Possibilities to use the meta model and classical approaches to evaluate the impact of hydraulic conditions in prediction of the critical submergence depth ratio

Ramin Vafaei Poursorkhabi and Roghayeh Ghasempour

ABSTRACT

One of the hydraulic phenomena that mainly occurs during the water withdrawal process of channels is the formation of vortices that can cause many problems for the hydro-mechanical facilities of intakes. In the current study, classical models and meta model approaches (i.e. Support Vector Machine and Gene Expression Programming) were applied to evaluate the impact of pipe diameter and hydraulic condition changes in prediction of the critical submergence depth ratio in horizontal intakes. In this regard, two types of critical submergence experiments, based on bottom clearance, were considered (i.e. \( c = 0 \) and \( c = \frac{d}{2} \), in which \( c \) and \( d \) are the bottom clearance and diameter of the intake, respectively). Different models were developed and tested using experimental data series. The results indicated that in modeling the critical submergence depth ratio, meta model approaches led to better predictions compared to the classical approaches. It was observed that the developed models for the state of \( c = \frac{d}{2} \) yielded better results. According to the outcome of sensitivity analysis, the ratio of velocities in the intake pipe and channel (\( \frac{V_i}{V_c} \)) had a key role in the modeling. It was also found that intake pipe diameter affected the critical submergence depth ratio in intake pipes. Increasing pipe diameter caused a decrease in model accuracy.

Key words | diameter, hydraulic condition, intake, possibilities, semi-empirical formula, submergence depth ratio, SVM

INTRODUCTION

Since water is transmitted from seas, lakes, rivers or simply reservoirs through intakes to be used in power generation, irrigation, domestic and industrial supply, improvements in the design criteria for the intakes have a great importance in minimizing cost and using water efficiently. In order to reduce the cost of construction, the intake must be placed as close to the water surface as possible. According to Figure 1, if the water depth above the pipe intake is not sufficient, this can lead to vortices and air entrainment. In Figure 1, the parameters \( V_c, V_i, d, c, \) and \( Sc \) represent channel velocity, intake velocity, pipe diameter, bottom clearance, and critical submergence depth, respectively. Air entraining vortices at intakes can lead to some serious problems, such as vibration, efficiency loss, structural damage, and flow reduction in hydroturbines and culverts. Therefore, the location of the intake should be arranged so that the water level is well above the intake to prevent the occurrence of air-entraining vortices.

In literature, many researchers have worked to solve vortex formation problems at intakes by comprehensive analytical and numerical methods. Gulliver & Rindels (1987) assessed the impact of Froude number variation on critical submergence of the intake. Jain et al. (1978) investigated the effects of viscosity and circulation on the critical submergence. Reddy & Pickford (1972) described a design criterion for preventing vortices in pump sumps and at horizontal intakes. They indicated that most free surface vortices occurred above the line \( Sc/d = Fr \), which indicates that the
dimensionless critical submergence should always be greater than the Froude number \( (Fr) \). They developed a formula considering that the Froude number was the only parameter affecting vortex formation.

\[
\frac{Sc}{d} = 1 + Fr
\]  

Ahmad et al. (2008) studied the critical submergence for a horizontal intake in an open channel flow. Analytical equations were developed based on the potential flow and critical spherical sink surface theories for two bottom clearances, as in Equations (2) and (3). In Equation (2), parameter \( g \) is acceleration due to gravity. Ahmad et al. (2008) indicated that the effects of the Froude number and the ratio of intake velocity and channel velocity are more pronounced in comparison to the other parameters.

\[
c = 0: \quad \frac{Sc}{d} = 0.36 \times Fr^{0.8} \times \left[ \frac{Vc}{\sqrt{gd}} \right]^{-0.9}
\]

\[
c = \frac{d}{2}: \quad \frac{Sc}{d} = 0.27 \times Fr^{0.039} \times \left[ \frac{Vc}{Vc} \right]^{1.02}
\]

