Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2014, Article ID 985659, 15 pages
http://dx.doi.org/10.1155/2014/985659

Research Article

Data Mining of the Thermal Performance of Cool-Pipes in Massive Concrete via In Situ Monitoring

Zheng Zuo,[1] Yu Hu,[1] Qingbin Li,[1] and Liyuan Zhang[2]

[1] *State Key Laboratory of Hydroscience and Engineering, Department of Hydraulic Engineering, Tsinghua University, Beijing 100084, China*
[2] *Planning and Design Institute of Water Transportation, Beijing 100007, China*

Correspondence should be addressed to Yu Hu; yu-hu@mail.tsinghua.edu.cn

Received 27 December 2013; Revised 30 January 2014; Accepted 4 February 2014; Published 5 May 2014

Academic Editor: Ting-Hua Yi

Embedded cool-pipes are very important for massive concrete because their cooling effect can effectively avoid thermal cracks. In this study, a data mining approach to analyzing the thermal performance of cool-pipes via in situ monitoring is proposed. Delicate monitoring program is applied in a high arch dam project that provides a good and mass data source. The factors and relations related to the thermal performance of cool-pipes are obtained in a built theory thermal model. The supporting vector machine (SVM) technology is applied to mine the data. The thermal performances of iron pipes and high-density polyethylene (HDPE) pipes are compared. The data mining result shows that iron pipe has a better heat removal performance when flow rate is lower than 50 L/min. It has revealed that a turning flow rate exists for iron pipe which is 80 L/min. The prediction and classification results obtained from the data mining model agree well with the monitored data, which illustrates the validness of the approach.

1. Introduction

Temperature is an important factor that affects massive concrete structures' health [1]. When the volume of a concrete structure is so large that it generates excessive heat and associated expansions [2], which may lead to a risk of temperature cracks, we can name it "massive concrete" [3, 4]. Postcooling measures, such as embedding cool-pipes, are commonly included in the design to mitigate such effects.

The first significant application case of cool-pipes was in a small experimental section of the Owyhee Dam in Oregon in 1931, which was the tallest dam in the world at the time of its completion [5]. The Bureau of Reclamation believed the Owyhee experiment was successful and extended its application in the construction of the Hoover Dam in 1936. In every slab coil of 1-inch-thick-walled steel pipe running water was embedded. More than 582-miles- (937 km) long cool-pipes were placed within the concrete [5–7]. The success of bedding steel cool-pipes in the Hoover Dam made the method as famous as the dam itself. The method has thus been utilized in a number of large-scale massive concrete projects [8–10] over the last 70 years.

Several researchers have carried out studies on the thermal effect of cool-pipes [7, 8, 11–16]. The Bureau of Reclamation first studied a theory solution of long iron pipes [7]. Zhu and Cai have made a great effort to conduct calculation for cool-pipes by using finite element methods [12, 14]. Much research has focused on the calculation model and method of cool-pipes based on assumptions (i.e., the pipe is straight; heat transfer between pipe and concrete is constant) that make the problem simple and solvable. However, during the construction of a massive concrete structure, the heat transfer process is very sophisticated. The cool-pipes are embedded in multiple U-shapes and the heat transfer between fluid (water) and solid (pipe) is very complicated. A recent study pointed out that the convection coefficient cannot be considered constant [13]. So study of in situ thermal performance of cool-pipes is needed, which can help us acknowledge the thermal mechanism of cool-pipes well.

To the knowledge of the authors, few studies have reported the statistics of the thermal performance of cool-pipes in actual concrete engineering studies, based on real measured data, especially for large-scale projects. An important reason for this scarcity is the lack of adequate monitoring

measurements and detailed field data analysis. Fortunately, a 285 m high arch dam project under construction gives a perfect opportunity to conduct the work. During the construction period of the dam, the temperature and the flow of every cool-pipe in every slab are monitored. A large amount of data is produced; we implement data mining of thermal performance of cool-pipes in this paper.

Many mathematical methods can be employed in data mining [17–27]. Support vector machine (SVM) is a novel machine learning method developed on the basis of statistical learning theory [28, 29]. It is one of the most popular methods for pattern recognition, regression estimation, and time series prediction problems [30]. A typical process of SVM training can be summarized into four steps as follows [31]: selecting kernel function, selecting the smoothing parameters of the kernel function, choosing penalty factor, and solving the quadratic problem. It has been shown that compared with other machine learning methods, such as artificial neutral network, SVM not only is easier to use [32] but also leads to higher accuracy and robustness. For specific cases, it converges 10 to 100 times faster in training [33, 34].

Our paper is organized as follows: Section 2 describes the delicate thermal monitoring program, including the temperature measurement of concrete, inlet-outlet temperature, and flow of cool-pipes. Section 3 shows the establishment of the thermal model to guide the data mining. Section 4 describes the SVM method and its relevant concepts. Section 5 presents the details of the monitoring database and a developed automatic data mining program flow. Section 6 provides the data mining results, including the calibrated parameters, numerical model to predict outlet cool-pipe temperature, the evaluation of the thermal performance, and pipe material classification. And Section 7 provides our conclusions.

2. In Situ Monitoring

2.1. Engineering Description. Xiluodu Dam is located in the middle of the Jinsha River, Yunnan Province in Southwest China. It is designed as an arch dam with a maximum height of 285.5 m and a crest length at the top of 680 m [8]. Based on the project plan, the amount of concrete utilized was estimated to be 1315×10^4 m^3 during the construction period from 2009 to 2014. The rear view of the dam being constructed is shown in Figure 1.

Across the river flow direction, the dam is divided by 31 monoliths [35]. Vertically, each monolith consists of a number of slabs poured at different times. Figure 2 shows the slab zoning map. Each slab is individually poured in a formed hexahedron block with a vertical thickness of 3.0 m or 1.5 m. The first slab was poured on March 27, 2009, and 2158 slabs altogether had been placed by December 2013. These slabs are the main study objects of the research presented in this paper.

Cool-pipes were horizontally embedded in every slab to control the concrete temperature properly. Circulating water was supplied via two large cooling water units, which could stabilize the supply temperature at a constant level on demand. Considering that water temperature increases along the flow direction, the flow direction was changed twice a day to make the concrete temperature field more uniform. Iron pipes and high-density polyethylene (HDPE) pipes are employed in this project.

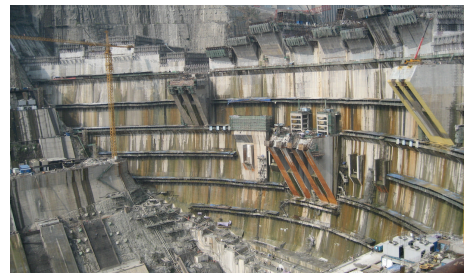


Figure 1: Picture of the dam under construction.

2.2. In Situ Thermal Monitoring Setup

2.2.1. Concrete Temperatures. Since the start of the dam construction, digital temperature sensors were installed during the pouring of every concrete slab. Staggered cool-pipes were vertically embedded every 1.5 m (i.e., two layers in a 3.0 m thick slab). The sensors were vertically positioned in the middle of two cool-pipe layers, as shown in Figure 3. Concrete temperature was measured and recorded for four or five times per day.

2.2.2. Cooling Water Temperatures and Flows. Figure 4 shows the schematic layout of the cool-pipes in one slab. Faucets were connected to both ends of each cool-pipe. By letting water run from the tap, the inlet and outlet temperatures of the cool-pipes were measured using a mercury thermometer and recorded four or five times per day. At one end of every pipe, one water meter was installed (Figure 4). The cooling water flow rate was measured and recorded.

3. Thermal Model

A simplified thermal model of cool-pipe and concrete is sketched in Figure 5. Water flows through a pipe whose wall clings to the concrete. Heat transfer per unit time is discussed below.

The heat energy loss from water per unit length, which is denoted as Q_w, can be expressed as follows:

$$Q_w = \rho_w C_{p,w} dT_w u A_p, \tag{1}$$

where ρ_w is the density of the water, $C_{p,w}$ is the specific heat capacity of water, T_w is the water temperature, dT_w is the temperature difference along the pipe (per unit length), u is the water velocity, and A_p is cross section area of the pipe.

The heat energy income for the concrete per unit length, which is denoted as Q_c, can be calculated by Newton's law of cooling, which is also called the Robin boundary condition

$$Q_c = k\left(T_c - T_w\right) dA_c, \tag{2}$$

where k is the heat transfer coefficient between water and concrete, T_c is the concrete temperature, and A_c is the contact area of pipe wall per unit length.

