



Study on the Dispersion Mechanism Based on Rough Set Theory

Zhilei Sun¹, Yaosheng Zhang¹, Kai Xu^{1*}, Leimeng Li²

¹Geotechnical Engineering Department, Nanjing Hydraulic Research Institute, Nanjing, Jiangsu, 210024, China

²PowerChina Huadong Engineering Corporation Limited, Hangzhou, Zhejiang, 311122, China

*kxu@nhri.cn

Abstract. Dispersive soil poses significant risks in civil engineering due to their susceptibility to erosion and structural instability, particularly in embankments, dams, and infrastructure. Since the discovery of dispersive soils, their dispersion mechanism has been a widely discussed and researched topic. This study introduces rough set theory to systematically analyze the dispersion mechanism of soils, aiming to identify the key factors influencing soil dispersion and their interrelationships. By analyzing eleven crucial parameters, including plastic limit, liquid limit, plasticity index, clay content, sodium ion (Na⁺), pH, and others, and validating the results through dispersion tests, the study finds that sodium ion and pH have the most significant influence on soil dispersion, with weights of 28.6% each. These are followed by plastic limit (19%) and liquid limit (14.3%), while clay content has a smaller effect (9.5%). These findings are critical for enhancing the stability and safety of infrastructure, especially in vulnerable regions.

Keywords: dispersive soil; dispersivity factors; rough set theory; key influencing factor

1 Introduction

Dispersive soils are a special type of soil in which the cohesive forces between soil particles significantly decrease or completely disappear under low salinity or pure water conditions, causing the soil to spontaneously break down into individual primary particles. The dispersive characteristics of these soils are influenced by both the inherent properties of the soil and external environmental factors. In particular, dispersive soils pose significant risks to the stability and safety of hydraulic structures, dams, and infrastructure.

Since the 1960s, the destructive impact of dispersive soils on civil facilities has attracted widespread attention, leading to the rapid development of research in this field. Australian researchers^[1] are the first to study the dispersive characteristics of these soils, and subsequently, American scholars^[2] delved into the relationship between soil dispersion and the physical, chemical, and electrochemical properties of soil particle surfaces. These research findings have significant implications for the civil engineering

field, particularly in the design and maintenance of flood control systems, dams, and embankments, where the behavior of dispersive soils directly affects their effectiveness, stability, and lifespan.

Recent studies have focused on the influence of pH and sodium ions on soil dispersion. J.L. Sherard^[3] and Edgar^[4] revealed that the dispersion of soil is closely related to the electrochemical properties of soil particles, particularly the content of exchangeable sodium ions, which are key factors influencing soil dispersion. These factors provide important conditional attributes for the decision system in rough set theory. In addition, Liu Jie^[5] and others argued that high pH alkaline environments promote the dispersion of dispersive soils, emphasizing the importance of maintaining water purity. Chen Hua^[6] notes that the pH level of the soil has a significant impact on dispersion. It also provides important conditional attributes for the decision system in rough set theory.

Traditional research methods often rely on empirical and factor-based approaches, analyzing problems by assessing the impact of a few selected variables. These methods typically require extensive testing to address the complex interactions among different factors, and the inherent empirical assumptions may limit the objectivity and accuracy of the analysis. In contrast, this study adopts a more systematic, data-driven approach by utilizing rough set theory^[7] to analyze a dataset comprising 20 soil samples with diverse physical, chemical, and mineralogical properties. The research results provide scientific and accurate theoretical support for engineering challenges in fields such as dam construction, flood management, and soil stabilization.

2 Data Source

This study collects 20 soil samples for dispersion identification from regions including Xinjiang, Ningxia, Gansu, Shandong, and Heilongjiang. The samples are taken through drilling, with a depth range from 1.5 meters to 5 meters, removing the surface planting soil layer and selecting the underlying loam. To further analyze the dispersion characteristics of the soil, six physical properties are selected: particle liquid limit, plastic limit, plasticity index, curvature coefficient (C_u), and uniformity coefficient (C_e), along with five chemical properties: calcium ions, magnesium ions, sodium ions, potassium ions, and pH value. Soil dispersion is determined using a comprehensive identification method, with the F-value serving as the key criterion: soils with $F < 1.66$ are classified as dispersive, those with $F \geq 2.55$ as non-dispersive, and soils with $1.66 \leq F < 2.55$ as transitional. These selected physical and chemical indicators provide reliable data support for analyzing soil dispersion revealing the interplay between the physical-chemical properties and dispersion, and offering a theoretical basis for further research into soil dispersion mechanisms and their applications.