Gurbuzdal (2009) conducted a series of experiments in horizontal intakes to study the possible scale effects on the formation of air-entraining vortices. The experiments were conducted on four intake pipes with different diameters located in a large reservoir. Parameters of the vortex formation, Froude number, Reynolds number \( (Re) \) and side wall clearance \( (b) \) were chosen as governing parameters and an empirical formula was derived based on these parameters as Equation (4). Gurbuzdal (2009) indicated that side-wall clearance ratio is not effective on the critical submergence ratio after it exceeds about 6.

\[
\frac{Sc}{d} = Fr^{0.865} \times \left[ \frac{b}{d} \right]^{-0.565} \times Re^{0.0424}
\]  

Also, Swaroop (1973) and Amphlett (1978) represented very simple formulas (Equations (5) and (6), respectively) for the estimation of critical submergence depth ratio based on Froude number:

\[
\frac{Sc}{d} = 1.5 + Fr
\]

\[
\frac{Sc}{d} = 3.95Fr^{0.5} - 0.5
\]

However, existing critical submergence equations rely on a limited database, untested model assumptions and a general lack of filed data, and do not show the same results under variable flow conditions. Therefore, applications of these models are often questionable and it is extremely critical to utilize methods that are capable of predicting the critical submergence of intakes.

In the recent years, application of intelligent approaches (e.g. Artificial Neural Networks (ANNs), Neuro-Fuzzy models (NF), Gene Expression Programming (GEP) and Support Vector Machine (SVM)) in water resources engineering has become viable, leading to numerous publications in this field. SVM and GEP approaches have been applied in modeling various components of water resources systems, including prediction of the side weir discharge coefficient (Azamathulla et al. 2016), prediction of suspended sediment concentration Kisi (2012), long-term prediction of lake water levels (Khan & Coulibaly 2006), modeling bed load transport in circular channels (Roushangar & Ghasempour 2017), modeling bridge pier scour (Azamathulla et al. 2010), and predicting total bedload (Chang et al. 2012).

This study aimed to assess the capability of SVM and GEP models for modeling the critical submergence depth ratio of horizontal intakes and investigate the impact of pipe diameter and hydraulic conditions on the prediction process. In this regard, three experimental data sets were considered and all data were divided into two cases regarding the bottom clearance (i.e. \( c = 0 \) and \( c = d/2 \)). In order to find the most appropriate input combination, two different SVM models were developed and tested using the experimental data. Also, an attempt was made to develop new explicit equations to predict the critical submergence depth ratio in intake pipes. Therefore, the best models of...
SVM in both cases of $c = 0$ and $c = d/2$ were selected and rerun using the GEP model. Then, the capabilities of the proposed equations were compared with several existing semi-empirical equations.

**MATERIALS AND METHODS**

**The data sets**

In this study, three kinds of data presented by Ahmad et al. (2008), Gurbuzdal (2009) and Baykara (2013) were used for prediction goals. The ranges of various parameters used in the experiments are listed in Table 1. Ahmad et al. (2008) examined the critical submergence for horizontal intakes in open channel flow. The experiments were conducted in a flume of 10 m length, 0.37 m width and 0.6 m depth. Three pipes with different diameters and two bottom clearance were used during experiments. Gurbuzdal (2009) conducted some experiments to assess vortices at horizontal intakes. The experiments were performed on four pipes with different diameters located in a large reservoir. Baykara (2013) collected a series of tests in an experimental setup composed of a reservoir with dimensions of $3.10 \times 3.10 \times 2.20$ m and a pump connected to the intake pipe. Six pipes with different diameters of 5, 10, 14.4, 19.4, 25 and 30 cm were used during experiments.

