

## HEURÍSTICA ADAPTATIVA DE INICIALIZAÇÃO DO MÉTODO FUZZY K-MEANS PARA O PROBLEMA DE AGRUPAMENTO DE DESCONTINUIDADES EM FAMÍLIAS EM MACIÇOS ROCHOSOS

André Monteiro Klen <sup>1</sup>; Pedro Gabriel de Carvalho <sup>2</sup>.

<sup>1</sup> Orientador: Pesquisador do IFMG, Campus Ouro Preto; andre.klen@ifmg.edu.br

<sup>2</sup> Bolsista (IFMG), Engenharia de Controle e Automação (UFOP); pedro.gabriel@aluno.ufop.edu.br

### RESUMO

Descontinuidade é uma superfície que separa internamente uma massa rochosa de outra. A identificação de grupos de descontinuidades permite identificar importantes parâmetros acerca da estabilidade de maciços rochosos, como mecanismos de ruptura e permeabilidade. Contudo, nem sempre é possível definir satisfatoriamente tais parâmetros utilizando a interpretação pessoal do diagrama de polos (método clássico), uma vez que esse método depende da experiência do geotécnico e pode ainda levar a múltiplas interpretações subjetivas. No contexto desse problema, é proposto a implementação de um algoritmo numérico com o objetivo de minimizar os fatores de subjetividade e seus erros e defeitos inerentes. O algoritmo é baseado no método *Fuzzy K-means* e permite agrupar as descontinuidades em famílias sem a influência das fontes de erros do método clássico, contudo, o mesmo apresenta sensibilidade à seleção dos centros iniciais dos agrupamentos, ou seja, a escolha errada dos centros iniciais pode conduzir ao empate entre os agrupamentos, podendo conduzir a uma execução demorada, convergência para vales de mínimos locais, por fim produzindo resultados errôneos. Este trabalho traz como principal contribuição um novo método de inicialização do *Fuzzy K-means* fundamentado na distribuição de probabilidades de Fisher para dados esféricos, que possibilita o algoritmo operar de maneira insensível à escolha dos centros iniciais. Para medir a qualidade dos agrupamentos identificados, foram adaptadas medidas de validação do espaço euclidiano para o espaço esférico que auxiliam tanto na definição do número de famílias quanto na identificação de regiões de sobreposição e de *outliers*. Os resultados do algoritmo se mostram satisfatórios tanto em termos de sucesso na inicialização quanto em termos dos agrupamentos gerados. Os agrupamentos gerados foram comparados tanto em relação aos resultados do método clássico quanto aos resultados de outros métodos, e para tal comparação, são utilizados bases de dados que foram explorados por diversos autores para a validação dos algoritmos de agrupamento de descontinuidades. O produto desse projeto será um *software* desenvolvido na linguagem de programação *Scilab* que incorpora tanto a heurística de inicialização quanto o agrupamento das descontinuidades, que pode ser empregado tanto no setor de mineração quanto na construção civil.

### INTRODUÇÃO

“Descontinuidade” é um termo geral que denota qualquer separação no interior de uma maciço rochoso que tenha nula ou baixa resistência à tração (ZHANG, 2005). O agrupamento de descontinuidades em conjuntos e a identificação de suas orientações médias são tarefas importantes na engenharia geotécnica, uma vez que os conjuntos de descontinuidades definem o sistema que controla o comportamento mecânico e hidráulico dos maciços rochosos (SONG *et al.*, 2016); (KLEN & LANA, 2014).

O reconhecimento de conjuntos de descontinuidade aplicando diagramas de densidade de polos é amplamente empregado. Consiste na interpretação de gráficos de contorno de densidade computados pela contagem do número de polos que caem dentro de um círculo de contagem o qual é usado para calcular a frequência dos polos.

No entanto, esse método é limitado por fatores subjetivos e leva a diferentes conjuntos de descontinuidades. Os principais fatores que influenciam o método são: experiência pessoal para interpretar o diagrama de densidade de polos, o tamanho do círculo de contagem, a sobreposição entre conjuntos e outliers (JIMENEZ & SITAR, 2006); (GAO *et al.*, 2019).