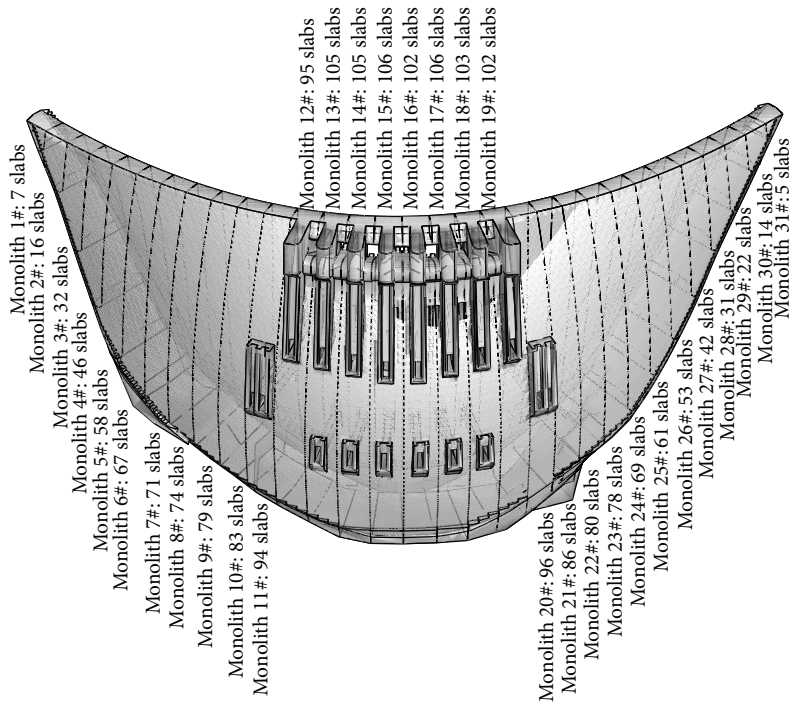


Figure 2: Zoning map of slabs.

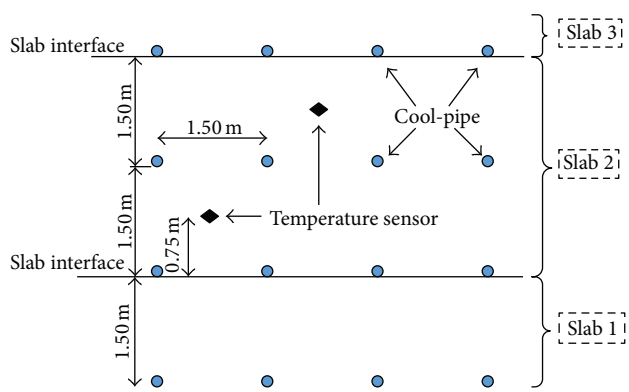


Figure 3: Staggered embedding configuration of cool-pipes and temperature sensor installed locations.

A_p and A_c can be further expressed by the geometry property of the pipe

$$A_p = \frac{1}{4}\pi D^2$$
$$dA_c = \pi D dL, \tag{3}$$

where D is the diameter of the pipe and L is the length of the pipe.

Propose that T_c is constant along the pipe and let ΔT represent $(T_c - T_w)$. The differential of T_w can be regarded as the differential of ΔT:

$$d\Delta T = d\left(T_c - T_w\right) = -dT_w. \tag{4}$$

Based on the conservation of energy, $Q_w = Q_c$. We can obtain that

$$\frac{1}{4}\pi D^2 C_{p,w}\rho_w u\left(-d\Delta T\right) = k\Delta T\pi D dL. \tag{5}$$

Integrating (5) along the whole pipe gives

$$\ln\frac{\Delta T_{\text{out}}}{\Delta T_{\text{in}}} = \frac{-4k}{C_{p,w}\rho_w D}\frac{L}{u}, \tag{6}$$

where subscripts in and out represent the inlet and outlet of the pipe, respectively. If the inlet status of the water pipe is known, the outlet water temperature can be determined by the following equation:

$$\left(T_c - T_{w\text{-out}}\right) = \left(T_c - T_{w\text{-in}}\right)\exp\left(\frac{-4k}{C_{p,w}\rho_w D}\frac{L}{u}\right). \tag{7}$$

Temperature rise along the pipe $T_{w\text{-out}} - T_{w\text{-in}}$ is an important indicator which directly gives the total energy loss of water. It can be obtained by (7):

$$\begin{aligned} T_{w\text{-out}} - T_{w\text{-in}} &= -\left(T_c - T_{w\text{-in}}\right) + \left(T_c - T_{w\text{-out}}\right) \\ &= \left(T_c - T_{w\text{-in}}\right)\left[\exp\left(\frac{-4k}{C_{p,w}\rho_w D}\frac{L}{u}\right) - 1\right]. \end{aligned} \tag{8}$$

The physical properties of the water, $C_{p,w}$ and ρ_w , are constant in usual environment conditions. k is determined by the pipe material. In engineering cases, the diameter of a pipe material is always fixed. In China's dam projects D for HDPE pipe and iron pipe are 32 mm. Let T_w^* denote

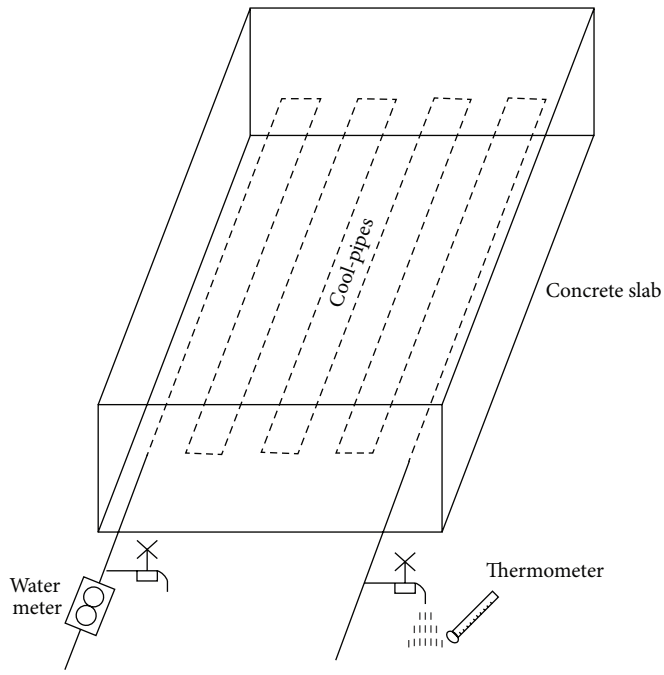


Figure 4: Schematic diagram of the cool-pipes layout.

$T_{w\text{-out}} - T_{w\text{-in}}$ and T_w^* can be seen as the function of some parameter as follows:

$$T_w^* = f\left(\Delta T_{\text{in}}, u, k, L\right). \tag{9}$$

With T_w^* known, the absorbed heat power by cool-pipe, which is denoted as P, can be calculated by the following equation:

$$P = \rho_w C_{p,w} T_w^* q_w, \tag{10}$$

where $q_w = uA_p$ is the flow rate of the cooling water.

Actually, the thermal process between the concrete and the cool-pipe is very complex, which has some differences with the theory model proposed above.

(1) Different pipe materials have different k values. The value of k is hard to determine and may change with flow states.

(2) Concrete temperature is not constant along the pipe.

(3) Pipe is embedded in multiple U-shapes like a long crawling snake, not in a straight line.

But (9) can give a good guide for us to conduct data mining, which tells us that T_w^* is relevant with the inlet temperature difference (ΔT_{in}), cooling water flow speed (u), pipe length (L), and the pipe material (k).

4. Data Mining Model

4.1. SVM Classification. For a classification problem of two classes, a set of sample data S is given as

$$S = \left\{\left(\mathbf{x}_1, y_1\right), \left(\mathbf{x}_2, y_2\right), \left(\mathbf{x}_3, y_3\right), \ldots, \left(\mathbf{x}_n, y_n\right)\right\}, \tag{11}$$

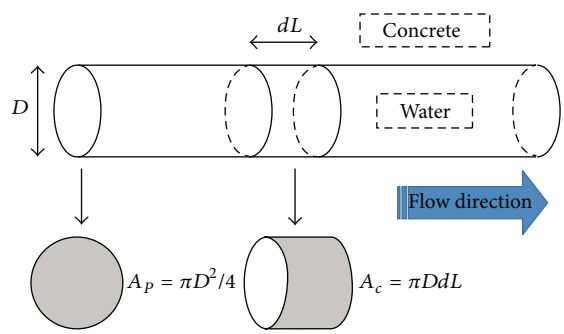


Figure 5: Schematic diagram of cool-pipe.

where $\mathbf{x}_i \in \mathbf{R}^d$ $(i = 1, 2, \ldots n)$ is a d-dimensional input vector and $y_i \in \{-1, 1\}$ $(i = 1, 2, \ldots n)$ is the classification label. The object is to find a function $f(x) \in \mathbf{R}^d$ that can separate the two classes with a margin as large as possible. Linear estimation function to achieve the goals can be expressed as follows:

$$f(\mathbf{x}) = \mathbf{w} \cdot \mathbf{x} + b \quad \text{with } \mathbf{w} \in \mathbf{R}^d,\ b \in \mathbf{R}, \tag{12}$$

where $\mathbf{w}$ and b are weight factors to be determined. Many possible linear classifiers exist that can separate the data. As shown in Figure 6, H^* is the classifier plane, and $H1$ and $H2$ are the planes that cross the nearest data point and parallel to the classifier plane (the sample points on $H1$ and $H2$ can be called support vector). The basic idea of SVM is to find an optimal plane that maximizes the classification margin

