Deformation behavior analysis of an arch dam during initial impoundment based on clustering and panel data regression

Rujiu Zhang, Wenyu Zhuang
State Key Laboratory of Hydroscience and Engineering, Tsinghua University, Beijing, China

Yaoru Liu (corresponding author)
State Key Laboratory of Hydroscience and Engineering, Tsinghua University, Beijing, China

Jianjun Xu, Liang Yin, Haining Wei
POWERCHINA Huadong Engineering Corporation Limited, Hangzhou, China

ABSTRACT: Deformation behavior analysis is crucial for ensuring safe operation of the arch dam. This study investigates the deformation distribution and influencing factors of an arch dam in southwest China based on *k*-means clustering analysis and panel regression models. Results show that dam deformation is closely related to the impounding process and reservoir water level is the main influencing factor. The deformation of some monitoring points near the dam foundation are also affected by aging component. The effects of temperature are negligible in the short term. In general, the studied arch dam is in a condition for safe operation; the abnormal deformation of some monitoring points is basically affected by unbalanced hydraulic thrust, whose magnitude is small and tends to be stable. Analysis results also validate the applicability of cluster analysis and panel regression model in the evaluation of arch dam deformation behavior during initial impoundment.

Keywords: arch dam, deformation behavior, panel data, k-means cluster, initial impoundment.

1 INTRODUCTION

During the initial impoundment of reservoir, the complex coupling effects of water pressure load, temperature change, aging and other factors significantly affect the stress, deformation and structural safety of arch dam. It is necessary to investigate the deformation trend and corresponding influencing factors, thus evaluating the working behavior and safety of arch dam under different conditions.

Since the deformation monitoring data sets during the impoundment period of the arch dam are typical panel data with two-dimensional space-time evolution characteristics, many studies have focused on using panel data model to analyze dam deformation behavior. Shi et al. (2016) introduced dummy variables and established variable-intercept panel models to analyze deformation behavior of different dam zones, which shows high explanatory power. Wang et al. (2020) studied the deformation behavior and influencing factors of Jinping-I arch dam based on variable-intercept panel models. Wang et al. (2021) developed the mixed-coefficient panel model combined with spatial clustering to reveal overall and local deformation behavior of Jinping-I arch dam. Liu et al. (2022) used panel data clustering theory and constructed clustering and zoning model which reflects spatiotemporal characteristics of dam deformation.

Panel models consider spatial heterogeneity and overcome the problem of multicollinearity, and more engineering practice are needed. Based on clustering algorithm and panel regression model, this paper analyzes the deformation behavior of an arch dam in southwest China, which provides reference for dam safety assessment during the initial impoundment.

2 METHODOLOGY

2.1 K-means clustering

The *k*-means cluster analysis method was first proposed by Hartigan & Wong (1979). For a data set $X = \{x_1, x_2, ..., x_i, ..., x_n\}$ that needs to be divided into *K* subsets, the main concept of *K*-means cluster analysis is to use Euclidean distance as the similarity and distance judgment criterion, and calculate the sum of squares of distances between each point in each class c_k and the cluster center μ_i :

$$J(c_k) = \sum_{x_i \in c_k} \|x_i - \mu_k\|^2 \tag{1}$$

The goal of clustering is to minimize the sum of squares of the total distances of each class:

$$J(C) = \sum_{k=1}^{K} J(c_k) = \sum_{k=1}^{K} \sum_{x_i \in c_k} \|x_i - \mu_k\|^2 = \sum_{k=1}^{K} \sum_{i=1}^{n} d_{ki} \|x_i - \mu_k\|^2 \tag{2}$$

where

$$d_{ki} = \begin{cases} 1, & \text{if } x_i \in c_i \\ 0, & \text{if } x_i \notin c_i \end{cases} \tag{3}$$

For analysis of spatial distribution of dam deformation, accumulative deformation and deformation increment are generally selected to conduct the double-index cluster analysis (Zhuang et al. 2023). The accumulative deformation is the observed value of the current date; deformation increment is the difference between the observed value of the current date and the starting date.

