^{\,0.5} \tag{14}$$

Romero (2018) utilized the results of two ANN models to evaluate the correlation between transport parameter and the compensated dimensionless shear stress, as proposed by Ota (1999). Due to the limitations of the ANN method to extrapolation, it was proposed another equation to determine the limit of deposition in terms of volumetric concentration based on the ANN results, presented in Equation 15.

$$C_v = 2.88\xi^{2,6}\left(\frac{R_h}{d_{50}}\right)^{1.1}\left(\frac{S}{\left(S_s-1\right)}\right)^{2.1}\sqrt{\frac{\lambda_0}{8}}\left(\frac{d_{50}}{k}\right)^{1.732} \tag{15}$$

Through analyzing the studies presented in this section, is possible to notice that, with exception of Ota (1999), all the equations are defined as a best-fit equation model. Equation 11 is a semi-theoretical model and it is explained by Ota & Nalluri (2003) in detail.

Montes et al. (2020a) presented a study which evaluated the performance of several empirical equations using different datasets. Among the utilized dataset Montes et al. (2020a) develop a new dataset using a larger pipe with 10.5 m long and 595 mm

diameter. The experimental data gathered and presented in their work covers the self-cleaning condition with and without deposition (clean bed). Through several sensibility analysis, Montes et al. (2020a) concluded that the existing models are overfitted and tends to overestimate the threshold of self-cleaning velocities. Montes et al. proposed another best-fit equation to determine the limit of deposition for larger pipes, i.e., 500mm and above, presented in Equation 16, which outperformed others equations comparing with the new dataset.

$$\frac{V_c}{\sqrt{g\,d_{50}\left(S_s-1\right)}} = 4.79\,\lambda^{0.058}\,C_v^{\,0.209}\left(\frac{d_{50}}{R_h}\right)^{-0.593} \tag{16}$$

Tafarojnoruz & Sharafati (2020) presented another equation capable to determine the Froude number at the limit of deposition, obtained through the generalized likelihood uncertainty estimation (GLUE). Using data from Ab Ghani (1993) and Vongvisessomjai et al. (2010), Tafarojnoruz & Sharafati (2020) presented the Equation 17, which was compared with the equations of Ab Ghani (1993), Vongvisessomjai et al. (2010) and Ebtehaj et al. (2014), showed more accurate predictions.

$$\frac{V_c}{\sqrt{g\,d_{50}\left(S_s-1\right)}} = 3.39C_v^{\,0.2}\left(\frac{R_h}{d_{50}}\right)^{0.62} \tag{17}$$

Table 1 summarizes all presented equations to determine the limit of deposition presented in this section.

Artificial neural networks

An artificial neural network (ANN) can be described, as a computational system capable of simulating the decision processes of the nervous system of a human being (Graupe, 2013). The correlation with the human nervous system is made not only by the definitions but also by the processing units of the model and connections between those units, called neurons and synaptic connections. The artificial neurons are the processing units of an ANN. These units receive information, process it, and then pass the resulting information to other neurons via weights (or artificial synaptic connections). Through the interaction between the artificial neurons over time, the computational system can extract and repeat patterns (Haykin, 1999; Braga et al., 2000; Silva et al., 2017).

According to Braga et al. (2000), the application of the ANN in solving mathematical problems is very attractive because of the parallelism of the model and its ability to achieve superior performance than other mathematical models. Bishop (1995) emphasized that an artificial neural network with two hidden layers can approximate any mathematical function. Another positive point of the model is its capability to learn and reproduce patterns observed in the data and generalize a problem. Generalization is the ANN's capacity to produce one acceptable output for a given input dataset that was not shown during the training phase (or the fitting phase) (Haykin, 1999; Silva et al., 2017).

There are several types of ANN. The most used and known types of ANN can be classified as feedforward and recurrent ANN. The difference between them lies in the connections of the artificial

Table 1. Self-cleaning models used to determine the limit of deposition inside storm sewers.

Author	Model	Equation
Mayerle et al. (1991)	$\dfrac{V_c}{\sqrt{g\,d_{50}\,(S_s-1)}} = 4.32\,C_v^{0.23}\left(\dfrac{d}{R_h}\right)^{-0.68}$	(1)
Ab Ghani (1993)	$C_v = 0.03\left(\dfrac{V_c}{\sqrt{g\,d_{50}\,(S_s-1)}}\right)^{3.54} D_{gr}^{-0.11}\left(\dfrac{d_{50}}{R_h}\right)^{2.47}\lambda_s^{0.57}$	(2)
May et al. (1996)	$C_v = 0.0303\left(\dfrac{D^2}{A}\right)\left(\dfrac{d_{50}}{D}\right)^{0.6}\left(1-\dfrac{V_t}{V}\right)^4\left(\dfrac{V^2}{g\,D\,(S_s-1)}\right)^{1.5}$	(5)
Ota (1999)	$C_v = 2.9\,\xi^{2.7}\left(\dfrac{R_h}{d_{50}}\right)^{1.2}\left(\dfrac{S}{(S_s-1)}\right)^{2.2}\sqrt{\dfrac{\lambda_0}{8}}\left(\dfrac{d_{50}}{k}\right)^{1.8}$	(11)
Vongvisessomjai et al. (2010)	$\dfrac{V_c}{\sqrt{g\,d_{50}\,(S_s-1)}} = 4.31\,C_v^{0.226}\left(\dfrac{d_{50}}{R_h}\right)^{-0.616}$	(12)
Ebtehaj et al. (2014)	$\dfrac{V_c}{\sqrt{g\,d_{50}\,(S_s-1)}} = 4.49\,C_v^{0.21}\left(\dfrac{d_{50}}{R_h}\right)^{-0.54}$	(13)
Najafzadeh et al. (2017)	$\dfrac{V_c}{\sqrt{g\,d_{50}\,(S_s-1)}} = 0.404\left(\dfrac{R_h}{d_{50}}\right)^{0.5} + 23.25\left(\dfrac{R_h}{d_{50}}\right)^{0.5}C_v^{0.5}$	(14)
Romero (2018)	$C_v = 2,88\,\xi^{2.6}\left(\dfrac{R_h}{d_{50}}\right)^{1.1}\left(\dfrac{S}{(S_s-1)}\right)^{2.1}\sqrt{\dfrac{\lambda_0}{8}}\left(\dfrac{d_{50}}{k}\right)^{1.732}$	(15)
Montes et al. (2020a)	$\dfrac{V_c}{\sqrt{g\,d_{50}\,(S_s-1)}} = 4,79\,\lambda_0^{0.058}\,C_v^{0.209}\left(\dfrac{d_{50}}{R_h}\right)^{-0.593}$	(16)
Tafarojnoruz & Sharafati (2020)	$\dfrac{V_c}{\sqrt{g\,d_{50}\,(S_s-1)}} = 3.39 C_v^{0.2}\left(\dfrac{R_h}{d_{50}}\right)^{0.62}$	(17)

neurons. While the recurrent type exhibits loop on the synaptic connections, the feedforward does not (Jain et al., 1996; Haykin, 2001; Silva et al., 2017). To develop the present research, the multilayer perceptron (MLP), one kind of feedforward ANN, was used.

The ANN-MLP can be characterized by the number of hidden (or processing) layers of neurons. The neuron is the smallest processing unit in the ANN-MLP architecture. It receives the information, processes it by calculating the sum of the products between the weights and the received data, aggregate a bias value and, applies this sum to one activation function, which generates the output. Figure 2 shows the representation of one artificial neuron.

In Figure 2, the input signals ($x_1, x_2, ..., x_n$) are multiplied by a synaptic weight ($w_1, w_2, ..., w_k$), after which they are summed, including one bias value (θ). The sum then feeds one activation function (φ) that produces one output signal (y). Haykin (2001) presented Equations 18 and 19 to describe the behavior of a single neuron mathematically.

$$u = \sum_{j=1}^{n} w_n x_n - \theta \tag{18}$$

$$\varphi(u) = y \tag{19}$$

Figure 3 shows the architecture of one ANN-MLP. Starting at the input layer, the data feeds the first hidden layer of neurons, which are processed following Equations 18 and 19. The output of the neurons in the first hidden layer feeds the neurons in the second hidden layer, continuing if there are hidden layers in the architecture until the signal reaches the output layer. The output produced in the output layer is the result of the ANN (Silva et al., 2017).

The process to obtain one trained ANN model capable of reproducing patterns and generalizing one desired problem consists of repeating two phases, called the forward and the backward phases, over a specific number of epochs (loop). In the forward phase, several input data are presented to the model to produce the outputs. In the backward phase, these outputs are compared with the desired value for the given input data. Using the difference between the output of the ANN and the desired data, the weights and biases are then updated (Bishop, 1995; Braga et al., 2000; Silva et al., 2017). The difference between the desired and given output signals of the ANN can be expressed as various kinds of metrics, such as Mean Squared Error (MSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE). The backward
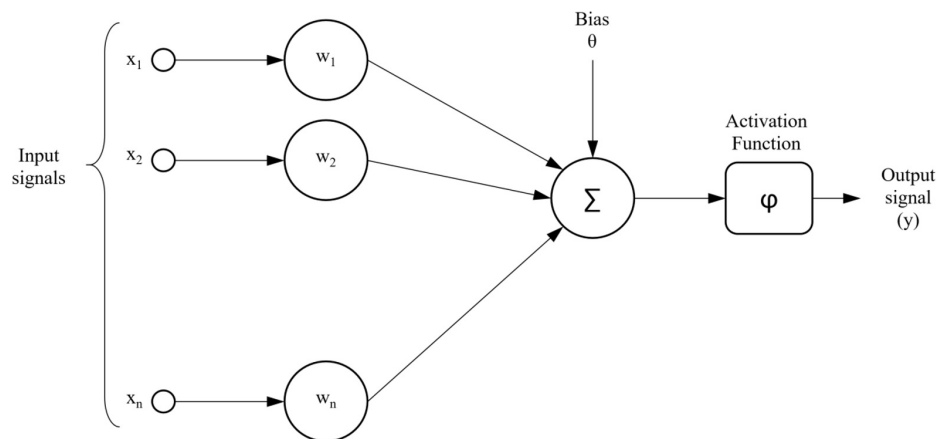
Figure 2. Representation of connections of one artificial neuron.
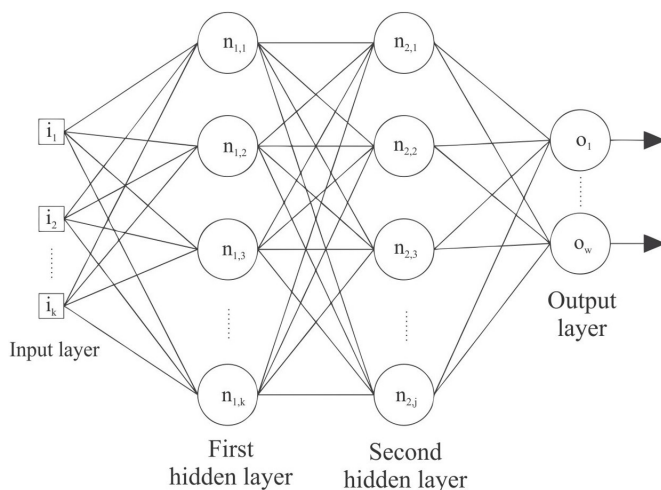**Source:** Adapted from Silva et al. (2017).

Figure 3. Architecture of typical multilayer perceptron neural network.
Source: Adapted from Silva et al. (2017).

phase is also called the training phase. During this phase, the update of the weights occurs following the backpropagation algorithm.

At the training phase of the ANN, one should always look for overfitting and underfitting situations. In overfitting, the ANN results will be very close to the values for the training set, however, for input data different from those used during the training phase, the outputs given by the ANN will not be satisfactory, diverging from the desired values. For underfitting situations, the ANN are not able to achieve a satisfactory level of results because of the low number of epochs, neurons, or hidden layers (Silva et al., 2017; Haykin, 2001).

Sarle (1995) and Bishop (1995) stated that one way to avoid the overfitting of a neural network consists of stopping the training phase before all the training epochs finish. This method is known as early stopping. To perform this method, the first step is to separate the dataset into two sets, a training set, and validation set, use small random initial values, use a slow learning rate, compute the validation error periodically during the training phase, and stop the training when the validation error starts to rise with the epochs.

For further details and explanations of the functioning of the ANN MLP, see Bishop (1995) and Silva et al. (2017).

Artificial intelligence and the limit of deposition in storm sewers

Artificial intelligence and soft computing methods started to be employed in sediment transport issues since the beginning of the 2000's. One of the first published works in the field was the Nagy et al. (2002), which aimed to evaluate the model capability to estimate the total discharge of sediments into streams and natural rivers utilizing ANN. Nagy et al. (2002) observed that several inputs variables can lead to a worse generalization capability of the ANN. Through a discrepancy analysis and comparing the deviations between experimental dataset and the calculated values, Nagy et al. (2002) showed that ANN can be successfully applied for estimating the sediment discharge in natural rivers, due to capability of the ANN accept any number of input parameters, without omission or simplifications (as empirical equations), the ANN model could be applied by field engineers to estimate the sediment discharge without knowing all the processes and theories involved in sediment transport and the parallelism of the model can overcome stochastic nature of the sediment movement more than other equation. In addition, Nagy et al. (2002) comment that one limitation of the ANN models is the restriction for data extrapolation, but this can be easily overcome by feeding new patterns in the learning process of the model.