3 Using Rough Set Theory to Assess the Influence of Parameters on Soil Dispersion

3.1 Rough Set Theory Overview

Rough set theory automatically identifies and eliminates factors with minimal influence on soil dispersion, thereby streamlining the analysis process and significantly reducing the complexity of testing. Its unsupervised nature allows the research to directly extract key influencing factors from the data, avoiding the interference of empirical assumptions and enhancing the objectivity and accuracy of the analysis. Through this methodology, the study offers a more precise and efficient approach to identifying the factors influencing soil dispersion, overcoming the limitations of traditional methods.

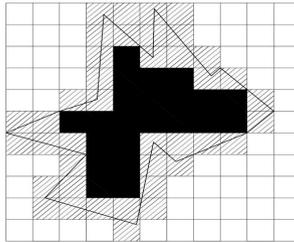


Fig. 1. Schematic diagram of rough set approximation.

An information system can be viewed as a data table consisting of objects and their attributes. In the context of a system for studying soil dispersion mechanisms, each soil sample is considered an object, while the physical and chemical factors potentially influencing soil dispersion are treated as attributes. To elucidate the concepts of upper and lower approximation sets in rough sets, a schematic diagram of rough set approximation is shown in Figure 1. In the diagram, the grid represents the equivalence classes partitioned by U ; the black solid-lined area delineates the rough set X ; the solid black squares signify the lower approximation set of X , where objects in this area belong to X ; the area outside the slashed shading denotes the negative domain, which consists of objects that do not belong to X ; and the slashed shaded area represents the boundary domain of X , encompassing objects whose ownership by X is ambiguous.

3.2 Raw Data Processing and Determination Information Table Formulation

Drawing from measurement results, the following factors are chosen as condition attributes for the dispersion determination system of dispersive soils: liquid limit, plastic limit, plasticity index, clay content, curvature coefficient (C_c), uniformity coefficient (C_u), K^+ , Ca^{2+} , Na^+ , Mg^{2+} , and pH of soil samples, denoted as C1 to C11. The determination attributes for the system are based on dispersive soil samples. A determination information table is formulated by amalgamating these conditions and determination attributes, as shown in Table 1.

Table 1. Determination information table for dispersion mechanism analysis.