2.2 Influencing factors of dam deformation

The deformation of arch dam during impoundment is generally affected by reservoir water level, air temperature, aging and other unmeasurable factors. The estimated deformation δ of single monitoring point can be expressed as (Zhuang et al. 2023):

$$\delta = \delta_H + \delta_T + \delta_\theta + C \tag{4}$$

where δ_H, δ_T and δ_θ are water pressure component, temperature component and aging component, respectively; *C* is a constant.

$$\delta_H = \sum_{i=1}^{4} \left[a_i \left(H^i - H_0^i \right) \right] \tag{5}$$

$$\delta_T = b_1 T + b_2 T_{(10)} + b_3 T_{(20)} + b_4 T_{(30)} \tag{6}$$

$$\delta_\theta = c_1(\theta - \theta_0) + c_2 \left[\ln(\theta + 1) - \ln(\theta_0 + 1) \right] \tag{7}$$

where *H* and H_0 the water level of observation date and initial measurement date; *T*, $T_{(10)}$, $T_{(20)}$ and $T_{(30)}$ represent the current air temperature and 10 days, 20 days and 30 days ahead, respectively; θ and θ_0 are the current date and initial measurement date; a_i, b_i and c_i are fitting coefficients.

2.3 Regression model for panel data

Regression models for panel data include pooled model, variable-intercept model, and variable-coefficient model. The pooled model estimates a single regression equation and ignores the heterogeneity between monitoring points, which is inappropriate for dam deformation analysis. Variable-coefficient model provides different regression coefficients for different monitoring points, reflecting the difference of deformation mechanism. Stepwise regression method was adopted to avoid multicollinearity when using variable-coefficient model in this study. Variable-intercept model is a compromise, which can simultaneously consider the individuality and generality of dam deformation. The basic form of variable-intercept model is (Shao et al. 2017):

$$y_{it} = \beta_1 x_{1it} + \beta_2 x_{2it} + \cdots + \beta_k x_{kit} + \alpha_i + \varepsilon_{it} \; (i = 1,2,\cdots,n\,; t = 1,2,\cdots,T) \tag{8}$$

where y_{it} is the monitoring deformation data; i is the number of monitoring points; t is the number of different time points; x_{1it}, x_{2it},..., x_{kit} are explanatory variables (influencing factors); k is the number of explanatory variables (k=10); β_1, β_2,..., β_k are fixed regression coefficients (a_i, b_i and c_i in Section 2.2), α_i is individual-specific intercept, ε_{it} is the random error that conforms to normal distribution.

3 CASE STUDY

3.1 Project overview

Yangfanggou Hydropower Station is located in Muli County, Sichuan Province, China, which is the sixth cascade hydropower station in the middle reaches of the Yalong River. The normal reservoir water level is 2094 m and dead water level is 2088 m. The water retaining structure is a concrete double-curvature arch dam with a height of 155 m. The dam began to impound on December 30, 2020. The reservoir water level rose to 2040 m in early March 2021 and reached the normal water level of 2094 m on July 19, 2021. Fig.1 shows the layout of on-site monitoring of the arch dam.