Regarding the limit of deposition inside storm sewers, the artificial intelligence and soft computing methods started to be used in the 2010's for conditions with deposited sediments and without deposited sediments (clean bed). Azamathulla et al. (2012) presented a study using data from Ab Ghani (1993) and Vongvisessomjai et al. (2010) in an ANFIS based model to predict the Froude number at the limit of deposition in clean pipes. The dataset utilized was divided into two groups, a training set and a testing set, which correspond of 85% and 15% of the total amount of used experiments, respectively. Through analyzing several input parameters in the model, Azamathulla et al. (2012) concluded that the developed ANFIS model outperformed the analyzed equation of Vongvisessomjai et al. (2010). In all input parameters combination utilized, the proposed ANSIS model by Azamathulla et al. (2012) overcome the Vongvisessomjai et al. (2010) equation. To demonstrate that, Azamathulla et al. (2012)

used the Root Mean Squared Error (RMSE) and the determination coefficient (R^2). The best statistical measures observed in the proposed model was $R^2 = 0.98$ and RMSE = 0.185 for the training set and $R^2 = 0.94$ and RMSE = 0.174 for the testing dataset.

Safari et al. (2013) compared three types of ANN architecture, i. e., neural networks with a radial basis function, generalized regression networks, and the multilayer perceptron (MLP), to determine the limit of deposition using Loveless's (1991) data. To train the models, Safari et al. (2013) used 80% of the experimental data obtained from literature, and the remaining 20% were used for evaluating the proposed models. Through analyzing the R^2 and the Mean Squared Error (MSE), Safari et al. (2013) demonstrates that ANN-MLP model outperforms the other two ANN architectures analyzed, with values of R^2 equal to 0.91 and MSE equal to 0.004. Despite the better performance of the ANN-MLP, Safari et al. (2013) concludes in addition that all the analyzed ANN architectures could be used when determining the limit of deposition, with a better performance than the empirical equations.

Ebtehaj & Bonakdari (2013) presented a comparison between the ANN-MLP and ANFIS model in the determination of the Froude Number at the limit of deposition in storm sewers. To perform the analysis, Ebtehaj & Bonakdari (2013) used the data of Ab Ghani (1993) and Vongvisessomjai et al. (2010) and five different input variables combinations. In addition, Ebtehaj & Bonakdari (2013) presented the performance of the equations of Novak & Nalluri (1975), Mayerle et al. (1991), Ab Ghani (1993), May et al. (1996), and Vongvisessomjai et al. (2010) in determining the Froude Number at the limit of deposition inside storm sewers. It was showed that ANN-MLP outperforms all the ANFIS and the empirical equations analyzed, presenting values of 0.98 and 0.0011 for R^2 and RMSE, respectively.

Developing six different ANFIS models with different input variables, Ebtehaj & Bonakdari (2014a) analyzed the performance of estimating the Froude Number at the limit of deposition in rigid boundaries (without deposition – Clean pipes). The experimental data of Ab Ghani (1993) and Vongvisessomjai et al. (2010) were used to training and evaluating the ANFIS models. To determine which model presented better performances it was evaluated the R^2, Mean Relative Error (MRE), Mean Absolute Relative Error (MARE), Mean Squared Relative Error (MSRE), Mean Error (ME), Mean Absolute Error (MAE) and the RMSE. In addition, the ANFIS model was compared with three empirical equations: Ab Ghani (1993), Azamathulla et al. (2012), and Ebtehaj et al. (2014). It was showed that ANFIS model outperformed the existing empirical equations to determine the Froude Number at the limit of deposition.

Ebtehaj & Bonakdari (2014b) evaluated the performance of Genetic Algorithm model (GA), Imperialist Competitive Algorithm (ICA) and an ANN model with the equations of May et al. (1996) and Ebtehaj et al. (2014). The developed analysis was performed using the data of Ab Ghani (1993) and Vongvisessomjai et al. (2010), which were divided into two datasets: training (80% of the data) and evaluation (20% of the data). Applying the RMSE and the Mean Absolute Percentile Error (MAPE), Ebtehaj & Bonakdari (2014b) showed that the best model evaluated was

the ICA, with values of RMSE and MAPE equal to 0.037 and 3.90%, respectively.

One comparison between two evolutionary algorithms was performed by Ebtehaj & Bonakdari (2016). In order to predict the limit of deposition in rigid boundaries through Froude Number, Particle Swarm Optimization (PSO) and the ICA were calibrated using the Ab Ghani (1993) and Vongvisessomjai et al. (2010) dataset. Besides the comparison between these two models, Ebtehaj & Bonakdari (2016) also used the results of an ANN-MLP model, and the equations of May et al. (1996) and Vongvisessomjai et al. (2010). All the predictions were compared with the experimental dataset utilizing the MAPE and the RMSE, which showed better values for the PSO model, corresponding of 3.5% and 0.007, respectively.

Ebtehaj & Bonakdari (2015) evaluated three different training algorithms applied to an ANN-MLP model to predict the Froude Number at limit of deposition. It was analyzed the variable learning rate, resilient back-propagation and the Levemberg-Marquardt algorithm. The predictions of the ANN models were compared with the results provided by equations of May et al. (1996), Azamathulla et al. (2012) and Ebtehaj et al. (2014) using the experimental data from Ab Ghani (1993) and Vongvisessomjai et al. (2010). To select the model with best performance, it was used the R^2, MAPE and RMSE. For each training algorithm, six different models with different input parameter were evaluated. With values of R^2, RMSE and MAPE equal to 0.98, 0.025 and 5.78%, respectively, the ANN-MLP trained with Levemberg-Marquardt algorithm showed the best performance, outperforming the empirical equations analyzed.

Najafzadeh & Bonakdari (2017) develop a study with an application of a neuro-fuzzy based group method of data handling associated with particle swarm optimization (NF-GMDH-PSO) to predict the velocity at the limit of deposition in clean pipes. To train and test the models, the authors utilized the Ab Ghani's (1993) dataset, with 221 experimental data points, which was separated into training dataset (used to train the model) with 75% of the total amount of data and testing dataset (used to evaluate the generalization capability of the model) with 25% of the total amount of data. The authors compared the model predictions with the results provided by the empirical equations of Novak & Nalluri (1975), Mayerle et al. (1991), Ab Ghani (1993), Vongvisessomjai et al. (2010) and Azamathulla et al. (2012). Using the correlation coefficient (R), RMSE, MAPE, scatter index (SI) and BIAS to evaluate the performance of the equations and the proposed model, they concluded that the NF-GMDH-PSO model surpasses the empirical equations performances, when determining the velocity at the limit of deposition.

Ebtehaj et al. (2017a) implemented a hybrid model using decision tree (DT) and two kinds of ANN: Multilayer Perceptron (MLP) and Radial Basis Function (RBF). It was employed data from Ab Ghani (1993), Vongvisessomjai et al. (2010) and, Ota & Nalluri (1999), where 70% of the total amount of data were used in the training and the remaining 30% were used to evaluate the model. To verify how close the model predictions are from experimental data, Ebtehaj et al. (2017a) used the R^2, RMSE, MARE, SI and BIAS. Through comparing predicted values obtained by the proposed model with experimental data utilized,

Ebtehaj et al. (2017a) concluded that the implementation of DT along with MLP improve the performance of the model in order to predict the Froude Number at the limit of deposition in clean pipes, presenting better results than the existing empirical equations.

Ebtehaj et al. (2017b) used data from Ab Ghani (1993), Ota & Nalluri (1999) and Vongvisessomjai et al. (2010) to train and develop a support vector machine model (SVM) coupled with a firefly algorithm (SVM-FFA) to predict the Froude number at the limit of deposition. The performance of the developed model was compared with the equations of May et al. (1996), Azamathulla et al. (2012) and Ebtehaj et al. (2014), using the R^2, MAPE, RMSE, BIAS and SI. In addition, the SVM-FFA model was also compared with the performance of SVM, GP and ANN models. It was concluded that the SVM-FFA provided the better predictions among the models analyses, with values of R^2 equal to 0.986, MAPE of 4.6%, RMSE of 0.235, BIAS of 0.046 and SI of 0.058.

As Safari et al. (2013), Wan Mohtar et al. (2018) analyzed different architectures of ANN to determine the incipient motion in sewers with deposited bed, using data from Yalin & Karahan (1979), Kuhnle (1993), Shvidchenko (2000), Salem (1998) and Bong (2013). Despite the study regard conditions with deposited beds, Wan Mohtar et al. (2018) showed the same conclusion as presented in Safari et al. (2013), i. e., the ANN which provided better results was the MLP. The performances of the evaluated models were analyzed using the R^2, RMSE and the relative error (RE).

Ebtehaj et al. (2019) published a comparison between the ANFIS and a hybrid model ANFIS-PSO to predict the Froude number at the limit of deposition using data from Ab Ghani (1993), Ota & Nalluri (1999) and Vongvisessomjai et al. (2010). In the research, the performance of these models was compared with equations of May et al. (1996), Azamathulla et al. (2012) and Ebtehaj et al. (2014) using R^2, RMSE, MAPE, BIAS and SI. It is showed that the hybrid ANFIS-PSO model outperforms all the equations and the ANFIS model, with values of R^2 equal to 0.982, RMSE of 0.262, MAPE of 5.932%, BIAS of -0.009 and SI of 0.064. Ebtehaj et al. (2019) also performed a sensitivity analysis, which indicate that the volumetric concentration is the most important variable for the correct estimation of the Froude number.

Safari (2019) presented a comparation using three different numerical models: Decision Tree (DT), Generalized Regression (GR) and Multivariate Adaptative Regression Splines (MARS) to determine the densimetric Froude number of to flow to guarantee the self-cleaning design concept. Safari (2019) analyzed his results with the estimations provided by the equations of Mayerle et al. (1991), Ab Ghani (1993) and Vongvisessomjai et al. (2010), using the experimental dataset of Mayerle (1988), May (1993), Ab Ghani (1993) and Vongvisessomjai et al. (2010). Through comparing the results of RMSE, MAPE and concordance coefficient (CC), Safari (2019) showed that all the numerical models presented provided better estimation than the best-fit equations, with the MARS presenting the best estimations, showing values of RMSE equal to 0.79, MAPE of 13.97% and CC of 0.92.

Kargar et al. (2019) analyzed a neuro-fuzzy (NF) and a GEP model to estimate the Froude Number in storm sewers at the limit of deposition. The models were compared with the equations of Mayerle et al. (1991), Ab Ghani (1993) and Vongvisessomjai et al. (2010). The GEP model developed by Ab Ghani & Azamathulla (2011) was also evaluated. To train and to evaluate the performance of the predicted models with the equations, Kargar et al. (2019) used experimental data from Mayerle (1988), May (1993), Ab Ghani (1993) and Vongvisessomjai et al. (2010). The comparisons between the presented models were performed using the RMSE and the CC, which showed that NF model presented better performances among all the analyzed models.

Mehr & Safari (2020) presented the application of multigene genetic programing (MGGP), Gene Expression Programming (GEP) and a multilayer perceptron artificial neural network (ANN-MLP) in order to evaluate non deposition criteria with deposited bed. It was used four experimental datasets from El-Zaemey (1991), Perrusquia (1992, 1993), May (1993) and, Ab Ghani (1993). All the numerical models evaluate presented better predictions than the empirical equations analyzed in the work. Among the models the MGGP presenting slightly better predictions than the other models evaluated. To evaluate the models, Mehr & Safari (2020) used the Nash-Sutcliffe coefficient of efficiency (NSE) and the RMSE.

Montes et al. (2020b) developed a numerical model using the Evolutionary Polynomial Regression-Multi-Objective Genetic Algorithmic (EPR-MOGA) using their own experimental data, carried out a 242 mm diameter acrylic pipe with 11.8 m long and longitudinal slope varying from 1.5% to 1.6%, at the Universidad de los Andes, Colombia. The EPR-MOGA model was trained to predict the Froude Number at the limit of deposition, validated with their own experimental data aggregated another data found in the literature. Using the sum of squared errors (SSE), R^2 and the Akaike Information Criterion (AIC), Montes et al. (2020b) concluded that the EPR-MOGA model performs better than the equations analyzed in their work, which showed values of SSE equal to 0.06, R^2 of 0.98 and the lower AIC observed. One reason attributed to that is because the analyzed models do not considerate the longitudinal slope as a dependent variable, and, the EPR-MOGA have improved generalization capability.

Montes et al. (2021) developed a random forest (RF) model to predict the non-deposition sediment transport in sewers. To train and evaluate the model, it was used data from Mayerle (1988), Ab Ghani (1993), Ota (1999), Vongvisessomjai et al. (2010) besides their own experimental data, regarding experiments on a 242mm acrylic pipe and on a 595mm PVC pipe, using sediment mean diameter of 0.35mm to 2.6mm. The RF model was trained to predict the Froude number at the limit of deposition in clean pipes and then compared with the following models: Montes et al. (2020a), Kargar et al. (2019), Safari (2019), May et al. (1996), Safari & Aksoy (2020), Ebtehaj et al. (2019) and Ebtehaj et al. (2020). To analyze the models, it was used the R^2, RMSE and MAPE. As stated by Montes et al. (2021), the presented RF model outperformed all other evaluated models with a low risk of overfitting, this was achieved due the capability of the model to better capture the complex interactions between input variables when compared to other models. Montes et al. (2021) also highlighted the importance in using larger datasets, once this can lead to better predictions as well. In addition, Montes et al. (2021) pointed that the volumetric sediment concentration at the limit of deposition is the most important input variable when predicting the self-cleaning velocity

in sewers, and a good estimation of this parameter can improve the design of new self-cleaning sewers.

Ebtehaj et al. (2020) presented a sensitivity analysis using several combinations of input parameters to calculate the Froude number at the limit of deposition. Through an extensive analyzes it was showed that the volumetric concentration and the relationship between the grain diameter and the hydraulic radius are the most important parameters, once these variables are present in all optimum models evaluated.