Assim, a interpretação visual dos diagramas de densidade de polos não é satisfatória em alguns casos, e conduz ao desenvolvimento de técnicas matemáticas e estatísticas para a identificação automática de conjuntos de descontinuidades (JIMENEZ & SITAR, 2006); (SONG *et al.*, 2016).

Dentre as técnicas numéricas, certamente a análise de *cluster* é a ferramenta mais adequada. Essa técnica é capaz de agrupar dados em clusters sem informações *a priori*. Para agrupar as descontinuidades é necessário um algoritmo de agrupamento e os mais utilizados para este fim são os pertencentes a família *K-means* e sua variação *Fuzzy K-means* (FKM) (JIMENEZ & SITAR, 2006); (LIU *et al.*, 2017).

Principalmente, a partir dos estudos desenvolvidos por Hammah & Curran (1998, 1999, 2000), o *Fuzzy K-means* e o *K-means* são amplamente utilizados como algoritmos para definir conjuntos de descontinuidades. No entanto, devido à sua natureza de decida de gradiente, este algoritmo é altamente sensível a seleção inicial dos centros dos agrupamentos.

A inicialização inadequada inclui clusters vazios, convergência mais lenta e maior chance de travar em mínimos locais incorretos (LIU *et al.*, 2017); (CELEBI *et al.*, 2012). Além disso, esses eventos são mais pronunciados nos casos em que os limites entre os clusters não são claros (HAMMAH & CURRAN, 1998); (XU *et al.* 2012); (SONG *et al.*, 2016).

No entanto, Celebi *et al.* (2012) sugere que o problema da inicialização pode ser remediado com o uso de métodos de inicialização adaptativos ou de acordo com Xu *et al.* (2013), Li *et al.* (2015), Song *et al.* (2017) e Cui & Yan (2020) a aplicação de algoritmos de otimização estocástica.

Por outro lado, várias pesquisas apontam para novos algoritmos que melhoram o desempenho da família *K-means* e do método *Fuzzy K-means* baseado em otimização estocástica que possibilita a otimização global para classificar os conjuntos de descontinuidades (CUI & YAN, 2020).

Assim como Xu *et al.* (2012), Li *et al.* (2014) propuseram um método de busca global baseado na otimização por enxame de partículas que emprega um algoritmo *K-means* (KPSO) para identificar os conjuntos de descontinuidade e remediar o defeito do *Fuzzy K-means*.

Song *et al.* (2017) propuseram o algoritmo *Fuzzy K-means* baseado em otimização por enxame de partículas quânticas (QPSO). Este método ajuda a controlar a velocidade e a posição das partículas. Além disso, oferece uma rápida otimização global.

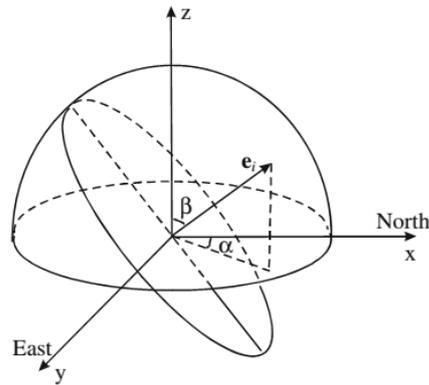
Cui & Yan (2020) introduziram um método que combina o *Fuzzy K-means* com o algoritmo genético de *string* de comprimento variável (VGA). Este método evita especificar o número de clusters, é robusto, atinge o ótimo global e é totalmente automático.

Embora os algoritmos baseados em otimização estocástica sejam capazes de determinar conjuntos de descontinuidade razoáveis e remediar o defeito de inicialização do *Fuzzy K-means*, às vezes os resultados de agrupamento são ligeiramente influenciados pelos centros iniciais dos agrupamentos e o algoritmo falha em convergir em alguma instância (LIU *et al.* 2017).

