



A spatiotemporal identification method for deformation characteristics of expansive soil canal slope based on spectral clustering

DOI:

[10.1016/j.eswa.2023.120108](https://doi.org/10.1016/j.eswa.2023.120108)

Document Version

Accepted author manuscript

[Link to publication record in Manchester Research Explorer](#)

Citation for published version (APA):

Li, X., Ma, F., Hu, J., Jivkov, A., & Chu, D. (2023). A spatiotemporal identification method for deformation characteristics of expansive soil canal slope based on spectral clustering. *Expert Systems with Applications*, 225, Article 120108. <https://doi.org/10.1016/j.eswa.2023.120108>

Published in:

Expert Systems with Applications

Citing this paper

Please note that where the full-text provided on Manchester Research Explorer is the Author Accepted Manuscript or Proof version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version.

General rights

Copyright and moral rights for the publications made accessible in the Research Explorer are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Takedown policy

If you believe that this document breaches copyright please refer to the University of Manchester's Takedown Procedures [<http://man.ac.uk/04Y6Bo>] or contact openresearch@manchester.ac.uk providing relevant details, so we can investigate your claim.



1 A spatiotemporal identification method for deformation
2 characteristics of expansive soil canal slope based on spectral
3 clustering

4 Xing Li^{1*}, Fuheng Ma^{1*}, Jiang Hu¹, Andrey P Jivkov², Dongdong Chu³

5 (1. State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering,
6 Nanjing Hydraulic Research Institute, Nanjing 210029, China

7 2. Department of Mechanical, Aerospace and Civil Engineering, The University of
8 Manchester, Manchester M13 9PL, UK

9 3. Jiangsu Water Conservancy Science Research Institute, Nanjing 210017, China)

10 **Abstract:** Structural health diagnosis of expansive soil slopes requires timely analysis
11 of deformation monitoring data. A method for spatiotemporal clustering of monitoring
12 data for health diagnosis is proposed. First, the deformation time series is upgraded to
13 a panel time series, which includes spatial positions and temporal variations, and
14 similarity characteristics of spatiotemporal deformation are discussed. Second, a
15 similarity distance indicator is defined using three deformation variables: weighted
16 absolute distance, weighted increment distance, and weighted growth rate distance.
17 Third, a spatiotemporal clustering model of the deformation of expansive soil slope
18 based on a spectral clustering algorithm is developed, together with a scoring algorithm
19 for determining optimal clusters. The method analyses and diagnoses the deformation
20 behaviour of the expansive soil slope structure of China's South-to-North Water
21 Diversion Project central line. The advantage of the proposed method is demonstrated
22 by comparing its results with results obtained by the commonly used temporal
23 clustering method. It is further shown how the new method can be used to identify
24 abnormal regions of expansive soil slope deformation.

25 **Keywords:** expansive soil slope; deformation; spatiotemporal clustering; weighed
26 clustering indicators; weighted comprehensive distance; anomaly identification

27 **Introduction**

28 Expansive soil is a geological body formed in natural geological processes. It can
29 undergo significant expansion, contraction, and fissure development. These processes
30 are susceptible to environmental changes, especially in humidity and temperature: the

*Corresponding author.

E-mail addresses: xingli@nhri.cn (X. Li), fhma@nhri.cn (F. Ma), huj@nhri.cn (J. Hu),
Andrey.Jivkov@manchester.ac.uk (A. Jivkov), bwhtxq@163.com (D. Chu)

1 soils expand when absorbing water and contract when water evaporates. Geological
2 surveys on slope destruction in expansive soils (Hou et al., 2013; Lu et al., 2020; Xie
3 et al., 2020; T. Li et al., 2021) indicate that irrespective of the slope formation – a natural
4 slope, an excavated slope with undisturbed expansive soil, or a filled slope constructed
5 with compacted expansive soil – it is easy to scour and peel off from the local slope,
6 i.e., the soil can suffer overall collapse, landslide or other instabilities under the action
7 of various geological stresses and environmental conditions. This behaviour often
8 causes problems in engineering construction. According to statistics, expansive soil has
9 been found in more than 40 countries worldwide, covering six continents. The most
10 prominent countries that reported expansive soil engineering accidents include the
11 USA, Australia, South Africa, India, Canada, Israel, and China. Examples include the
12 Pittsburgh earth dam landslide and the collapse of the St. Francis arch dam (De
13 Wrachien, 2009). Expansive soil channels of many irrigation areas and water transfer
14 projects in China encountered landslide problems during the operation period, e.g., 55
15 landslides in the slope treatment of the expansive soil canal in Hubei's Ebei Gangdi,
16 195 landslides in the main canals in the 30 years before the operation of the Huaishihang
17 Irrigation District, over 70 landslides in the expansive soil slope during the seven years
18 operation of the South-to-North Water Diversion Project (Xiao et al., 2019). The
19 uncertainty in the safe operating period affects the exploitation of the channel and may
20 cause severe social, economic, and environmental impacts and even threaten life and
21 safety. Currently, China is in the process of large-scale construction of irrigation areas
22 and water diversion projects. Ensuring the long-term security and stability of the
23 channels in the expansive soil area is critical for this ambitious program. Therefore, it
24 is essential to analyse the deformation characteristics of expansive soil slopes and
25 identify deformation anomalies.

26 Methods for expansive soil slope stability analysis can be categorized into four
27 types. The first type is the traditional limit equilibrium calculation method (Zhu et al.,
28 2003). It provides a convenient quantitative assessment of a slope stability safety factor
29 but involves assumptions and restrictions, including judgment based on experience,
30 which is highly subjective. The second type includes numerical calculation methods,
31 such as the strength reduction method (Qi and Vanapalli, 2016). These involve
32 convenient calculations but require many material parameters whose experimental
33 determination is often challenging, and the accuracy of the judgment also depends on
34 engineering experience. The third type includes methods for predicting deformation

1 trends by artificial intelligence algorithms (Alimohammadlou et al., 2014; Y. Chen et
2 al., 2022; Li et al., 2018; Li et al., 2019; Miao et al., 2017). These are limited to
3 analysing the time scale of the deformation monitoring data. Due to the particular
4 property of expansive soil, it has been challenging to obtain reliable predictions, and
5 such methods are rarely used for expansive soil slopes. The fourth type is based on
6 engineering analogy, such as fuzzy comprehensive evaluation, grey cluster, reliability,
7 and neural network evaluation methods (Li et al., 2017; Zhang et al., 2018, 2022). Their
8 advantage is that they can capture the uncertainties affecting the stability of the slope.
9 However, the current development of these methods, with few exceptions, is still
10 limited to cluster analysis on the time scale, and the effect of space dimension is
11 ignored. Some researchers (B. Chen et al., 2022; Chen et al., 2019; Salazar et al., 2017;
12 Zhang et al., 2023b, 2023a) have studied the spatiotemporal clustering of arch dams but
13 have only considered radial displacements and have not captured the collective effect
14 of multiple deformation indicators from the monitoring system. Currently, the main
15 issues in data clustering analysis for monitoring are twofold. Firstly, the weight of the
16 indicators is not considered, meaning that there is no distinction between the importance
17 of individual indicators when determining the distance during sample clustering. This
18 leads to an inadequate reflection of the significance of different monitoring data indices
19 in differentiating categories. Secondly, the weight of time is not taken into account,
20 despite some researchers examining its impact on clustering. As a result, they are unable
21 to develop a distance function that accurately captures the effect of different time points
22 on category differentiation. Given that the failure of expansive soil canal slopes is
23 repetitive, seasonal, and spatially unbalanced, and could cause significant harm to
24 people and property in case of a landslide, a comprehensive spatiotemporal clustering
25 method utilizing various monitoring data can be suggested to reflect the importance of
26 different indicators and time points on clustering. Such a method can identify abnormal
27 deformation of expansive soil canal slopes, laying a strong foundation for their safe and
28 long-term operation.

29 This work presents an analysis of slope deformation development based on long-
30 term monitoring data of expansive soil slopes. The analysis uses clustering methods
31 from spatiotemporal data mining and extracts the similarity characteristics of
32 deformation sequences. Considering the effects of different deformation monitoring
33 data, such as internal horizontal displacement, horizontal surface displacement, and
34 vertical displacement, three similarity indicators of "weighted absolute distance,"

1 "weighted incremental distance," and "weighted growth rate distance" for different
2 observation point deformation sequences of expansive soil slope are proposed. Based
3 on the entropy weight method, the corresponding "weighted comprehensive distance "
4 index is obtained to quantitatively analyse the similarity of the expansive soil slope
5 from the time section and the spatial section. Based on the spectral clustering algorithm,
6 a scoring algorithm is proposed to determine the number of clusters and divide the
7 deformation period and its response deformation area of the expansive soil slope.
8 Finally, a spatiotemporal clustering analysis model of the expansive soil slope based on
9 various deformation data is developed to identify the abnormal deformation area and to
10 issue an early warning of the anomalous region. The proposed method is applied to the
11 expansive soil slope of the Taocha section of the central line of the South-to-North
12 Water Diversion Project. The deformation behaviour of the expansive soil slope in
13 different stages is analysed and diagnosed. The advantage of the proposed method is
14 demonstrated by the comparative analysis of the diagnostic results obtained by the
15 common and spatiotemporal clustering.

16 1 Spatiotemporal evolution characteristics of deformation and failure of 17 expansive soil slope

18 Expansive soils are clay-rich materials with expansion, contraction, fissuring, and
19 super-consolidation characteristics requiring special failure law. The shallow surface
20 region of expansive soil is subjected to climatic stress. It forms a swelling-shrinking
21 zone or a weathering-affected zone of a certain depth. A weak surface is formed
22 between the swelling-shrinking region and the underlying expansive soil. The depth of
23 this soft surface is within 6m. An expansive soil slope can therefore experience repeated
24 landslides, potentially easily along the soft surface (shallow sliding) and possibly along
25 surfaces at more considerable depths. These can be either gradual and multiple sliding
26 at different locations or repeated at one location. The latter refers to the case of
27 expansive soil slopes that have undergone initial sliding and have formed a relatively
28 stable slope, which is reactivated by changes in the local shape of the landform and
29 meteorological and hydrological conditions. This mechanism can continue to produce
30 new landslides in many subsequent years. For example, most canal slope landslides
31 occur in the rainy season, especially in the first rainy season after a long-term drought,
32 with clear seasonal and intermittent patterns.

33 Landslide failure of expansive soil slopes has temporal and spatial characteristics.

1 The slope can change its volume, shape, or macroscopic continuity. Changes that do
2 not involve discontinuous motion are referred to as deformation. In safety monitoring,
3 deformation refers to the displacement readings such as horizontal and vertical
4 displacement – which is different from the established terminology in continuum
5 mechanics. Discontinuous motion or sliding is an element of slope failure. Spatially,
6 landslide failure does not occur globally at once but involves several phenomena:
7 localization of deformation, stress transfer, deformation increase, and surface
8 penetration (Qin et al., 2002; Reichenbach et al., 2018).

9 In the traditional channel slope deformation analysis, the deformation data is often
10 given by a one-dimensional array representing the time history of deformation at a
11 single monitoring point. This representation omits the spatial dimension of
12 deformation. The observation points need to be arranged in a two-dimensional grid
13 covering the slope, and the temporal changes provide the third dimension to capture the
14 spatiotemporal characteristics of slope deformation. Abnormal deformations in the
15 slope can be obtained by analysing such 3D records instead of the traditional method.
16 Monitoring the deformation of expansive soil slopes through various monitoring means
17 and providing early warnings for abnormal areas will strongly support the prevention
18 of large-scale deformations and landslides.

19 To fully represent the spatiotemporal feature of the expansive soil slope
20 monitoring data, the layout of the observation point is regarded as a two-dimensional
21 panel, with time processing as the third dimension.

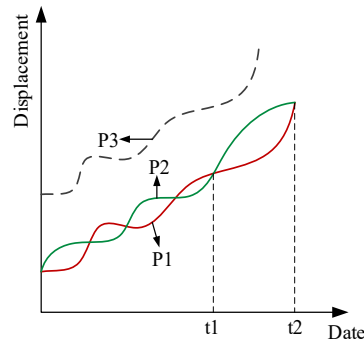
22 2 Spatiotemporal similarity indicators of deformation

23 2.1 Similarity feature selection

24 Expansive soil canal slope deformation monitoring is a systematic project, and
25 observation points at different positions have different deformation data. The essence
26 of clustering analysis for these deformation observation points is to extract similarity
27 features between different deformation sequences and then cluster the similar series. To
28 achieve a better clustering effect, it is necessary to determine which deformation
29 features to extract.