**Support vector machine**

Support vector machines (SVM) as an intelligence approach are essentially used in information categorization and data set classification. This approach, which was developed by Vapnik (Vapnik 1995), is known as structural risk minimization (SRM), which minimizes an upper bound on the expected risk, as opposed to the traditional empirical risk (ERM) which minimizes the error on the training data. The SVM method is based on the concept of an optimal hyper plane that separates samples of two classes by considering the widest gap between them (Gunn 1998). SVM was originally developed for binary decision problems and it can be used as a binary classifier. It is based on statistical learning theory, which deals with the problem of finding a predictive function based on data. This classification method has also been extended to solve prediction problems. For classification, support vector machines (SVMs) have recently been introduced and quickly became the state of the art. Now, the incorporation of prior knowledge into SVMs is the key element that allows the increase in the performance in many applications. The particular forms of prior knowledge can be presented in two main groups: class-invariance and knowledge on the data. Support Vector Regression (SVR) is an extension of SVM regression. The aim of SVR is to characterize a kind of function that has at most $\epsilon$ deviation from the actually obtained objectives for all training data $y_i$, and at the same time would be as flat as possible. SVR formulation is as follows:

$$f(x) = w\phi(x) + b$$  \hspace{1cm} (7)

where $\phi(x)$ denotes a nonlinear function in a feature of the input $x$, $b$, which is called the bias, and the vector $w$ is known as the weight.

<table>
<thead>
<tr>
<th>Bottom clearance</th>
<th>Researcher</th>
<th>$d$ (cm) (pipe diameter)</th>
<th>$V_i$ (m/s) (pipe velocity)</th>
<th>$Sc$ (cm) (submergence depth)</th>
<th>$Re \times 10^5$ (Reynolds number)</th>
<th>$Fr$ (Froude number)</th>
<th>$We$ (Weber number)</th>
<th>No. of data</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c = 0$</td>
<td>Ahmad et al. (2008)</td>
<td>4.25–10.16</td>
<td>0.45–4.23</td>
<td>1.99–21.05</td>
<td>0.45–1.8</td>
<td>0.45–6.56</td>
<td>282–10,425</td>
<td>162</td>
</tr>
<tr>
<td></td>
<td>Baykara (2013)</td>
<td>5–30</td>
<td>0.502–7.48</td>
<td>1.49–26.94</td>
<td>1.23–6.2</td>
<td>0.30–10.68</td>
<td>842–59,251</td>
<td>308</td>
</tr>
<tr>
<td>$c = d/2$</td>
<td>Ahmad et al. (2008)</td>
<td>4.25–10.16</td>
<td>0.37–4.3</td>
<td>0.86–25.13</td>
<td>0.376–1.83</td>
<td>0.37–6.66</td>
<td>190–10,800</td>
<td>162</td>
</tr>
<tr>
<td></td>
<td>Gurbuzdal (2009)</td>
<td>1.597–5.147</td>
<td>0.486–3.124</td>
<td>1.68–24.77</td>
<td>0.263–2.89</td>
<td>0.509–4.032</td>
<td>198–10,700</td>
<td>41</td>
</tr>
</tbody>
</table>
The coefficients of Equation (7) are predicted by minimizing the regularized risk function as expressed below:

$$R_{\min} = C \frac{1}{N} \sum_{i=1}^{n} L_x(t_i, y_i) + \frac{1}{2} \|w\|^2$$

(8)

where

$$L_x(t_i, y_i) = \begin{cases} 0 & |t_i - y_i| \leq \epsilon \\ |t_i - y_i| - \epsilon & \text{Otherwise} \end{cases}$$

(9)

The constant $C$ is the cost factor and represents the trade-off between the weight factor and approximation error. $\epsilon$ is the radius of the tube within which the regression function must lie. The $L_x(t_i, y_i)$ represents the loss function in which $y_i$ is the forecasted value and $t_i$ is the desired value in period $i$. Since some data may not lie inside the $\epsilon$-tube, the slack variables $(\xi_i, \xi^*)$ must be introduced. These variables represent the distance from the actual values to the corresponding boundary values of $\epsilon$-tube. Therefore, it is possible to transform Equation (8) into:

$$R_{\min} = C \sum_{i=1}^{n} (\xi_i, \xi^*) + \frac{1}{2} \|w\|^2$$

(10)