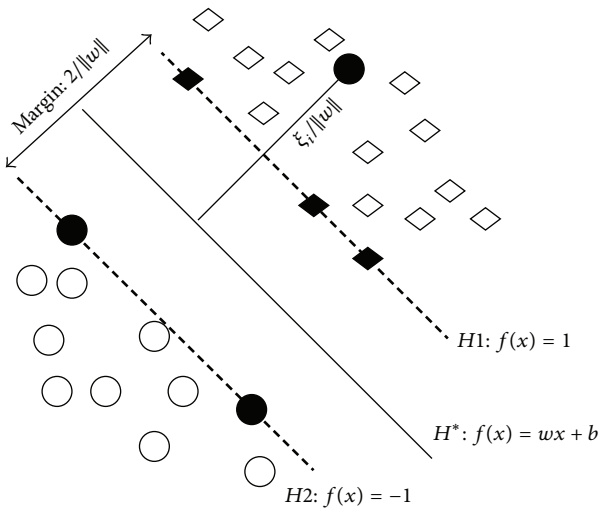


Figure 6: Optimal hyperplane for classification.

between $H1$ and $H2$. One way to ensure this is to minimize the norm of $\mathbf{w}$ [36], that is, $\|\mathbf{w}\|^2$:

$$\begin{aligned} &\text{Minimize} \quad \frac{1}{2}\|\mathbf{w}\|^2 \\ &\text{subject to:} \quad \begin{cases} f(\mathbf{x}_i) > 0, & \text{when } y_i = 1,\ i = 1,2,\ldots,n \\ f(\mathbf{x}_i) < 0, & \text{when } y_i = -1,\ i = 1,2,\ldots,n. \end{cases} \end{aligned} \tag{13}$$

The constraint condition of (13) can be expressed as

$$y_i f(\mathbf{x}_i) \geq 1, \quad i = 1,2,\ldots,n. \tag{14}$$

In particular cases, data of different labels cannot be clearly separated. A set of slack variable $\{\xi_i\}$ is introduced that measure the amount by which the constraints are violated, as shown in Figure 6. Equation (13) is then transformed into the following form:

$$\begin{aligned} &\text{Minimize} \quad \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n}\xi_i \\ &\text{subject to:} \quad \begin{cases} y_i f(\mathbf{x}_i) \geq 1 - \xi_i, & i = 1,2,\ldots,n \\ \xi_i \geq 0, & i = 1,2,\ldots,n, \end{cases} \end{aligned} \tag{15}$$

where C is a penalty constant chosen a priori, which determines the cost of constraint violation. As is well known, (15) is a quadratic programming optimization problem. The method of Lagrange multipliers is a good strategy to find the solution of an optimization problem considering the constraint conditions. Solving (15) can be converted into finding the saddle point of the following Lagrange function:

$$\begin{aligned} L(\mathbf{w}, b, \alpha) = {} & \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n}\xi_i \\ & - \sum_{i=1}^{n}\alpha_i\left\{\left[y_i f(\mathbf{x}_i)\right] - 1 + \xi_i\right\} - \sum_{i=1}^{n}\eta_i\xi_i, \end{aligned} \tag{16}$$

where $\alpha_i \geq 0, \eta_i \geq 0$ is the Lagrange factor. The saddle points can be obtained by

$$\begin{aligned} \frac{\partial L}{\partial \mathbf{w}} &= 0 \Longrightarrow \mathbf{w} = \sum_{i=1}^{n}\alpha_i y_i \mathbf{x}_i \\ \frac{\partial L}{\partial b} &= 0 \Longrightarrow \sum_{i=1}^{n}\alpha_i y_i = 0 \\ \frac{\partial L}{\partial \xi_i} &= 0 \Longrightarrow C - \alpha_i - \eta_i = 0 \\ \frac{\partial L}{\partial \alpha_i} &= 0 \Longrightarrow \left[y_i f(\mathbf{x}_i)\right] - 1 + \xi_i = 0. \end{aligned} \tag{17}$$

Substituting (17) into (16) yields the dual optimization problem:

$$\begin{aligned} &\max_{\alpha_i} \quad Q(\alpha_i) = \sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(\alpha_i y_i \mathbf{x}_i)^T \cdot (\alpha_j y_j \mathbf{x}_j) \\ &\text{subject to:} \quad 0 \leq \alpha_i \leq C, \quad i = 1,2,\ldots,n \\ &\qquad \sum_{i=1}^{n}\alpha_i y_i = 0. \end{aligned} \tag{18}$$

The solution α_i^* for (18) can determine the optimal parameter $\mathbf{w}^*$:

$$\mathbf{w}^* = \sum_{i=1}^{n}\alpha_i^* y_i \mathbf{x}_i. \tag{19}$$

With $\mathbf{w}^*$, the optimal parameter b^* can be obtained from the boundary condition in (17), which can be written as follows:

$$\left[y_i(\mathbf{w}^*\mathbf{x}_i + b^*)\right] - 1 + \xi_i = 0. \tag{20}$$

The optimal hyperplane decision function can be finally written as

$$f(\mathbf{x}) = \operatorname{sgn}(\mathbf{w}^* \cdot \mathbf{x} + b^*). \tag{21}$$

4.2. SVM Regression. Regression can also be performed by SVM. The goal is to find a function $f(\mathbf{x})$ that can achieve the targets for all data points and is as flat as possible. By flat, we mean a small $\mathbf{w}$, which can generate a problem as follows:

$$\begin{aligned} &\text{Minimize} \quad \frac{1}{2}\|\mathbf{w}\|^2 \\ &\text{subject to:} \quad \left|y_i - f(\mathbf{x})\right| \leq \varepsilon, \quad i = 1,2,\ldots,n, \end{aligned} \tag{22}$$

where ε is a loss precision, which is employed to create an intensive band near the regression line, as shown in Figure 7. A double-set of slack variable $\{\xi_i^{\text{up}}\}\{\xi_i^{\text{down}}\}$ is introduced to

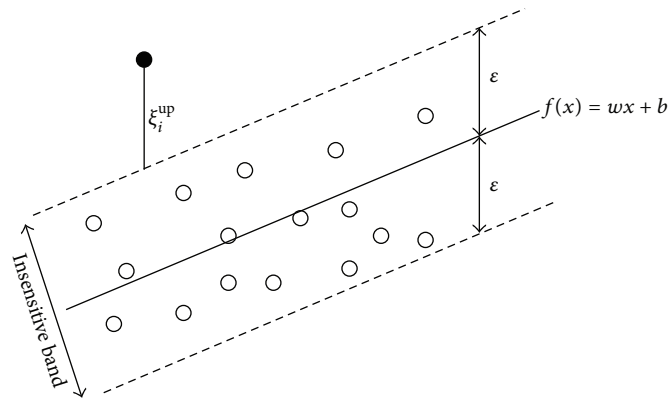


Figure 7: Intensive band of regression.

make the constraint more feasible. Equation (13) is then transformed into the following form:

$$\begin{aligned} &\text{Minimize} \quad \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n}\left(\xi_i^{\text{up}} + \xi_i^{\text{down}}\right) \\ &\text{subject to:} \quad \begin{cases} y_i - f(\mathbf{x}) \le \varepsilon + \xi_i^{\text{up}}, & i = 1, 2, \ldots, n \\ f(\mathbf{x}) - y_i \le \varepsilon + \xi_i^{\text{down}}, & i = 1, 2, \ldots, n \\ & \xi_i^{\text{up}}, \xi_i^{\text{down}} > 0. \end{cases} \end{aligned} \tag{23}$$

Similar to classification, (23) can be converted into a Lagrange function

$$\begin{aligned} L(\mathbf{w}, b, \alpha) = {} & \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n}\left(\xi_i^{\text{up}} + \xi_i^{\text{down}}\right) \\ & - \sum_{i=1}^{n}\alpha_i^{\text{up}}\left\{\xi_i^{\text{up}} + \varepsilon - \left[y_i - f(\mathbf{x})\right]\right\} \\ & - \sum_{i=1}^{n}\alpha_i^{\text{down}}\left\{\xi_i^{\text{down}} + \varepsilon - \left[f(\mathbf{x}) - y_i\right]\right\} \\ & - \sum_{i=1}^{n}\left(\eta_i^{\text{up}}\xi_i^{\text{up}} + \eta_i^{\text{down}}\xi_i^{\text{down}}\right), \end{aligned} \tag{24}$$

where $\alpha_i^{\text{up}} \ge 0$, $\alpha_i^{\text{up}} \ge 0$, $\eta_i^{\text{up}} \ge 0$, $\eta_i^{\text{down}} \ge 0$ is the Lagrange factor. The saddle point of (24) is the solution of (23), which can be obtained by

$$\begin{aligned} \frac{\partial L}{\partial \mathbf{w}} = 0 &\Longrightarrow \mathbf{w} = \sum_{i=1}^{n}\left(\alpha_i^{\text{down}} - \alpha_i^{\text{up}}\right)\mathbf{x}_i \\ \frac{\partial L}{\partial b} = 0 &\Longrightarrow \sum_{i=1}^{n}\left(\alpha_i^{\text{up}} - \alpha_i^{\text{down}}\right) = 0 \\ \frac{\partial L}{\partial \xi_i^{\text{up}}} = 0 &\Longrightarrow C - \alpha_i^{\text{up}} - \eta_i^{\text{up}} = 0 \\ \frac{\partial L}{\partial \xi_i^{\text{down}}} = 0 &\Longrightarrow C - \alpha_i^{\text{down}} - \eta_i^{\text{down}} = 0 \\ \frac{\partial L}{\partial \alpha_i^{\text{up}}} = 0 &\Longrightarrow \xi_i^{\text{up}} + \varepsilon - \left[y_i - f(\mathbf{x})\right] = 0. \end{aligned} \tag{25}$$

Substituting (25) into (24) yields the dual optimization problem

$$\begin{aligned} \max_{\alpha_i^{\text{up}}, \alpha_i^{\text{down}}} \quad & Q(\alpha_i) = -\frac{1}{2}\sum_{i=1}^{n}\sum_{i=1}^{n}\left(\alpha_i^{\text{up}} - \alpha_i^{\text{down}}\right) \\ & \times \left(\alpha_j^{\text{up}} - \alpha_j^{\text{down}}\right)\mathbf{x}_i^T\mathbf{x}_j \\ \text{subject to:} \quad & \alpha_i^{\text{up}}, \alpha_i^{\text{down}} \in [0, C] \\ & \sum_{i=1}^{d}\left(\alpha_i^{\text{up}} - \alpha_i^{\text{down}}\right) = 0. \end{aligned} \tag{26}$$

The solution $\alpha_i^{\text{up},*}$ and $\alpha_i^{\text{down},*}$ for (25) can determine the optimal parameter $\mathbf{w}^*$,

$$\mathbf{w}^* = \sum_{i=1}^{n}\left(\alpha_i^{\text{up},*} - \alpha_i^{\text{down},*}\right)\mathbf{x}_i. \tag{27}$$

With $\mathbf{w}^*$, the optimal parameter b^* can be calculated from the boundary condition in (25). Then the optimal regression function can be finally obtained.