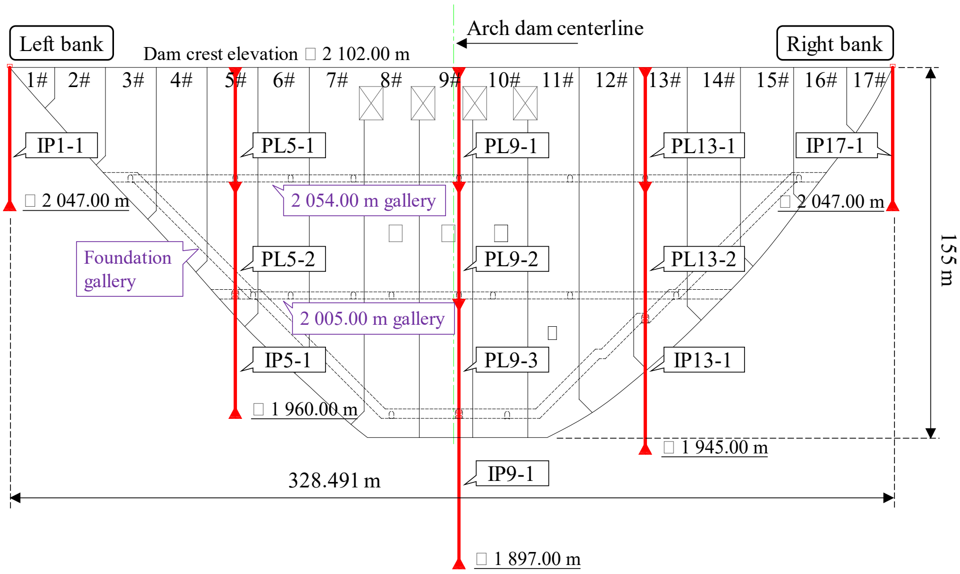
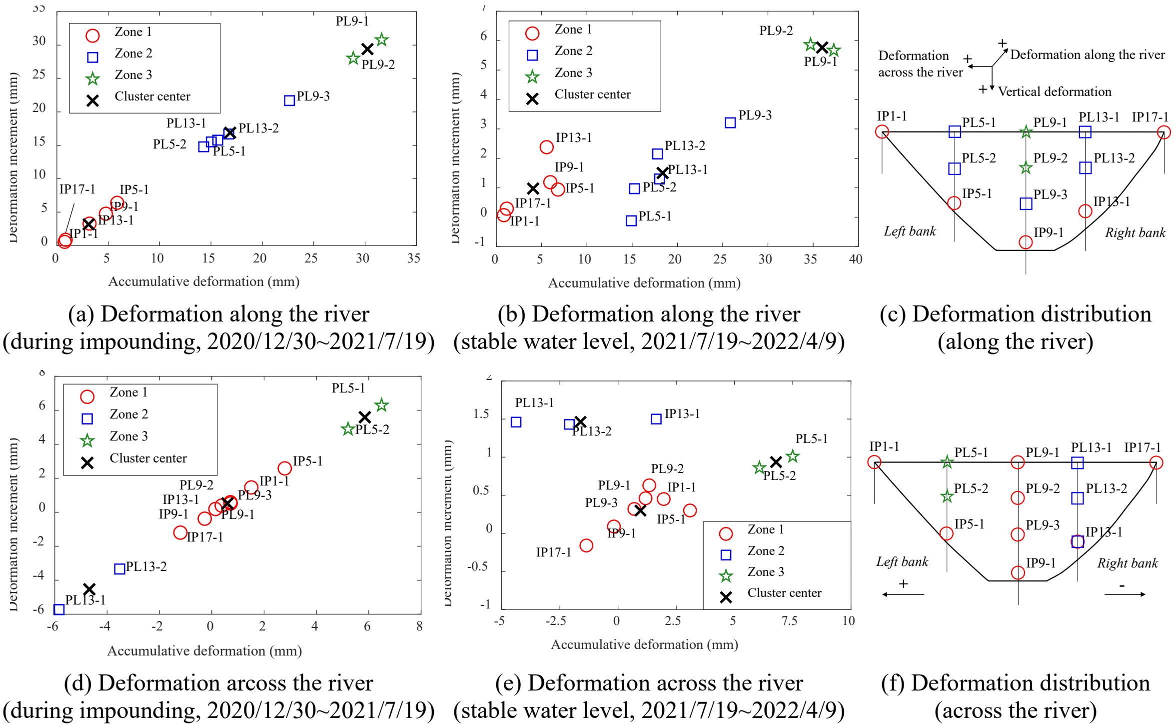


Figure 1. Layout of on-site monitoring and size parameters of Yangfanggou arch dam.

3.2 Spatial characteristics of dam deformation

The clustering results of dam deformation is shown in Fig.2. Monitoring points were divided to three zones based on their total and changed deformation. It can be clearly seen that the arch dam generally has the trend of deformation towards the downstream direction and mountains on both banks, and deformation distribution is basically symmetrically distributed along the axis of dam body, which conforms to the general law of arch dam deformation. Due to the constraint of the foundation surface, the closer the monitoring point is to dam foundation, the smaller the displacement change is. When the water level is stable at 2094 m, PL13-1, PL13-2 and IP13-1 move away from right bank, which is probably the result of slight asymmetry of arch dam shape.



(a) Deformation along the river (during impounding, 2020/12/30~2021/7/19)

(b) Deformation along the river (stable water level, 2021/7/19~2022/4/9)

(c) Deformation distribution (along the river)

(d) Deformation arcoss the river (during impounding, 2020/12/30~2021/7/19)

(e) Deformation across the river (stable water level, 2021/7/19~2022/4/9)

(f) Deformation distribution (across the river)

Figure 2. Double-index clustering distribution of dam deformation.