In general, in all the presented studies, one can notice that artificial intelligence and soft computing models were developed in order to determine the limit of deposition using the Froude Number. In addition, with exception of Montes et al. (2020b), the presented researches use a limited experimental data, restraining the range of the observed variables in the training and evaluation dataset. Also, the conclusions and assumptions of previous studies indicates that, equations present worse predictions of the Froude Number than the proposed models evaluated, but there are no discussions of which model should be used in extrapolations.

Due the importance of volumetric concentration at limit of deposition conditions inside storm sewers, as stated by Montes et al. (2021) and Ebtehaj et al. (2020), this study applies an ANN-MLP with a larger experimental dataset gathered in the literature, in order to evaluate if the proposed model can outperform the existing equations when determining the volumetric concentration at the limit of deposition, raising a discussion of which model should be used in the extrapolations of the data. With this approach, better estimations of volumetric concentrations would be possible, improving the self-cleaning determination condition when designing new storm sewers.

Also, it is investigated if the same parameters used in the determination of Froude Number at the limit of deposition is sufficient to produce results with same performance as observed in literature when determining the volumetric concentration.

Table 2 presents a summary relating the cited works with the condition analyzed, numerical model evaluated, dataset utilized, metrics used to evaluate the model, parameters used and, the existing equations evaluated in each work.

MATERIAL AND METHODS

This section presents the dataset utilized to train and evaluate the model, as the architectures and the metrics used to compare the results of the artificial neural network with the other proposed equations.

Experimental dataset

The dataset used in this work was composed of experiments available in the literature. The objective of using the datasets already developed was to use the same information that was used to obtain the equations presented in the literature review to train and evaluate the ANN model.

The entire set of experimental data regarding the definition of the volumetric concentration at the limit of deposition inside storm sewers consisted of those in experiments performed by

May (1993), Mayerle (1988), Ab Ghani (1993), Ota (1999), and Montes et al. (2020a) totaling 544 experimental data. The dataset was caried out by the researchers to study the conditions of the flow at the limit of deposition inside pipes in clean bed and it was used to define the proposed equations presented in the literature review. In those data, the following variables were observed: water depth, pipe roughness, pipe diameter, sediment particle size, velocity, longitudinal gradient, hydraulic radius, and the volumetric concentration. Table 3 shows the ranges for each variable measured in the experiments, and the number of experiments performed by each author.

The quality of experimental dataset used was verified using the Manning equation. If the Manning equation was satisfied through the presented experimental data, then the data was validated. Also, the Manning's coefficient obtained was compared with theoretical values presented by Henderson (1966). In addition, the relation between the depth and the pipe diameter was also verified. Experiments with this relation under 0.05 were not considered, once these conditions do not represent a storm sewer operational condition. The experimental dataset of May (1993), Mayerle (1988), Ab Ghani (1993) and Ota (1999), are widely used in the sediment transport studies. When used aggregated, these datasets provide good ranges of the observed variables. In addition, as exposed in Table 3, the Montes et al. (2020a) experiments were carried out with higher volumetric concentrations, slopes and pipe diameter than the other datasets, which were evaluated following the same abovementioned criteria and used in order to improve the used dataset.

Using the experimental data, the dimensionless sediment diameter (D_{gr}), the relationship between mean shear stress and shear stress at the bottom of the pipe (ξ), the shear velocity (u^*), the Reynolds number of the particle (Re^*) and the dimensionless shear stress (ψ) were calculated according to Equations 4, 7, 20, 21 and 22.

$$u^* = \xi\sqrt{gR_hS} \tag{20}$$

$$Re^* = \frac{u^*d}{\upsilon} \tag{21}$$

$$\psi = \frac{u^{*2}}{g\,d\left(S_s - S\right)} \tag{22}$$

Definition of the training and evaluating datasets

The experimental data were separated into two datasets (training and evaluation datasets). The training dataset, representing 75% of the data (408 experiments), was utilized to train the ANN, and the evaluation dataset with 25% of the data (136 experiments), was used to evaluate the generalization capacity of the ANN. The ratio of data splitting was settled following values found in previous studies.

The training dataset was defined ensuring that 75% of the experiments of each author data were present inside of it, not interfering in the representativeness of the points of each

Table 2. Summary of numerical models evaluated in previous works regarding the definition of self-cleaning condition inside storm sewers.

Author	Bed condition	Best model evaluated	Metrics used to evaluate the models	Dataset used	Parameters used	Equations evaluated
Azamathulla et al. (2012)	Clean bed (without deposit)	ANFIS	R² RMSE	Ab Ghani (1993) Vongvisessomjai et al. (2010)	$F_r = f(d_{50}, R_h, \lambda_0, D_{gr}, C_v)$	Vongvisessomjai et al. (2010)
Safari et al. (2013)	Clean bed (without deposit)	ANN-MLP	R² MSE	Loveless (1991)	$V = f(Q, C_v, y, S_0, d_{50}, R_h)$	None
Ebtehaj & Bonakdari (2013)	Clean bed (without deposit)	ANN	R² RMSE	Ab Ghani (1993) Vongvisessomjai et al. (2010)	$F_r = f(d_{50}, R_h, \lambda_0, D_{gr}, C_v)$	Novak & Nalluri (1975) Mayerle et al. (1991) Ab Ghani (1993) May et al. (1996) Vongvisessomjai et al. (2010)
Ebtehaj & Bonakdari (2014a)	Clean bed (without deposit)	ANFIS	R² MRE MARE MSRE ME MAE RMSE	Ab Ghani (1993) Vongvisessomjai et al. (2010)	$F_r = f(d_{50}, R_h, \lambda_0, D_{gr}, C_v)$	May et al. (1996) Ebtehaj et al. (2014) Azamathulla et al. (2012)
Ebtehaj & Bonakdari (2014b)	Clean bed (without deposit)	ICA	RMSE MAPE	Ab Ghani (1993) Vongvisessomjai et al. (2010)	$F_r = f(d_{50}, R_h, \lambda_0, D_{gr}, C_v, D)$	May et al. (1996) Ebtehaj et al. (2014)
Ebtehaj & Bonakdari (2016)	Clean bed (without deposit)	ICA	RMSE MAPE	Ab Ghani (1993) Vongvisessomjai et al. (2010)	$F_r = f(d_{50}, R_h, \lambda_0, D_{gr}, C_v)$	May et al. (1996) Vongvisessomjai et al. (2010)
Ebtehaj & Bonakdari (2015)	Clean bed (without deposit)	ANN-MLP-LM	R² RMSE MAPE	Ab Ghani (1993) Vongvisessomjai et al. (2010)	$= f(d_{50}, R_h, \lambda_0, D_{gr}, C_v, D)$	Novak & Nalluri (1975) May et al. (1996) Azamathulla et al. (2012) Ebtehaj et al. (2014)
Najafzadeh & Bonakdari (2017)	Clean bed (without deposit)	Neuro-Fuzzy GMDH Model	R RMSE MAPE SI BIAS	Ab Ghani (1993)	$F_r = f(d_{50}, R_h, \lambda_0, D_{gr}, C_v)$	Novak & Nalluri (1975) Mayerle et al. (1991) Ab Ghani (1993) Vongvisessomjai et al. (2010) Azamathulla et al. (2012)
Ebtehaj et al. (2017a)	Clean bed (without deposit)	ANN-DT-MLP	R² RMSE MARE SI BIAS	Ab Ghani (1993) Vongvisessomjai et al. (2010) Ota & Nalluri (1999)	$F_r = f(d_{50}, R_h, D, C_v)$	Azamathulla et al. (2012) May et al. (1996) Ebtehaj et al. (2014)
Ebtehaj et al. (2017b)	Clean bed (without deposit)	SVM-FFA	R² MAPE RMSE BIAS SI	Ab Ghani (1993) Vongvisessomjai et al. (2010) Ota & Nalluri (1999)	$F_r = f(d_{50}, R_h, \lambda_0, D_{gr}, C_v, D)$	Azamathulla et al. (2012) May et al. (1996) Ebtehaj et al. (2014)
Wan Mohtar et al. (2018)	Deposited bed	ANN-MLP	R² RMSE RE	Yalin & Karahan (1979) Kuhnle (1993) Shvidchenko (2000) Salem (1998) Bong (2013)	$V = f(W_b, y, d_{50})$	Novak & Nalluri (1975) El-Zaemey (1991) Salem (1998) Bong (2013)

Table 2. Continued...

Author	Bed condition	Best model evaluated	Metrics used to evaluate the models	Dataset used	Parameters used	Equations evaluated
Ebtehaj et al. (2019)	Clean bed (without depositi)	ANFIS-PSO	R², RMSE, MAPE, BIAS, SI	Ab Ghani (1993), Ota & Nalluri (1999), Vongvisessomjai et al. (2010)	$F_r = f(d_{50}, R_h, \lambda_0, C_v, D)$	Azamathulla et al. (2012), May et al. (1996), Ebtehaj et al. (2014)
Safari (2019)	Clean bed (without deposit)	MARS	RMSE, MAPE, CC	Mayerle (1988), May (1993), Ab Ghani (1993), Vongvisessomjai et al. (2010)	$F_r = f(d_{50}, R_h, \lambda_0, D_{gr}, C_v)$	Mayerle et al. (1991), Ab Ghani (1993), Vongvisessomjai et al. (2010)
Kargar et al. (2019)	Clean bed (without deposit)	NF	RMSE, CC	Mayerle (1988), May (1993), Ab Ghani (1993), Vongvisessomjai et al. (2010)	$F_r = f(d_{50}, R_h, \lambda_0, D_{gr}, C_v)$	Mayerle et al. (1991), Ab Ghani (1993), Vongvisessomjai et al. (2010)
Mehr & Safari (2020)	Deposited bed	MGGP	RMSE, NSE	El-Zaemey (1991), Perrusquia (1992, 1993), May (1993)	$F_r = f(d_{50}, R_h, \lambda_0, D_{gr}, C_v)$	Ab Ghani (1993), Nalluri et al. (1997), Safari & Shirzad (2019)
Montes et al. (2020b)	Clean bed (without deposit)	EPR-MOGA	SSE, R², AIC	Ab Ghani (1993), Mayerle (1988), Ab Ghani (1993), Ota (1999), Vongvisessomjai et al. (2010), Montes et al. (2020b)	$F_r = f(d_{50}, R_h, \lambda_0, D_{gr}, C_v)$	Craven (1953), Mayerle et al. (1991), Ab Ghani (1993), May et al. (1996), Vongvisessomjai et al. (2010), Ebtehaj et al. (2014), Najafzadeh et al. (2017), Safari et al. (2017), Montes et al. (2020b), Kargar et al. (2019), Safari (2019), May et al. (1996)
Montes et al. (2021)	Clean bed (without deposit)	RF	R², RMSE, MAPE	Mayerle (1988), Ab Ghani (1993), Ota (1999), Vongvisessomjai et al. (2010), Montes et al. (2020b)	$F_r = f(d_{50}, R_h, \lambda_0, D_{gr}, C_v)$	Safari & Aksoy (2020), Ebtehaj et al. (2019), Ebtehaj et al. (2020)

Table 3. Variation range of the main parameters of the experimental data set used.

Author	Number of experiments	y (m)	k (m)	D (mm)	d_{50} (mm)	V (m/s)	S (m/m)	Rh (m)	Cv (ppm)
May (1993)	97	0.058-0.299	0.00035-0.00070	76.7-299	0.57-7.9	0.429-1.498	0.0007-0.0237	0.019-0.090	0.31-2110
Mayerle (1988)	106	0.028-0.122	0.00035	152	0.5-8.74	0.374-1.104	0.0014-0.0056	0.017-0.046	20.4-1275
Ab Ghani (1993)	256	0.024-0.338	0.00035-0.001	154-450	0.46-8.3	0.24-1.216	0.0004-0.0056	0.0145-0.14	0.759-1450
Ota (1999)	49	0.055-0.189	0.001	225-305	0.714-5.61	0.39-0.756	0.002	0.034–0.086	4.193–77.99
Montes et al. (2020a)	36	0.003-0.218	0.00035	595	0.35-2.6	0.58-1.41	0.008-0.014	0.019–0.119	20-3410

author. Moreover, it was ensured that each variable's maximum and minimum values were contained in the set. This procedure was taken to ensure that the evaluation dataset do not hold values outside the limits of the ones presented during the training phase.

The evaluation set was composed of the remaining 25% of the experimental data, where, as in the training set, it was ensured that 25% of the data of each author were inserted in this set, without affecting the representativeness of the data. This set

contains information that was not presented to the ANN in the training stage. This information is presented after the training phase to evaluate the capacity of the ANN to generalize the determination of the volumetric concentration at the limit of deposition.

ANN topologies

To observe if the number of neurons in the hidden layers and the combination of the input variables produce significant changes in the generalization capability for the determination of the volumetric concentration at the limit of deposition, several topologies were analyzed. In addition, evaluate different topologies is one way to avoid overfitting of the ANN. Another procedure adopted to avoid overfitting is to stop the training phase before all the training epochs occur, as suggested by Sarle (1995) and Bishop (1995). The metric observed for this procedure was the error function calculated over the epochs in the training phase; when the error ceased decreasing significantly over the epochs, the training was stopped. For the initial epoch, the synaptic weight matrix was randomly initialized. The method adopted to prevent local minima of the error function was restarting the model with different values of the weight matrix.

The ANN architecture used in this study was the MLP, presented in Figure 3. In the input layer, where the data is presented to the ANN, the number of neurons is equal to the number of variables shown to the model, followed by the first and second hidden layers, and finally the output layer, which has only one neuron, whose output value corresponds to the volumetric concentration at the limit of deposition.