Este trabalho propõe uma nova heurística adaptativa de inicialização para o algoritmo *Fuzzy K-means* para identificação de conjuntos de descontinuidade em maciços rochosos. Este método produz resultados consistentes selecionando os centros iniciais mais apropriados. Entre suas vantagens, este método não requer ajustes de parâmetros ou a inserção de quaisquer outros métodos de otimização no FKM, ou seja, os recursos de implementação e simplicidade do FKM são preservados.

## METODOLOGIA

As descontinuidades de um maciço rochoso, geralmente, são consideradas como sendo uma estrutura planar e suas orientações espaciais podem ser expressas em termos de *dip direction*, direção de mergulho ( $0^{\circ} \leq \alpha \leq 360^{\circ}$ ), e *dip angle*, mergulho ( $0^{\circ} \leq \beta \leq 90^{\circ}$ ). A orientação de uma descontinuidade *i* qualquer também pode ser representada por um vetor normal unitário, o qual pode ser determinado pelo cosseno diretor  $e_i = (x_i, y_i, z_i)$ . A Figura 1 mostra o vetor normal unitário da descontinuidade *i* no sistema de coordenadas cartesianas. O eixo positivo *x* é horizontal à direção norte, o eixo positivo *y* é horizontal a direção oeste e o eixo positivo *z* é direcionado para cima (CUI *et al.*, 2020); (GAO *et al.*, 2019); (LI *et al.*, 2014).



**Fig.1:** Representação da orientação da descontinuidade. Adaptado de Li *et al.* (2014).

Onde:

$$e_i = \begin{cases} x_i = \cos\alpha \sin\beta \\ y_i = \sin\alpha \sin\beta \\ z_i = \cos\beta \end{cases} \quad (1)$$

O método FKM é sensível aos centros iniciais, ou seja, diferentes centros iniciais podem conduzir a agrupamentos finais também diferentes. Para remediar esse defeito, existem diversos algoritmos como por exemplo, dos trabalhos de Jimenez e Sitar (2006), Klen (2015), Li *et al.* (2015) entre outros.

O método de inicialização proposto parte do conceito de que os centros iniciais dos agrupamentos devem ser gerados aleatoriamente e bem separados (CELEBI *et al.*, 2012). Assim, o método gera K descontinuidades expressas por seus vetores normais unitários para representar os centros iniciais. Para isso, emprega o método de simulação de Monte Carlo para orientações de descontinuidade baseado na distribuição de Fisher (1987), desenvolvido por Zheng *et al.* (2014).

A distribuição de Fisher é uma distribuição simétrica em torno da orientação média em uma esfera unitária e pode ser definida da seguinte forma:

$$f(\theta) = \frac{k \sin\theta e^{k \cos\theta}}{e^k - e^{-k}} \quad (2)$$

Onde  $\theta$  é o ângulo de desvio da orientação média e  $k$  é chamado de parâmetro de concentração e controla a dispersão direcional dos dados sobre a média. Para  $k$  grandes, os dados são concentrados em uma pequena porção da esfera em torno da orientação média e para valores próximos a zero os dados são uniformemente distribuídos (ZHENG *et al.*, 2014).