3.3 *Influencing factors and mechanism of dam deformation*

Estimation results of regression coefficients are shown in Table 1 and Table 2. The monitoring points in Table 1 were divided into three groups based on R-squared; coefficients in bold in Table 2 indicate p>0.05. The regression and component curves of Zone 1 and 2 (deformation along the river) and typical monitoring points (deformation across the river) are shown in Fig.3, Fig.4, and Fig.5.

Table 1. Coefficients of influencing factors using variable-coefficient model (deformation across the river).

	a_1	a_2	a_3	a_4	b_1	b_2	b_3	b_4	c_1	c_2	R^2
PL5-1	2.081	-3.239×10^{-2}	2.168×10^{-4}	-5.116×10^{-7}	**0**	**0**	-2.329×10^{-2}	-1.732×10^{-2}	4.598×10^{-3}	-3.455×10^{-1}	0.996
PL5-2	1.325	-2.125×10^{-2}	1.455×10^{-4}	-3.492×10^{-7}	**0**	**0**	-1.554×10^{-2}	-1.620×10^{-2}	3.203×10^{-3}	-8.345×10^{-2}	0.996
IP5-1	0.486	-8.569×10^{-3}	6.462×10^{-5}	-1.697×10^{-7}	**0**	**0**	-1.304×10^{-2}	**0**	9.839×10^{-4}	4.039×10^{-2}	0.992
IP1-1	0.899	-1.449×10^{-2}	9.864×10^{-5}	-2.386×10^{-7}	4.934×10^{-3}	-1.040×10^{-2}	**0**	6.119×10^{-3}	1.923×10^{-3}	**0**	0.986
PL13-1	-3.320	5.209×10^{-2}	-3.501×10^{-4}	8.407×10^{-7}	**0**	**0**	**0**	**0**	6.190×10^{-3}	-1.565×10^{-1}	0.964
PL13-2	-3.267	5.111×10^{-2}	-3.435×10^{-4}	8.309×10^{-7}	**0**	**0**	**0**	-1.632×10^{-2}	5.432×10^{-3}	**0**	0.911
IP17-1	**0**	**0**	5.913×10^{-8}	**0**	1.023×10^{-2}	**0**	**0**	3.823×10^{-3}	1.466×10^{-3}	-3.683×10^{-1}	0.901
PL9-2	**0**	1.933×10^{-4}	-9.167×10^{-7}	**0**	-5.467×10^{-3}	**0**	-6.564×10^{-3}	-5.790×10^{-3}	2.355×10^{-3}	-1.129×10^{-1}	0.887
IP9-1	0.067	7.510×10^{-4}	-2.734×10^{-6}	**0**	**0**	2.862×10^{-3}	**0**	**0**	**0**	**0**	0.866
IP13-1	-2.177	3.585×10^{-2}	-2.522×10^{-4}	6.411×10^{-7}	**0**	**0**	-1.852×10^{-2}	-2.563×10^{-2}	6.757×10^{-3}	-1.570×10^{-1}	0.820
PL9-3	**0**	5.164×10^{-4}	-4.981×10^{-6}	1.376×10^{-8}	**0**	**0**	**0**	-6.502×10^{-3}	1.615×10^{-3}	-1.575×10^{-1}	0.772
PL9-1	1.622	-2.586×10^{-2}	1.779×10^{-4}	-4.440×10^{-7}	**0**	1.072×10^{-2}	**0**	1.942×10^{-2}	**0**	-8.405×10^{-2}	0.606

Table 2. Coefficients of influencing factors using variable-intercept model (deformation along the river).

	a_1	a_2	a_3	a_4	b_1	b_2	b_3	b_4	c_1	c_2	R^2
Zone 1	**0.807**	**-1.297×10^{-2}**	**8.828×10^{-5}**	**-2.110×10^{-7}**	**1.842×10^{-3}**	**-6.026×10^{-3}**	**-3.034×10^{-3}**	**-6.852×10^{-5}**	3.575×10^{-3}	**1.027×10^{-1}**	0.869
Zone 2	6.995	-1.024×10^{-1}	6.292×10^{-4}	-1.344×10^{-6}	**-2.102×10^{-2}**	**-1.924×10^{-2}**	**-8.604×10^{-3}**	-4.621×10^{-2}	1.003×10^{-2}	**-2.474×10^{-1}**	0.977
Zone 3	13.106	-1.903×10^{-1}	1.161×10^{-3}	-2.470×10^{-6}	-4.205×10^{-2}	-4.618×10^{-2}	**-1.521×10^{-2}**	-6.819×10^{-2}	2.776×10^{-2}	-6.264×10^{-1}	0.988