Combinations of input variables analyzed

To verify if and how the results of the ANN models are affected by changing the input parameters, several combinations of input variables were considered. These variables were obtained from the experimental data and also calculated using Equations 4, 7, 20, 21 and 22. Eighteen combinations of different input variables presented in Table 4 were analyzed as input parameters, defining one ANN model. In models ANN-XI to ANN-XIV, the input parameters refer to those observed in the literature which belong to the best models evaluated using numerical models, as presented in Table 2 and, the models ANN-XV to ANN-XVIII corresponds with the parameters of equations used to determine the limit of deposition inside storm sewers, presented in Table 1.

The objective of analyzing different combinations of input parameters was to determine if the number and combinations of these variables could lead to a better or worse prediction by the ANN models in defining the volumetric concentration at the limit of deposition. Additionally, this was also needed to compare if the same parameters used in defining this condition through the other models, presented in Table 1 and Table 2, would reproduce the same performance regarding the determination of volumetric concentration at the limit of deposition.

Table 4. Combinations of the input variables analyzed.

ANN Model	Input layer parameters
ANN-I	D, y/D, d_{50}, k, R_h, S, V, Re^*, ψ, u^*, ξ
ANN-II	D, y/D, d_{50}, k, R_h, V, ψ, u^*, ξ
ANN-III	D, y/D, d_{50}, k, R_h, V, Re^*, ψ, ξ
ANN-IV	D, y/D, d_{50}, k, R_h, S, ψ, Re^*
ANN-V	D, y/D, d_{50}, k, R_h, S_0, V, ψ, u^*, ξ
ANN-VI	D, y/D, d_{50}, k, S_0, V, Re^*, ψ, u^*, ξ
ANN-VII	D, y/D, d_{50}, S, Re^*, u^*
ANN-IX	D, y/D, d_{50}, V, S, Re^*, u^*, ξ
ANN-X	D, y/D, d_{50}, k, S, Re^*, u^*, ξ
ANN-XI	d_{50}, R_h, V, D_{gr}, λ_0
ANN-XII	D, d_{50}, R_h, V
ANN-XIII	D, d_{50}, R_h, V, D_{gr}, λ_0
ANN-XIV	D, d_{50}, R_h, V, λ_0
ANN-XV	d_{50}, R_h, V
ANN-XVI	D, d_{50}, A, V, y
ANN-XVII	d_{50}, k, R_h, S, λ_0, ξ
ANN-XIII	d_{50}, R_h, V, λ_0

Number of neurons in the hidden layers and other considerations regarding the training of the ANN models

For each ANN model presented in Table 4, 200 different topologies of ANN were analyzed to obtain the best performance to predict the volumetric concentration at the limit of deposition in sewers.

All the ANN models presented in Table 4 correspond to an ANN-MLP model, with one input layer, two internal processing layers, and one output layer. The number of neurons in the input layer corresponded to the number of input variables in each model presented in Table 4. For all combinations and topologies, the number of neurons in the output layer is equal to one, representing the desired output data, i.e., the volumetric concentration at the limit of deposition. The difference of the 200 analyzed topologies was the variation in the number of neurons in the processing (hidden) layers, which vary from 5 to 25 neurons in the first hidden layer and from 5 to 15 neurons in the second hidden layer. For all neurons, the activation function was hyperbolic tangent, the learning rate was equal to 0.01, and the maximum training epochs was 10,000.

Comparison between experimental and calculated data

After training the 200 topologies for each ANN model presented in Table 4, every architecture with different combinations of input variables and number of neurons in the two hidden layers was evaluated using the training dataset, evaluation dataset and with all the data collected from literature. The objective in calculating and comparing the volumetric concentration using the evaluation

data set was to evaluate the capability of the ANN to provide results as good as for the training dataset, avoiding overfitting and showing the capability of the model to generalize the definition of the volumetric concentration at the limit of deposition.

The equations of Mayerle et al. (1991) (Equation 1), Ab Ghani (1993) (Equation 2), May et al. (1996) (Equation 5), Ota (1999) (Equation 11), Vongvisessomjai et al. (2010) (Equation 12), Ebtehaj et al. (2014) (Equation 13), Najafzadeh et al. (2017) (Equation 14), Romero (2018) (Equation 15), Montes et al. (2020a) (Equation 16) and Tafarojnoruz & Sharafati (2020) (Equation 17) were also evaluated using all the experimental data collected in the literature.

The equations of Mayerle et al. (1991), Vongvisessomjai et al. (2010), Ebtehaj et al. (2014), Najafzadeh et al. (2017), Montes et al. (2020a) and Tafarojnoruz & Sharafati (2020) (Equation 17) were algebraically modified to set the volumetric concentration as the dependent variable of the equation.

The comparison between calculated and desired (experimental) data was evaluated by calculating the coefficient of determination (R²), the Mean Absolute Percentile Error (MAPE), the root mean squared error (RMSE), the mean absolute error (MAE), the mean bias error (MBE), scatter index (SI) and, the Akaike information criterion (AIC). These statistical measures are commonly used in other studies developed to determine the limit of deposition in sewers using artificial intelligence and soft computing methods and give good overview of the generalization capability of the model.

The coefficient of determination (R²) indicates which the total variation is common to the elements that constitute the pairs analyzed, where the closer to 1.0 the greater is the proximity to the experimental values. The R² is calculated though Equation 23, where the r_i is the real (experimental) value, p_i is the predicted (calculated) value and $\bar{r}$ is the mean value of the real data.

$$R^2 = 1 - \frac{\Sigma(r_i - p_i)^2}{\Sigma(r_i - \bar{r})^2} \tag{23}$$

The MAPE is directly related to the accuracy of the values given by the ANN in the face of the experimental values. This parameter quantifies the dispersion of the calculated observations in relation to the values observed experimentally in mean percentile terms, where lower values indicate better predictions. The MAPE can be calculated through Equation 24, where N is the number of samples.

$$MAPE = \frac{100}{N}\Sigma\left|\frac{r_i - p_i}{p_i}\right| \tag{24}$$

The RMSE, defined by the Equation 25, can measure how large the residuals are. It is another method to measure the fit of the predicted values with the real values, where lower values indicate better predictions. Due to the squared term, the RMSE is sensitive to outliers in the sample.

$$RMSE = \sqrt{\frac{\Sigma(r_i - p_i)^2}{N}} \tag{25}$$

Through the absolute difference between the calculated and desired values, the MAE describes the mean magnitude of the residuals, independently if they are positives or negatives. As the MAPE and the RMSE, lower MAE values indicate better predictions. The MAE can be calculated by Equation 26.

$$MAE = \frac{1}{N}\Sigma|r_i - p_i| \tag{26}$$

Differently of R², MAPE, RMSE and MAE, the MBE is not used as a measure of the model fit, once low MBE can produce predictions with high residuals. The result of MBE is used to analyze if the model is underestimating or overestimating the predictions based if it is negative or positive. The MBE can be calculated through Equation 27.

$$MBE = \frac{1}{N}\Sigma(p_i - r_i) \tag{27}$$

The SI, calculated through the relation of the RMSE and the mean value of the observed data, and provides a normalized measure of the errors. As reported by Bryant et al. (2016), lower values are an indication of better model performance. The SI can be calculated through Equation 28.

$$SI = \frac{RMSE}{\bar{r}} \tag{28}$$

Finally, the AIC was calculated as a measure of trade-off with the goodness of the fit and the model parsimony (Montes et al. 2020a). Lower values of the AIC indicate a better model. The AIC is calculated through Equation 29, where k is the number of predictors.

$$AIC = N \ln\left[\frac{1}{N}\Sigma(r_i - p_i)^2\right] + 2k \tag{29}$$

For the selection of the best ANN topology for each input dataset shown in Table 4, it was observed which number of neurons in the two hidden layers provided the most satisfactory behavior when comparing the volumetric concentration given by the ANN and the three experimental datasets defined, i. e., training dataset, evaluation dataset and all the data, using the statistical measures presented.

The same analysis was repeated to identify the equation which provide the best estimations of the volumetric concentration at the limit of deposition, however, for this analysis, it was used the entire experimental dataset.

Finally, a similar analysis with the MAPE, RMSE and MAE was performed grouping the dataset into four ranges of volumetric concentration, defined by the quartiles of the volumetric concentration in experimental data. The purpose of this analysis was to evaluate how the model performances could be affected by the range of the volumetric concentration magnitude, indicating ranges where the model predictions could have worse accuracy and how large the magnitude of the residuals are.

RESULTS AND DISCUSSIONS

The results presented in this chapter are divided in four sections. In the first section, the results obtained by the ANN is presented and compared with the experimental data. The second regards the comparison between the results provided by the equations and the experimental data, in order to point the equation that better generalize the definition of the volumetric concentration at the limit of deposition. The third section presents a performance evaluation of the ANN model and the equations separating the experimental dataset into quartiles. The Fourth section presents a comparison of the results obtained by the ANN and the equation with the best performance.

Determination of the volumetric concentration at the limit of deposition in storm sewers using ANN

The capacity to determine the volumetric concentration at the limit of deposition using the ANN model was evaluated considering different combinations of input variables through eighteen ANN models presented in Table 4. For each ANN model, 200 different topologies were analyzed with different numbers of neurons in the hidden layers. The number of topologies was needed to verify whether the number of neurons in the hidden layers could improve, or not, the goodness of the fit of model comparing the calculated and the experimental volumetric concentration at the limit of deposition. This procedure was also used to avoid overfitting in addition with the early stopping method.

For each combination of the input variables shown in Table 4, the 200 topologies were trained and evaluated using the statistical measures presented.

The best topology for each ANN model is presented in Table 5, which also presents the topology (number of neurons) that provided the best performances. The statistical measures presented in Table 5 is separated for the training dataset, evaluation dataset and for all experimental data gathered from the literature . The best values for each statistical parameter and dataset are highlighted in bold and underlined.

Analyzing Table 5, it is possible to observe that all evaluated models had better performance for the training dataset, than the

Table 5. Topology of the best evaluated ANN model for each combination of input data with the related statistical measures.

		I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV	XV	XVI	XVII	XVIII
Neurons in first layer*		5	11	9	18	15	7	22	23	23	15	8	14	22	21	17	21	16	10
Neurons in second layer*		12	5	7	8	5	9	5	5	7	5	9	10	6	6	10	11	9	12
Numbers of input variables		11	9	9	8	10	10	6	7	8	8	5	4	6	5	3	5	6	4
R2	Training dataset	0.78	0.88	0.88	**0.92**	0.86	0.88	0.72	0.80	0.80	0.85	0.73	0.82	0.76	0.82	0.74	0.81	0.91	0.76
	Evaluation dataset	0.70	0.76	0.57	**0.90**	0.65	0.85	0.77	0.77	0.70	**0.90**	0.45	0.60	0.41	0.50	0.42	0.42	0.89	0.35
	All data	0.78	0.88	0.89	**0.92**	0.86	0.87	0.72	0.80	0.80	0.86	0.73	0.82	0.76	0.81	0.74	0.81	0.90	0.75
MAPE	Training dataset	39.47	41.56	45.27	34.24	40.57	37.66	66.73	49.30	52.08	**34.20**	75.72	64.22	49.41	51.22	66.79	60.79	41.17	61.86
	Evaluation dataset	57.52	47.45	51.48	**39.63**	52.80	47.65	168.41	91.64	88.30	44.97	77.11	57.58	73.78	58.76	64.21	54.78	65.22	66.06
	All data	41.14	44.90	47.76	**35.09**	42.28	39.06	77.88	57.53	58.81	35.99	81.23	67.43	60.30	54.11	68.75	64.52	45.24	69.30
RMSE	Training dataset	175.25	176.26	222.97	**143.73**	171.55	148.22	182.74	164.61	170.56	152.19	353.27	224.25	217.99	233.40	353.72	231.94	253.26	349.14
	Evaluation dataset	238.11	218.29	305.84	**167.11**	225.04	180.77	218.32	218.65	227.00	187.43	462.04	303.92	301.40	317.52	463.95	323.74	330.70	456.39
	All data	184.84	183.35	231.38	**146.32**	180.99	153.16	190.90	172.71	179.89	159.61	374.41	230.61	224.59	240.70	376.19	240.96	277.42	370.01
MAE	Training dataset	70.40	71.29	83.11	**57.18**	70.15	60.76	83.91	70.40	74.40	62.66	134.79	94.20	90.41	94.87	137.92	92.53	92.58	135.64
	Evaluation dataset	112.23	101.43	140.23	**73.85**	106.40	86.40	111.46	107.69	111.83	89.97	205.34	143.17	146.19	150.62	208.69	150.17	133.98	205.72
	All data	77.44	76.51	89.94	**59.84**	76.34	64.89	90.16	76.09	80.94	68.54	146.91	98.82	96.11	100.83	150.14	98.48	104.05	147.91
MBE	Training dataset	39.70	33.25	47.68	**13.61**	36.78	23.00	43.00	32.60	30.52	32.30	97.76	37.47	47.13	51.00	102.99	48.42	52.91	100.00
	Evaluation dataset	95.89	86.89	131.27	**49.91**	88.72	63.69	61.75	82.54	83.51	66.05	193.22	126.25	134.21	135.29	194.97	142.91	104.68	190.74
	All data	50.12	43.09	57.98	**20.47**	46.62	31.87	49.79	41.10	39.31	41.58	110.95	46.40	55.80	59.51	116.67	58.17	66.87	112.43
SI	Training dataset	0.65	0.65	0.82	**0.53**	0.63	0.55	0.67	0.61	0.63	0.56	1.30	0.83	0.80	0.86	1.30	0.85	0.93	1.29
	Evaluation dataset	0.78	0.72	1.00	**0.55**	0.74	0.59	0.72	0.72	0.75	0.62	1.52	1.00	0.99	1.04	1.52	1.06	1.09	1.50
	All data	0.66	0.66	0.83	**0.52**	0.65	0.55	0.68	0.62	0.64	0.57	1.34	0.82	0.80	0.86	1.35	0.86	0.99	1.32
AIC	Training dataset	4238	4238	4430	**4070**	4218	4099	4262	4179	4210	4117	4798	4425	4406	4459	4795	4454	4528	4786
	Evaluation dataset	1511	1483	1575	**1408**	1493	1434	1477	1479	1492	1439	1679	1563	1565	1576	1676	1582	1590	1674
	All data	5701	5688	5941	**5441**	5676	5494	5726	5619	5665	5535	6457	5928	5903	5976	6458	5977	6133	6442

*Number of neurons which produced the best results for each combination of input variables.

evaluation dataset. This behavior was already expected once the training dataset were known by the model, differently from the evaluation dataset, which is composed with data that was not presented to the ANN during the training phase.