Using Lagrangian multipliers in Equation (10) thus yields the dual Lagrangian form:

$$\text{Max} l(a_i, a^*_i) = -\epsilon \sum_{i=1}^{n} (a_i + a^*_i) + t_i \sum_{i=1}^{n} (a_i - a^*_i) - \frac{1}{2}$$

\begin{align*}
\times \sum_{i=1}^{n} \sum_{j=1}^{n} (a_i - a^*_i)(a_j - a^*_j) - K(x_i, x_j)
\end{align*}

(11)

where $a_i$ and $a^*_i$ are Lagrange multipliers and $l(a_i, a^*_i)$ represents the Lagrange function. $K(x_i, x_j)$ is a kernel function to yield the inner products in the feature space $\varphi(x_i)$ and $\varphi(x_j)$. In general, there are several types of kernel function, namely linear, polynomial, radial basis function (RBF) and sigmoid functions. It should be noted that the biggest limiting factor for SVMs is that the performance of the SVM is highly dependent on the choice of kernel as well as the kernel and cost parameters. In fact, SVM prediction accuracy depends on a good setting of meta-parameters $C$ and $\epsilon$ and the kernel parameters. A deeper understanding of the kernel map would be useful to choose appropriate kernels for a specific task (e.g. by incorporating prior knowledge (Smola & Schölkopf 2004)).

**Gene expression programming (GEP)**

GEP was developed by Ferreria (2001) using fundamental principles of Genetic Algorithms (GA) and Genetic Programming (GP). One strength of the GEP approach is that the creation of genetic diversity is extremely simplified as genetic operators work at the chromosome level. Another strength of GEP consists of its unique, multigenic nature, which allows the evolution of more complex programs composed of several subprograms. GEP, as GA, mimics biological evolution to create a computer program for simulating a specified phenomenon. A GEP algorithm begins by selecting the five elements such as the function set, terminal set, fitness function, control parameters, and stopping condition. There is comparison between predicted values and actual values in the subsequent step. When the desired results in accordance with the error criteria initially selected are found, the GEP process is terminated. If the desired error criteria cannot be found, some chromosomes are chosen by a method called roulette wheel sampling and they are mutated to obtain new chromosomes. After the desired fitness score is found, this process terminates and then the chromosomes are decoded for the best solution to the problem.

**Performance criteria**

In the current study, the model’s performance was evaluated using three statistical parameters: Correlation Coefficient ($R$), Determination Coefficient (DC), and Root Mean Square Errors (RMSE) as depicted in Equation (12). The smaller the RMSE and the greater the DC and $R$, the higher the accuracy of the model will be.

$$DC = 1 - \frac{\sum_{i=1}^{N} (l_o - l_p)^2}{\sum_{i=1}^{N} (l_o - \bar{l_o})^2},$$

$$R = \frac{\sum_{i=1}^{N} (l_o - \bar{l_o}) \times (l_p - \bar{l_p})}{\sqrt{\sum_{i=1}^{N} (l_o - \bar{l_o})^2 \times (l_p - \bar{l_p})^2}},$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (l_o - l_p)^2}$$

(12)
where \( l_o, l_p, l_{\bar{m}}, l_p, N \) respectively are the measured values, predicted values, mean measured values, mean predicted values and the number of data samples.

**Simulation and models development**

**Input variables**

Appropriate selection of input parameters is an important step in the modeling process. Experimental studies (Yıldırım & Kocabas 2002; Ahmad et al. 2008) show that the critical submergence \((Sc)\) can be a function of the parameters:

\[
Sc = f(d, Vi, Vc, b_1, b_2, c, \mu, \rho, g, \sigma)
\]

where \( d \): diameter of intake, \( Vi \): velocity in intake, \( Vc \): velocity in channel, \( b_1 \) and \( b_2 \): left and right-side-wall distance of the reservoir to the intake center line, \( c \): bottom clearance, \( \rho \): water density, \( \mu \): dynamic viscosity, \( \sigma \): surface tension and \( g \): acceleration due to gravity. From dimensional analysis, these parameters can be expressed as follows:

\[
\frac{Sc}{d} = f\left(\frac{Vi}{\sqrt{g d}}, \frac{Vi}{Vc}, \frac{\rho V_i d}{\mu}, \frac{\rho V_i^2 d}{\sigma}, \frac{b_1}{d}, \frac{b_2}{d}, \frac{c}{d}\right)
\]  