4.3. Kernel Function. Actually the algorithms in Sections 4.1 and 4.2 are specified in linear space. Either in classification or regression problem, the input data $\mathbf{x}$ is projected into y by function f, which is built during training. Furthermore, $f(\mathbf{x}) = \mathbf{w}\cdot\mathbf{x} + b$ can be expanded into $f(\mathbf{x}) = \mathbf{w}\cdot\varphi(\mathbf{x}) + b$ to solve more complicated problems, such as nonlinear

```
Create mesh grid of ε, C, γ.
for i, j, k in the searching grid do
    for m = 1 to V do
        make S_m as the testing data set, and S_i (i = 1, ..., m − 1, m + 1, ..., n) as the training data set
        do SVM with parameters of (ε_i, C_j, γ_k), calculate the accuracy as a_m
    end loop m
    obtain the highest a_m, and record it as A_ijk
end loop i, j, k
obtain the highest A_ijk
```

Algorithm 1: Grid-search V-fold CV method.

problems. Thus, formulas (18) and (26) can be rewritten as follows:

$$\max_{\alpha_i} \quad Q(\alpha_i) = \sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\varphi(\mathbf{x}_i)^T \cdot \varphi(\mathbf{x}_j)$$

$$\text{subject to:} \quad 0 \le \alpha_i \le C, \quad i = 1, 2, \ldots, n \tag{28}$$

$$\sum_{i=1}^{d}\alpha_i y_i = 0.$$

$K(x_i, x_j) = \varphi(\mathbf{x}_i)^T \cdot \varphi(\mathbf{x}_j)$ is defined as the kernel function. A kernel function provides a way to project data into a higher dimensional space by manipulating in its original space. The kernel function is chosen beforehand; then, a subset of vectors $\mathbf{x}_i$ is determined during training, which is called a "support vector." The most commonly adopted kernel functions include polynomial function, radial basic function (RBF), exponential function, and ANOVA function. The RBF is chosen in this paper for it is suitable for nonlinear problems, which can be written as

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\gamma\left\|\mathbf{x}_i - \mathbf{x}_j\right\|^2\right). \tag{29}$$

4.4. Parameter Calibration. As can be deduced, parameters γ in RBF and ε and C in SVM greatly influence the classification/regression result. Proper parameters are significant in obtaining satisfying results; thus, cross validation (CV) is introduced. CV was first introduced to fix the overoptimistic result [37] derived from training an algorithm and evaluating its statistical performance on the same data [38, 39].

V-fold CV is a common method used in many statistics problems, such as biometrics [40], computational physics [41], and artificial intelligence [42]. In this method, the data set is divided into several subsets:

$$S = S_1, \ldots, S_V, \tag{30}$$

where $V \geqslant 1$ is a integer. Every subset becomes the test set once, so that the training results get closer to the distribution of original data without much increase in computing cost.

Combined with "grid-search," V-fold CV is adopted in this paper to identify good parameter combination of (ε, C, and γ) that makes the SVM model work well. The process can be summarized in Algorithm 1.

5. Data Processor and Programming

5.1. Database. We established a database to store quantities of monitoring data. Microsoft SQL Server (in Version 10.50.1600) is applied as the database engine. All the monitored data introduced in Section 2 were stored in the database.

5.2. Data Cleaning. Data cleaning is carried out as a preprocedure of data mining; some work is done as follows.

(1) Monitored data that has a character of instrument failure is filtered out.

(2) The flow direction is changed every day, and each pair of $T_{w\text{-out}}$ and $T_{w\text{-in}}$ are automatically compared to guarantee that $T_{w\text{-out}} > T_{w\text{-in}}$.

(3) Pipe length is not studied as a variable in the data mining process, for most pipes are 220 m long in the engineering. So for all the pipes, the length is considered as a constant and not selected during the data mining procedure.

(4) HDPE pipe's diameter is identical as, in the iron pipe in the engineering, flow rate q_w is adopted in data mining instead of the water speed u for convenience.

5.3. Data Extractor. A set of programs is developed in Python language to seek the database and export data samples. Given the massive stored data, the program is developed in parallel to accelerate the processing speed.

Taking data mining on T_w^* as an example, the relevant data include pipe material, inlet and outlet water temperature, water flow, and the concrete temperature. They are stored in three different data tables, the structures of which are shown in Figure 8. Data extraction is carried out based on Algorithm 2.

5.4. Data Mining Implementation. The SVM method is developed in Python programs to model the thermal performance. Based on the integration of database, data extractor, and SVM model engine, the whole program flow of data mining is shown in Figure 9.

```
(1)  Select* from TbPipeInfo where Material = "HDPE", record them as Pipes.
(2)  for each Pipe in Pipes do:
(3)          Select* from TbSensorInfo where SlabId = Pipe.SlabId, record them as Sensors.
(4)          Select* from TbCoolRecord where Id = pipe.PipeId, record them as Rows.
(5)          for each Row in Rows do:
(6)                  for each Sensor in Sensors do:
(7)                          Select Temperature from TbTemRecord
                                     where TemTime nearest* to Row.CoolTime
                                           and SensorId = Sensor.SensorId
(8)                  end loop Sensors
(9)                  T_c = average(sum(Temperature))
(10)                 output T_c, T_w-in, T_w-out, Flow
(11)         end loop Rows
(12) end loop Pipes
```

Note *: all the temperature sensors' reading in the slab are collected. If the nearest time difference is larger than a threshold, the data of the time will not be used.

Algorithm 2: Data extraction flow (HDPE pipes as examples).

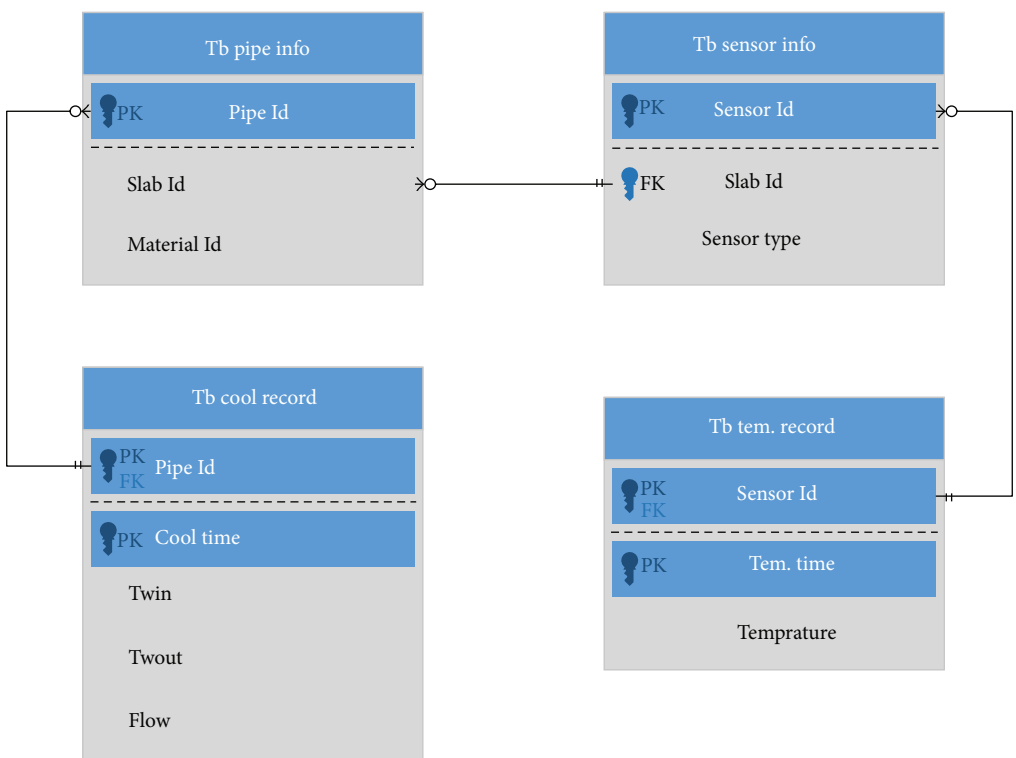


Figure 8: Data structures.