30 Fig. 1 is used to illustrate a selection of features (Chen et al., 2019). The curves
31 represent displacement time histories measured at three locations on the expansive soil
32 slope: P1, P2, and P3. The displacement histories of points P1 and P3 are similar, but
33 their initial values differ, so the final displacement values differ. The initial and final

1 values of points P1 and P2 are the same, but their time histories differ in different
 2 periods. For example, in the period from t1 to t2, the displacements at both points are
 3 increasing, but P1 is accelerating while P2 is decelerating. The most significant
 4 difference in magnitude and time history is between points P2 and P3. If the criterion
 5 for clustering is by similar time histories, points P1 and P3 should be classified into one
 6 cluster. If the criterion for clustering is by similar initial and final values, points P1 and
 7 P2 should be classified into one category. If the criterion for clustering increases the
 8 value of the monitored parameter, all points should be clustered into one category. A
 9 rational clustering method should consider several different deformation
 10 characteristics, including the change amount and the rate at each observation point
 11 during the entire period.



12
 13 Fig. 1 Development process lines of different deformation data

14 In clustering analysis, the distance measurement methods used mainly Euclidean
 15 distance, Canberra distance, and Mahalanobis distance (Ercanoglu et al., 2004). The
 16 square of the Euclidean distance, referred to here as SED, is more convenient to work
 17 with as it avoids the calculation of square roots. Since the roles of each indicator and
 18 deformation variable are different, it is necessary to calculate the weight of each
 19 indicator and deformation variable to obtain a weighted distance indicator. Therefore,
 20 the commonly used SED is replaced in this work with a "weighted distance function,"
 21 which captures the characteristics of the spatiotemporal deformation sequence.

22 Different measurement methods use different scales for a variable, e.g., interval
 23 scale, nominal scale, and ordered scale. The deformation monitoring frequency is
 24 usually once a week or once a day, and the deformation unit is generally millimetres.
 25 Therefore, the analysis of spatiotemporal deformation data in this work is based on the
 26 interval scale.

27 Let δ_{it} ($i = 1, 2, \dots, N; t = 1, 2, \dots, T$) denote a deformation value at measuring
 28 point i at time t , and d_{ij} denote the similarity of the deformation at observation points

1 i and j , also known as the "distance" between the points (BONZO and HERMOSILLA,
2 2002). This is defined by:

$$3 \quad d_{ijt}(SED) = (\delta_{it} - \delta_{jt})^2. \quad (1)$$

4 Large d_{ij} indicates a significant difference between deformations at points i and j ;
5 small d_{ij} indicates similar deformations at the two observation points.

6 The spatiotemporal clustering of the deformation requires an indicator system for
7 time series division and spatial observation point clustering. All similarity indicators in
8 the following text are based on the SED.

9 2.2 Temporal clustering indicators

10 The time series of deformation has a variety of modes, and the deformation can be
11 divided into different stages under different external environmental conditions. If the
12 deformation time series can be effectively divided, the deformation state and stage can
13 be identified, which is conducive to judging the safety state of the canal slope. Several
14 major similarity indicators for time series clustering are studied below.

15 2.2.1 Basic time indicators

16 To consider the role of different similarity features and define a temporal similarity
17 index, three fundamental similarity indicators based on SED are introduced: cross-
18 section "weighted absolute distance," cross-section "weighted incremental distance,"
19 and cross-section "weighted growth distance."

20 (1) The cross-section "weighted absolute distance" between time instance i and j
21 is denoted as $d_{ij}^T(AD)$ and given by:

$$22 \quad d_{ij}^T(AD) = \sum_{n=1}^N \sum_{m=1}^M WX_m [\mathbf{x}_{nm}(i) - \mathbf{x}_{nm}(j)]^2 \quad (2)$$

23 where WX_m is the weight of the m -th deformation variable x_m ; $\mathbf{x}_{nm}(i)$ is the
24 deformation value of m -th variable at observation point n at time instance i , ($n =$
25 $1, 2, \dots, N$, $m = 1, 2, \dots, M$). Here, $\mathbf{x}_{nm}(i) = \delta_{nm}(i)$; $\mathbf{x}_{nm}(j) = \delta_{nm}(j)$. Since
26 several deformation variables will be considered, WX_m is used to weigh the importance
27 of the m -th deformation variable x_m .

28 The value of $d_{ij}^T(AD)$ measures the distance between time instance i and j of the
29 M deformation variables at the N observation points. The closer the deformation values
30 at the two-time instances are, the smaller the value of $d_{ij}^T(AD)$ is, indicating similar
31 deformation at the two-time instances.

32 (2) The cross-section "weighted increment distance" between time instance i and
33 j is denoted by $d_{ij}^T(ID)$ and given by:

$$34 \quad d_{ij}^T(ID) = \sum_{n=1}^N \sum_{m=1}^M WX_m [\mathbf{y}_{nm}(i) - \mathbf{y}_{nm}(j)]^2 \quad (3)$$

1 where $\mathbf{y}_{nm}(i) = \mathbf{x}_{nm}(i) - \mathbf{x}_{nm}(i-1)$; $\mathbf{y}_{nm}(j) = \mathbf{x}_{nm}(j) - \mathbf{x}_{nm}(j-1)$.
 2 The value of $d_{ij}^T(ID)$ measures the distance between the deformation increments of the
 3 M variables at the N measuring points during a time interval. The smaller the value of
 4 $d_{ij}^T(ID)$, the more similar the deformation increments during the interval are.

5 (3) The cross-section "weighted growth rate distance" between time instance i and
 6 j is denoted by $d_{ij}^T(GRD)$ and given by:

$$7 \quad d_{ij}^T(GRD) = \sum_{n=1}^N \sum_{m=1}^M WX_m [\mathbf{z}_{nm}(i) - \mathbf{z}_{nm}(j)]^2 \quad (4)$$

$$8 \quad \mathbf{z}_{nm}(i) = \frac{\mathbf{y}_{nm}(i)}{\mathbf{x}_{nm}(i-1)} \quad (5)$$

$$9 \quad \mathbf{z}_{nm}(j) = \frac{\mathbf{y}_{nm}(j)}{\mathbf{x}_{nm}(j-1)} \quad (6)$$

10 The value of $d_{ij}^T(GRD)$ measures the distance between the relative deformation
 11 increments of the M variables at the N measuring points during a time interval. The
 12 smaller the value of $d_{ij}^T(GRD)$, the more similar the relative deformation increments
 13 during the interval are.

14 2.2.2 Cross-section "weighted comprehensive distance"

15 A temporal similarity index is defined by integrating the three similarity indicators.
 16 The index measures the overall similarity of the deformation at observation points at
 17 different times. It is referred to as the cross-section "weighted comprehensive distance"
 18 between time instance i and j , and denoted by $d_{ij}^T(CD)$:

$$19 \quad d_{ij}^T(CD) = \alpha_1 \cdot d_{ij}^T(AD) + \alpha_2 \cdot d_{ij}^T(ID) + \alpha_3 \cdot d_{ij}^T(GRD) \quad (7)$$

20 where $\alpha_1, \alpha_2, \alpha_3$ are the weights of the three fundamental time similarity indicators,
 21 respectively, $\alpha_1 + \alpha_2 + \alpha_3 = 1$, $\alpha_i > 0$ ($i = 1, 2, 3$).

22 2.3 Spatial clustering indicators

23 A clustering analysis of observation points with similar deformation sequences is
 24 applied to divide the deformation area. Fundamental similarity indicators and a spatial
 25 similarity index corresponding to spatial observation data clustering are defined.

26 2.3.1 Basic spatial indicators

27 Analogously to the fundamental temporal similarity indicators, three basic spatial
 28 similarity indicators are defined: the full-time "weighted absolute distance," the full-
 29 time "weighted increment distance," and the full-time "weighted growth rate distance."

30 (1) The full-time "weighted absolute distance" between observation points k and
 31 l is denoted by $d_{kl}^S(AD)$ and given by:

$$32 \quad d_{kl}^S(AD) = \sum_{m=1}^M \sum_{t=1}^T WX_m [\mathbf{x}_{mt}(k) - \mathbf{x}_{mt}(l)]^2 \quad (8)$$

33 where WX_m is the weight of the m -th deformation variable x_m ; $\mathbf{x}_{mt}(k)$ is the value of

1 the m -th deformation variable of the observation point k at time section t ($m =$
2 $1, 2, \dots, M, t = 1, 2, \dots, T$) ; $\mathbf{x}_{mt}(k) = \delta_{mt}(k)$; $\mathbf{x}_{mt}(l) = \delta_{mt}(l)$. The value of
3 $d_{kl}^S(AD)$ measures the distance between deformations at observation points k and l at
4 a given time instance. The closer the deformation values of the two observation points
5 are, the smaller the value of $d_{kl}^S(AD)$ is, indicating similar deformation at the two
6 observation points at the time instance.

7 (2) The full-time "weighted increment distance" between the observation point k
8 and l is denoted by $d_{kl}^S(ID)$ and given by:

$$9 \quad d_{kl}^S(ID) = \sum_{m=1}^M \sum_{t=1}^T WX_m [\mathbf{y}_{mt}(k) - \mathbf{y}_{mt}(l)]^2 \quad (9)$$

10 where $\mathbf{y}_{mt}(k) = \mathbf{x}_{mt}(k) - \mathbf{x}_{m,t-1}(k)$; $\mathbf{y}_{mt}(l) = \mathbf{x}_{mt}(l) - \mathbf{x}_{m,t-1}(l)$. The value
11 $d_{kl}^S(ID)$ measures the distance between the deformation increments at observation
12 points k and l at a given time instance from the last time. The smaller the value of
13 $d_{kl}^S(ID)$, the more similar the deformation increments at the two observation points are.

14 (3) The full-time "weighted growth rate distance" between the observation point k
15 and l is denoted by $d_{kl}^S(GRD)$ and given by:

$$16 \quad d_{kl}^S(GRD) = \sum_{m=1}^M \sum_{t=1}^T WX_m [\mathbf{z}_{mt}(k) - \mathbf{z}_{mt}(l)]^2 \quad (10)$$

$$17 \quad \mathbf{z}_{mt}(k) = \frac{y_{mt}(k)}{x_{m,t-1}(k)} \quad (11)$$

$$18 \quad \mathbf{z}_{mt}(l) = \frac{y_{mt}(l)}{x_{m,t-1}(l)} \quad (12)$$

19 The value of $d_{kl}^S(GRD)$ measures the distance between the relative deformation
20 increments at observation points k and l at a given instance from the last time. The
21 smaller the value of $d_{kl}^S(GRD)$, the more similar the relative deformation increments of
22 the two observation points are.

23 2.3.2 Full-time "weighted comprehensive distance"

24 Integrating the three fundamental spatial similarity indicators defines a spatial
25 similarity index. The index measures the overall similarity of the deformation at
26 different observation points. It is referred to as the full-time "weighted comprehensive
27 distance" between observation points k and l is denoted by $d_{kl}^S(CD)$:

$$28 \quad d_{kl}^S(CD) = \beta_1 \cdot d_{kl}^S(AD) + \beta_2 \cdot d_{kl}^S(ID) + \beta_3 \cdot d_{kl}^S(GRD) \quad (13)$$

29 where $\beta_1, \beta_2, \beta_3$ are the weights of the three fundamental spatial similarity indicators,
30 respectively, $\beta_1 + \beta_2 + \beta_3 = 1, \beta_i > 0 (i = 1, 2, 3)$.

31 2.4 Standardized methods for eigenvalues

32 The fundamental temporal and spatial indicators and the corresponding composite
33 similarity indices have inconsistent dimensions and magnitudes with potentially
34 significant differences, affecting the clustering results. Therefore, all the indicators are

1 standardized before clustering. The Z-score standardized method (Mohamad and
 2 Usman, 2013) can transform data into a normal distribution with zero means and unit
 3 variance and remove the effect of the feature dimension. This method is applied here
 4 for similarity indicators standardization. Consider N observation points and T time
 5 sections with recorded deformations. The Z-score standardized formula is given by:

$$6 \quad Z(x_{it}) = \frac{x_{it} - \mu_x}{\sigma_x} \quad (14)$$

$$7 \quad \mu_x = \frac{\sum_{i=1}^N \sum_{t=1}^T x_{it}}{N \cdot T} \quad (15)$$

$$8 \quad \sigma_x = \sqrt{\frac{\sum_{i=1}^N \sum_{t=1}^T (x_{it} - \mu_x)^2}{N \cdot T - 1}} \quad (16)$$

9 where μ_x is the arithmetic mean (mathematical expectation) of all x_{it} , and σ_x is the
 10 standard deviation of all x_{it} . The range of the standardized data is limited, which can
 11 effectively avoid the influence of different dimensions and value ranges for clustering
 12 performance.