With an exception of the MAPE for the training dataset, the ANN-IV showed the best performance for all the calculated statistical measures, including the lowest AIC, showing that was the best model evaluated. The best evaluated topology configuration of the ANN-IV was with 18 neurons in the first hidden layer and 8 in the second hidden layer. The input parameters used in the model were the D, y/D, d_{50}, k, R_h, S, ψ and Re^{*}.

The ANN-IV model presented good R^2 values, corresponding to 0.92, 0.90 and 0.92 for the training, evaluation and all data, respectively. The RMSE values obtained for the training, evaluation and all the data were 143.73, 167.11 and, 146.32 respectively, the lowest among all the others ANN models evaluated. In terms of MAPE, the ANN model presented a value equal to 35.09%, considering all the evaluated data. This means that the predictions performed by the model present, an average deviation of 35.09 % from the real data. In terms of ppm, this can be evaluated with MAE, which show us a value of 59.84 ppm, for all the data. Through MBE values, it is possible to notice that for all the datasets the ANN model tends to overestimate the volumetric concentration at the limit of deposition, since the values were all positives (13.61 for the training dataset, 49.91 for the evaluation dataset and 20.47 using all the data). ANN-IV showed SI values of 0.53, 0.55 and 0.52 for the training, evaluation and all the data.

In addition, Table 5 show how the topology and the input parameters can be decisive. Is interesting to notice that, more neurons are not correlated with better predictions and, with exception of ANN-I, ANN-VI, ANN-XI and ANN-XVIII, all the models presented better performance with more neurons in the first hidden layer than in the second.

In terms of the number of input variables, through Table 5 it is possible to observe that the models with less than 7 showed worse performance in relation of the other models, indicating that the ANN model can provide better results with more predictors than less.

In addition, with the performed analysis it was possible to notice that ANN-XI to ANN-XIV do not showed good predictions of the volumetric concentration at the limit of deposition. In these models it was used the same parameters applied in other researches that estimates the Froude number at the limit of deposition, presented in Table 2. Between the evaluated models that used same variables as the numerical models presented in Table 2, the ANN-XIII showed the best performance. This model had 6 input variables (D, d_{50}, R_h, V, D_{gr}, λ_s) which correspond with the same variables adopted by Ebtehaj & Bonakdari (2014b, 2015), with the exception of the volumetric concentration, once it was the dependent variable in analyzed case. For all the data, ANN-XIII showed values of R^2 equal to 0.76, MAPE of 60.3%, RMSE of 224.59 ppm, MAE of 96.11 ppm, MBE indicating the overestimation of the predictions equal to 55.8 and an SI indicating an expected error of 80%.

The models ANN-XV to ANN-XVIII evaluated the volumetric concentration at the limit of deposition using input variables which corresponded with the same parameters used in equations presented in the literature review and summarized in Table 1. It was interesting to observe that models ANN-XVI and ANN-XVII presented good results, even with only 5 and 6 variables in the input layer, respectively. The parameters used as input variables in the model ANN-XVI was D, d_{50}, A, V, y, and in the model ANN-XVII was d_{50}, k, R_h, S, λ_0, ξ, corresponding with the same parameters presented by May et al. (1996) (Equation 5), Ota (1999) (Equation 11) and Romero (2018) (Equation 15). The performance of both models was close. For the entire dataset ANN-XVI showed $R^2 = 0.81$, MAPE = 54.82%, RMSE = 240.96 ppm, MAE = 98.48 ppm, MBE = 58.17 and SI = 0.86. As ANN-XVII presented $R^2 = 0.90$, MAPE = 45.24 %, RMSE = 277.42 ppm, MAE = 104.05 ppm, MBE = 66.87 and SI = 0.99. These results may indicate that the equations of May et al. (1996), Ota (1999) and Romero (2018) could have the better performances when evaluating the volumetric concentration at the limit of deposition.

Figure 4 presents the correlation between the experimental data and the values obtained by ANN-IV (the best evaluated ANN model) and by the models ANN-XVI and ANN-XVII,
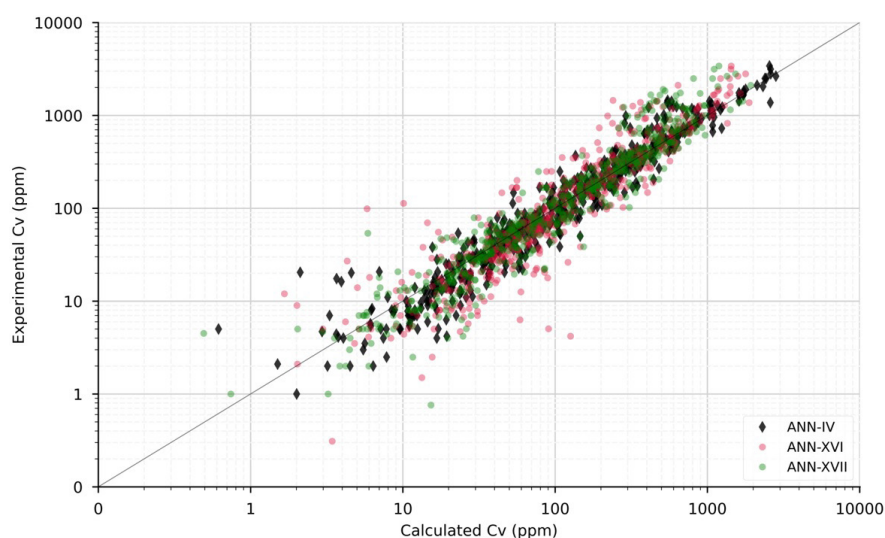

Figure 4. Comparison between the volumetric concentration at the limit of deposition obtained by the experimental data with the ANN-IV (best ANN model), ANN-XVI and ANN-XVII.

which shows the best performance using the same predictors as the equations presented in Table 1. From Figure 4 it is possible to notice that, highest deviations in terms of percentage between calculated and experimental value lies within Cv below 30 ppm, a condition that the experiment is very difficult to perform and is not interested in real storm sewers.

Determination of the volumetric concentration at the limit of deposition in storm sewers using the equations presented in the literature

After analyzing the performance of the ANN models in determining the volumetric concentration at the limit of deposition in storm sewers, another analysis was performed using the results provided by the equations of Mayerle et al. (1991), Ab Ghani (1993), May et al. (1996), Ota (1999), Vongvisessomjai et al. (2010), Ebtehaj et al. (2014), Najafzadeh et al. (2017), Romero (2018), Montes et al. (2020a) and Tafarojnoruz & Sharafati (2020). The data used in this analysis were the same for the training and evaluating the performance of the ANN.

The objective of this analysis was to evaluate the performance of the equations using a larger and different set of data, and determine which of the equations presents the best performance when comparing the results provided and the experimental data set.

Thus, as for the neural networks, after calculating the volumetric concentration at the limit of deposition using the abovementioned equations, statistical quantities were calculated for each one. These values are listed in Table 6, where the best values for each statistical parameter is highlighted in bold and underlined. Figure 5 and Figure 6 shows the performance of each equation in determining the volumetric concentration at the limit of deposition, by comparing the values calculated with the values obtained experimentally in log-log scale which was used to calculate the R² presented.

By analyzing the results presented in Table 6, Figure 5 and Figure 6, it can be observed that the results obtained by Equations 5 and 11 showed better performances, with close values of R², MAPE, RMSE, and MAE, confirming the observation raised at the performance evaluation of the ANN models. Equation 15 showed a slightly worse results than Equation 5 and 11, but still presented a good result between the evaluated equations.

Equation 11 presented a coefficient of determination higher than all the other equations, corresponding to 0.88, lower MAPE, RMSE and MAE, equal to 53.69 %, 328.38 ppm and 130.75 ppm, respectively. The MBE however resulted in 14.11, indicating an overestimation of the predicted data, higher than the MBE presented by Equation 2, of -6.32. For Equation 11 the SI corresponded to 1.17, and the lower value of AIC between analyzed models, indicating that this equation present better estimations of the volumetric concentration at the limit of deposition than the other equations evaluated, and should be used.

Besides the better values of the statistical measures analyzed, the better performance of Equation 11 can be associated with the stronger theoretical basis used to obtain it, regarding the relationship between the shear stress at the bottom of the pipe and the mean shear stress (ξ) and the influence of the pipe roughness (k), expressed in terms of the relationship with the sediment grain diameter (d_{50}).

Equation 5 presented a R² equal to 0.84, MAPE of 56.53 %, RMSE of 342.28 ppm and MAE of 133.42 ppm. As Equation 11, the MBE was also positive, indicating an overestimation of the volumetric concentration at the limit of deposition, however, a higher value, equal to 66.95. The SI was equal to 1.22, indicating worse predictions than Equation 11.

The equations that were algebraically modified to calculate the volumetric concentration at the limit of deposition as a dependent variable, showed worse predictions, indicating that these equations can be good estimators for the Froude number, but when used to predict the volumetric concentration at the limit of deposition, these equations do not show the same performance observed in the determination of the Froude number.

Evaluation of ANN model and equations performances into quartiles

Additionally, the performance of the equations and the ANN-IV model was analyzed separating the experimental dataset into 4 subsets, defined by the volumetric concentration quartiles. The importance of this analysis relies on the need to observe how the models perform with a limited range of volumetric concentrations, and if there is a range of volumetric concentration that other equation outperforms Equation 11 and the ANN. For this analysis, MAPE, MAE and RMSE was used, which are summarized in

Table 6. Statistical quantities obtained comparing the results of the equations and the experimental data.

| Author	R²	MAPE	RMSE	MAE	MBE	SI	AIC
Mayerle et al. (1991)[*] – Equation 1	0.61	59.22	367.13	163.34	113.17	1.31	6435
Ab Ghani (1993) – Equation 2	0.83	79.71	352.43	144.50	**-6.32**	1.26	6395
May et al. (1996) – Equation 5	0.84	56.53	342.28	133.42	66.95	1.22	6363
Ota (1999) – Equation 11	**0.88**	**53.69**	**328.38**	**130.75**	14.11	**1.17**	**6318**
Vongvisessomjai et al. (2010)* – Equation 12	0.82	69.95	484.42	173.56	-32.03	1.73	6735
Ebtehaj et al. (2014)* – Equation 13	0.72	119.63	901.50	258.50	-158.30	3.22	7411
Najafzadeh et al. (2017)* – Equation 14	0.38	228.29	463.09	243.46	-175.08	1.66	6686
Romero (2018) – Equation 15	0.86	73.16	352.44	151.24	-33.99	1.26	6395
Montes et al. (2020a)* – Equation 16	0.80	84.01	727.45	220.22	-113.24	2.60	7181
Tafarojnoruz & Sharafati (2020)* – Equation 17	0.78	75.73	804.04	231.52	-118.36	2.87	7286

*Cv as dependent variable.
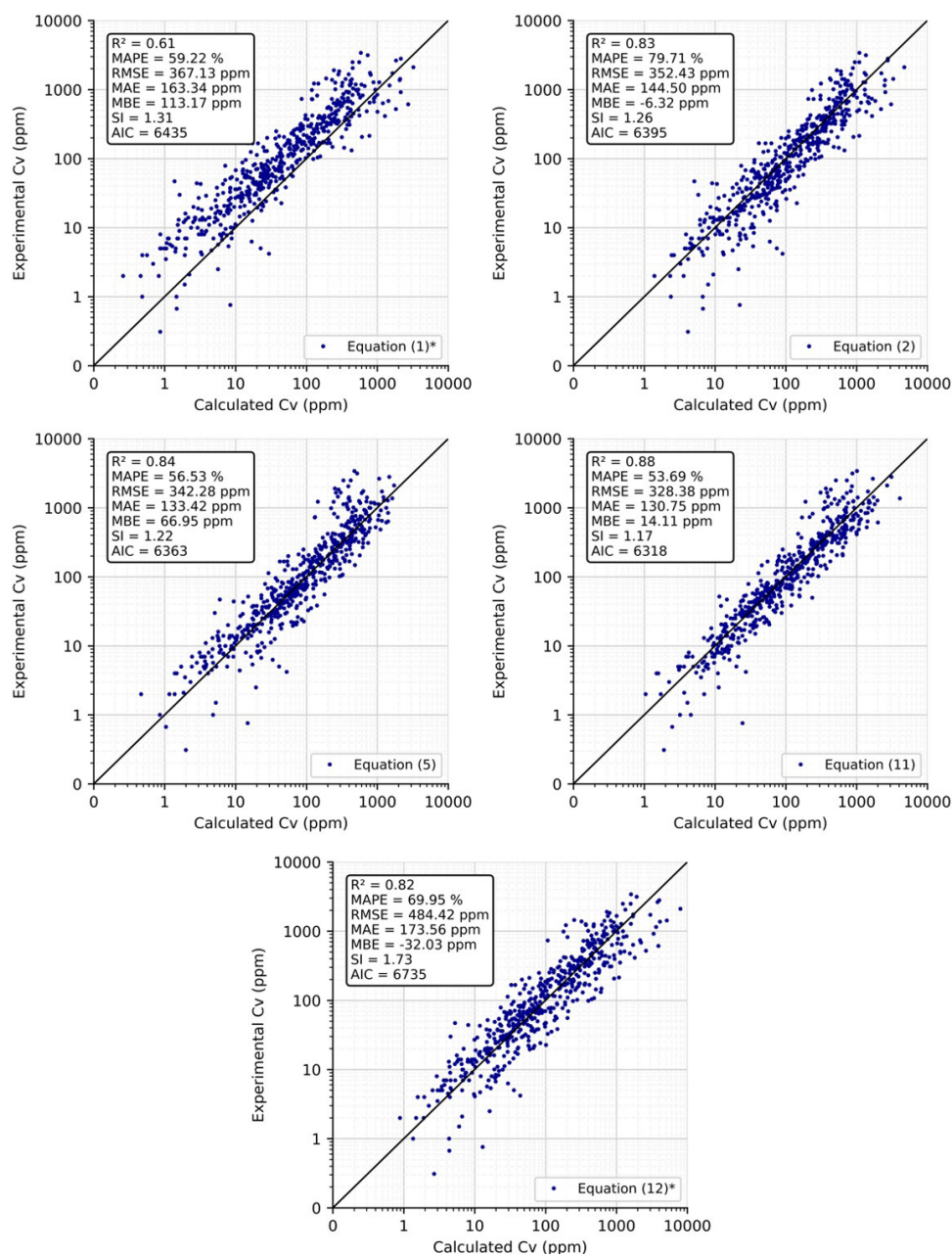
Figure 5. Correlation between calculated and experimental volumetric concentrations at the limit of deposition for Equations 1, 2, 5, 11 and 12.