6. Results and Discussions

6.1. Parametric Analysis. The SVM coefficients including ε, C, and γ need to be determined before utilizing the method. To the knowledge of authors, no determinate function exists that directly obtains the optimal coefficients for a model. So, parametric analysis is carried out to calibrate the model.

Taking prediction of outlet temperature as a study case, T_w^* is employed as the target variable whereas ΔT_{in} and q_w are adopted as the observed variables. The data of cool-pipes embedded in monolith 15# is used as the training data set, and monolith 16# is used as the testing data set. The grid search method integrated with V-fold CV is used here to find optimal parameters, as mentioned in Section 4.4. The mean squared error is applied to evaluate the model performance, which is expressed as

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left[f\left(\mathbf{x}_i\right) - y_i\right]^2, \tag{31}$$

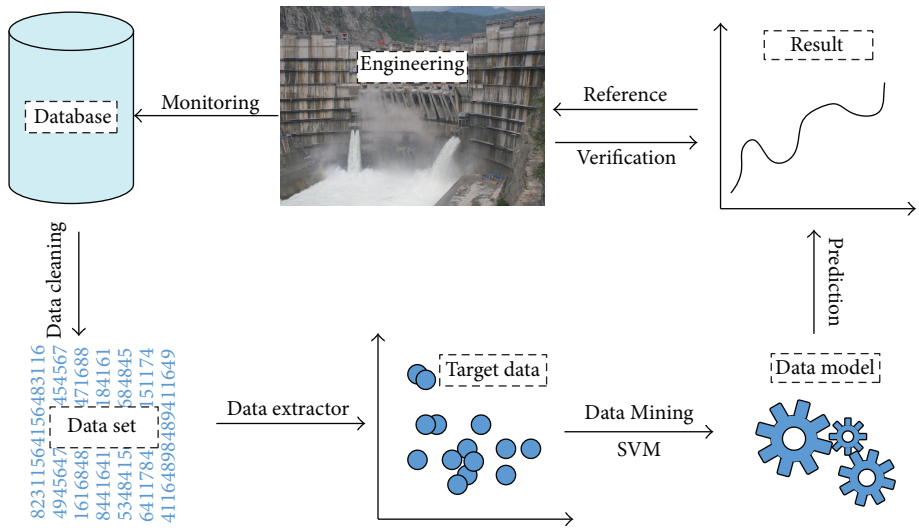


Figure 9: Program flow chart.

Table 1: Optimal SVM coefficients.

Material	ε	Training data set			Testing data set		
		C	γ	MSE	C	γ	MSE
HDPE	0.1	32	32	0.00915854	0.25	8	0.006821
	0.01	32	32	0.00917	2	8	0.006481
	0.001	16	32	0.00916964	2	8	0.006485
	0.0001	32	32	0.00917025	8	8	0.006485
Iron	0.1	32	32	0.00616273	8	0.5	0.003472
	0.01	32	32	0.00585564	0.25	4	0.003306
	0.001	16	32	0.00586528	0.5	4	0.003299
	0.0001	16	32	0.00587013	0.0625	8	0.003301
Optimal results:							
HDPE		$\varepsilon = 0.01$		$C = 2$		$\gamma = 8$	
Iron		$\varepsilon = 0.001$		$C = 0.5$		$\gamma = 4$	

where $\mathbf{x}$ stands for the combination of ΔT_{in} and q_w and y stands for T_w^*. A small MSE indicates a good fitness of the numerical model. The ranges of C and γ are both set to be $(2^{-5} \sim 2^5)$. The results of MSE of different ε are drawn in logarithmic plots with $\log_2(C)$ and $\log_2(\gamma)$ as the x- and y-axis, respectively.

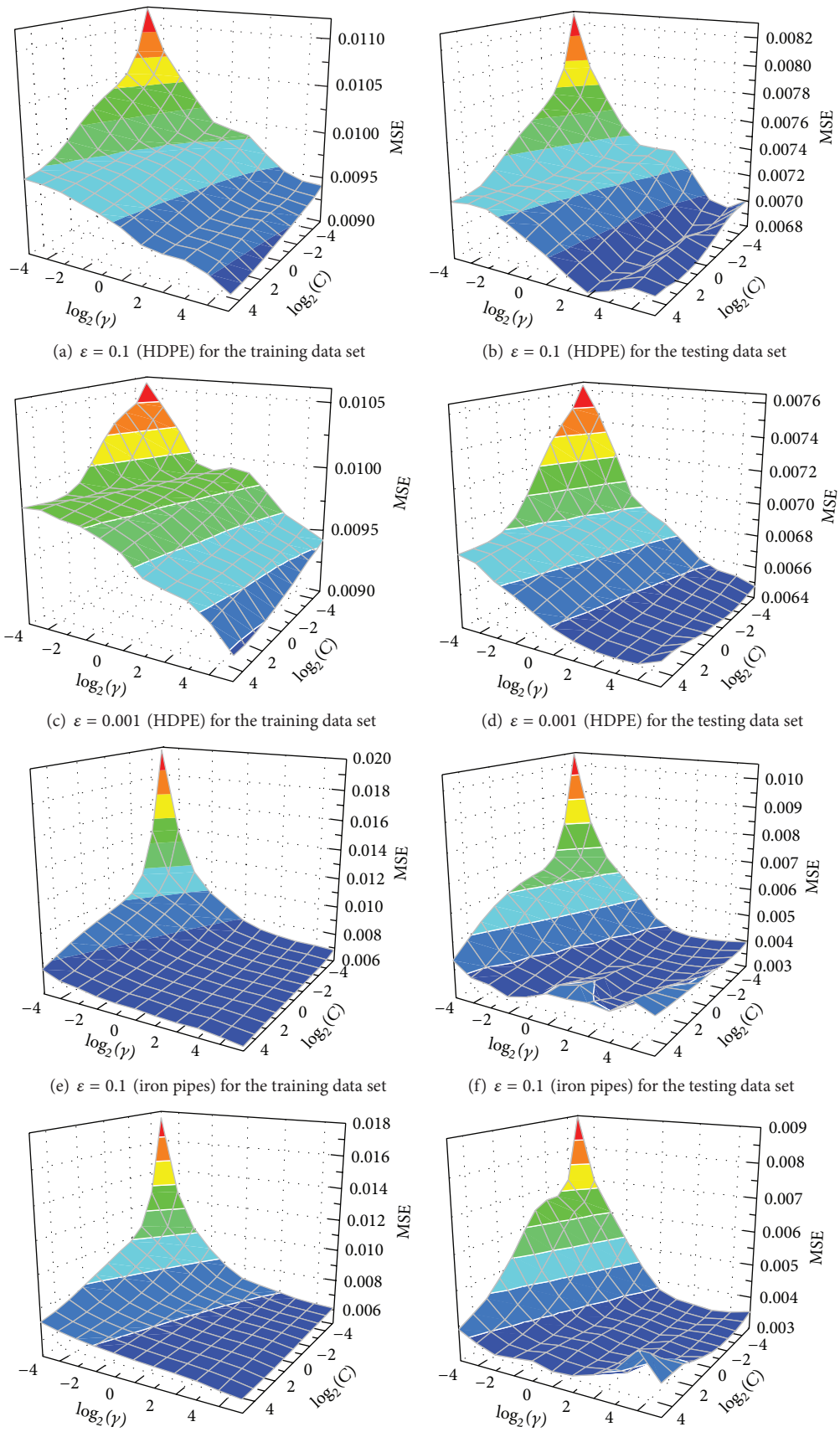
As shown in Figure 10, the model has a very well fit effect when γ and C are large for training data set. But it performs relatively poor for the testing data set. This situation is called the overtraining phenomenon. Optimal SVM coefficients are chosen as the one that fits the testing data set best, in order to avoid the phenomenon. The optimal results are listed in Table 1.

6.2. Prediction of Water Temperature Rise. As written in Section 3, the water temperature rise of cool-pipe, T_w^*, is relevant to the inlet temperature difference and cooling flow rate. We employ the data of monolith 15# as the training data set to model the cooling performance. All the other slabs are used as testing cases. The predicted outlet temperature agrees well with the monitoring result. Four slabs are taken as examples in Figure 11.

6.3. Thermal Performance Comparison. The pattern of thermal performance of cool-pipes can be found in the numerical model that has mined the data mentioned above. The numerical results of T_w^* at different combination of ΔT_{in} and q_w are listed in Table 2. Three-dimensional visualization results of HDPE and iron pipes are also compared in Figure 12. From the results, conclusions can be given as follows.

(1) Outlet water temperature of HDPE pipes decreases with the water flow rate in a lower flow range, which is less than 30 L/min. But in a higher flow range, say more than 40 L/min, the outlet temperature is intensive with the flow rate.