Equation (14) can be expressed as follows:

\[
\frac{Sc}{d} = f\left(\frac{Fr}{Vc}, \frac{Vi}{Vc}, \frac{Re}{d}, \frac{We}{d}, \frac{b_1}{d}, \frac{b_2}{d}, \frac{c}{d}\right)
\]

where \( Fr \): intake Froude number, \( Re \): intake Reynolds number and \( We \): intake Weber number. In the experimental setup the used bottom clearance \( c \) was always equal to zero or \( d/2 \); therefore, the \( c/d \) ratio can be removed from Equation (15). Also, the intake pipe was placed midway between the left and right side-walls so that \( b \) was used instead of \( b_1 \) and \( b_2 \). After these modifications, Equation (15) can be expressed as follow:

\[
\frac{Sc}{d} = f\left(\frac{Fr}{Vc}, \frac{Vi}{Vc}, \frac{Re}{d}, \frac{We}{d}, \frac{b}{d}\right)
\]

The developed models for predicting the critical submergence depth ratio of horizontal intakes are listed in Table 2. It should be noted that for both cases of \( c = 0 \) and \( c = d/2 \), 75% of the whole data set were used for training goals and the remaining 25% of data were used for testing goals (for both cases training and testing sets were from the same set of fixed \( c/d \)). The order of the data sets was selected in a way such that the training data set contains a representative sample of all the behavior in the data in order to obtain a model with higher accuracy. One method for finding a good training set which can give good accuracy both in training and testing sets, is an instance exchange that starts with a random selected training set (Bolat & Yıldırım 2004).

**SVM models development**

For determining the best performance of SVM and selecting the best kernel function, the M5 model was predicted via SVM using various kernels. Table 3 shows the results of statistical parameters of different kernels for this model in the case of \( c = d/2 \). The results of Table 3 revealed that using a model with a kernel function of RBF led to better prediction accuracy in comparison to the other kernels. Therefore, the RBF kernel

---

**Table 2** | Developed models in the study

<table>
<thead>
<tr>
<th>Model</th>
<th>Input variables</th>
<th>Output variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>Fr, ( Vi/Vc )</td>
<td>( Sc/d )</td>
</tr>
<tr>
<td>M2</td>
<td>Fr, We, ( Vi/Vc )</td>
<td>( Sc/d )</td>
</tr>
<tr>
<td>M3</td>
<td>Fr, Re, ( Vi/Vc )</td>
<td>( Sc/d )</td>
</tr>
<tr>
<td>M4</td>
<td>Fr, Re, We</td>
<td>( Sc/d )</td>
</tr>
<tr>
<td>M5</td>
<td>Fr, Re, We, ( Vi/Vc )</td>
<td>( Sc/d )</td>
</tr>
<tr>
<td>M6</td>
<td>Fr, Re, We, ( b/d )</td>
<td>( Sc/d )</td>
</tr>
</tbody>
</table>