13 2.5 Entropy weight method

14 The entropy weight method is a method for determining indicators' weights.
 15 According to information theory (Boer et al., 2005), information measures the order in
 16 a system, i.e., a negative entropy. The smaller the entropy of an indicator, the greater
 17 the information provided by the indicator. Hence an indicator with smaller entropy
 18 should be assigned a more considerable weight. The steps for determining the weight
 19 of an indicator by the entropy weight method (ZOU et al., 2006) are as follows:

20 Step1: Calculate the ratio between the value of the variable x_m at observation point
 21 n and all observation points

$$22 \quad q_{nm} = \frac{x_{nm}}{\sum_{n=1}^N x_{nm}}, m = 1, 2, \dots, M \quad (17)$$

23 Step 2: Calculate the entropy of the variable x_m .

$$24 \quad e_m = -k \sum_{n=1}^N q_{nm} \ln(q_{nm}), m = 1, 2, \dots, M \quad (18)$$

25 where $k = \frac{1}{\ln N}$ is assumed in this work. The smaller the e_m , the greater the effect of the
 26 variable x_m .

27 Step 3: Determine the variable's weight after standardizing the entropy value.

28 Since e_m is a reverse indicator after processing it positively and standardizing, the
 29 weight coefficient of the variable x_m is obtained.

$$30 \quad WX_m = \frac{1 - e_m}{\sum (1 - e_m)}, m = 1, 2, \dots, M \quad (19)$$

31 The value of $\alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2$ and β_3 in the temporal and spatial similarity
 32 indices are also obtained by these steps.

1 3 The spatiotemporal weighted clustering model

2 3.1 Principles of spectral clustering

3 Spectral clustering (SC) is based on spectral division theory (Luxburg, 2007). This
4 method uses eigenvectors of similar data matrices for clustering to make the algorithm
5 independent of the dimension of data points but only related to the number of data
6 points. Compared with other methods, SC is simple in idea, easy to implement, does
7 not easily fall into optimal local solutions, and can identify non-convex distribution
8 clustering, which can be applied to many practical application problems.

9 The SC algorithm first defines an affinity matrix describing the similarity of paired
10 data points according to the given sample dataset, then calculates the eigenvalues and
11 eigenvectors of the matrix, and finally selects the appropriate eigenvectors to cluster
12 different data points. SC was first applied to computer vision (Yu and Shi, 2003), VLSI
13 design (Weiss, 1999), and other fields. It is currently widely studied in behaviour
14 recognition (Mohamad and Usman, 2013) and text image segmentation (Boer et al.,
15 2005), which is one of the research hotspots in machine learning. Specific SC
16 algorithm-related content can be found in the literature (Luxburg, 2007) and will not be
17 described in this paper.

18 3.2 Validation index of cluster number

19 It is necessary to conduct a comparative analysis through some indexes to evaluate
20 the clustering effect and determine the appropriate number of clusters. The main
21 indexes used are the Davies-Bouldin index (DB index), Dunn index, Rousseeuw's
22 silhouette value (Silhouette value), and Calinski-Harabaz index (CH index).

23 (1) DB index

24 The definition of the DB index (Karo et al., 2017) is as follows:

$$25 \quad DB = \frac{1}{M} \sum_{i=1}^m \max_{i \neq j} (d_{ij})_{j=1,2,\dots,M} \quad (20)$$

$$26 \quad d_{ij} = \frac{\sigma_i + \sigma_j}{d(c_i, c_j)} \quad (21)$$

27 where M is the number of clusters, d_{ij} is the distance between the two samples, σ_i is
28 the average distance from all patterns in cluster i to the center cluster c_i , and $d(c_i, c_j)$
29 is the distance between cluster centers c_i and c_j . If the clusters i and j are compact and
30 their centres are far from each other, the d_{ij} value will be small. Therefore, the DB
31 index gets a good partition by looking for the minimum value. Smaller DB values
32 indicate a better clustering effect.

33 (2) Dunn index

1 The Dunn index (Arbelaitz et al., 2013) can be calculated by Eq. (22):

$$2 \quad D = \frac{d_{min}}{d_{max}} \quad (22)$$

3 where d_{min} is the shortest distance between two elements in different categories, and
4 d_{max} is the farthest distance between the above two elements. The more significant the
5 Dunn index, the better the clustering effect.

6 (3) Silhouette value

7 Silhouette value (Struyf et al., 1996) can be calculated by Eq. (23):

$$8 \quad s(i) = \frac{b(i)-a(i)}{\max\{a(i),b(i)\}} \quad (23)$$

9 where $a(i)$ is the average distance of i to all other nodes in cluster a , while $b(i)$ is the
10 average minimum distance from i to all nodes in clusters other than cluster a . The
11 larger the Silhouette value, the better the clustering effect.

12 (4) CH index

13 CH index (Krasnov and Sen, 2019) is a widely used non-labelled clustering effect
14 evaluation index. The higher the value, the closer the cluster itself, the more dispersed
15 the cluster, and the better the performance of the clustering algorithm. Cluster cohesion
16 is calculated based on the distance of each data in the cluster to the centroid. The cluster
17 separation degree is calculated based on the distance from the cluster's centroid to the
18 global centroid. The index can be calculated by the formula (24):

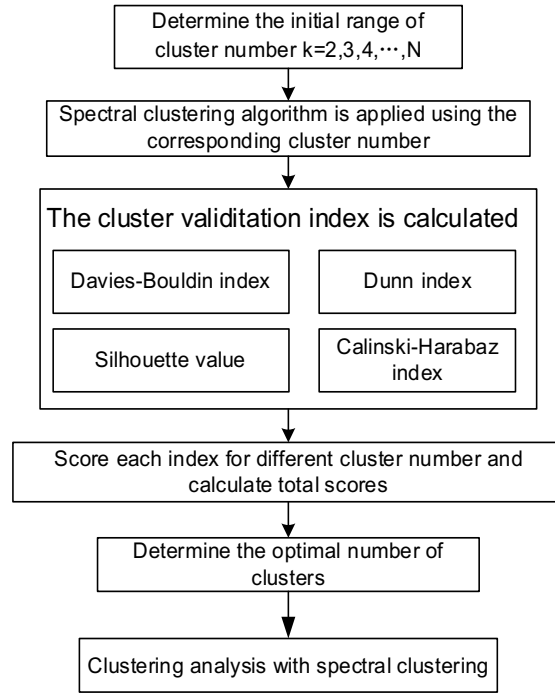
$$19 \quad \lambda(N_c) = \frac{\min D_{ij}}{\max D_j} \quad (24)$$

20 where D_{ij} is the distance between the centres of the two classification clusters, and D_j
21 is the distance set within each cluster. The optimal cluster obtained when $\lambda(N_c)$ reaches
22 a peak, and when $\lambda(N_c)$ reaches a global maximum, N_c is the optimal value. The value
23 of the CH index has an extensive range, so all the obtained CH indexes will be
24 standardized by the Z-score standardized method to facilitate subsequent analysis of the
25 results.

26 3.3 Scoring algorithm

27 The elbow method is generally applied to determine the optimal number of spectral
28 clustering clusters (Syakur et al., 2018). However, the number of clusters determined
29 by one index is not necessarily referenced. This study determines the optimal number
30 of clusters by calculating the above four indexes and then using a voting scoring
31 algorithm (X. Li et al., 2021). The voting scoring algorithm is described as follows: for
32 a single index, the score of the optimal number of clusters determined by the index is
33 P . The number of clusters that are second to the effect is scored $P - 1$, and so on, to

1 determine the score of each cluster. Then add the scores of each cluster of the above
 2 four indexes to get the total score of each cluster. The highest-scoring number of clusters
 3 is selected as the final cluster number, and the result obtained by clustering these data
 4 is the optimal clustering result. This scoring algorithm effectively combines the
 5 advantages of different clustering effectiveness indexes and weakens the limitations of
 6 individual indexes. The implementation flowchart of the clustering analysis is shown
 7 in Fig. 2.



8
9

Fig. 2 Spectral clustering analysis implementation flow chart

10 3.4 Implementation steps of weighted clustering analysis

11 A weighted clustering model is established based on the spatial-temporal similarity
 12 indicators and SC. First, the deformation sequences are divided into several time
 13 sections. Then the spatial observation points are clustered to realize the effective
 14 clustering of the deformation area based on the division of the time sections. The model
 15 establishment and analysis process are shown in Fig. 3. The basic steps are as follows:

16 Step 1 Standardize the deformation data to obtain pre-processed data.

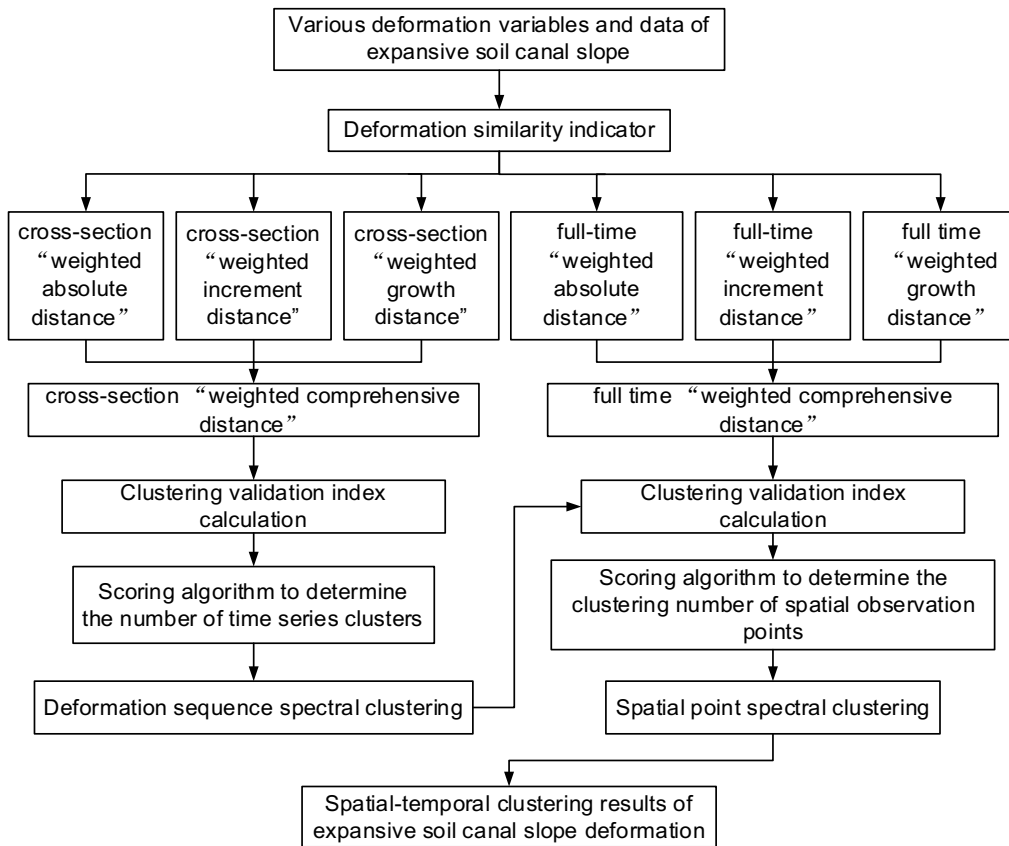
17 Step 2 Calculate the indicator weights by Eqs. (17) ~ (19).

18 Step 3 Use Eqs. (2) ~ (4) to calculate $d_{ij}^T(AD)$, $d_{ij}^T(ID)$, $d_{ij}^T(GRD)$.

19 Step 4 Calculate the weights of the indicators $d_{ij}^T(AD)$, $d_{ij}^T(ID)$ and $d_{ij}^T(GRD)$
 20 after standardization.