Soil Samples	C ₁	C ₂	C ₃	C ₄	C ₅	C ₆	C ₇	C ₈	C ₉	C ₁₀	C ₁₁	D
	Plastic	Liquid	Plasticity	Clay	Curvature	Uniformity	K ⁺	Ca ²⁺	Na ⁺	Mg ²⁺		Dispersibility
	Limit %	Limit %	Index	Content %	Coefficient	Coefficient	1/n	1/n	1/n	1/n	pH	Determination
					C _c	C _u	mmol/L	mmol/L	mmol/L	mmol/L		Value
1	12.05	24.93	12.88	34.27	0.62	10.07	0.06	7.65	2.07	0.86	8.00	1.39
2	12.18	24.09	11.91	32.43	0.66	11.07	0.22	5.29	5.52	0.72	8.58	1.94
3	12.69	23.08	10.39	10.21	0.65	10.12	0.19	6.67	8.83	0.65	8.64	1.75
4	13.98	24.30	10.33	28.80	0.89	9.91	0.22	6.36	2.66	1.13	7.48	1.39
5	14.40	30.25	15.84	25.39	1.15	10.96	0.06	1.89	21.00	0.65	9.55	2.75
6	18.66	26.91	8.25	40.58	0.81	5.23	0.20	7.21	4.17	0.15	8.29	1.67
7	13.96	27.50	13.54	17.73	2.29	10.85	0.05	2.43	18.49	0.63	8.78	2.75
8	16.24	28.57	12.33	18.33	2.31	10.12	0.05	2.75	17.38	0.73	8.60	3.00
9	15.48	26.20	10.73	20.18	1.09	9.25	0.05	1.69	19.58	0.61	8.77	2.46
10	13.99	27.97	13.98	6.47	0.93	9.07	0.04	3.45	22.03	1.06	9.15	2.61
11	12.63	35.44	22.81	42.71	0.75	6.44	0.06	3.51	16.34	0.19	9.72	3.00
12	13.72	28.51	14.79	39.84	0.75	7.84	0.05	2.39	11.48	0.98	9.91	3.00
13	17.93	27.93	10.00	10.43	1.28	9.66	0.05	2.55	23.12	1.02	9.34	2.52
14	18.24	27.69	9.46	18.04	1.54	11.31	0.05	3.69	24.54	0.18	8.96	2.72
15	11.98	27.07	15.09	29.82	0.83	7.35	0.14	10.01	3.00	1.19	7.81	1.66
16	12.66	24.98	12.32	35.28	0.84	6.33	0.21	7.54	6.69	0.51	8.26	1.94
17	13.18	25.50	12.32	14.39	1.44	9.72	0.20	1.16	23.88	0.24	9.20	2.72
18	11.78	26.07	14.29	35.94	0.77	7.88	0.28	8.39	7.34	0.31	7.49	1.66
19	14.07	25.27	11.20	34.13	0.84	6.64	0.23	7.19	2.75	1.27	8.80	1.67
20	14.25	28.40	14.15	31.54	0.79	7.64	0.22	9.13	4.81	0.18	8.37	1.64

Table 2. Level division of condition and determination attributes.

Level Classification	Plastic Limit	Liquid Limit	Plasticity Index	Clay Content	Curvature Coefficient	Uniformity Coefficient
1	C ₁ ≤13%	C ₂ <25.5%	C ₃ ≤12	C ₄ ≤25%	C ₅ ≤0.8	C ₆ ≤7
2	13%<C ₁ ≤15%	25.5%≤C ₂ <27%	12<C ₃ ≤13	25%<C ₄ ≤35%	0.8<C ₅ ≤1	7<C ₆ ≤10
3	15%<C ₁	27%≤C ₂	13<C ₃	35%<C ₄	1<C ₅	10<C ₆

Table 3. Level division of condition and determination attributes.

Level Classification	K ⁺	Ca ²⁺	Na ⁺	Mg ²⁺	pH	Dispersivity
1	C ₇ ≤0.05	C ₈ <3	C ₉ <5	C ₁₀ <0.3	C ₁₁ <8.4	D<1.66
2	0.05<C ₇ <0.2	3≤C ₈ <7	5≤C ₉ <17	0.3≤C ₁₀ <0.8	8.4≤C ₁₁ <9	1.66≤D<2.55
3	0.2≤C ₇	7≤C ₈	17≤C ₉	0.8≤C ₁₀	9≤C ₁₁	2.55≤D

Before conducting attribute reduction analysis, the first task is to establish a determination information table. This required consideration of the size, range, and frequency distribution of experimental values for each factor in the raw data. Based on these considerations, specific discretization criteria are devised to convert continuous

data into discrete values, thereby facilitating more streamlined and effective attribute reduction analysis. These criteria are shown in Table 2.

Table 4. Discretized determination information table for dispersion mechanism analysis.

