SEEPAGE AND DAM DEFORMATION ANALYSES WITH STATISTICAL MODELS: SUPPORT VECTOR REGRESSION MACHINE AND RANDOM FOREST

Infiltration et analyse des déformations d'un barrage par modèles statistiques: machines à vecteur support et de forêt aléatoire

BELMOKRE Ahmed. Laboratoire Mobilisation et Valorisation des Ressources en Eau (MVRE), Ecole Nationale Supérieure d’Hydraulique (ENSH), Blida, Algeria. a.belmokre@ensh.dz

MIHOUBI Mustapha Kamel. Laboratoire Mobilisation et Valorisation des Ressources en Eau (MVRE), Ecole Nationale Supérieure d’Hydraulique (ENSH), Blida, Algeria. mihkam@ensh.dz

SANTILLAN David. Dep't de Ingeniería Civil: hidráulica, Energía y Medio Ambiente, Universidad Politécnica de Madrid, Spain. david.santillan@upm.es

Abstract: Dam monitoring and their safety are the main concern of dam engineers. The collected data of seepage are indicators of structure behavior; they are influenced by various environmental actions such as air temperature, water temperature and water level variation. For this, the analysis of collected data is an operation of big importance. Generally, statistical models are used to analyze data. In this work, support vector regression machine and random forest models are used to interpret data of seepage at different points of dam and identify the environmental variables which affect dam's safety.

Keywords: Analysis, dam, deformation seepage, temperature, random forest.

Résumé: La surveillance des barrages et leur sécurité sont la principale préoccupation des gestionnaires. Les données collectées sur les infiltrations sont un indicateur du comportement de la structure. Elles sont influencées par diverses actions environnementales telles que la température de l'air, la température de l'eau et la variation du niveau de la retenue. Pour cela, l'analyse des données d’auscultation est une opération inévitable. Généralement, les modèles statistiques sont utilisés pour l’analyse. Dans ce travail, les modèles de machines à vecteurs de support et de forêt aléatoire sont utilisés pour interpréter les données d'infiltration en différents points du barrage et
identifier les variables environnementales qui affectent la sécurité du barrage

**Mots clés**: Analyse, barrage, déformation, infiltration, température, forêt aléatoire,

**INTRODUCTION**

Displacements and seepage measurements are the important performance indicators for Roller compacted concrete (RCC) dams (Rubertis, 2018). Flow rate could provide an indicator of physical changes in the structure of dam. Different environmental parameters could influence seepage flow rate like: water level variation, ambient and water temperatures. For that, the prediction and analysis of seepage measurements is an essential operation in dam monitoring. Generally, it is done by deterministic or statistical methods (Santillan, 2011). Recently machine learning technique has become more common for time series data analysis (Salazar et al. 2017).

In this paper, we present a methodology for predicting seepage through Beni Haroun RCC dam via two models: Random forest (RF) model and Support vector regression machine (SVR) model. Water temperature, water level variation and time effect are considered as input. The accuracy of each model was assessed by comparing results with real data. Finally, we estimate the importance of each variable in seepage mechanism.

**METHODOLOGY**

**Mechanisms of seepage in concrete dam**

Seepage is the movement of water through the voids of soil and rock (Li et al. 2008). Seepage analysis of RCC materials is based on Darcy’s law (Li et al. 2015):

\[ \nu = \frac{Q_s}{A} = -k_s \frac{dh}{dl} = k_s J \]  \hspace{1cm} (1)

Where: \( \nu \) is the average velocity \( Q_s \) is the seepage flow, \( A \) is the cross-sectional area, \( k_s \) piezometric head, \( l \) is the seepage path length, and \( J \) is the seepage gradient.

**Random forest method**

Random forest was proposed in 2001 by Leo Breiman. They are part of machine learning techniques. This algorithm combines the concepts of random subspaces and bagging.
The random forest algorithm performs learning on multiple decision trees driven on slightly different subsets of data (Breiman, 2001).