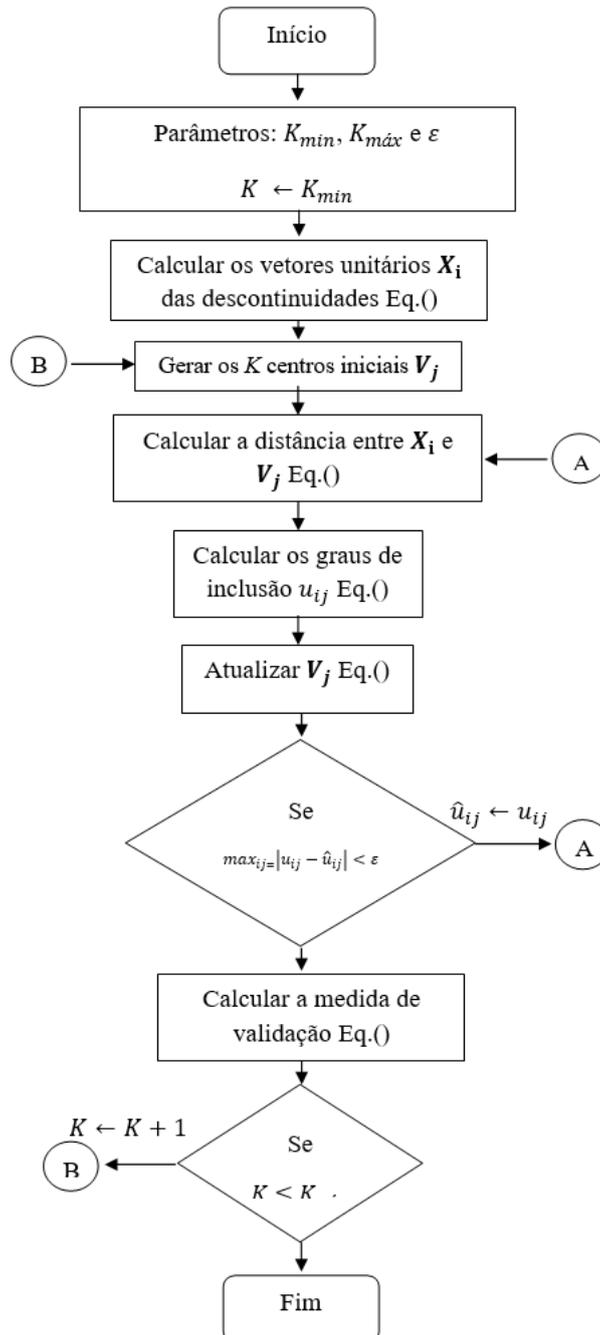
Para cada conjunto gerado é calculada a mínima distância entre os K vetores que o compõe. Em seguida, os conjuntos são comparados em relação à menor distância e aquele que obtiver a máxima distância entre as menores, tem seus vetores escolhidos para representar os centros iniciais dos K agrupamentos. Quando é adotado o critério da máxima distância entre as menores, se evita que entre dois centros bem separados exista um terceiro que esteja próximo o suficiente de um deles e possa então comprometer o particionamento do conjunto de dados.

Para se determinar os centros dos agrupamentos é necessário definir os autovalores e autovetores da matriz de orientação  $S_j$  de cada agrupamento.

$$S_j = \begin{bmatrix} \sum_{i=1}^N (u_{ij})^m x_i x_i & \sum_{i=1}^N (u_{ij})^m x_i y_i & \sum_{i=1}^N (u_{ij})^m x_i z_i \\ \sum_{i=1}^N (u_{ij})^m x_i y_i & \sum_{i=1}^N (u_{ij})^m y_i y_i & \sum_{i=1}^N (u_{ij})^m y_i z_i \\ \sum_{i=1}^N (u_{ij})^m x_i z_i & \sum_{i=1}^N (u_{ij})^m y_i z_i & \sum_{i=1}^N (u_{ij})^m z_i z_i \end{bmatrix} \quad j = 1, 2 \dots K. \quad (3)$$

Onde  $(x_i, y_i, z_i)$  é o cosseno diretor da  $i$ -ésima descontinuidade  $i$ .

Os autovalores de  $S_j$  representados por  $(\tau_{j1}, \tau_{j2}, \tau_{j3})$  são ordenados de tal forma que  $(\tau_{j1} < \tau_{j2} < \tau_{j3})$  estão relacionados aos seus respectivos autovetores normalizados  $[\xi_{j1}, \xi_{j2}, \xi_{j3}]$ . Portanto, o autovetor associado ao maior autovalor, neste caso  $\tau_{j3}$ , torna  $\xi_{j3}$  um excelente estimador do centro do agrupamento  $K$  (HAMMAH e CURRAN, 1998); (CUI e YAN, 2020). Segue abaixo o fluxograma explicativo do algoritmo implementado.



**Fig. 2:** Fluxograma de execução do algoritmo desenvolvido.

## RESULTADOS E DISCUSSÕES

Para avaliar o desempenho do algoritmo, utiliza-se uma base de dados artificial com 437 descontinuidades, gerada a partir da distribuição de Fisher para dados esféricos, Eq.(2). Esse conjunto é formado por cinco agrupamentos com diferentes concentrações e quantidades de elementos. Os parâmetros de cada agrupamento estão descritos na tabela 1.