Table 7, Table 8, and Table 9, respectively. Figure 7, Figure 8 and Figure 9 presents the result of these statistical measures for all the equations and the ANN for each quartile.

The values presented in Table 7 to Table 9 and in Figure 7 to Figure 9 can give a better understanding of the residuals between the calculated and the experimental volumetric concentration at the limit of deposition using the quartiles of the experimental data. As already observed when analyzing the entire dataset, the Equations 5 and 11 presented better performance than the other equations.

Regarding the MAPE, analyzing the first quartile, Equation 1 presented the best value for MAPE, equal to 71.84 %. Equation 11 showed the lower percentiles deviations in the predictions for the second and third quartile, with values of 90.68 %, 39.10 % and 40.44 %, respectively. In the fourth quartile,

Equation 5 showed better results for MAPE, resulting in a mean percentile error of 40.97 %.

For the MAE, Equation 11 also showed a better performance in the first, third and fourth quartile, with values of 6.82 ppm, 73.51 ppm and 422.20 ppm. In the second quartile, Equation 5 presented a MAE of 20.42 ppm.

For the RMSE the better results were divided between Equation 5 and 11. Equation 5 showed the lower RMSE in the second and third quartile of 29.06 ppm and 111.01 ppm, respectively, as Equation 11 presented the lower values for the first and fourth quartile, resulting in RMSE values of 10.06 ppm and 646.29 ppm, respectively.

Is interesting to observe that equations obtained from algebraic modifications still presented the worse results when analyzing the
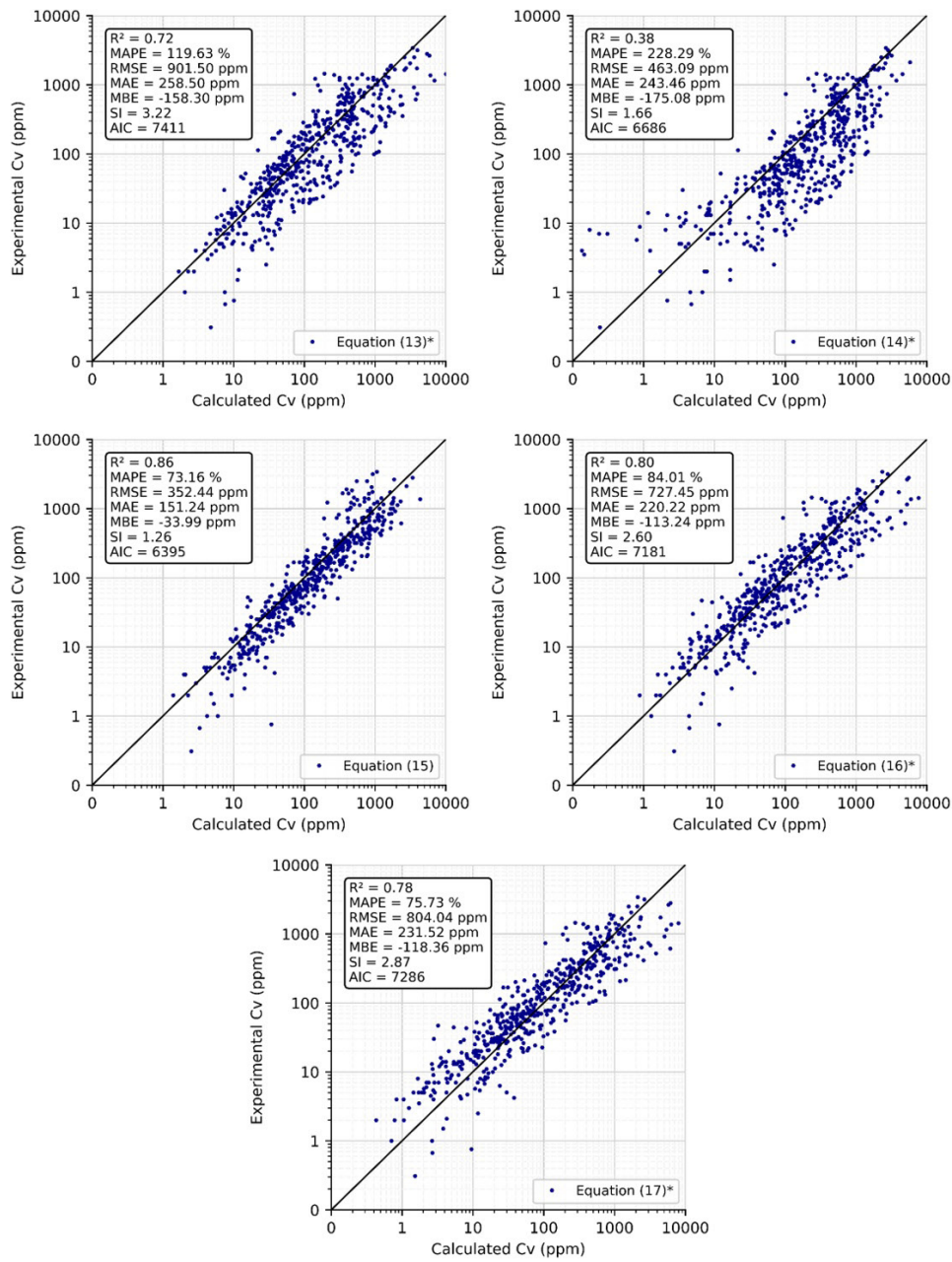
Figure 6. Correlation between calculated and experimental volumetric concentrations at the limit of deposition for Equations 13, 14, 15, 16 and 17.

Table 7. MAPE analyzed for each quartile of the experimental data.

Equation	First Quartile (0.31 ppm − 28.82 ppm)	Second Quartile (29 ppm − 86.9 ppm)	Third Quartile (87.3 ppm − 320 ppm)	Fourth Quartile (290 ppm − 3,410 ppm)
(1)	**71.84**	53.48	55.45	56.09
(2)	162.00	51.74	54.94	50.18
(5)	99.25	40.51	45.39	**40.97**
(11)	90.68	**39.10**	**40.44**	44.52
(12)	102.39	48.78	65.92	62.69
(13)	168.77	100.28	118.84	90.62
(14)	383.57	269.10	187.72	72.75
(15)	132.51	52.10	54.30	53.75
(16)	105.44	62.12	86.77	81.70
(17)	87.16	51.22	77.47	87.05
ANN-IV	**66.00**	**29.81**	**24.82**	**19.71**

Table 8. MAE analyzed for each quartile of the experimental data.

Equation	First Quartile (0.31 ppm – 28.82 ppm)	Second Quartile (29 ppm – 86.9 ppm)	Third Quartile (87.3 ppm – 320 ppm)	Fourth Quartile (290 ppm – 3,410 ppm)
(1)	7.63	28.20	102.19	515.35
(2)	12.53	26.45	95.75	443.26
(5)	8.84	**20.42**	80.36	424.06
(11)	**6.82**	20.47	**73.51**	**422.20**
(12)	8.73	25.04	114.78	545.68
(13)	16.60	50.14	200.15	767.10
(14)	49.37	132.76	306.94	484.78
(15)	9.95	27.52	98.58	468.91
(16)	9.71	31.59	150.48	689.12
(17)	8.47	26.50	136.38	754.71
ANN-IV	**6.93**	**15.44**	**43.84**	**173.16**

Table 9. RMSE analyzed for each quartile of the experimental data.

| Equation	First Quartile (0.31 ppm – 28.82 ppm)	Second Quartile (29 ppm – 86.9 ppm)	Third Quartile (87.3 ppm – 320 ppm)	Fourth Quartile (290 ppm – 3,410 ppm)
(1)	9.54	31.70	125.70	722.66
(2)	22.34	35.38	154.94	686.34
(5)	15.76	**29.06**	**111.01**	674.70
(11)	**10.06**	33.16	111.57	**646.29**
(12)	14.85	36.16	176.54	951.82
(13)	31.12	90.74	329.58	1770.01
(14)	88.90	216.00	469.71	763.30
(15)	14.82	45.10	149.36	687.25
(16)	17.59	50.48	243.17	1433.44
(17)	12.90	36.31	227.00	1591.5
ANN-IV	**9.84**	**22.50**	**60.60**	**285.25**
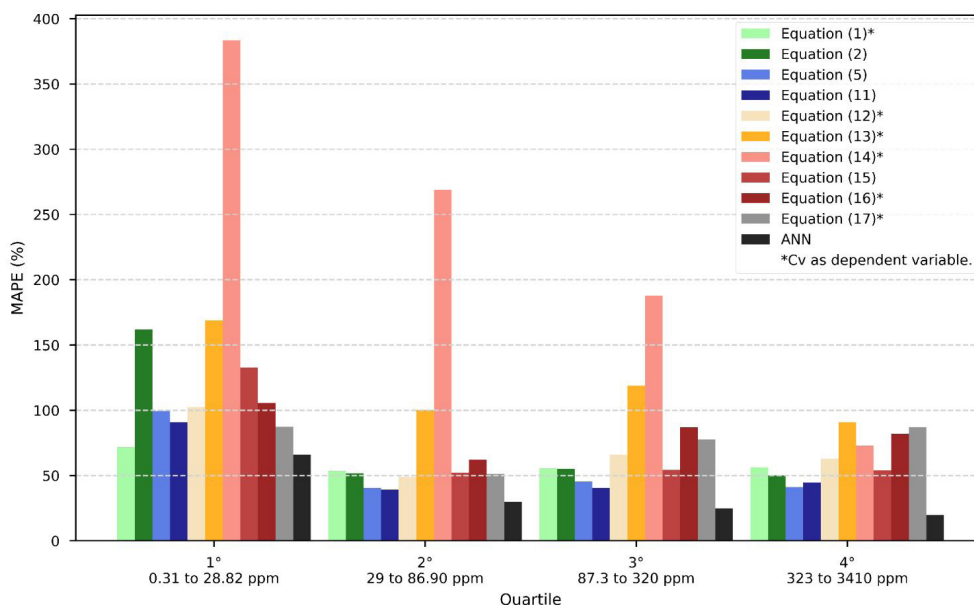
Figure 7. MAPE calculated separately in the quartiles of the experimental dataset.

dataset separated in quartiles, reinforcing that they should not be used for estimating the volumetric concentration at the limit of deposition.