**Table 3** | The statistical parameters of SVM method with different kernel functions; model M5

<table>
<thead>
<tr>
<th>Kernel function</th>
<th>Train</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( R )</td>
<td>( DC )</td>
</tr>
<tr>
<td>Linear</td>
<td>0.855</td>
<td>0.791</td>
</tr>
<tr>
<td>Polynomial</td>
<td>0.945</td>
<td>0.898</td>
</tr>
<tr>
<td>RBF</td>
<td>0.996</td>
<td>0.987</td>
</tr>
<tr>
<td>Sigmoid</td>
<td>0.587</td>
<td>0.183</td>
</tr>
</tbody>
</table>
was selected as the core tool of SVM which was applied for the rest of the models. It should be noted that implementation of SVM requires the selection of three parameters, which are the constant $C$, $\varepsilon$, and kernel parameter $\gamma$, where $\gamma$ is a constant parameter of the RBF kernel. In the current study, according to Cherkassky & Ma (2004), optimization of these parameters has been done by a systematic grid search of the parameters using cross-validation on the training set. First, optimized values of $C$ and $\varepsilon$ for a specified $\gamma$ were obtained and then $\gamma$ was changed. Statistical parameters were used to find optimums. The Statistics parameters, via $\gamma$ values to find SVM optimums of the testing set for the model M5 of the case of $c = d/2$, are shown in Figure 2. In the same way, optimal parameters were obtained for all models.

Explicit equation developing for critical submergence depth ratio using GEP model

SVM models typically do not really represent the physics of a modeled process, and are just a device used to capture relationships between the relevant input and output variables; therefore, in this part, the GEP method was used to develop explicit equations for the critical submergence depth ratio. In this regard, for each bottom clearance state (i.e. $c = 0$ and $c = d/2$), the superior model (model M5) was run with the GEP model and the results were compared with the SVM model. As can be seen from Table 4, the results of the SVM models are more accurate than GEP models. However, the important point of the GEP model is that it is able to give the explicit expression of the relationship between the variables. The mathematical expressions of GEP for the best models are as follows:

\[
\begin{align*}
    c = 0: \quad & \frac{Sc}{d} = 0.16 \times \frac{Vi}{Vc} + \left(\frac{2.97 + Fr - 0.34 \times \frac{Vi}{Vc}}{0.34 \times Fr - 2.99 \times \frac{Vi}{Vc}}\right) \\
    & + \frac{Re - 2 \times We \left(\frac{Vi}{Vc}\right)^2}{Re} \\
    c = \frac{d}{2}: \quad & \frac{Sc}{d} = 0.1 \times Fr + \frac{2.26 \times \frac{Vi}{Vc}}{5.92 + Fr} + \frac{2 \times Fr \times (We + 4.99)}{0.387 \times Re + \frac{Vi}{Vc} - 4.94}
\end{align*}
\]

Figure 2 | Statistics parameters via $\gamma$ values to find SVM optimums of the testing set for model M5 for $c = d/2$.

Table 4 | Statistical parameters of the SVM and GEP models; model M5

<table>
<thead>
<tr>
<th>Condition</th>
<th>Method</th>
<th>R</th>
<th>DC</th>
<th>RMSE</th>
<th>R</th>
<th>DC</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c = 0$</td>
<td>SVM</td>
<td>0.991</td>
<td>0.981</td>
<td>0.183</td>
<td>0.985</td>
<td>0.971</td>
<td>0.199</td>
</tr>
<tr>
<td></td>
<td>GEP</td>
<td>0.985</td>
<td>0.973</td>
<td>0.203</td>
<td>0.966</td>
<td>0.941</td>
<td>0.233</td>
</tr>
<tr>
<td>$c = d/2$</td>
<td>SVM</td>
<td>0.996</td>
<td>0.987</td>
<td>0.181</td>
<td>0.988</td>
<td>0.976</td>
<td>0.197</td>
</tr>
<tr>
<td></td>
<td>GEP</td>
<td>0.988</td>
<td>0.973</td>
<td>0.197</td>
<td>0.977</td>
<td>0.959</td>
<td>0.228</td>
</tr>
</tbody>
</table>
RESULTS AND DISCUSSION