21 Step 5 According to Eq. (7), the cross-section "weighted comprehensive distance"
 22 $d_{ij}^T(CD)$ is calculated, and the scoring algorithm determines the number of time series

1 clusters. The deformation sequence spectral clustering is performed.
 2 Step 6 Based on the time series clustering results, using Eqs. (8) ~ (10) to calculate
 3 $d_{kl}^S(AD)$, $d_{kl}^S(ID)$, $d_{kl}^S(GRD)$.
 4 Step 7 Calculate the full-time "weighted comprehensive distance" $d_{kl}^S(CD)$ of the
 5 observation points according to Eq. (13). Determine the number of spatial measuring
 6 points clusters by scoring algorithm, perform spectral clustering of the spatial
 7 measuring points, and obtain the clustering analysis results.
 8 Through the spatiotemporal clustering analysis, the similarity period and area of
 9 the deformation can be clustered to realize the state analysis of the expansive soil canal
 10 slope deformation.



11
 12 Fig. 3 Implementation process of spatiotemporal clustering

13 4 Engineering application and discussion

14 This section takes the deformation of the expansive soil canal slope in the middle
 15 line of the China South-to-North Water Diversion Project as an example to evaluate the
 16 validity of the spatiotemporal clustering analysis model for deformation.

17 4.1 Engineering background

18 Expansive soils are distributed in Nanyang, Shahe, Handan, and other areas in the

1 central canal of China's South-to-North Water Diversion Project. The cumulative length
 2 of the expansive soil slope is 346.85km, accounting for about 27% of the total length
 3 of the central canal (1266.495km). Due to the unique properties of the expansive soil,
 4 part of the expansive soil canal slope has encountered some problems during
 5 construction and operation. Therefore, it is of great significance to identify the
 6 deformation anomaly area for the long-term operation of the project. Since the project
 7 was put into operation on December 12, 2014, a large amount of canal slope monitoring
 8 data has been accumulated after seven years of operation.

9 The expansive soil canal slope of the left bank section with stakes 9+120~9+363
 10 in Nanyang is selected for research. The cross-section structure of this canal section is
 11 shown in Fig. 4. The bottom width of the channel is 13.5m, the slope ratio of the water
 12 crossing section is 1:3.0, and the width of the first berm is 5m. The other berm is set up
 13 every 6m above, and the slope ratio between the first-grade berm and the fourth-grade
 14 berm is 1:2.5. The whole section of the canal slope is refilled with cement-modified
 15 soil, of which the thickness of cement-modified soil in the water crossing section is
 16 1.5m, and above the first berm is 1.0m. The concrete arch skeleton and grass planting
 17 in the arch protect the slope. Longitudinal drainage ditches are arranged on all grades
 18 of berms, and horizontal drainage ditches are constructed above the first-grade slope
 19 and berm. The canal section construction was completed in December 2013 and
 20 implemented in December 2014.

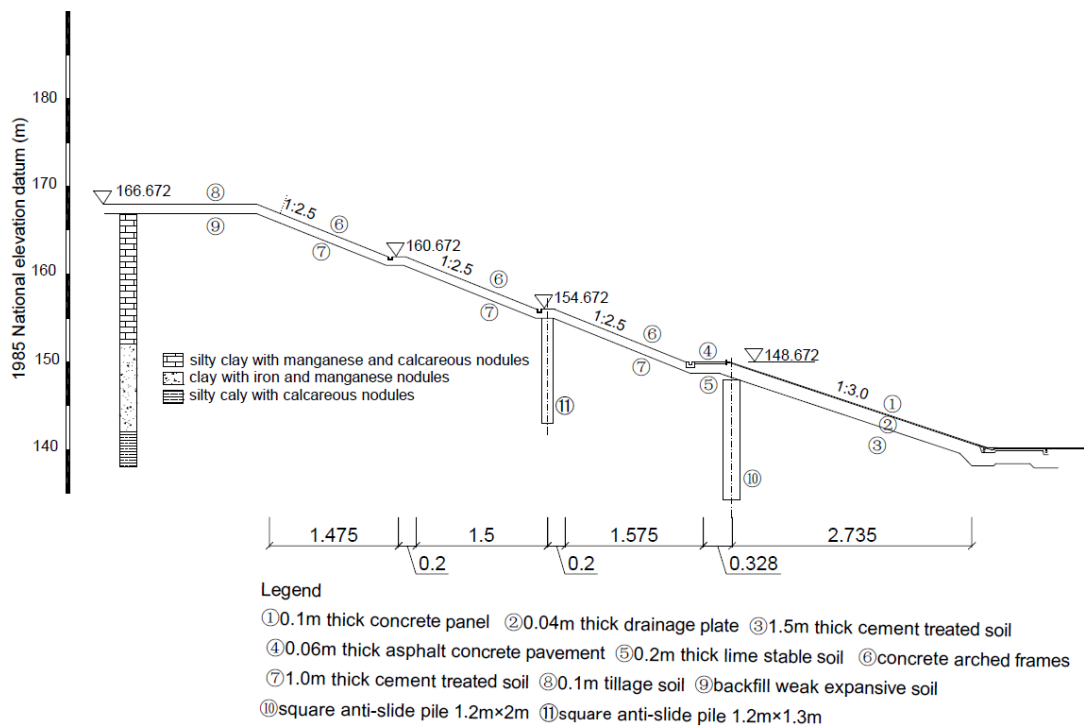


Fig. 4 9+120~9+363 canal slope cross-section structure

21
 22

1 This section of the canal slope is mainly composed of Quaternary Pleistocene silty
2 clay and calcareous nodule silty clay. The layer description is as follows:

3 Quaternary Middle Pleistocene (a1~p1Q2): the first layer is silty clay, brown and
4 brownish yellow, with no uniform boundary between soil colors, rigid plastic,
5 containing ferromanganese and calcareous concretions. The content of calcareous
6 nodules is 5~15%, enriched in local lumps. This layer is distributed at the elevation of
7 152.00m~167.00m, and the thickness generally exceeds 15m. The second layer
8 predominantly consists of clay, grayish yellow, light brownish yellow, hard plastic, iron
9 and manganese plaques, and occasionally ginger stones. These layers distributed in a
10 lenticular manner, distributed between 152.00m and 135.00m in elevation, and the
11 thickest part of the clay layer is located at stake 9+065. The third layer is mainly
12 composed of calcareous nodule silty clay, the overall color is yellowish brown, the
13 content of calcareous nodules is between 50-60%, the particle size is generally 0.01-
14 0.04m, and the fine-grained soil is silty clay, hard plastic. It is distributed between
15 142.00m and 138.20m in elevation, generally 4.0~6.5m thick.

16 The stratification of fissure development is described as follows: the first layer is
17 weak to moderate expansion. The top vertical fissures are developed, the micro, small
18 and large fissures are developed, and the growing fissures are developed—the fissures
19 towards the NE in this layer. The second layer is moderate expansion. Micro and small
20 fissures are developed, and large and growing fissures are less developed. The third
21 layer is moderately expanded, and no fissures are developed.

22 The section of the canal slope is equipped with 15 inclinometer pipes, 15 surface
23 horizontal displacement monitoring points, and 15 vertical displacement monitoring
24 points. The specific layout of deformation monitoring instruments is shown in [Fig. 5](#).
25 The displacement measured by the inclinometer tube is divided into A and B directions.
26 The A direction (internal horizontal displacement) is perpendicular to the channel water
27 flow direction. The B direction is the channel water flow direction. The deformation
28 amount in the B direction is small. Therefore, the displacement in the B direction is not
29 analysed here. The A direction displacement is positive when pointing to the hollow
30 surface of the canal slope and negative otherwise; the horizontal surface displacement
31 is positive towards the channel's centreline and negative towards the outside of the
32 channel; the vertical displacement is positive for sinking and negative for rising.

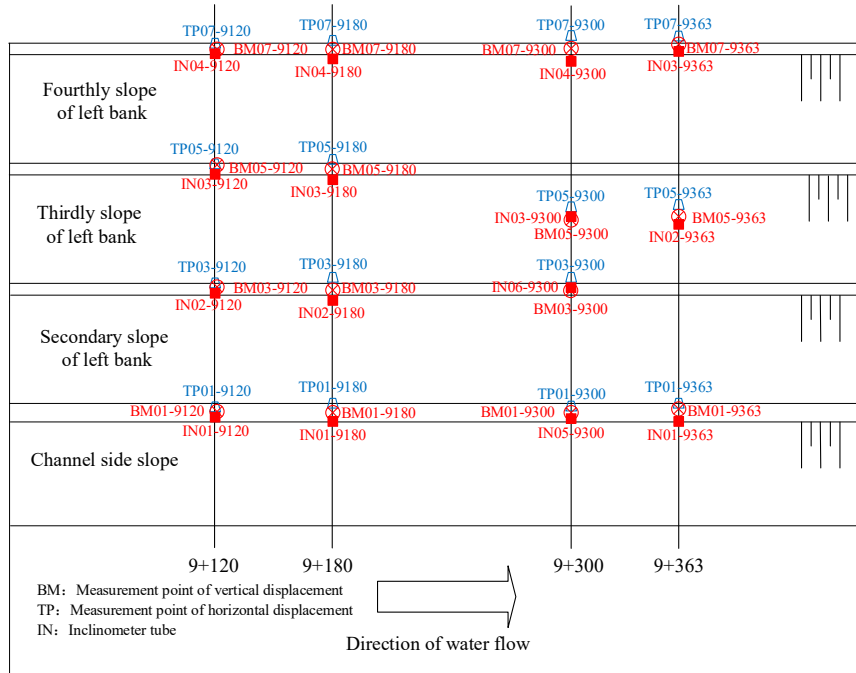
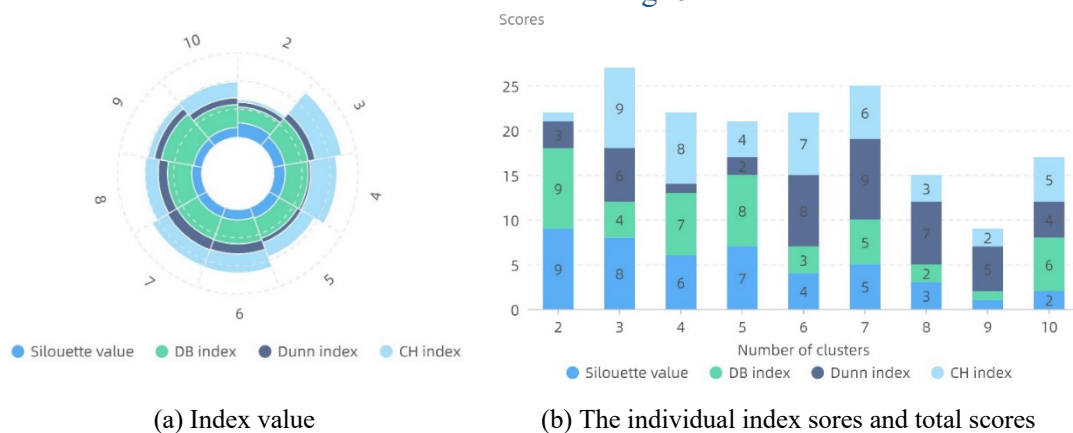


Fig. 5 Schematic diagram of the layout of deformation monitoring instruments

The deformation monitoring data and related standardized derivative similarity indicators of the four sections of 9+120, 9+180, 9+300, and 9+363 of the past four years from October 10, 2017, to May 21, 2021, were analysed. The process line of the deformation values (where Fig. 7 shows the maximum internal horizontal displacement of the inclination tube) of all the observation points in the canal section are shown in Fig. 7~9.

4.2 Time section division of deformation sequence

Firstly, the deformation period is divided. The range of the number of clusters is set to [2, 10], and the DB index, Silhouette value, Dunn index, and CH index for different cluster numbers are calculated, respectively. Then use the scoring algorithm mentioned in Section 3.3 to calculate the scores for all numbers of clusters. The calculated index values and scores are shown in Fig. 6.



1 Fig. 6 Evaluation of clustering indexes in the scoring algorithm

2 In Fig. 6(a), the numbers surrounding the circle are the number of clusters, and the
3 radius is the corresponding index value. In Fig. 6(b), the number in the graph represents
4 the score of each index at the corresponding number of clusters, and the total height of
5 the bar chart is the total score.

6 As shown in Fig. 6(a), the Silhouette value and DB index indicate that the
7 clustering effect is optimal when the number of clusters is 2, while the CH index leads
8 to the exact opposite conclusion. The CH index gets the maximum value when the
9 number of clusters is 3. Using a single index to determine the number of clusters proves
10 unreliable. As can be obtained from Fig. 6(b), when the number of clusters is 3, the
11 total score of all the indexes is the highest, and the highest score is 27. So, the optimal
12 number of clusters for the time section division is 3. This result further divides and
13 analyses the deformation in different time sections.