(2) Outlet water temperatures for iron pipes have similar phenomenon with HDPE pipes in a lower flow rate range. And an obvious turning point can be found in the result. T_w^* increases with q_w when q_w is



(a) $\varepsilon = 0.1$ (HDPE) for the training data set

(b) $\varepsilon = 0.1$ (HDPE) for the testing data set

(c) $\varepsilon = 0.001$ (HDPE) for the training data set

(d) $\varepsilon = 0.001$ (HDPE) for the testing data set

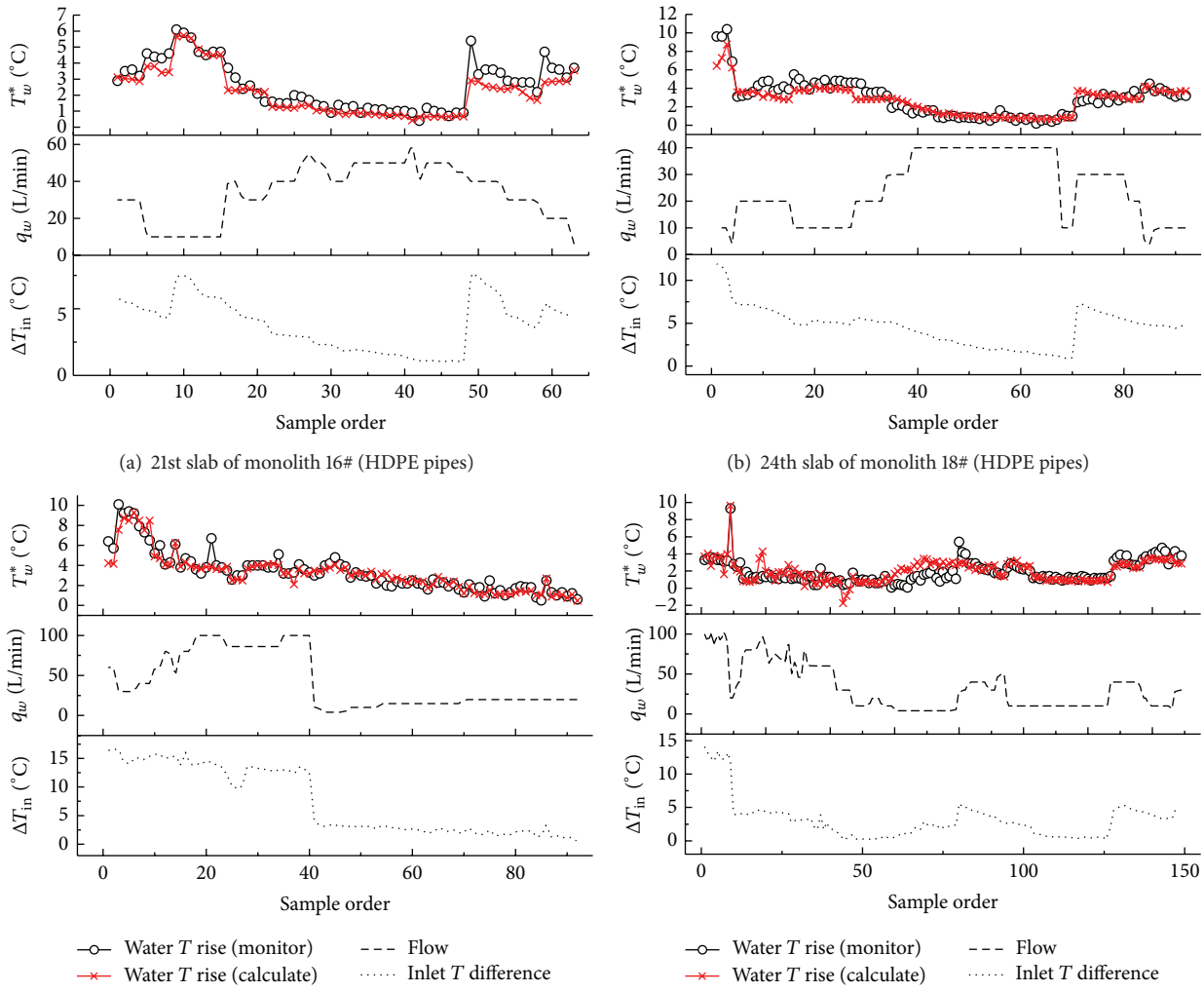
(e) $\varepsilon = 0.1$ (iron pipes) for the training data set

(f) $\varepsilon = 0.1$ (iron pipes) for the testing data set

(g) $\varepsilon = 0.001$ (iron pipes) for the training data set

(h) $\varepsilon = 0.001$ (iron pipes) for the testing data set

Figure 10: Mean squared errors of different parameter combinations.



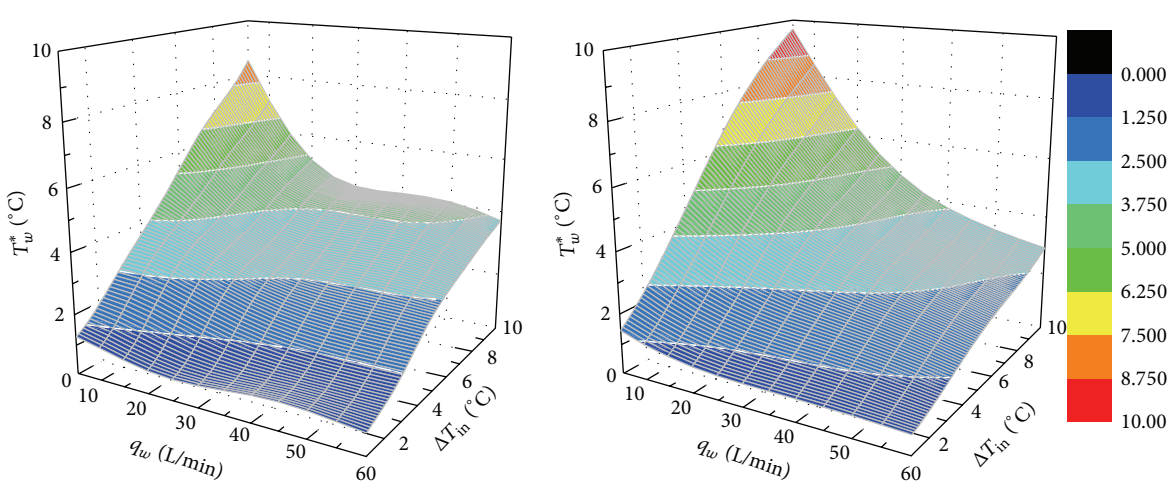
(a) 21st slab of monolith 16# (HDPE pipes)

(b) 24th slab of monolith 18# (HDPE pipes)

(c) 4th slab of monolith 16# (iron pipes)

(d) 7th slab of monolith 14# (iron pipes)

Figure 11: Prediction outlet temperature and monitoring temperature.



(a) Thermal model of HDPE pipes

(b) Thermal model of iron pipes

Figure 12: 3D results of the numerical model.

Table 2: Result of T_w^* (°C) output by the numerical model.

q_w (L/min)	ΔT_{in} (°C)			
	2.0	4.0	6.0	8.0
5	1.69 (2.28)*	3.36 (4.43)	5.11 (6.65)	6.76 (8.41)
10	1.58 (1.93)	3.24 (3.82)	4.41 (5.74)	5.65 (7.25)
20	1.17 (1.47)	2.75 (2.93)	3.36 (4.29)	4.17 (5.30)
30	0.86 (1.27)	2.23 (2.44)	2.90 (3.38)	3.73 (3.99)
40	0.79 (1.18)	1.93 (2.15)	2.83 (2.86)	3.75 (3.23)
50	0.80 (1.05)	1.83 (1.87)	2.81 (2.46)	3.61 (2.77)
60	0.79 (0.87)	1.80 (1.50)	2.64 (2.02)	3.18 (2.38)
70	0.87 (0.70)	1.72 (1.12)	2.29 (1.58)	2.63 (2.03)
80	1.02 (0.72)	1.59 (0.93)	1.91 (1.31)	2.23 (1.81)
90	1.01 (1.03)	1.45 (1.08)	1.78 (1.37)	2.15 (1.84)
100	0.76 (1.62)	1.45 (1.58)	2.03 (1.78)	2.31 (2.16)

* The number enclosed in brackets is the iron result, and the number outside of the brackets is the HDPE result.

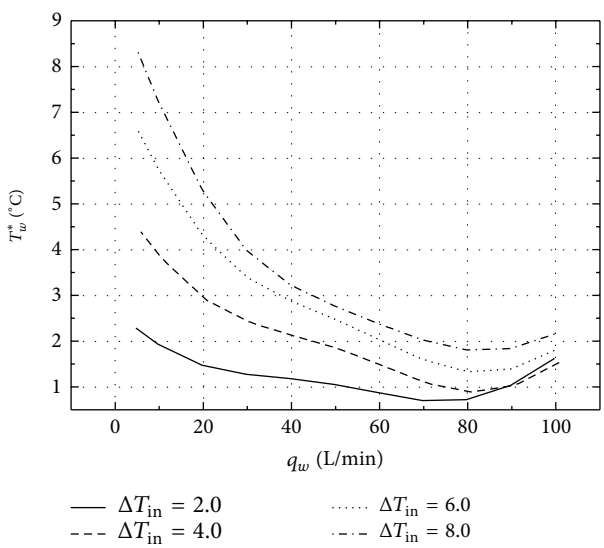


Figure 13: Turning point of iron pipe.

higher than this turning point, which can be seen in Figure 13. The turning point for iron pipe is at 80 L/min.

(3) Higher ΔT_{in} results in higher T_w^*, which is obvious in low flow rate range. When $q_w = 5$ L/min, T_w^* is almost ΔT_{in}. When q_w is large, the difference of ΔT_{in} does not take obvious change in T_w^*.