The SVM models

The obtained results of the SVM models are listed in Table 5 and shown in Figure 3(a). According to Table 5, it can be inferred that in both cases of \( c = 0 \) and \( c = d/2 \), the model M5 with input parameters \( Fr, Re, We, Vi/Vc \) led to the best results. Also, it can be seen that the model M3 with parameters \( Fr, Re, Vi/Vc \) approximately showed the same results. Considering the results of these two models, it seems that the Weber number did not have a significant impact on increasing the accuracy of the models. The results obtained from models M1 and M2 also clarified this issue. These two models showed that adding the \( We \) parameter to input combinations did not significantly improve the model accuracy. From the \( RMSE, R \), and \( DC \) viewpoints (i.e. the highest \( R \) and \( DC \) and lowest \( RMSE \)) it was found that adding the parameters \( Re \) and \( Vi/Vc \) caused an increment in model efficiency. Based on the results of Table 5, it can be seen that the developed models for the case of \( c = d/2 \) yielded better results in comparison with the case of \( c = 0 \). However, the SVM method in both cases performed successfully in modeling the critical submergence depth ratio in horizontal intakes. To investigate the impacts of different parameters of the SVM-best model on critical submergence depth, a sensitivity analysis was performed.

The significance of each parameter was evaluated by eliminating it from the input set. Figure 3(b) illustrates the results of sensitivity analysis of the model M5. This figure clearly shows that in predicting the critical submergence depth ratio, \( Vi/Vc \) is the most efficient parameter. The second important parameter is \( Fr \).

Investigating the impact of pipe diameter on critical submergence depth ratio

In the present research, the impact of pipe diameter was investigated for the critical submergence depth ratio using the data set of each intake pipe. The M3 and M5 models were selected for this assessment because of their performance in prediction of the critical submergence depth ratio. As can be seen from Table 6, increasing the pipe diameter from 4.25 to 30 cm decreased the model efficiency and the pipe with the results for the largest diameter (30 cm) showed the least accuracy.

Comparison of the best SVM and GEP models with the critical submergence classical equations

The experimental data of the test series were used to assess the capability of several existing formulas for the critical submergence depth ratio of horizontal intakes.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Model</th>
<th>Performance criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Train RMSE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( R )</td>
</tr>
<tr>
<td>( c = 0 )</td>
<td>M1</td>
<td>0.977</td>
</tr>
<tr>
<td></td>
<td>M2</td>
<td>0.976</td>
</tr>
<tr>
<td></td>
<td>M3</td>
<td>0.996</td>
</tr>
<tr>
<td></td>
<td>M4</td>
<td>0.745</td>
</tr>
<tr>
<td></td>
<td>M5</td>
<td>0.991</td>
</tr>
<tr>
<td></td>
<td>M6</td>
<td>0.913</td>
</tr>
<tr>
<td>( c = d/2 )</td>
<td>M1</td>
<td>0.986</td>
</tr>
<tr>
<td></td>
<td>M2</td>
<td>0.984</td>
</tr>
<tr>
<td></td>
<td>M3</td>
<td>0.998</td>
</tr>
<tr>
<td></td>
<td>M4</td>
<td>0.828</td>
</tr>
<tr>
<td></td>
<td>M5</td>
<td>0.996</td>
</tr>
<tr>
<td></td>
<td>M6</td>
<td>0.913</td>
</tr>
</tbody>
</table>
Three evaluation criteria (R, DC, and RMSE), were used as indications of the accuracy of the equations. A comparison was also performed among the best SVM and GEP models and those equations. The results of the comparisons are plotted in Figure 4. From the obtained results, among all equations, the Ahmad et al. (2008) equation provided a reasonable fit to the experimental data for both cases of $c = 0$ and $c = d/2$. It should be noted that Ahmad’s equation was derived for specific cases of $c = 0$ and $c = d/2$; therefore, under similar conditions it performed more successfully than other equations. The agreement between models and measurements was rather good for the limiting submergence depth ratio. Also it can be seen that Reddy & Pikford (1972), Swaroop (1973) and the Amphlett (1978) equations overestimated and Gurbuzdal’s (2009) equation underestimated the critical submergence depth data in most cases. However, the results obtained by the best SVM...
and GEP models were close to the measured data. The meta model approaches had the highest $R$ and $DC$ and the lowest $RMSE$ among all models, and this confirms the applicability of these methods as efficient approaches in modeling of the critical submergence depth ratio in horizontal intakes.