14 The time section division results of the internal horizontal, surface horizontal, and
15 vertical surface displacement are shown in Figs. 7-9. The first stage is the deformation
16 onset stage from June 27, 2017, to June 9, 2018. In this stage, the slope's horizontal
17 direction surface and vertical surface displacement grow slowly and steadily.

18 The second stage is from June 20, 2018, to September 13, 2020, which is the
19 deformation rapid development stage, in which the deformation shows an increasing
20 trend. The vertical surface displacement grows relatively smooth without jumping
21 increases. Some observation points' internal and horizontal surface displacements have
22 a significant jump.

23 The third stage is from November 9, 2020, to May 21, 2021, which is the
24 deformation convergence stage. No displacement has changed significantly during this
25 stage, and the deformation tends to converge. The horizontal surface displacement still
26 shows an increasing trend, but the increase rate is substantially lower than in the second
27 stage. This slope is in a safe state at stage 3.

28 The time-section division results of this model are close to the actual slope
29 deformation state, which shows the reasonability and the ability to distinguish the
30 deformation development stage by dividing the time section of the monitoring sequence
31 through the proposed model.

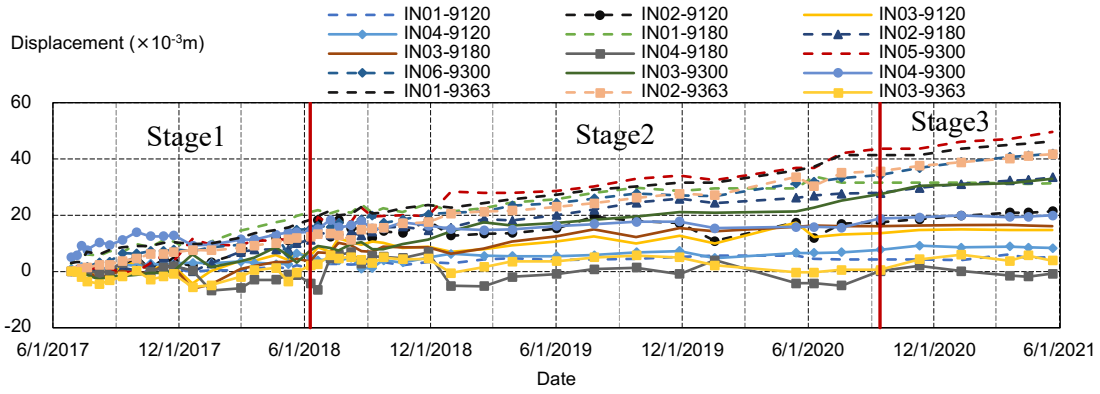


Fig. 7 The process line of the maximum internal horizontal displacement

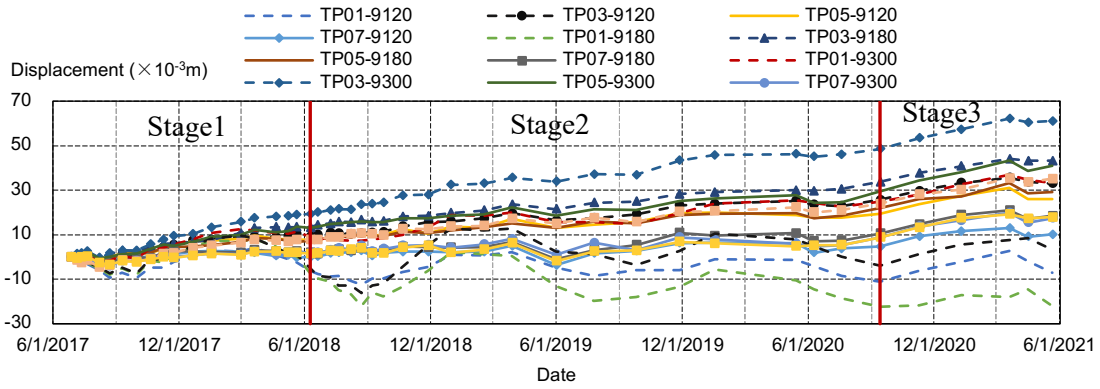


Fig. 8 The process line of horizontal surface displacement

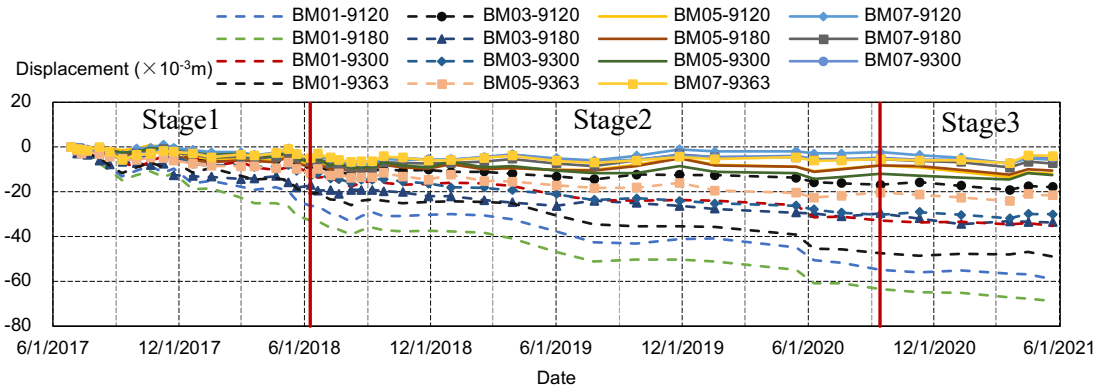


Fig. 9 The process line of vertical surface displacement

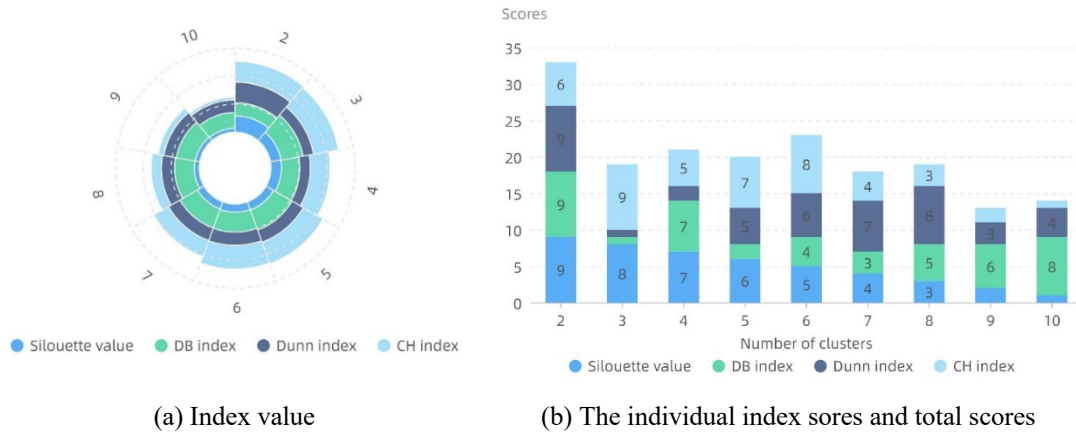
4.3 Clustering of spatial observation points

According to the time section division results, spatial clustering analysis was carried out of the observation points in the three stages.

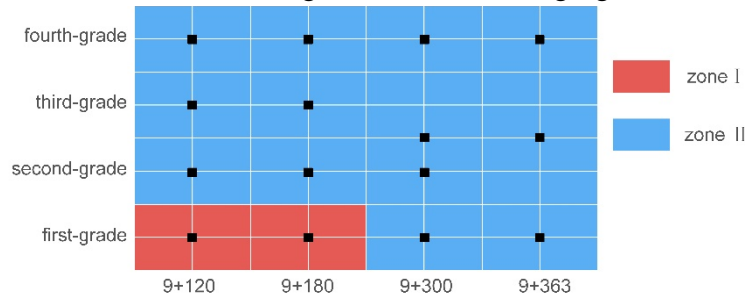
4.3.1 Deformation onset stage

Perform spatial clustering on the deformation features of the first stage and set the range of the number of clusters to [2, 10]. The calculated different index values and scores are shown in Fig.10. The CH index shows the optimal number of clusters is 3, while all other indexes indicate that the clustering effect is supreme when the number

1 of clusters is 2. The total score ranks first when the number of clusters is two can be
 2 obtained from Fig. 10(b), so the best number of clusters is used in the following
 3 clustering analysis.



4 Fig. 10 Evaluation of clustering indexes in the scoring algorithm of stage 1



5
 6 Fig. 11 Clustering results in deformation onset stage

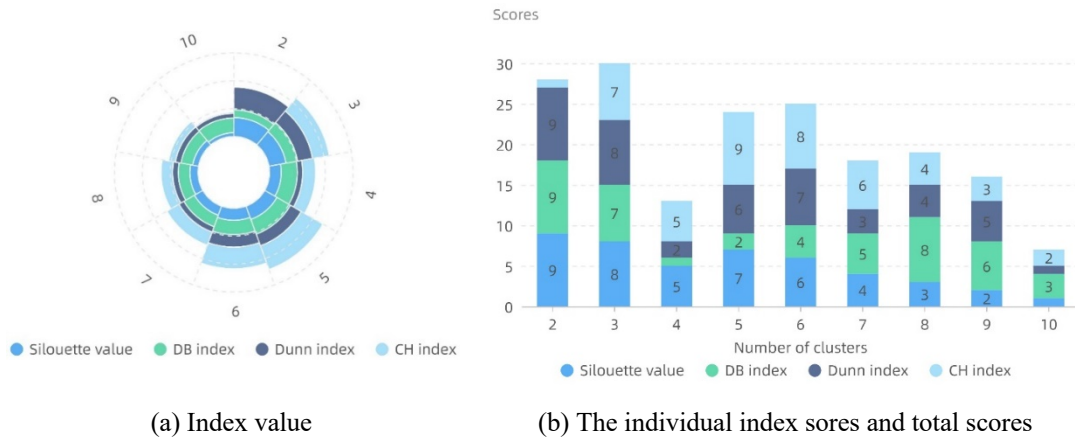
7 The clustering result obtained after clustering the spatial observation points is
 8 shown in Fig. 11. The black dots in the figure represent the corresponding observation
 9 points in Fig. 5. The first grade in the figure means the first-grade bridle path, and so
 10 on. Areas of different colours represent the clustering results.

11 It can be seen from Fig. 11 that in the first stage, the expansive soil canal slope
 12 section is divided into two regions with similar deformation, among which the first-
 13 grade bridle path at the 9+120 and 9+180 sections are similar to zone I. The remaining
 14 are the similarity zone II.

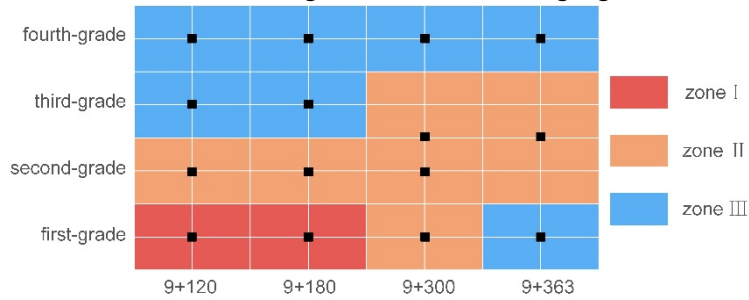
15 4.3.2 Deformation rapid development stage

16 Perform spatial clustering of the deformation rapid development stage, set the
 17 range of the number of the cluster to [2, 10], and different index values and scores are
 18 calculated and shown in Fig. 12. The Silhouette index, DB index, and Dunn index
 19 indicate that clustering effect is optimum when the number of clusters is 2. In contrast,
 20 the CH index leads to the exact opposite conclusion. The number of 3 clusters achieved
 21 the highest total score. Therefore, when clustering the deformed feature spaces at this

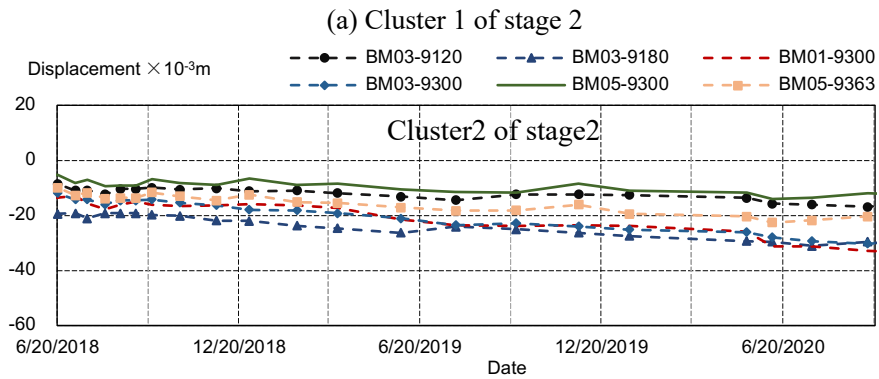
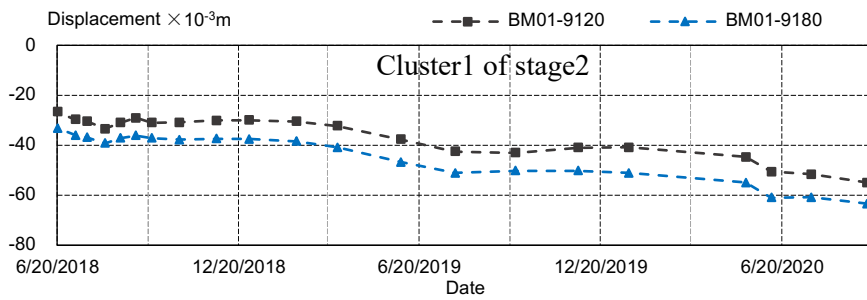
1 stage, the spaces are divided into 3 clusters. The spatial clustering results are shown in
 2 Fig. 13, and the vertical displacement process lines in different regions of this stage are
 3 shown in Fig. 14.