(4) Absorbed energy rate by cool-pipe can be calculated by (10), which is an important index for the design of cooling procedure. For the same flow rate, higher T_w^* means that more heat can be released by the cool-pipe. The absorbed power by HDPE and iron cool-pipes is plotted in Figure 14. It can be seen that power increases with the flow rate and has a saddle point when q_w is high.

(5) The turning point mentioned in (2) is an interesting point. It can be seen that the absorbed power's saddle point locates near the point. We can name the point as the turning flow rate. For iron pipes, the turning flow rate is 80 L/min, the detailed reason of which will be deep studied in our future researches.

(6) Iron pipe has a better performance when flow rate is lower than 40 L/min. When flow rate is higher than 60 L/min, HDPE pipe can have a better performance of absorbing heat.

(7) When $\Delta T_{in} = 2$°C and $q_w > 80$ L/ min, HDPE pipe has a strange decline curve, which is not consistent with other situations. The reason has not been explored by the authors by now and will be studied in detail.

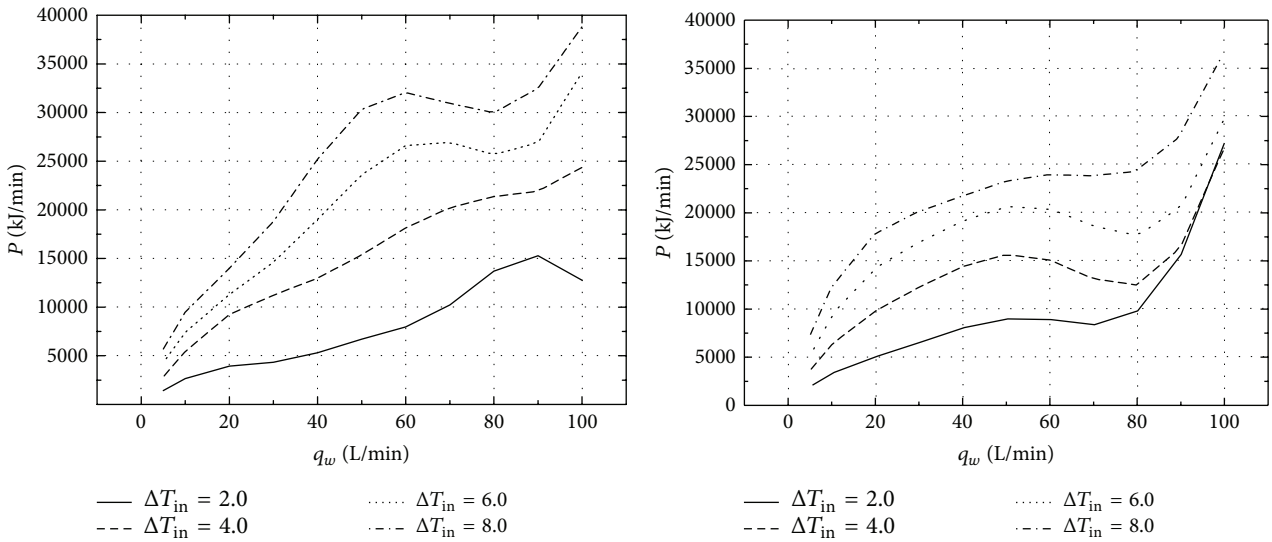
6.4. Classification of Different Pipe Materials. The pipe material determines the value of k in Section 3. In order to correctly classify the pipe materials, ΔT_{in}, T_w^*, and q_w are adopted as the variables, and iron/HDPE is used as the classification label. The data of monolith 15# is used as the training data set and 108 slabs are used as the testing data set.

A large number of data exist for a slab as the product of the monitoring program. The labels of HDPE pipes are defined as 0 and iron ones are defined as 1. Every day's data of the slab is tested and labeled by the SVM classifier, and the following equation is used to estimate the pipe material of the slab:

$$\text{Material} = \begin{cases} \text{HDPE} & \text{when avg} < 0.4 \\ \text{iron} & \text{when avg} > 0.6 \\ \text{alternative} & \text{when otherwise,} \end{cases} \quad (32)$$

where avg is the average value of the label values. The classification result is shown in Figure 15. The detailed classification results of two slabs are plotted as examples, as shown in Figure 16.

From the classification result, we can see that the SVM model can classify the different pipe materials well, with accuracy at 83%. The HDPE pipes have higher accuracy than iron pipes. The result shows that the classification by SVM method is effective.



(a) Absorbed power curve of HDPE

(b) Absorbed power curve of iron pipes

Figure 14: Absorbed power performance.

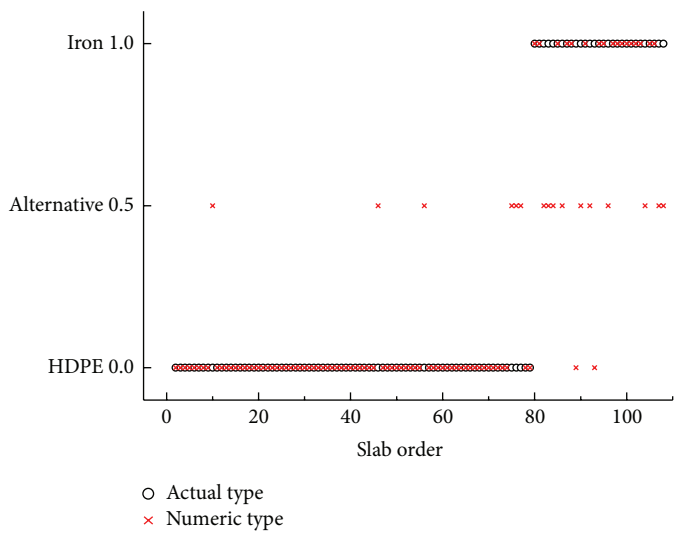


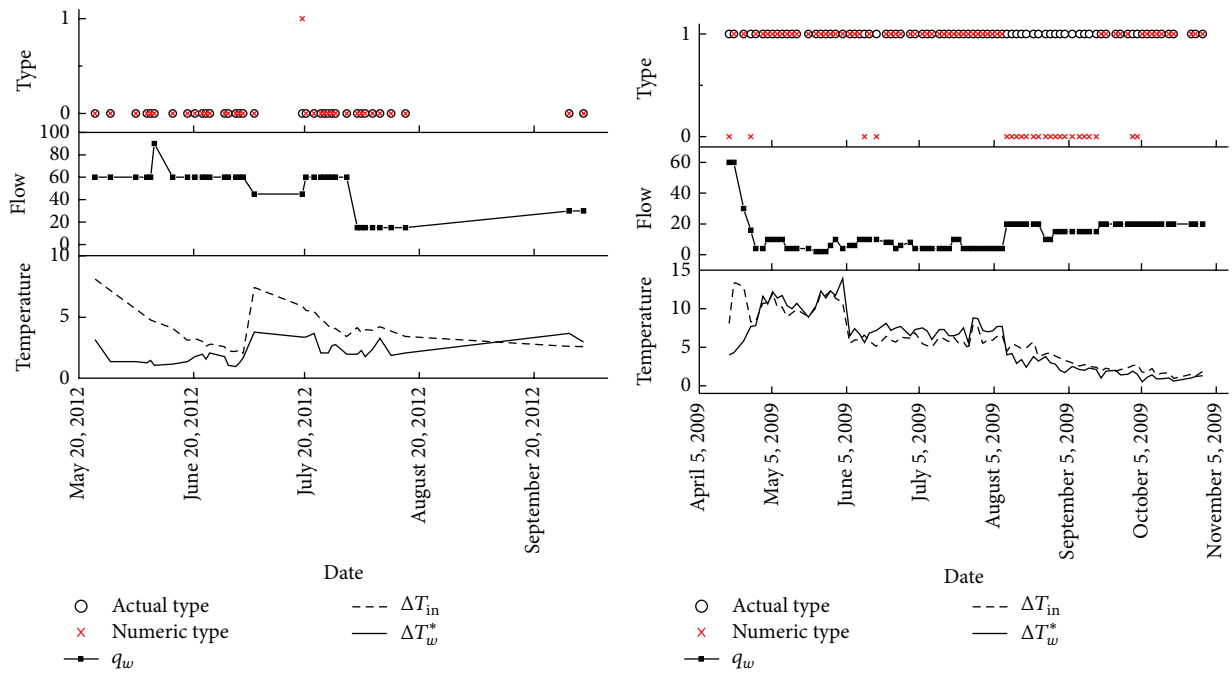
Figure 15: Classification result.

7. Conclusions

The heat transfer process inside the cool-pipe embedded concrete is very complex. Most current researches focus on the theory calculations. Statistical study on the actual measurement is limited. Based on a delicate monitoring program during the construction period of a super high arch dam, data mining on the thermal performance of the cool-pipes is carried out. SVM is applied as the data mining method. With the direction of thermal model built in this paper, the relative factors of outlet temperature of cool-pipes are determined and the relationship is numerically mined. The prediction result has a good agreement with the monitoring data, which verifies the validation of the approach proposed in this paper. The thermal performances between iron pipes and HDPE pipes are also analyzed in detail in this paper. Iron pipes have better performance when flow rate is low, and HDPE have better performance when flow rate is large. Iron pipes have a turning flow rate of 80 L/min. The classification of pipe material is also conducted in this paper, which has correctly automatically distinguished HDPE pipes and iron pipes from monitoring data.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.



(a) 68th slab of monolith 16# (HDPE)

(b) 1st slab of monolith 17# (iron)

Figure 16: Detailed classification result.