**CONCLUSION**

Modeling critical submergence depth has great importance since it affects the operation of hydraulic structures such as intake pipes. In the present study, the capability of the SVM, GEP and classical approaches were verified for predicting the critical submergence depth ratio in intakes with different hydraulic and intake pipe conditions. Two different data sets of laboratory experiments on horizontal intakes in the cases of $c = 0$ and $c = d/2$ were used. For each case of bottom clearance, several models were developed and tested with the SVM method. According to the results, it was found that in both cases, of $c = 0$ and $c = d/2$, the model with input variables $Fr$, $Re$, $We$, $Vi/Vc$ led to the best results. Also, it was observed that the model with input parameters $Fr$, $Re$, $Vi/Vc$ approximately represented the same results. This result showed that using the $We$ parameter as the input variable did not significantly improve the model accuracy. It was found that using the parameters $Re$ and $Vi/Vc$ caused an increment in model efficiency. Based on the results, it was observed that the developed models for the case of $c = d/2$ yielded better results compared to the case of $c = 0$. According to the results obtained from sensitivity analysis, it was found that $Vi/Vc$ had the most important role in prediction of critical submergence depth ratio compared to other parameters. Also, an explicit equation was developed for each cases of $c = 0$ and $c = d/2$.

A comparison was also made between the SVM, GEP and some critical submergence semi-empirical equations from the literature. The obtained results proved the superior performance of the meta model approaches over all of the semi-empirical equations in prediction of the critical submergence depth ratio in intakes. Pipe diameter also affected the critical submergence depth ratio of intakes. Results revealed that increasing the pipe diameter decreased the efficiency of the models and a pipe with 42.5 mm diameter led to the best results. Since SVM and GEP are data-driven models, it is suggested to investigate the sufficiency of the proposed models via SVM and GEP for data ranges outside this study to find the merits of the models to estimate critical submergence depth ratio in horizontal intakes.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Model</th>
<th>Train</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$R$</td>
<td>$DC$</td>
</tr>
<tr>
<td>$d = 4.25$ cm</td>
<td>M3</td>
<td>0.996</td>
<td>0.960</td>
</tr>
<tr>
<td></td>
<td>M5</td>
<td>0.991</td>
<td>0.971</td>
</tr>
<tr>
<td>$d = 6.25$ cm</td>
<td>M3</td>
<td>0.981</td>
<td>0.942</td>
</tr>
<tr>
<td></td>
<td>M5</td>
<td>0.976</td>
<td>0.956</td>
</tr>
<tr>
<td>$d = 10$ cm</td>
<td>M3</td>
<td>0.964</td>
<td>0.920</td>
</tr>
<tr>
<td></td>
<td>M5</td>
<td>0.959</td>
<td>0.943</td>
</tr>
<tr>
<td>$d = 14.4$ cm</td>
<td>M3</td>
<td>0.804</td>
<td>0.785</td>
</tr>
<tr>
<td></td>
<td>M5</td>
<td>0.810</td>
<td>0.794</td>
</tr>
<tr>
<td>$d = 25$ cm</td>
<td>M3</td>
<td>0.744</td>
<td>0.737</td>
</tr>
<tr>
<td></td>
<td>M5</td>
<td>0.741</td>
<td>0.746</td>
</tr>
<tr>
<td>$d = 30$ cm</td>
<td>M3</td>
<td>0.759</td>
<td>0.700</td>
</tr>
<tr>
<td></td>
<td>M5</td>
<td>0.759</td>
<td>0.734</td>
</tr>
</tbody>
</table>
REFERENCES


---

**Figure 4** Comparison of prediction from proposed equations and the SVM and GEP for $c = 0$ and $c = d/2$; (a) scatter plot and (b) performance criteria.


Swaroop, R. 1973 *Vortex Formation at Intakes*. ME dissertation, CED, University of Roorkee, India.


First received 14 April 2018; accepted in revised form 29 August 2018. Available online 10 September 2018.