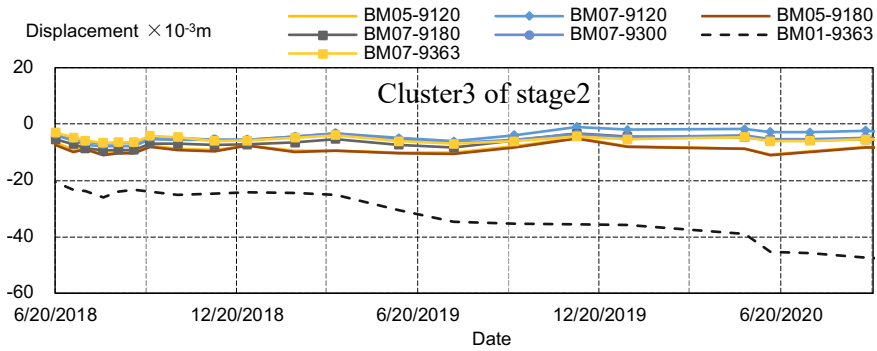
4 Fig. 12 Evaluation of clustering indexes in the scoring algorithm of stage 2



5
 6 Fig. 13 Clustering results in deformation rapid development stage



7
 8
 9
 10



(c) Cluster 3 of stage 2

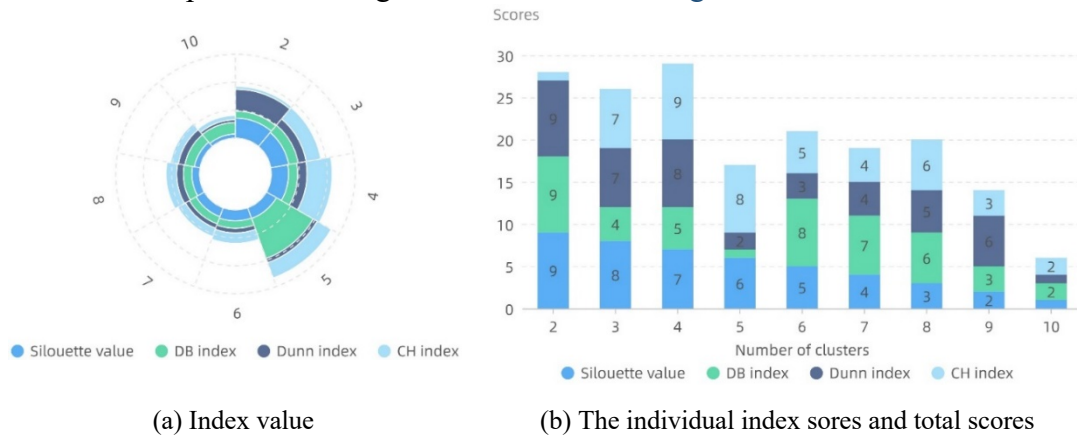
Fig. 14 Vertical displacement clustering results in stage 2

It can be seen from Fig. 13 that the canal slope is divided into three regions with similar deformation in the second stage, among which the first-grade bridle path at the 9+120 and 9+180 sections is still zone I. The 9+120 second-grade bridle path, the 9+180 second-grade bridle path, the 9+300 first and second berm, and the third-grade slope of the 9+300 and 9+363 sections formed zone II. The remaining area is zone III.

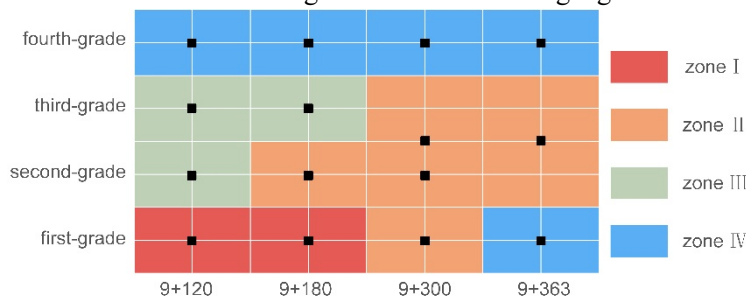
It can be seen from the vertical displacement process line in Fig. 14 that there is a $6.5 \times 10^{-3} \text{m}$ displacement difference between the observation points in zone I. However, the changing trend is almost the same, so the two points are still clustered into similar areas. By September 13, 2020, the vertical displacement range of the observation point in zone I is $-5.5 \times 10^{-2} \text{m} \sim -6.5 \times 10^{-2} \text{m}$. The development trend in zone II is relatively close. The displacement growth rate in zone II is lower than that in zone I. By September 13, 2020, the vertical displacement range in zone II is $-1.0 \times 10^{-2} \text{m} \sim -3 \times 10^{-2} \text{m}$. Except for BM01-9363, the vertical displacement has less growth in zone III. The vertical displacement of BM01-9363 has increased significantly, and it seems that the clustering result is not reasonable in terms of vertical displacement results alone. Since the clustering results comprehensively consider the three deformation variables of the canal slope, the internal horizontal displacement of the third cluster in the second stage and the horizontal surface displacement are comprehensively considered. It is found that the deformation trend and deformation amount of the internal and surface horizontal displacement of BM01-9363 are closer to those of the observation point in zone III. Therefore, the BM01-9363 finally belongs to zone III. The clustering results can reflect the similar degree of deformation values, the similarity of growth trends, and the comprehensive effect of various deformation variables, reflecting the advantages of the indicators proposed.

1 4.3.3 Deformation convergence stage

2 Perform spatial clustering on the deformation features of the third stage, and set
 3 the range of clusters to [2, 10]. The calculated different index values and scores are
 4 shown in Fig. 15. The Silhouette index, DB index, and Dunn index all show the
 5 maximum number of clusters is 2, while the CH index shows the optimal one is 4.
 6 Combining the total scores of all the indexes, the number of 4 clusters got the highest
 7 score 29. The spatial clustering results are shown in Fig. 16.



8 Fig. 15 Evaluation of clustering indexes in the scoring algorithm of stage3



9
 10 Fig. 16 Clustering results in the deformation convergence stage

11 It can be seen from Fig. 16 that in the stage of deformation tending to converge,
 12 the slope of the expansive soil canal is clustered into 4 clusters in the spatial area. The
 13 9+120 and 9+180 first-grade bridle paths are still in zone I. Zone II consists of the
 14 9+180 second-grade bridle path, 9+300 first and second berm, 9+300 and 9+363 third-
 15 grade slope. Zone III is composed of 9+120 second-grade, third-grade, and 9+180 third-
 16 grade bridle path. The remaining area is zone IV.

17 4.4 Discussion

18 4.4.1 Comparative analysis of spatiotemporal clustering results with engineering

19 Through on-site inspection, it was found that the drainage pipes had water seepage
 20 in the second slope of the 9+180 (Fig. 17) and 9+300 (Fig. 18) sections. The concrete

1 lining plate had arched in the first slope of the 9+295 (Fig. 19) and the 9+320 (Fig. 20)
2 sections. It is inferred that the groundwater in the abovementioned area is high, resulting
3 in drainage pipe seepage and lining arching. In addition, the above regions are all
4 located in zone II, reflecting the rationality of the spatial clustering.

5 The results of SIR-3000 geological radar detection on different canal slopes show
6 that there are significant irregular scattering waves, firm reflection surfaces, abnormal
7 vertical section waves, and irregular radar reflection wave waveforms in the
8 9+115~9+180 section of the first berm (Fig. 21). It is inferred that the soil in this area
9 is a soft interlayer zone, and the water content is abnormally high. These areas are all
10 located in Zone I, indicating the rationality of spatial clustering.



Fig. 17 Second slope in the 9+180



Fig. 18 Second slope in the 9+300

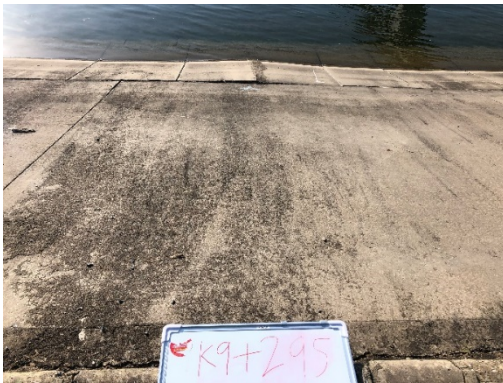
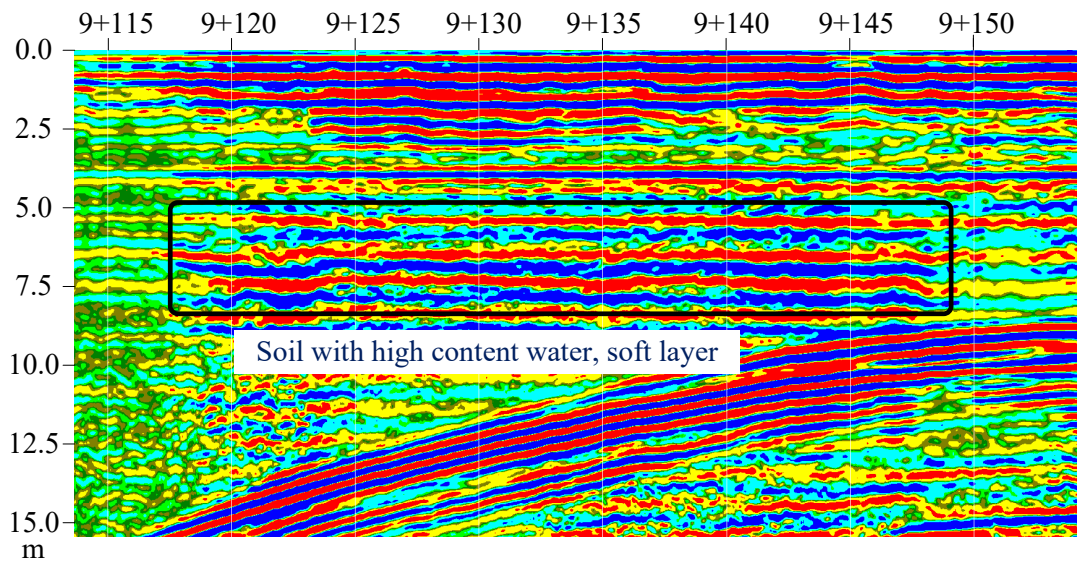


Fig. 19 Lining plate in the 9+295

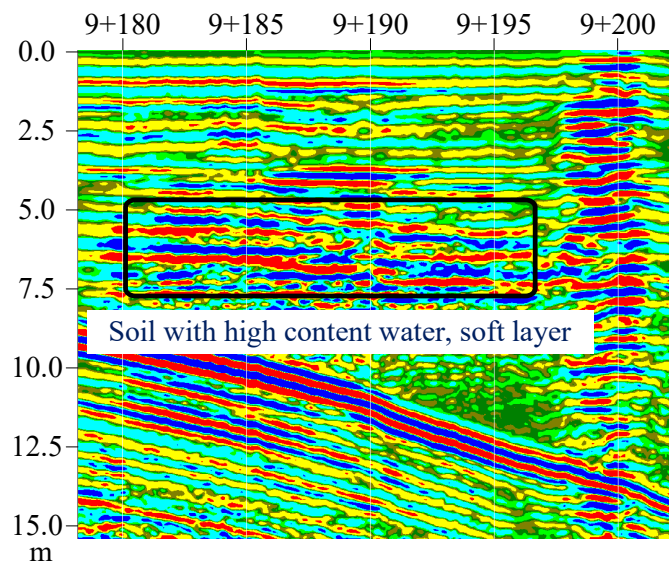


Fig. 20 Lining plate in the 9+320



1
2

(a) The first berm of 9+115~9+150



3
4

(b) The first berm of 9+180~9+200

Fig. 21 Geological radar interpretation map

6
7
8
9
10

It can be seen from the above clustering results that the spatiotemporal weighted clustering method proposed comprehensively analyses the deformation in the different periods and different zones based on proposed similarity indicators. It can identify the areas with large and rapidly growing deformation and provide early warning for the expansive soil canal slope landslide.