Observing the ANN-IV results it was clear that the model outperformed all the equations in determining the volumetric concentration at the limit of deposition. For all the analyzed statistical measures and quartiles, the ANN showed lower deviations from the experimental data than the equations, except for the MAE in the first quartile, where the ANN-IV presented a value of 6.93 ppm and Equation 11 presented a slightly lower value, equal to 6.82 ppm.
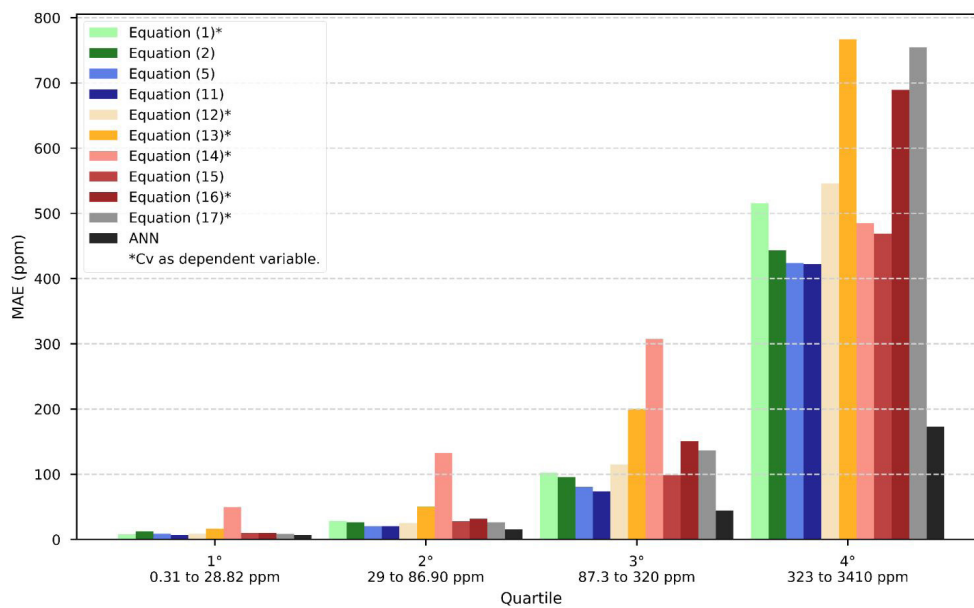
Figure 8. MAE calculated separately in the quartiles of the experimental dataset.
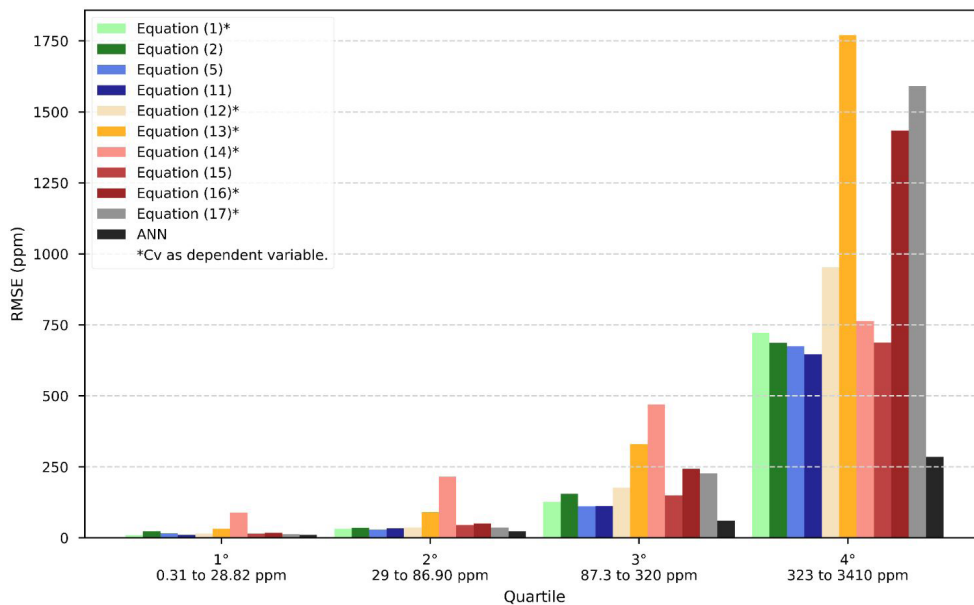

Figure 9. RMSE calculated separately in the quartiles of the experimental dataset.

In addition, the ranges of volumetric concentrations in the first quartile do not match with real operational conditions observed in a storm sewer. In this range of volumetric concentration the experiments are difficult to perform due to the low transport capacity of the flow. This is the reason of the higher MAPE values observed in the results in this quartile.

The increasing residuals pointed by the MAE and RMSE are related with the increased values of the volumetric conccentration in each quartile. These increasing deviations, do not reflect in terms of percentage, as observed in the MAPE results.

It is difficult to point just to one equation that provides the best results when compared with the experimental dataset. Equations 11 and 5 presented the best results in terms of MAPE and R² when analyzing the entire dataset, as well when analysing the dataset separeted into quartiles.

However, owning to the theoretical basis utilized to obtain Equation 11, it is reasonable affirm that this equation can produce better estimations of the volumetric concentration at the limit of deposition than the others equations analyzed. In addition, due to the same reason, Equation 11 is also the most indicated model to use for calculating the limit of deposition in extrapolations.

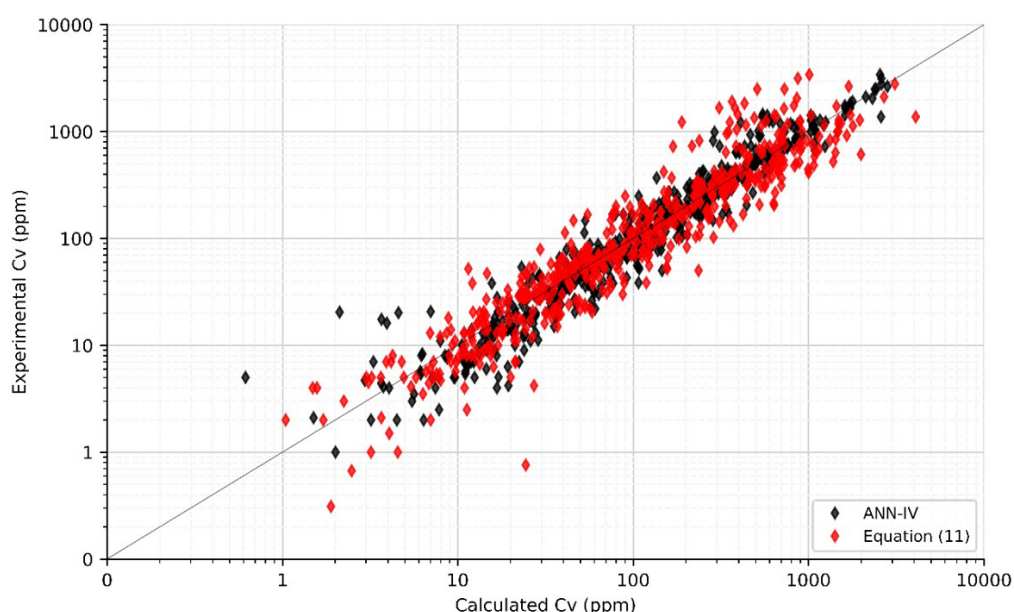## Comparison between volumetric concentration at the limit of deposition obtained by the ANN-IV and by Equation 11

The best model analyzed was the ANN-IV, with 8 input variables (D, y/D, d_{50}, k, R_h, S, ψ, Re^*), 18 neurons in the first hidden layer and 8 in the second. When analyzing the proposed

Table 10. Summary of the statistical measures presented by the ANN-IV and Equation 11 using the entire dataset.

Model	R2	MAPE	RMSE	MAE	MBE	SI	AIC
ANN-IV	0.92	35.09	146.32	59.84	20.47	0.52	5441
Equation 11	0.88	53.69	328.38	130.75	14.11	1.17	6318

Table 11. Summary of the statistical measures calculated for the ANN-IV and Equation 11 for the experimental data quartiles.

Model	Variable	First Quartile	Second Quartile	Third Quartile	Fourth Quartile
ANN-IV	MAPE	66	29.81	24.82	19.71
	MAE	6.93	15.44	43.84	173.16
	RMSE	9.84	22.5	60.6	285.25
Equation 11	MAPE	90.68	39.1	40.44	44.52
	MAE	6.82	20.47	73.51	422.2
	RMSE	10.06	33.16	111.57	646.29

Figure 10. Comparison between the results obtained through the ANN analysis and Equation 11.

equations, Equation 11 showed the best performance in determining the volumetric concentration at the limit of deposition inside storm sewers, leading to the indication that this equation is the most recommended equation for the generalization and extrapolation when determining the volumetric concentration at the limit of deposition. Table 10 and Table 11 presents a summary of the statistical measures obtained by ANN-IV and Equation 11 analysing all the data and separating the data into quartiles.

Analysing both of the models it is clear that the ANN model can outperform the Equation 11 in all aspects. But, due to the limitations of the method, the ANN is not indicated in extrapolation cases. For extrapolation, Equation 11 is recommended. Figure 10 shows a comparison between the volumetric concentration values obtained experimentally with the values calculated by ANN-IV and those calculated by Equation 11. It can be noted that there is a greater dispersion in the volumetric concentrations calculated by Equation 11 when compared to the values calculated by ANN-IV, more frequently in regions with volumetric concentrations around 1,000 ppm. It is also observed that both methods lead to larger dispersions for smaller volumetric concentrations, less than 30 ppm, it

although as commented, this range of volumetric concentration not match with the operational conditions of storm sewer.

CONCLUSIONS

The determination of the volumetric concentration at the limit of deposition inside storm sewers is highly important, since it defines a sediment rate which the flow can transport without deposition, thus, avoiding obstructions and reducing the discharge capacity of the pipes. In addition, as stated by Montes et al. (2021) and Ebtehaj et al. (2020), the volumetric concentration is the most important variable used in defining when predicting the self-cleaning velocity in sewers, and a good estimation of this parameter can improve the design of new self-cleaning sewers.

In the present research, it was demonstrated using seven different statistical measures (R^2, MAPE, MAE, RMSE, MBE, SI and AIC) that ANN outperformed existing equations to determine the volumetric concentration at the limit of deposition. Through analyzing the results obtained by the ANN models, it

was clear that the best predictions was achieved using a larger set of input variables, since the evaluated ANN models with less than 7 predictors showed, in general, worse predictions. This observation shows that the volumetric concentration can be harder to define than the Froude number at the limit of deposition, since the ANN and other numerical models presented in the literature that define it use less than 6 variables. Several equations presented in the literature, developed to determine the Froude number, uses only d_{50}, and C_v, and yet present satisfactory predictions.

However, the ANN-XVI and ANN-XVII models with 5 and 6 input parameters respectively shown good results. Is interesting notice that, these ANN models had the same input parameters than the equations of May et al. (1996) (Equation 5), Ota (1999) (Equation 11) and Romero (2018) (Equation 15), the equations which showed the best predictions.

The best results regarding the evaluated ANN models were achieved with ANN-IV, which had D, y/D, d_{50}, k, R_h, S, ψ, Re^* as input parameters, and showed R² = 0.92, MAPE = 35.95 %, RMSE = 146.32 ppm, MAE = 59.84 ppm, MBE = 20.47 ppm, SI = 0.52 and the lower AIC between the analyzed ANN models. Is interesting to note that reports from the literature review in researches that used artificial intelligence and soft computing methods to predict the Froude number, show MAPE values ranging from 3% to 14% approximately. But, for the volumetric concentration, the minimum MAPE achieved was 35.95%, confirming the higher difficulty associated in determining it.

The study also showed that the equations obtained to predict the Froude number do not show a satisfactory performance when used to calculate the volumetric concentration at the limit of deposition. This could be seen also in the ANN analysis, where the ANN models that used the same parameters of models used to predict the Froude Number do not had a good performance as well.

Sustaining the hypothesis presented, Ota's (1999) model (Equation 11) showed the best performance among all the presented equations. The better performance of Equation 11 in determining the volumetric concentration at the limit of deposition could be associated due to the stronger theoretical basis used to obtain it. The Equation 11 presented a R² = 0.88, MAPE = 53.69 %, RMSE = 328.38 ppm, MAE = 130.75 ppm, MBE = 14.11 ppm, SI = 1.17 and the lower AIC between the evaluated equations. When analyzing the performance at the volumetric concentration quartiles, the performance of Equation 11 was in most of the cases the best one, followed by Equation 5, of May et al. (1996), confirming the observed when analyzing the ANN models. These results suggest that Equation 11 should be used for extrapolations.

In addition, through the quartiles analysis, it was observed that all methods converged to better results in higher ranges of concentration. This analysis demonstrate that it is hard to ensure the volumetric concentrations at the limit of deposition for low values, despite the method applied. The higher MAPE observed in this range of volumetric concentration is related with the more difficulties involved to perform the experiments.

Fortunately, this condition is not important for real storm sewer design.

REFERENCES

Ab Ghani, A. (1993). *Sediment transport in sewers* (PhD thesis). University of Newcastle upon Tyne, Newcastle, England.

Ab Ghani, A., & Azamathulla, H. M. (2011). Gene-expression programming for sediment transport in sewer pipe systems. *Journal of Pipeline Systems Engineering and Practice*, *2*(3), 102-106. http://dx.doi.org/10.1061/(ASCE)PS.1949-1204.0000076.

American Society of Civil Engineers – ASCE. (2007). *Gravity sanitary sewer design and construction* (2nd ed.). Reston: ASCE.

Azamathulla, H. M., Ab Ghani, A., & Fei, S. Y. (2012). ANFIS-based approach for predicting sediment transport in clean sewer. *Applied Soft Computing*, *12*(3), 1227-1230. http://dx.doi.org/10.1016/j.asoc.2011.12.003.

Bishop, C. M. (1995). *Neural networks for pattern recognition*. Oxford: Clarendon Press.

Bong, C. H. J. (2013). *Self-cleansing urban drain using sediment flushing gate based on incipient motion* (Phd thesis). University Sans Malaysia, Penang, Malaysia.

Braga, A. P., Ludemir, T. B., & Carvalho, A. C. P. L. F. (2000). *Artificial neural networks: theory and applications* (1st ed.). Rio de janeiro: LTC.

Bryant, M. A., Hesser, T. J., & Jensen, R. E. (2016). *Evaluation statistics computed for the Wave Information Studies (WIS)* (ERDC/CHL, No. CHETN-I-91). Washington: U. S. Army Corps of Engineers.

Butler, D., & Davies, J. W. (2004). *Urban drainage* (2nd ed.). London: Spon Press. . http://dx.doi.org/10.1201/b12810.

Craven, J. (1953). The transportation of sand in pipes: I. Full-pipe flow. In *Proceedings of the fifth hydraulics Conference* (pp. 67-76), Iowa City, United States of America.

Ebtehaj, I., & Bonakdari, H. (2013). Evaluation of sediment transport in sewer using artificial neural network. *Engineering Applications of Computational Fluid Mechanics*, *7*(3), 382-392. http://dx.doi.org/10.1080/19942060.2013.11015479.