Soil Samples	C ₁	C ₂	C ₃	C ₄	C ₅	C ₆	C ₇	C ₈	C ₉	C ₁₀	C ₁₁	D
1	1	1	2	2	1	3	1	3	1	3	1	1
2	1	1	2	2	1	3	3	2	2	2	2	2
3	1	1	1	1	1	3	2	2	2	2	2	2
4	2	1	1	2	2	2	3	2	1	3	1	1
5	2	3	3	2	3	3	1	1	3	2	3	3
6	3	2	1	3	2	2	3	3	1	1	1	2
7	2	3	3	1	3	3	1	1	3	2	2	3
8	3	3	2	1	3	3	1	1	3	2	2	3
9	3	2	1	1	3	2	1	1	3	2	2	2
10	2	3	3	1	2	2	1	2	3	3	3	3
11	1	3	3	3	1	2	1	2	3	1	3	3
12	2	3	3	3	1	2	1	1	3	3	3	3
13	3	3	1	1	3	2	1	1	3	3	3	2
14	3	3	1	1	3	3	1	2	3	1	2	3
15	1	3	3	2	2	2	2	3	1	3	1	2
16	1	1	2	3	2	2	3	3	2	2	1	2
17	2	2	2	1	3	2	3	1	3	1	3	3
18	1	2	3	3	1	2	3	3	2	2	1	2
19	2	1	2	2	2	2	3	3	1	3	2	2
20	2	3	3	2	1	2	3	3	1	1	1	1

3.3 Attribute Reduction for Dispersion Influence Factors

Attribute reduction for the discretized determination system is executed using the Rosetta rough set software. The determination information table (Table 1) is imported into Rosetta in Excel form. Based on the level classification criteria for condition attributes specified in Table 2 and Table 3, discretization operations are executed within the software, as shown in Figure 2. The resulting discretized determination information table for dispersion mechanism analysis is shown in Table 4.

Within the rough set software, attribute reduction for the determination system is performed using Johnson’s algorithm, a novel reduction method. Johnson’s algorithm

iteratively selects the attribute with the highest frequency for inclusion in the reduction set. Compared to other algorithms, Johnson's algorithm offers significant advantages: it is highly efficient in handling sparse data, particularly suitable for scenarios with complex or incomplete data relationships, thus improving the accuracy of rough set analysis; it can handle negative edge weights, facilitating adaptability to diverse data types and rules, which enhances its capacity to process uncertain and fuzzy data; and its low memory consumption makes it more efficient for large-scale datasets, optimizing storage and computational performance in rough set analysis.

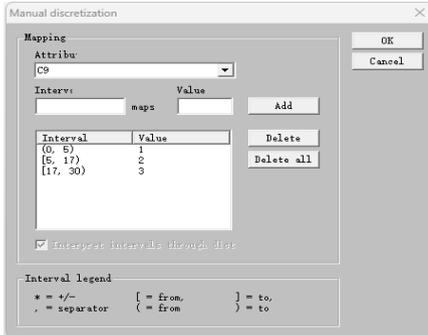


Fig. 2. Discretization of conditional attributes.

	Reduct	Support	Length
1	{C1, C2, C4, C9, C11}	100	5

Fig. 3. Attribute reduction result

These advantages make Johnson's algorithm well-suited for rough set software. As shown in Figure 3, the attribute reduction result obtained using Johnson's algorithm is {C1, C2, C4, C9, C11}. Consequently, the retained condition attributes are liquid limit, plastic limit, clay content, Na⁺, and pH. These attributes exhibit 100% support relative to the determination attribute set, indicating that these five condition attributes are sufficient for accurate classification of determination.

3.4 Calculation of the Weights of Dispersion Influence Factors

After the determination information table for the soil dispersion mechanism study is reduced using Johnson's algorithm, we derive a determination system comprising five condition attributes {C1, C2, C4, C9, C11} and the determination attribute D. To further analyze the influence of each factor on soil dispersion, we calculate the importance of each condition attribute in the determination classification process within the determination system. This enabled us to investigate the weights of plastic limit, liquid limit, clay content, Na⁺, and pH in influencing soil dispersion. However, rough set theory also has certain limitations. Anomalies or inconsistencies in the data may lead to incorrect attribute reduction or inaccurate weight distribution, which could affect the reliability of the conclusions.

$$K = \gamma_C D = \frac{|pos_C(D)|}{|U|} = 1 \quad (1)$$

D : Determination attribute; U : Non empty finite object set; C : Conditional attribute; “ γ_C ”: Dependency of determination attribute on D .

The importance and weight results of each condition attribute relative to the determination attribute D are shown in Table 4.

Table 5. Summary of the importance and weights of reduced attributes.