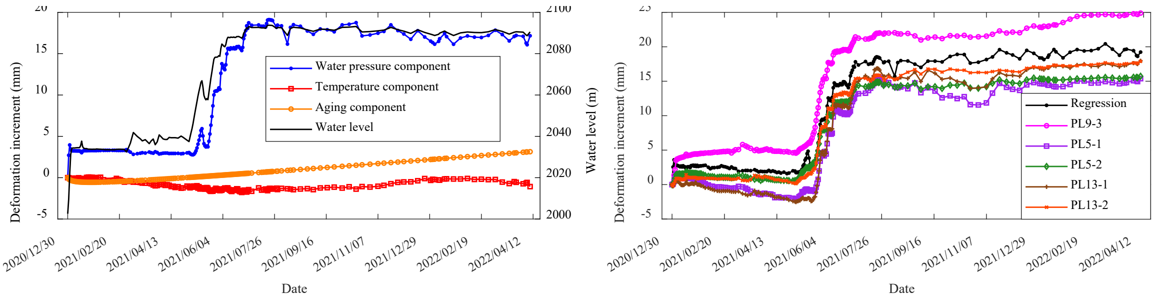


Figure 3. Regression curve, component curve and monitoring data in Zone 2 (deformation along the river).

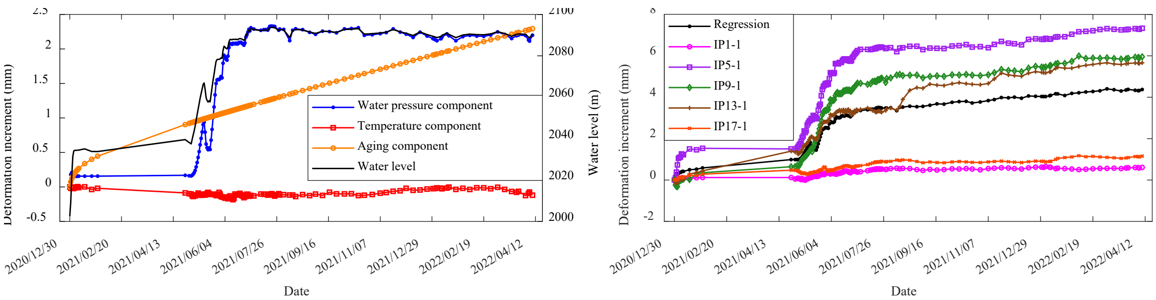
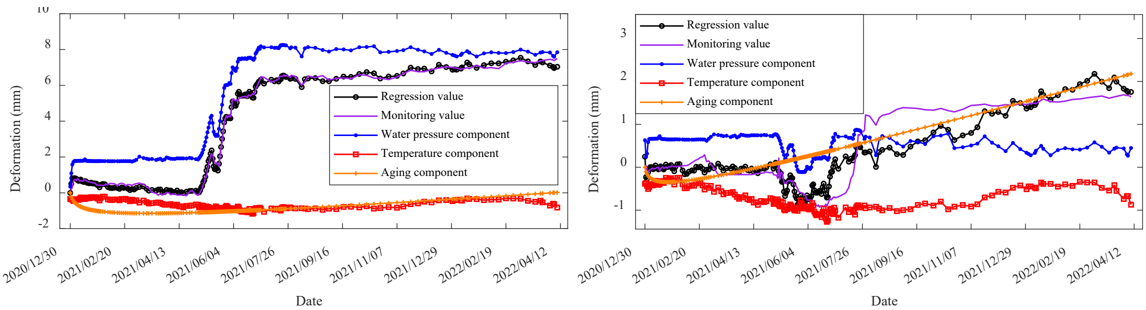


Figure 4. Regression curve, component curve and monitoring data in Zone 1 (deformation along the river).



(a) PL5-1 (R^2=0.996) (b) IP13-1 (R^2=0.820)

Figure 5. Regression curves & component curves of typical monitoring points (deformation across the river).