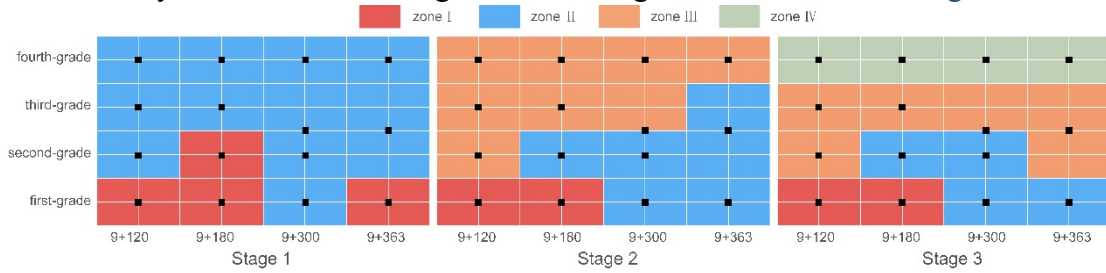
11

4.4.2 Comparative analysis of spatiotemporal weighted and common clustering

12
13
14

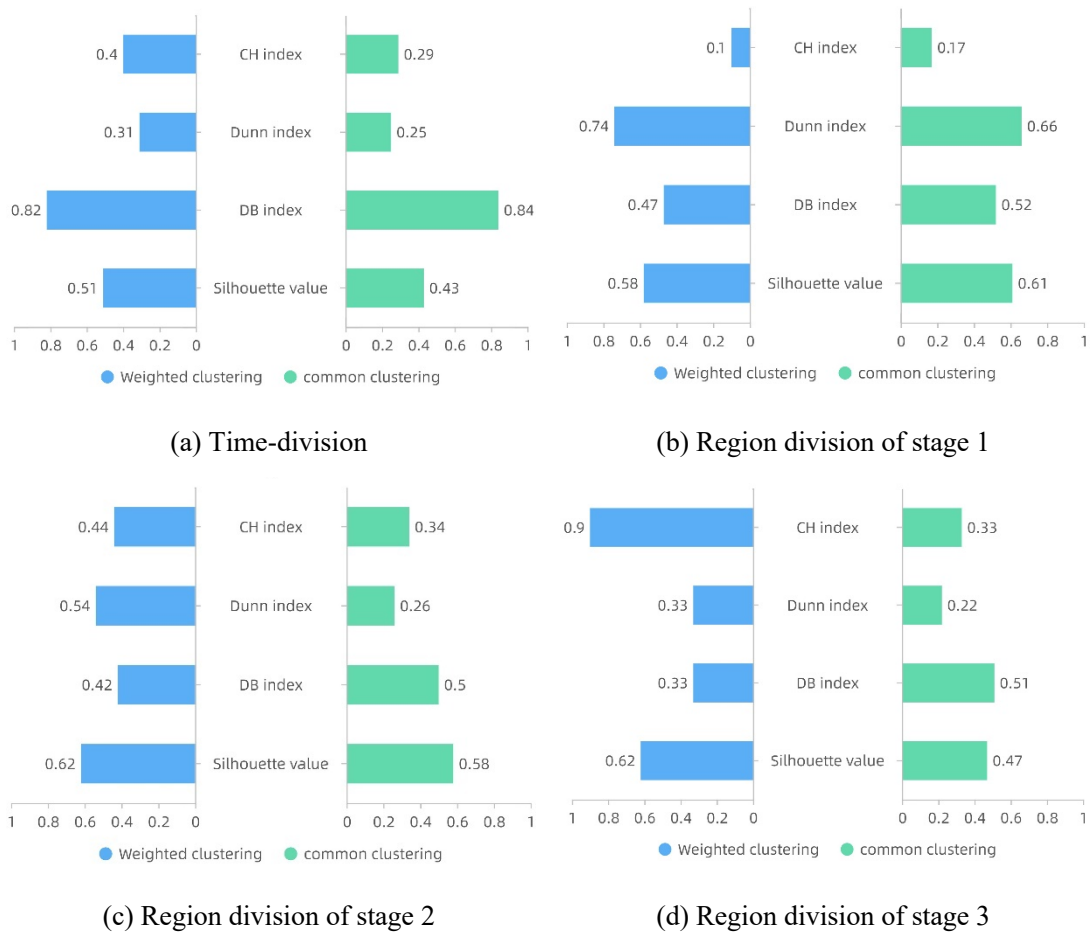
In the common clustering analysis of deformation data without considering the index weight and similarity indicators, the clustering analysis is carried out directly from the vertical displacement. Spectral clustering is carried out in the time section and

1 regional division of the above observation points. The time section division results are
 2 from June 27, 2017, to May 8, 2018, May 20, 2018, to January 17, 2020, and May 14,
 3 2020, to May 21, 2021, and the regional clustering results are shown in Fig. 22.



4
5 Fig. 22 Common clustering results

6 Both clustering methods divide the vertical displacement sequence into three-time
 7 sections. The expansive soil canal slope is divided into two similar deformation regions
 8 in the first stage, three similar deformation regions in the second stage, and four similar
 9 deformation regions in the third stage. The clustering effect evaluation index values of
 10 the proposed spatiotemporal weighted clustering and common clustering for time
 11 sections and region division for the divided time stage are shown in Fig. 23.



12
13 Fig. 23 Clustering effect evaluation index for the two methods

1 It can be seen from the effect evaluation indexes of the time division that the
2 Silhouette value, Dunn index, and CH index values of weighted clustering are more
3 significant than the common clustering. The DB index value is smaller than the
4 common clustering. The smaller the DB index value, the better the clustering effect.
5 The other three indexes are that the more significant the value, the better the clustering
6 effect. Therefore, the above index value indicates that the effect of weighted time
7 division clustering is better than common clustering.

8 In the first stage, the values of the Silhouette index and Dunn index of weighted
9 clustering are more significant than the common clustering, while the DB index is
10 smaller than the common clustering. The above three indexes show that the clustering
11 performance of weighted clustering is better than the common clustering at this stage,
12 but the CH index value of the weighted clustering is smaller than that of the common
13 clustering. It is speculated that the reason is that the BM03-9180 and BM01-9363 are
14 clustered into zone II of the weighted clustering, while the common clustering cluster
15 the above two observation points into the zone I.

16 The CH index value reflects the internal compactness of the clustering results.
17 From the common clustering results (Fig. 24), the deformation value of the zone where
18 the above two observation points are located is relatively large, which is close to the
19 other two observation points in zone I. However, combining all the evaluation index
20 results and the deformation variable clustering results shows that the effect of weighted
21 clustering is better than common clustering. The above phenomenon also indicates that
22 the clustering effect evaluation should comprehensively consider a variety of indexes.
23 If the clustering effect is evaluated only through a single index, it may cause a bias in
24 assessing the clustering effect.

25 In the second and third stages of weighted clustering, the Silhouette index, Dunn
26 index, and CH index values of weighted clustering are more significant than the
27 common clustering. In contrast, the DB index value is smaller than the common
28 clustering. All four indexes show that weighted clustering has a better effect than
29 common clustering.

30 The above results show that the performance of weighted clustering is better than
31 common clustering. It can also comprehensively reflect the magnitude and development
32 trend of the expansive soil canal slope deformation, which has obvious advantages for
33 identifying abnormal areas of canal slopes.

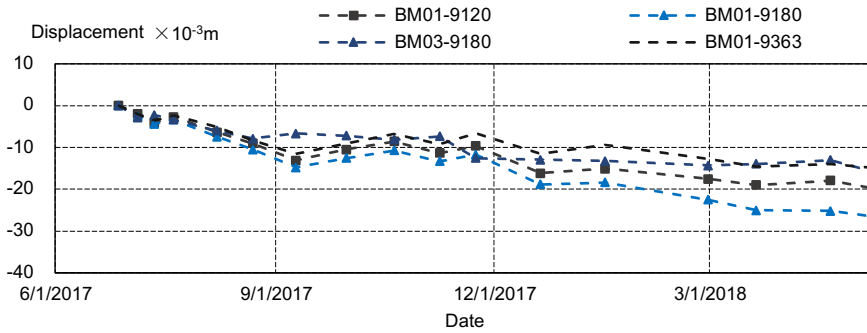


Fig. 24 Vertical displacement process line of common spatial clustering in area I

Furthermore, an evaluation of the time complexity for both algorithms was conducted using MATLAB on a central processing unit of Intel(R) Core(TM) i7-9700H with 16GB of RAM (Zhang et al., 2021b, 2021a). The resulting total time for both clustering methods on the monitoring data of the expansive soil canal slope is provided in Table 1. It can be observed from the table that the proposed spatiotemporal weighted clustering method takes more time compared to the common clustering method, as it has a higher computational cost. This is due to the need for pre-processing of the monitoring data, which results in a longer calculation time than the common clustering method.

Table 1 Execution time (in seconds) of the methods in the monitoring data set

Method	Times(s)
Spatiotemporal weighted clustering method	302.2
Common clustering method	98.1

5 Conclusions

According to the main characteristics of expansive soil canal slope deformation, weighted similarity indicators are constructed from the perspective of spatiotemporal clustering and consider the influence of various deformation monitoring data. The entropy method is used to determine the variable monitoring weight, the comprehensive spatiotemporal indicators, and the corresponding calculation method is proposed. The spectral clustering method offers a way for dividing the time section and clustering the spatial area of the observation point. A spatiotemporal clustering model for the expansive soil canal slope deformation is established. By the proposed method, the spatiotemporal clustering analysis of the expansive soil canal slope of the Taocha section of the middle line of the South-to-North Water Diversion Project is carried out, and the following conclusions are obtained.

(1) The similarity indicators and comprehensive similarity indicators proposed are

1 reasonable and can comprehensively reflect the state of expansive soil canal slope
2 deformation. The adopted spectral clustering method is quick and straightforward to
3 implement and can divide the deformation sequence into the onset stage, the high-speed
4 development stage, and the convergence stage. Moreover, it can realize the division of
5 similar regions in the corresponding stage to reflect the deformed state of the expansive
6 soil canal slope in different periods.

7 (2) By comparing the spatiotemporal and common clustering methods, it is found
8 that the proposed spatiotemporal weighted clustering method has better resolution
9 ability and can comprehensively reflect the magnitude and development trend of
10 various monitoring data in different stages.

11 (3) The collapse of expansive soil canal slopes has certain features of repeatability,
12 seasonality, and spatial imbalance, which are highly detrimental to the safety of
13 engineering. To prevent such occurrences, a technique has been suggested that can
14 accurately detect abnormal deformation in advance, and this approach is recommended
15 for implementation in relevant projects.

16 Author Contribution

17 **Xing Li** conceptualized the study, designed the methodology, prepared the
18 software, curated the data, conducted the formal analysis and investigation, validated
19 the results, visualized the data, wrote the original draft, reviewed, and edited the
20 manuscript. **Fuheng Ma** curated the data, acquired the funding, administered the
21 project, supervised the study, reviewed, and edited the manuscript. **Jiang Hu** curated
22 the data, conducted the formal analysis, acquired the funding, supervised the study,
23 reviewed, and edited the manuscript. **Andrey P. Jivkov** conducted the formal analysis,
24 acquired the funding, reviewed, and edited the manuscript. **Dongdong Chu** reviewed,
25 and edited the manuscript.