	Plastic Limit	Liquid Limit	Clay Content	Na ⁺	pH
Attribute importance	0.1	0.075	0.05	0.15	0.15
Weight	19%	14.3%	9.5%	28.6%	28.6%

Table 5 shows that the Na⁺ content and pH value are the most important factors for determining soil dispersion, followed by liquid limit, plastic limit, and clay content. When calculating the importance of other reduced attributes, the result is 0, indicating that attributes such as the curvature coefficient, uniformity coefficient, K⁺, and Mg²⁺ exert no direct influence on soil dispersion identification. Na⁺ and pH have the greatest influence among the primary factors affecting soil dispersion. Based on their attribute importance, the weight distribution of the factors influencing soil dispersion is as follows: Na⁺ and pH account for the highest proportion at 28.6%, followed by plastic limit at 19%, liquid limit at 14.3%, and clay content at 9.5%. The reason for the equal weight distribution of Na⁺ and pH lies in the fact that, during attribute reduction and weight calculation, the influence of these two factors on soil dispersion is very similar. Rough set theory, by analyzing the inherent correlations within the data, automatically assigns equal weights to these two factors, resulting in them occupying the same proportion in the final weight distribution.

This study's findings are crucial for improving soil testing methods and civil engineering practices. Na⁺ and pH, key factors influencing soil dispersion, can enhance the accuracy of soil monitoring. In practical applications like dam construction and flood management, these insights guide engineers in selecting treatment methods, controlling key factors, and improving soil stability, thus reducing risks and enhancing infrastructure safety and durability.

4 Conclusion

This study employs the data analysis methodology of rough set theory to statistically analyze the factors influencing the dispersion of 20 soil samples. Through attribute reduction, condition attributes with minimal influence on soil dispersion are excluded, leading to the identification of the primary factors influencing soil dispersion. The key research findings are as follows:

(1) The study demonstrates that among the 11 factors influencing soil dispersion, including plastic limit, liquid limit, plasticity index, clay content, curvature coefficient (C_c), uniformity coefficient (C_u), K⁺, Ca²⁺, Na⁺, Mg²⁺, and pH, Na⁺ and pH exhibit the most pronounced influence on soil dispersion.

(2) Based on rough set theory, the weights of factors influencing soil dispersion are calculated. The results indicate that Na⁺ and pH are the most important factors, ac-

counting for 28.6%, followed by plastic limit (19%), liquid limit (14.3%), and clay content (9.5%). Therefore, the order of influence among the primary factors is as follows: $\text{Na}^+ = \text{pH} > \text{plastic limit} > \text{liquid limit} > \text{clay content}$.

Acknowledgments

This work has been supported by the National Key Research and Development Program of China (2023YFB2604200) and the Central Public-interest Scientific Institution Basal Research Fund grant funded by the Chinese government (Y324003). This support is greatly appreciated. The authors also thank those who provided technical support in the laboratories from the State Key Lab of Hydrology-Water Resources and Hydraulic Engineering and Department of Civil Engineering, Nanjing Hydraulic Research Institute.

References

1. Cole D C H, Lewis J G. Piping failure of earthen dams built of plastic materials in arid climates[C]//*New Zealand Conference Soil Mechanics*. 1960: 93-99.
2. Decker R S, Dunnigan L P. Development and use of the soil conservation service dispersion test[M]//*Dispersive Clays, Related Piping, and Erosion in Geotechnical Projects*. ASTM International, 1977: 75-96.
3. Sherard J L and Norman R S D 1973 *Piping in Earth Dams of Dispersive Clay*. In Performance of Earth and Earth-Supported Structures.
4. Edgar H 1991 *Soil mechanics note No. 13: Dispersive clays*[J]. Soil Conservation Service, United States Department of Agriculture, Washington, DC: 11-12.
5. Liu J, Yang Y, Yao H, et al. 2020 Experimental study on treatment of dispersed clay based on different modification methods. *Geotechnical Mechanics*, (S1), 163-170.
6. Chen H, Fan H, Lu L, et al 2015 Dispersion discrimination and mechanism analysis of cohesive soil. *Journal of Hydroelectric Power Generation*, 34(02), 149-155.
7. Pawlak Z 1982 Rough Sets. *International Journal of Computer & Information Sciences*, 11:341-356.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