Acknowledgments

This work is supported by the National "973" Researching Project of China (no. 2013CB035902), Research Project of State Key Laboratory of Hydroscience and Engineering of Tsinghua University (no. 2012-KY-4), and National Nature Science Foundation of China (nos. 51279087 and 51339003).

References

[1] G.-D. Zhou and T.-H. Yi, "Thermal load in large-scale bridges: a state-of-the-art review," *International Journal of Distributed Sensor Networks*, vol. 2013, Article ID 217983, 17 pages, 2013.

[2] H.-S. Shanga and T.-H. Yi, "Behavior of HPC with fly ash after elevated temperature," *Advances in Materials Science and Engineering*, vol. 2013, Article ID 478421, 7 pages, 2013.

[3] F.-J. Ulm and O. Coussy, "What is a "massive" concrete structure at early ages? Some dimensional arguments," *Journal of Engineering Mechanics*, vol. 127, no. 5, pp. 512–522, 2001.

[4] Z. Zuo, Y. Hu, Q. Li et al., "User-friendly thermal-stress coupled simulating platform of mass concrete," *Chinese Journal of Computational Mechanics*, vol. 30, supplement 1, pp. 1–6, 2013 (Chinese).

[5] Wikipedia, "Hoover dam," 2011, http://en.wikipedia.org/wiki/Hoover_dam.

[6] D. o. t. I. U.S, "Concrete," 2005, http://www.usbr.gov/lc/hooverdam/History/essays/concrete.html.

[7] B. o. R. United States Department of the Interior, *Cooling of Concrete Dams: Final Reports*, Boulder Canyon project: Final Reports. Part VII, Cement and Concrete Investigations, United States Department of the Interior, Washington, DC, USA, 1949.

[8] Y. Hu, Z. Zuo, Q. Li, and Y. Duan, "Boolean-based surface procedure for the external heat transfer analysis of dams during construction," *Mathematical Problems in Engineering*, vol. 2013, Article ID 175616, 17 pages, 2013.

[9] L. Zhang, Y. Hu, Q. Li et al., "Hex-meshing method for gravity dam based on planar boolean operations," *Journal of Hydroelectric Engineering*, vol. 31, no. 5, pp. 209–215, 2012 (Chinese).

[10] H. Xie and Y. Chen, "Influence of the different pipe cooling scheme on temperature distribution in RCC arch dams," *Communications in Numerical Methods in Engineering*, vol. 21, no. 12, pp. 769–778, 2005.

[11] Z. Zuo, Y. Hu, Y. Duan, and J. Yang, "Simulation of the temperature field in mass concrete with double layers of cooling pipes during construction," *Journal of Tsinghua University*, vol. 52, no. 2, pp. 186–189, 2012 (Chinese).

[12] B. F. Zhu and J. B. Cai, "Finite element analysis of effect of pipe cooling in concrete dams," *Journal of Construction Engineering and Management*, vol. 115, no. 4, pp. 487–498, 1989.

[13] J. Yang, Y. Hu, Z. Zuo, F. Jin, and Q. Li, "Thermal analysis of mass concrete embedded with double-layer staggered heterogeneous cooling water pipes," *Applied Thermal Engineering*, vol. 35, no. 1, pp. 145–156, 2012.

[14] B. Zhu, "Effect of cooling by water flowing in nonmetal pipes embedded in mass concrete," *Journal of Construction Engineering and Management*, vol. 125, no. 1, pp. 61–68, 1999.

[15] T. Myers and J. Charpin, "Modelling the temperature, maturity and moisture content in a drying concrete block," *Mathematics-in-Industry Case Studies*, vol. 1, pp. 24–48, 2008.

[16] M. Jiaxuan, "A combined method of theoretic and numerical solutions for pipe cooling in concrete dams," *Journal of Hydroelectric Engineering*, vol. 29, no. 4, pp. 31–41, 1998 (Chinese).

[17] X. Wu, V. Kumar, J. R. Quinlan et al., "Top 10 algorithms in data mining," *Knowledge and Information Systems*, vol. 14, no. 1, pp. 1–37, 2008.

[18] D. H. Dan, Y. M. Zhao, T. Yang, and X.-F. Yan, "Health condition evaluation of cable-stayed bridge driven by dissimilarity measures of grouped cable forces," *International Journal of Distributed Sensor Networks*, vol. 2013, Article ID 818967, 12 pages, 2013.

[19] J. M. Ko and Y. Q. Ni, "Technology developments in structural health monitoring of large-scale bridges," *Engineering Structures*, vol. 27, no. 12, pp. 1715–1725, 2005.

[20] J. Ou and H. Li, "Structural health monitoring in mainland china: review and future trends," *Structural Health Monitoring*, vol. 9, no. 3, pp. 219–231, 2010.

[21] D. Dan, Y. Zhao, T. Yang et al., "Vibration measurement based cable transion identification method for cable-damper systerm," *Journal of Vibration and Shock*, vol. 41, no. 6, pp. 123–127, 2013 (Chinese).

[22] S. D. Glaser, M. Li, M. L. Wang, J. Ou, and J. Lynch, "Sensor technology innovation for the advancement of structural health monitoring: a strategic program of US-China research for the next decade," *Smart Structures and Systems*, vol. 3, no. 2, pp. 221–244, 2007.

[23] A. K. Choudhary, J. A. Harding, and M. K. Tiwari, "Data mining in manufacturing: a review based on the kind of knowledge," *Journal of Intelligent Manufacturing*, vol. 20, no. 5, pp. 501–521, 2009.

[24] S. J. Lee and K. Siau, "A review of data mining techniques," *Industrial Management & Data Systems*, vol. 101, no. 1, pp. 41–46, 2001.

[25] R. Baker and K. Yacef, "The state of educational data mining in 2009: a review and future visions," *Journal of Educational Data Mining*, vol. 1, no. 1, pp. 3–17, 2009.

[26] I. H. Witten, E. Frank, and M. A. Hall, *Data Mining: Practical Machine Learning Tools and Techniques*, Elsevier, San Diego, Calif, USA, 2011.

[27] X. Ye, L. Ran, T. H. Yi, and X. B. Dong, "Intelligent risk assessment for dewatering of metro-tunnel deep excavations," *Mathematical Problems in Engineering*, vol. 2012, Article ID 618979, 13 pages, 2012.

[28] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[29] Y. Liu, T.-H. Yi, and Z.-J. Xu, "Safety early warning research for highway construction based on case-based reasoning and variable fuzzy sets," *The Scientific World Journal*, vol. 2013, Article ID 178954, 10 pages, 2013.

[30] Y. B. Dibike, S. Velickov, D. Solomatine, and M. B. Abbott, "Model induction with support vector machines: introduction and applications," *Journal of Computing in Civil Engineering*, vol. 15, no. 3, pp. 208–216, 2001.

[31] V. Kecman, *Learning and Soft Computing: Support Vector Machines, Neural Networks, and Fuzzy Logic Models*, The MIT Press, Cambridge, Mass, USA, 2001.

[32] C.-W. Hsu, C.-C. Chang, and C.-J. Lin, "A practical guide to support vector classification," Tech. Rep., Department of Computer Science, National Taiwan University, Taipei, Taiwan, 2003.

[33] R. M. Balabin and E. I. Lomakina, "Support vector machine regression (SVR/LS-SVM)—an alternative to neural networks (ANN) for analytical chemistry? Comparison of nonlinear methods on near infrared (NIR) spectroscopy data," *Analyst*, vol. 136, no. 8, pp. 1703–1712, 2011.

[34] C. H. Q. Ding and I. Dubchak, "Multi-class protein fold recognition using support vector machines and neural networks," *Bioinformatics*, vol. 17, no. 4, pp. 349–358, 2001.

[35] Y. Hu, Z. Zuo, Q. Li et al., "Study on abrupt gape of transverse joints and its cause during construction of high arch dams," *Journal of Hydroelectric Engineering*, vol. 32, no. 5, pp. 218–225, 2013 (Chinese).

[36] H. Yu, J. Vaidya, and X. Jiang, "Privacy-preserving svm classification on vertically partitioned data," in *Advances in Knowledge Discovery and Data Mining*, vol. 3918 of *Lecture Notes in Computer Science*, pp. 647–656, Springer, Berlin, Germany, 2006.

[37] S. C. Larson, "The shrinkage of the coefficient of multiple correlation," *Journal of Educational Psychology*, vol. 22, no. 1, pp. 45–55, 1931.

[38] S. Geisser, "The predictive sample reuse method with applications," *Journal of the American Statistical Association*, vol. 70, no. 350, pp. 320–328, 1975.

[39] F. Mosteller and J. W. Tukey, *Data Analysis, including Statistics*, 1968.

[40] P. Burman, "A comparative study of ordinary cross-validation, v-fold cross-validation and the repeated learning-testing methods," *Biometrika*, vol. 76, no. 3, pp. 503–514, 1989.

[41] T. M. Khoshgoftaar and N. Seliya, "Fault prediction modeling for software quality estimation: comparing commonly used techniques," *Empirical Software Engineering*, vol. 8, no. 3, pp. 255–283, 2003.

[42] L. Yang and M. Song, "Coal mine safety evaluation with V-fold cross-validation and BP neural network," *Journal of Computers*, vol. 5, no. 9, pp. 1364–1371, 2010.