One aspect of the random forest approach is the ability to quantify the importance of the variables used to construct the model (Grömping, 2001). The widely used measures of the importance of a given variable, \( X^j \), is the mean increase in a tree's error in the forest when the observed values of this variable are randomly exchanged in the samples OOB. Denote by \( err_{OOB_t} \) the error of a single tree \( t \) on this \( OOB_t \) sample. Now, randomly permute the values of \( X^j \) in \( OOB_t \) to get a perturbed sample denoted by \( \overline{OOB_t}^j \) and compute \( err_{\overline{OOB_t}^j} \), the error of predictor on the perturbed sample. Variable importance of \( X^j \) is then equal to (Genuer et al. 2010):

\[
VI(X^j) = \frac{1}{q} \sum_{i=1}^{q} (err_{\overline{OOB_t}^j} - err_{OOB_t})
\]  

(2)

**Support vector machine method**

Support Vector Regression (SVR) is the most common application form of support vector machine (SVM) (Basak et al. 2007). It was developed by Vladimir Vapnik (1995):

\[
F_2(x, \hat{\omega}) = \sum_{i=1}^{N} (\alpha_i^* - \alpha_i)(v_i^t x + 1)^p + b
\]

(3)

\( F_2 \) is an expansion explicitly using the training examples. \( \alpha^* \) and \( \alpha \) are a multiplicative constant and his alternative. \( 2N + 1 \) values of \( \alpha_i \) \( \alpha_i^* \) and \( b \) are chosen. If we expand the term raised to the \( p \)'th power, we find \( f \) coefficients that multiply the various powers and cross product terms of the components of \( x \). So, in this sense \( F_1 \) looks very similar to \( F_2 \) in that they have the same number of terms. However, \( F_1 \) has \( f \) free coefficients while \( F_2 \) has \( 2N+1 \) coefficients that must be determined from the \( N \) training vectors (Drucker et al, 1995).

**RESULTS AND DISCUSSION**

With a height of 90 m and a capacity of 1 billion cubic meters, Beni Haroun is the biggest RCC dam in Algeria. Operated since 2002, it is equipped with flow meters installed at different points of the dam: rock gallery, concrete gallery and concrete rock abutment. In this work, the data collected by the GR175RD1 flowmeter located in the right abutment at elevation 175 m are chosen for this study.
The daily water temperature, the variation of water level and time effect are considered as inputs for RF and SVM models. 70% of the data will be used for training both models, and the rest to evaluate their performance (Fig.1 et 2).

![Water level variation](image1.png)

**Fig.1.** Water level variation

![Water temperature at elevation 175 m](image2.png)

**Fig.2.** Water temperature at elevation 175 m

The results of prediction by the two models are presented in (Fig.3). The accuracy of each model is evaluated by calculating root mean square error (RMSE) and mean absolute error (MAE). The RF model gives good predictions than SVM model, with RMSE = 0.1332 l/s and MAE = 0.0901 l/s compared to RMSE = 0.1556 l/s and MAE = 0.0964 l/s for SVM model.

The relationship between seepage and its influencing is nonlinear, RF model is able to detect nonlinearities between predictive variable, which could explains its robustness in seepage predictions (Table 1).
The variable importance score by random forest technique allowed us to determine which variable influences seepage flow evolution in dam body. Water temperature is the parameter that influences most water percolation. While time effect accompanied by the effect of water level variation is similar.

\begin{table}
\centering
\begin{tabular}{|c|c|c|}
\hline
Model & RMSE (l/s) & MAE (l/s) \\
\hline
R F & 0.1332 & 0.0901 \\
S V R & 0.1556 & 0.0964 \\
\hline
\end{tabular}
\caption{Performance of RF and SVM models}
\end{table}
Water temperature has a great impact on the viscosity of water, it is the function of the fluid temperature. The viscosity of water can be obtained by the following empirical equation (Philip, 1957):

\[ \mu = 0.01775/(1 + 0.033T + 0.000221T^2) \]  

The hydraulic conductivity of the concrete dam body is not only the function of the concrete property, but also the function of the property of the fluid flowing through the dam body (Junrui 2002). The hydraulic conductivity of the concrete dam body is inversely proportional to the viscosity of the fluid (Wu, 1995).

\[ K = \rho g k / \mu \]  

CONCLUSION

In this work, a study that aims to analyze and predict the seepage flow rate through an RCC dam by two models of machine learning: RF and SVR models is performed. Based on data from the flowmeter installed in the Beni Haroun dam, located in the east of Algeria. The results show that the RF model is suitable for seepage analysis in concrete than SVM model.

RF technique could be better handle noise in data which make it a powerful tool for analyzing time series data. Due to its direct impact on water viscosity and hydraulic conductivity of the concrete dam, water temperature is the most influential parameter in the seepage process. SVR technique are more sensitive to further tuning than RF model, they have more hyper-parameter and are more sensitive to the presence of correlated or informative inputs, as a consequence, there is a larger margin for improvement than RF model, which may be the object of future research.
REFERENCES

&&&&&