Ebtehaj, I., & Bonakdari, H. (2014a). Performance evaluation of adaptative neural fuzzy inference system for sediment transport in sewers. *Water Resources Management*, *28*(13), 4765-4779. http://dx.doi.org/10.1007/s11269-014-0774-0.

Ebtehaj, I., & Bonakdari, H. (2014b). Comparison of genetic algorithm and imperialist competitive algorithms in predicting bed load transport in clean pipe. *Water Science and Technology*, *70*(10), 1695-1701. PMid:25429460. http://dx.doi.org/10.2166/wst.2014.434.

Ebtehaj, I., & Bonakdari, H. (2015). Bed load sediment transport estimation in a clean pipe using multilayer perceptron with different training algorithms. *KSCE Journal of Civil Engineering*, *20*(2), 581-589. http://dx.doi.org/10.1007/s12205-015-0630-7.

Ebtehaj, I., & Bonakdari, H. (2016). Assessment of evolutionary algorithms in predicting non-deposition sediment transport. *Urban Water Journal*, *13*(5), 499-510. http://dx.doi.org/10.1080/15730 62X.2014.994003.

Ebtehaj, I., Bonakdari, H., & Sharifi, A. (2014). Design criteria for sediment transport in sewers based on self-cleaning concept. *Journal of Zhejiang University.*, *15*(11), 914-924. http://dx.doi.org/10.1631/jzus.A1300135.

Ebtehaj, I., Bonakdari, H., & Zaji, A. H. (2017a). A new hybrid decision tree method based on two artificial neural networks for predicting sediment transport in clean pipes. *Alexandria Engineering Journal*, *57*(3), 1789-1795.

Ebtehaj, I., Bonakdari, H., Shamshirband, S., Ismail, Z., & Hashim, R. (2017b). New approach to estimate velocity at limit of deposition in storm sewers using vector machine coupled with firefly algorithm. *Journal of Pipeline Systems Engineering and Practice*, *8*(2), 04016018. http://dx.doi.org/10.1061/(ASCE)PS.1949-1204.0000252.

Ebtehaj, I., Bonakdari, H., & Es-Haghi, M. S. (2019). Design of a hybrid ANFIS-PSO model to estimate sediment transport in open channels. *Iranian Journal of Science and Technology, Transactions of Civil Engineering*, *43*(4), 851-857.

Ebtehaj, I., Bonakdari, H., Safari, M. J. S., Gharabaghi, B., Zaji, A. H., Madavar, H. R., Khozani, Z. S., Es-Haghi, M. S., Shishegaran, A., & Mehr, A. D. (2020). Combination of sensitivity and uncertainty analyses for sediment transport modeling in sewer pipes. *International Journal of Sediment Research*, *35*(2), 157-170. http://dx.doi.org/10.1016/j.ijsrc.2019.08.005.

El-Zaemey, A. K. S. (1991). *Sediment transport over deposited bed sewers* (Phd thesis). University of Newcastle, Newcastle upon Tyne, England.

Fendrich, R., Obladen, N. L., Aisse, M. M., & Garcias, C. M. (1997). *Urban erosion drainage and control* (4th ed). Curitiba: Champagnat.

Graupe, D. (2013). *Principles of artificial neural networks* (3rd ed.). Chicago: World Scientific. . http://dx.doi.org/10.1142/8868.

Haykin, S. (1999). *Neural networks: a comprehensive foundation* (2nd ed.). Hamilton: Pearson Education.

Haykin, S. (2001). *Redes neurais: princípios e prática* (2nd ed.). Porto Alegre: Artmed.

Henderson, F. M. (1966). *Open channel flow* (1st ed.). New York: Macmillan.

Jain, A. K., Jianchang Mao, & Mohiuddin, K. M. (1996). Artificial neural networks: a tutorial. *Computer*, *29*(3), 31-44. http://dx.doi.org/10.1109/2.485891.

Kargar, K., Safari, M. J. S., Mohammadi, M., & Samadianfard, S. (2019). Sediment transport modeling in open channels using neuro-fuzzy and gene expression programming techniques. *Water Science and Technology*, *79*(12), 2318-2327. PMid:31411586. http://dx.doi.org/10.2166/wst.2019.229.

Kuhnle, R. A. (1993). Incipient motion of sand-gravel sediment mixtures. *Journal of Hydraulic Engineering*, *119*(12), 1400-1415. http://dx.doi.org/10.1061/(ASCE)0733-9429(1993)119:12(1400).

Loveless, J. H. (1991). *Sediment transport in rigid boundary channels with particular reference to the condition of incipient deposition* (PhD thesis). University of London, London.

May, R. W. P. (1982). *Sediment transport in sewers* (Technical Report, No. 222). Wallingford: Hydraulic Research Station.

May, R. W. P. (1993). *Sediment transport in pipes and sewers with deposited beds* (Report SR 320). Wallingford: HR Wallingford.

May, R. W. P., Ackers, J. C., Butler, D., & John, S. (1996). Development of design methodology for self-cleaning sewers. *Water Science and Technology*, *33*(9), 195-205. http://dx.doi.org/10.2166/wst.1996.0210.

Mayerle, R. (1988). *Sediment transport in rigid boundary channels* (PhD thesis). Newcastle Upon Tyne, Newcastle, England.

Mayerle, R., Nalluri, C., & Novak, P. (1991). Sediment transport in rigid bed conveyances. *Journal of Hydraulic Research*, *29*(4), 475-495. http://dx.doi.org/10.1080/00221689109498969.

Mays, L. W. (1999). *Hydraulic design handbook*. Michigan: McGraw-Hill.

Mays, L. W. (2004). *Stormwater collection systems design handbook* (1st ed.). Michigan: McGraw-Hill.

Mehr, A. D., & Safari, M. J. S. (2020). Application of soft computing techniques for particle Froude number estimation in sewers pipes. *Journal of Pipeline Systems Engineering and Practice*, *11*(2), 04020002. http://dx.doi.org/10.1061/(ASCE)PS.1949-1204.0000449.

Montes, C., Vanegas, S., Kapelan, Z., Berardi, L., & Saldarriaga, J. (2020a). Non-deposition self-cleaning models for large sewer pipes. *Water Science and Technology*, *81*(3), 606-621. PMid:32385214. http://dx.doi.org/10.2166/wst.2020.154.

Montes, C., Berardi, L., Kapelan, Z., & Saldarriaga, J. (2020b). Predicting bedload sediment transport of non-cohesive material in sewer pipes using Evolutionary Polynomial Regression-Multi-Objective Genetic Algorithm strategy. *Urban Water Journal*, *17*(2), 154-162. http://dx.doi.org/10.1080/157306 2X.2020.1748210.

Montes, C., Kapelan, Z., & Saldarriaga, J. (2021). Predicting non-deposition sediment transport in sewer pipes using Random Forest. *Water Research*, *189*(1), 116639. PMid:33227613. http://dx.doi.org/10.1016/j.watres.2020.116639.

Nagy, H. M., Watanabe, K., & Hirano, M. (2002). Prediction of sediment load concentration in rivers using artificial neural network model. *Journal of Hydraulic Engineering*, *128*(6), 588-595. http://dx.doi.org/10.1061/(ASCE)0733-9429(2002)128:6(588).

Najafzadeh, M., & Bonakdari, H. (2017). Application of a neuro-fuzzy GMDH model for predicting the velocity at limit of deposition in storm sewers. *Journal of Pipeline Systems Engineering and Practice*, *8*(1), 06016003. http://dx.doi.org/10.1061/(ASCE)PS.1949-1204.0000249.

Najafzadeh, M., Laucelli, D. B., & Zahiri, A. (2017). Application of Model Tree and evolutionary polynomial regression for evaluation of sediment transport in pipes. *KSCE Journal of Civil Engineering*, *21*(5), 1956-1963. http://dx.doi.org/10.1007/s12205-016-1784-7.

Nalluri, C., El-Zaemey, A. K., & Chan, H. L. (1997). Sediment transport over fixed deposited beds in sewers: an appraisal of existing models. *Water Science and Technology*, *36*(8-9), 123-128. http://dx.doi.org/10.2166/wst.1997.0654.

Novak, P., & Nalluri, C. (1975). Sediment transport in smooth fixed bed channels. *Journal of the Hydraulics Division*, *101*(9), 1139-1154. http://dx.doi.org/10.1061/JYCEAJ.0004412.

Ota, J. J. (1999). *Effect of particle size and gradation on sediment transport in storm sewers*. Newcastle: University of Newcastle Upon Tyne.

Ota, J. J., & Nalluri, C. (1999). Graded sediment transport at limit of deposition in clean pipe channel. In *28th International Association for Hydro-environment Engineering and Research*, Graz, Austria.

Ota, J. J., & Nalluri, C. (2003). Urban storm sewer design: approach in consideration of sediments. *Journal of Hydraulic Engineering*, *129*(4), 291-297. http://dx.doi.org/10.1061/(ASCE)0733-9429(2003)129:4(291).

Pedroli, R. (1963). *Bed load transportation in channels with fixed and smooth inverts* (PhD thesis). Scuola Politecnica Federale, Zurigo, Switzerland.

Perrusquia, G. (1992). *Sediment transport in pipe channels* (Postdoctoral experiments studies). Chalmers University of Technology, Göteborg, Switzerland.

Perrusquia, G. (1993). *An experimental study from flume to stream traction in pipe channels* (Report). Göteborg, Switzerland: Chalmers University of Technology.

Romero, A. (2018). *Determination of sediment deposition limit inside drainage pipes using artificial neural networks* (Master degree dissertation). Universidade Federal do Paraná, Curitiba.

Rushforth, P. J., Tait, S. J., & Saul, A. J. (2003). Use of a full-scale test facility to examine sewer-sediment mobility. *Water and Environment Journal : the Journal / the Chartered Institution of Water and Environmental Management*, *7*(1), 40-44.

Safari, M. J. S. (2019). Decision tree (DT), generalized regression neural network (GR) and multivariate adaptive regression splines (MARS) models for sediment transport in sewer pipes. *Water Science and Technology*, *79*(6), 1113-1122. PMid:31070591. http://dx.doi.org/10.2166/wst.2019.106.

Safari, M. J. S., & Shirzad, A. (2019). Self-cleansing design of sewers: definition of the optimum deposited bed thickness. *Water Environment Research*, *91*(5), 407-416. PMid:30731037. http://dx.doi.org/10.1002/wer.1037.

Safari, M. J. S., Aksoy, H., & Mohammadi, M. (2013). Application of ANN for the analysis of sediment incipient deposition in rigid boundary channels. In *Proceedings of the 16th International Conference Transport and Sedimentation of Solid Particles* (pp. 53-62), Rostock, Germany.

Safari, M. J. S., Mohammadi, M., & Ab Ghani, A. (2018). Experimental studies of self-cleansing drainage system design: a review. *Journal of Pipeline Systems Engineering and Practice.*, *9*(4), 04018017. http://dx.doi.org/10.1061/(ASCE)PS.1949-1204.0000335.

Safari, M., Aksoy, H., Unal, N., & Mohammadi, M. (2017). Non-deposition self-cleansing design criteria for drainage systems. *Journal of Hydro-Enviroment Research.*, *14*, 76-84. http://dx.doi.org/10.1016/j.jher.2016.11.002.

Safari, M. J., & Aksoy, H. (2020). Experimental analysis for self-cleansing open channel design. *Journal of Hydraulic Research*, *59*(3), 500-511.

Salem, A. M. (1998). *Incipient motion over loose deposited beds in a rigid rectangular channel* (MSc thesis). University Sains Malaysia, Penang, Malasya.

Sarle, W. S. (1995). Stopped training and other remedies for overfitting. In *Proceedings of the 27th Symposium on the Interface of Computing Science and Statistics*, Pittsburgh, USA.

Shvidchenko, A. B. (2000). *Incipient motion of streambeds* (PhD thesis). University of Glasgow, Glasgow, United Kingdom.

Silva, I. N., Spatti, D. H., Flauzino, R. A., Liboni, L. H. B., & Alves, S. F. R. (2017). *Artificial neural networks* (1st ed.). Cham: Springer International Publishing. . http://dx.doi.org/10.1007/978-3-319-43162-8.

Tafarojnoruz, A., & Sharafati, A. (2020). New formulations for prediction of velocity at limit of deposition in storm sewer based on a stochastic technique. *Water Science and Technology*, *81*(12), 2634-2649. PMid:32857749. http://dx.doi.org/10.2166/wst.2020.321.

Tucci, C. E. M. (2004). *Hydrology: science and application* (3rd ed.). Porto Alegre: UFRGS.

Vongvisessomjai, N., Tingsanchali, T., & Babel, M. S. (2010). Non-deposition design criteria for sewers with part-full flow. *Urban Water Journal*, *7*(1), 61-77. http://dx.doi.org/10.1080/15730620903242824.

Wan Mohtar, W. H. M., Afan, H., El-Shafie, A., Bong, C. H. J., & Ghani, A. A. (2018). Influence of bed deposit in the prediction of incipient sediment motion in sewers using artificial neural networks. *Urban Water Journal*, *15*(4), 296-302. http://dx.doi.org/10.1080/1573062X.2018.1455880.

Yalin, M. S., & Karahan, E. (1979). Inception of sediment transport. *Journal of the Hydraulics Division*, *105*(11), 1433-1443. http://dx.doi.org/10.1061/JYCEAJ.0005306.

Authors contributions

Adhemar Romero: Contributed in literature review, data analysis, programming, development of ANN models, and writing.

José Junji Ota: Contributed in reviewing the theoretical basis of the sediment transport, supported in writing, data analysis, literature review and datasets collecting.

Editor-in-Chief: Adilson Pinheiro

Associated Editor: Iran Eduardo Lima Neto