For the deformation along the river, the accuracy of panel model for Zone 2 and 3 is relatively high, which means regression results can effectively reflect the influencing factors and mechanism of this region. It can be found from Fig.3 that the variation of water pressure component is much higher than temperature and aging component, indicating that impounding is the main influencing factor of dam deformation along the river. Moreover, the closer the monitoring point is to the dam foundation, the greater the effect of aging on deformation (Fig.4). Because some monitoring points close to dam foundation (IP17-1) are less affected by water level, the fitting accuracy of variable-intercept model is reduced, which fails to describe the deformation behavior of all points in Zone 3. This indicates that the variable-intercept model is applicable to multiple monitoring points with same deformation mechanism. When there is a significant difference in deformation mechanisms between monitoring points, it is more appropriate to use the variable-coefficient model.

For the deformation across the river, due to the large difference in deformation behavior between monitoring points, the variable-coefficient model was used for regression to consider the characteristics of each monitoring point. The deformation of PL5-1, PL5-2, IP5-1, IP1-1, PL13-1, PL13-2 is mainly affected by water level; the deformation of IP17-1 is basically controlled by aging effects (see Table 1 and Fig.5a). The regression results of these monitoring points have a relatively high fitting accuracy (R^2>0.9). However, for other monitoring points with low accuracy (PL9-2, IP9-1, IP13-1, PL9-3, PL9-1), there may be some unobserved factors affecting their deformation. As shown in Fig.5b and Fig.2e, the transverse deformation of IP13-1 has a sudden change away from the right bank at a stable water level (2021.7.9), which is probably because the direction of hydraulic thrust force (across the river) points to the left bank. The deformation value of IP13-1 is not large, so temperature change or random error may also lead to this phenomenon. Furthermore, it is worth noting that the air temperature component has impacts on dam deformation to some extent, and it seems to have a lag effect of temperature on deformation, which can clearly be seen from Table 1.

In general, the dam deformation distribution conforms to the general law of arch dam deformation, and reservoir water level is the main influencing factor. The deformation evolution of dam body is basically steady and the arch dam is in a condition for safe operation.

4 CONCLUSIONS

Deformation behavior of an arch dam during its initial impoundment was analyzed based on clustering and panel data regression. The influencing factors and mechanism of dam deformation (along and across the river) were revealed. Results show that the dam deformation behavior is closely related to the impounding process; reservoir water level is the main influencing factor. As the distance from dam foundation decreases, the influence of aging on deformation of some monitoring points increases, showing the deformation characteristics of slopes to some extent. The effects of temperature are negligible in the short term. The dam deformation distribution conforms to the general law of arch dam deformation, which indicates that this arch dam is in good condition. The abnormal deformation of some monitoring points is basically affected by unbalanced hydraulic thrust, whose magnitude is small and tends to be stable. Moreover, analysis results validate the applicability of cluster analysis and panel regression model in the evaluation of arch dam deformation behavior.

ACKNOWLEDGEMENTS

This work was supported by the National Natural Science Foundation of China (Grant No. 41961134032) and the Swiss National Science Foundation (Grant No. 189882).

REFERENCES

Hartigan, J.A. & Wong, M.A. 1979. Algorithm AS 136: a k-means clustering algorithm. *Journal of the Royal Statistical Society* 28 (1), pp. 100-108.

Liu, Y.T., Zheng, D.J., Georgakis, C., Kabel, T., Cao, E.H., Wu, X. & Ma, J.J. 2022. Deformation analysis of an ultra-high arch dam under different water level conditions based on optimized dynamic panel clustering. *Applied Sciences* 12 (1), 481.

Shao, C.F., Gu, C.S., Yang, M., Xu, Y.X. & Su, H.Z. 2017. A novel model of dam displacement based on panel data. *Structural Control and Health Monitoring* 25 (1), e2037.

Shi, Z.W., Gu, C.S. & Qin, D. 2016. Variable-intercept panel model for deformation zoning of a super-high arch dam. *SpringerPlus* 5, 898.

Wang, J.M., Gu, C.S. & Zhang, C. 2020. Deformation behavior analysis of Jinping arch dam based on spatiotemporal model of variable intercept panel data. *Journal of Hydroelectric Engineering* 39 (11), pp. 21-30. (in Chinese)

Wang, S.W., Xu, C., Liu, Y. & Wu, B.B. 2021. Mixed-coefficient panel model for evaluating the overall deformation behavior of high arch dams using the spatial clustering. *Structural Control and Health Monitoring* 28 (10), e2809.

Zhuang, W.Y., Liu, Y.R., Zhang, R.J., Hou, S.K. & Yang, Q. 2023. Study on deformation mechanism and parameter inversion of a reservoir bank slope during initial impoundment. Acta Geotechnica.