26 Acknowledgments

27 The research is based upon the work supported by the National Natural Science
28 Foundation of China (Grant No.52209165, Grant No.52179138, and Grant No.
29 5187916), China Postdoctoral Science Foundation (2022M711667), Jiangsu Funding
30 Program for Excellent Postdoctoral Talent, Joint fund for scientific research of Yellow
31 River (Grant No. U2243223). Jivkov acknowledges the support of the Engineering and
32 Physical Sciences Research Council (EPSRC), UK, via Grant EP/N026131/1.

1 References

- 2 Alimohammadlou, Y., Najafi, A., Gokceoglu, C., 2014. Estimation of rainfall-induced landslides
3 using ANN and fuzzy clustering methods: A case study in Saeeen Slope, Azerbaijan province, Iran.
4 CATENA 120, 149–162. <https://doi.org/10.1016/j.catena.2014.04.009>
- 5 Arbelaitz, O., Gurrutxaga, I., Muguerza, J., Pérez, J.M., Perona, I., 2013. An extensive comparative
6 study of cluster validity indices. Pattern Recognition 46, 243–256. <https://doi.org/10.1016/j.patcog.2012.07.021>
- 7
8 Boer, P.-T. de, Kroese, D.P., Mannor, S., Rubinstein, R.Y., 2005. A Tutorial on the Cross-Entropy
9 Method. Annals of Operations Research 134, 19–67. <https://doi.org/10.1007/s10479-005-5724-z>
- 10 BONZO, D.C., HERMOSILLA, A.Y., 2002. Clustering panel data via perturbed adaptive simulated
11 annealing and genetic algorithms. Advances in Complex Systems 05, 339–360. <https://doi.org/10.1142/s0219525902000559>
- 12
13 Chen, B., Hu, T., Huang, Z., Fang, C., 2019. A spatio-temporal clustering and diagnosis method for
14 concrete arch dams using deformation monitoring data. Structural Health Monitoring 18, 1355–
15 1371. <https://doi.org/10.1177/1475921718797949>
- 16 Chen, B., Huang, Z., Liu, C., Wu, Z., 2022. Spatio-temporal data mining method for joint cracks in
17 concrete dam based on association rules. Structural Control and Health Monitoring 29, e2848.
18 <https://doi.org/10.1002/stc.2848>
- 19 Chen, Y., Xu, Y., Jamhiri, B., Wang, L., Li, T., 2022. Predicting uniaxial tensile strength of expansive
20 soil with ensemble learning methods. Computers and Geotechnics 150, 104904.
21 <https://doi.org/10.1016/j.compgeo.2022.104904>
- 22 De Wrachien, D., 2009. Dam-break problems, solutions and case studies. WIT press.
- 23 Ercanoglu, M., Gokceoglu, C., Asch, T.W.J.V., 2004. Landslide Susceptibility Zoning of North of
24 Yenice (NW Turkey) by Multivariate Statistical Techniques. Natural Hazards 32, 1–23.
25 <https://doi.org/10.1023/b:nhaz.0000026786.85589.4a>
- 26 Hou, T., Xu, G., Shen, Y., Wu, Z., Zhang, N., Wang, R., 2013. Formation mechanism and stability
27 analysis of the Houba expansive soil landslide. Engineering Geology 161, 34–43.
28 <https://doi.org/10.1016/j.enggeo.2013.04.010>
- 29 Karo, I.M.K., MaulanaAdhinugraha, K., Huda, A.F., 2017. A cluster validity for spatial clustering
30 based on davies bouldin index and Polygon Dissimilarity function, in: 2017 Second International
31 Conference on Informatics and Computing (ICIC). IEEE. <https://doi.org/10.1109/iac.2017.8280572>
- 32 Krasnov, F., Sen, A., 2019. The Number of Topics Optimization: Clustering Approach. Machine
33 Learning and Knowledge Extraction 1, 416–426. <https://doi.org/10.3390/make1010025>
- 34 Li, H., Xu, Q., He, Y., Deng, J., 2018. Prediction of landslide displacement with an ensemble-based
35 extreme learning machine and copula models. Landslides 15, 2047–2059. <https://doi.org/10.1007/s10346-018-1020-2>
- 36
37 Li, S., Laima, S., Li, H., 2017. Cluster analysis of winds and wind-induced vibrations on a long-
38 span bridge based on long-term field monitoring data. Engineering Structures 138, 245–259.
39 <https://doi.org/10.1016/j.engstruct.2017.02.024>
- 40 Li, T., Kong, L., Guo, A., 2021. The deformation and microstructure characteristics of expansive
41 soil under freeze–thaw cycles with loads. Cold Regions Science and Technology 192, 103393.
42 <https://doi.org/10.1016/j.coldregions.2021.103393>

1 Li, X., Chen, X., Jivkov, A.P., Hu, J., 2021. Assessment of damage in hydraulic concrete by gray
2 wolf optimization-support vector machine model and hierarchical clustering analysis of acoustic
3 emission. *Structural Control and Health Monitoring* 29. <https://doi.org/10.1002/stc.2909>
4 Li, X., Wen, Z., Su, H., 2019. An approach using random forest intelligent algorithm to construct a
5 monitoring model for dam safety. *Engineering with Computers* 37, 39–56. [https://doi.org/](https://doi.org/10.1007/s00366-019-00806-0)
6 [10.1007/s00366-019-00806-0](https://doi.org/10.1007/s00366-019-00806-0)
7 Lu, Y., Liu, S., Zhang, Y., Li, Z., Xu, L., 2020. Freeze-thaw performance of a cement-treated
8 expansive soil. *Cold Regions Science and Technology* 170, 102926. [https://doi.org/10.1016/j.](https://doi.org/10.1016/j.coldregions.2019.102926)
9 [coldregions.2019.102926](https://doi.org/10.1016/j.coldregions.2019.102926)
10 Luxburg, U. von, 2007. A tutorial on spectral clustering. *Statistics and Computing* 17, 395–416.
11 <https://doi.org/10.1007/s11222-007-9033-z>
12 Miao, F., Wu, Y., Xie, Y., Li, Y., 2017. Prediction of landslide displacement with step-like behavior
13 based on multialgorithm optimization and a support vector regression model. *Landslides* 15, 475–
14 488. <https://doi.org/10.1007/s10346-017-0883-y>
15 Mohamad, I.B., Usman, D., 2013. Standardization and Its Effects on K-Means Clustering Algorithm.
16 *Research Journal of Applied Sciences, Engineering and Technology* 6, 3299–3303. [https://doi.org/](https://doi.org/10.19026/rjaset.6.3638)
17 [10.19026/rjaset.6.3638](https://doi.org/10.19026/rjaset.6.3638)
18 Qi, S., Vanapalli, S.K., 2016. Influence of swelling behavior on the stability of an infinite
19 unsaturated expansive soil slope. *Computers and Geotechnics* 76, 154–169. [https://doi.org/10.](https://doi.org/10.1016/j.compgeo.2016.02.018)
20 [1016/j.compgeo.2016.02.018](https://doi.org/10.1016/j.compgeo.2016.02.018)
21 Qin, S., Jiao, J.J., Wang, S., 2002. A nonlinear dynamical model of landslide evolution.
22 *Geomorphology* 43, 77–85. [https://doi.org/10.1016/s0169-555x\(01\)00122-2](https://doi.org/10.1016/s0169-555x(01)00122-2)
23 Reichenbach, P., Rossi, M., Malamud, B.D., Mihir, M., Guzzetti, F., 2018. A review of statistically-
24 based landslide susceptibility models. *Earth-Science Reviews* 180, 60–91. [https://doi.org/10.](https://doi.org/10.1016/j.earscirev.2018.03.001)
25 [1016/j.earscirev.2018.03.001](https://doi.org/10.1016/j.earscirev.2018.03.001)
26 Salazar, F., Toledo, M.Á., González, J.M., Oñate, E., 2017. Early detection of anomalies in dam
27 performance: A methodology based on boosted regression trees. *Structural Control and Health*
28 *Monitoring* 24, e2012. <https://doi.org/10.1002/stc.2012>
29 Struyf, A., Hubert, M., Rousseeuw, P., 1996. Clustering in an Object-Oriented Environment. *Journal*
30 *of Statistical Software* 1. <https://doi.org/10.18637/jss.v001.i04>
31 Syakur, M.A., Khotimah, B.K., Rochman, E.M.S., Satoto, B.D., 2018. Integration K-Means
32 Clustering Method and Elbow Method For Identification of The Best Customer Profile Cluster. *IOP*
33 *Conference Series: Materials Science and Engineering* 336, 012017. [https://doi.org/10.1088/1757-](https://doi.org/10.1088/1757-899x/336/1/012017)
34 [899x/336/1/012017](https://doi.org/10.1088/1757-899x/336/1/012017)
35 Weiss, Y., 1999. Segmentation using eigenvectors: a unifying view, in: *Proceedings of the Seventh*
36 *IEEE International Conference on Computer Vision*. IEEE. [https://doi.org/10.1109/](https://doi.org/10.1109/iccv.1999.790354)
37 [iccv.1999.790354](https://doi.org/10.1109/iccv.1999.790354)
38 Xiao, R., Shi, H., He, X., Li, Z., Jia, D., Yang, Z., 2019. Deformation Monitoring of Reservoir Dams
39 Using GNSS: An Application to South-to-North Water Diversion Project, China. *IEEE Access* 7,
40 54981–54992. <https://doi.org/10.1109/access.2019.2912143>
41 Xie, C., Ni, P., Xu, M., Mei, G., Zhao, Y., 2020. Combined measure of geometry optimization and
42 vegetation for expansive soil slopes. *Computers and Geotechnics* 123, 103588. [https://doi.org/10.](https://doi.org/10.1016/j.compgeo.2020.103588)
43 [1016/j.compgeo.2020.103588](https://doi.org/10.1016/j.compgeo.2020.103588)

1 Yu, Shi, 2003. Multiclass spectral clustering, in: Proceedings Ninth IEEE International Conference
2 on Computer Vision. IEEE. <https://doi.org/10.1109/iccv.2003.1238361>

3 Zhang, J., He, P., Xiao, J., Xu, F., 2018. Risk assessment model of expansive soil slope stability
4 based on Fuzzy-AHP method and its engineering application. *Geomatics, Natural Hazards and Risk*
5 *9*, 389–402. <https://doi.org/10.1080/19475705.2018.1445664>

6 Zhang, Y., Li, W., Sun, W., Tao, R., Du, Q., 2023a. Single-source domain expansion network for
7 cross-scene hyperspectral image classification. *IEEE Transactions on Image Processing*. <https://doi.org/10.1109/TIP.2023.3243853>

9 Zhang, Y., Li, W., Tao, R., Peng, J., Du, Q., Cai, Z., 2021a. Cross-scene hyperspectral image
10 classification with discriminative cooperative alignment. *IEEE Transactions on Geoscience and*
11 *Remote Sensing* *59*, 9646–9660. <https://doi.org/10.1109/TGRS.2020.3046756>

12 Zhang, Y., Li, W., Zhang, M., Qu, Y., Tao, R., Qi, H., 2021b. Topological structure and semantic
13 information transfer network for cross-scene hyperspectral image classification. *IEEE Transactions*
14 *on Neural Networks and Learning Systems*. <https://doi.org/10.1109/TNNLS.2021.3109872>

15 Zhang, Y., Li, W., Zhang, M., Wang, S., Tao, R., Du, Q., 2022. Graph information aggregation cross-
16 domain few-shot learning for hyperspectral image classification. *IEEE Transactions on Neural*
17 *Networks and Learning Systems*. <https://doi.org/10.1109/TNNLS.2022.3185795>

18 Zhang, Y., Zhang, M., Li, W., Wang, S., Tao, R., 2023b. Language-aware Domain Generalization
19 Network for Cross-Scene Hyperspectral Image Classification. *IEEE Transactions on Geoscience*
20 *and Remote Sensing*. <https://doi.org/10.1109/TGRS.2022.3233885>

21 Zhu, D.Y., Lee, C.F., Jiang, H.D., 2003. Generalised framework of limit equilibrium methods for
22 slope stability analysis. *Géotechnique* *53*, 377–395. <https://doi.org/10.1680/geot.53.4.377.37322>

23 ZOU, Z., YUN, Y., SUN, J., 2006. Entropy method for determination of weight of evaluating
24 indicators in fuzzy synthetic evaluation for water quality assessment. *Journal of Environmental*
25 *Sciences* *18*, 1020–1023. [https://doi.org/10.1016/s1001-0742\(06\)60032-6](https://doi.org/10.1016/s1001-0742(06)60032-6)

26
27