Conforme Xu *et al.*(2012), o FKM tem uma dificuldade inerente em selecionar os centros iniciais e definir os agrupamentos, quando os limites entre estes não são claros e se tem um número grande de conjuntos. Por isso, os dados são gerados com sobreposição entre os agrupamentos para que o algoritmo possa ser testado em uma situação de maior complexidade operacional.

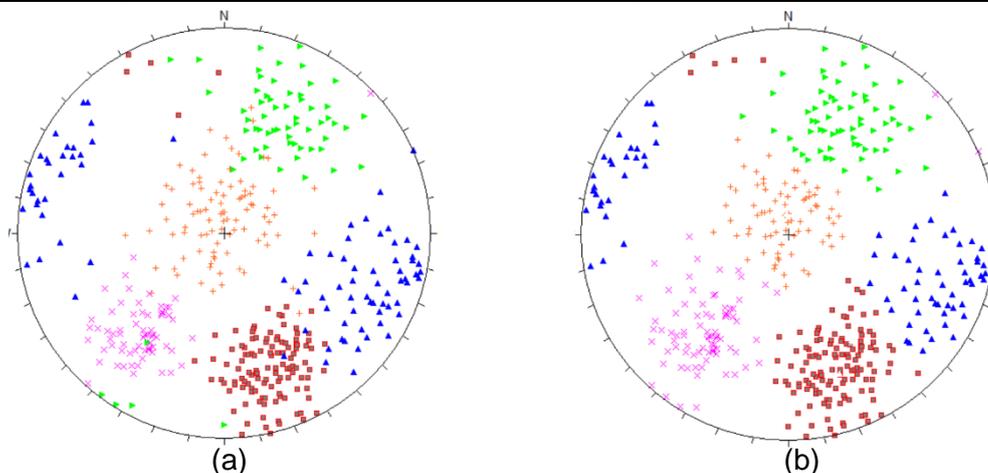
**Tab. 1:** Parâmetros para geração dos dados artificiais.

Agrupamentos	Número de descontinuidades	Orientação Média (Dip /Dip Direction)	Parâmetros (k)
1	122	71 / 341	27
2	89	82 / 290	13
3	63	68 / 208	17
4	93	10 / 173	10
5	70	65 / 40	35

Os resultados do particionamento realizado pelo algoritmo que constam na Tabela 2 e Figura 3, apontam que o FKM identifica os agrupamentos e suas orientações médias de forma consistente. As diferenças em relação à base de dados original ocorrem, porque os agrupamentos se sobrepõem e determinadas descontinuidades se situam mais próximas do centro do agrupamento que elas estão associadas do que do agrupamento original.

**Tab. 2:** Comparação dos resultados obtidos pelo algoritmo e a base de dados artificial.

Agrupamentos	Novo método		Dados artificiais	
	Dip / Dip direction	Número de descontinuidades	Dip / Dip direction	Número de descontinuidades
1 – (■)	71/340	131	71 / 341	122
2 – (▲)	85/288	79	82 / 290	89
3 – (►)	63/207	66	68 / 208	63
4 – (+)	12/175	76	10 / 173	93
5 – (x)	66/39	85	65 / 40	70



**Fig. 3:** Comparação de resultados: **(a)** Base de dados artificiais. **(b)** Resultados do algoritmo.

Para avaliar o número de agrupamentos, compacidade, e a separação dos conjuntos formados pelo algoritmo, é utilizada a medida de Xie-Beni (XB) (Tabela 3). Essa medida se trata de um critério de minimização que indica a qualidade dos agrupamentos gerados. O resultado do menor índice XB aponta que o Nébula identifica corretamente o número de conjuntos. Isso evidencia que o método consegue particionar adequadamente a base de dados apesar da presença de sobreposição e do elevado número de elementos.

**Tab. 3:** Índice Xie-Beni para os agrupamentos gerados pelo algoritmo aplicado a base de dados artificial.

Agrupamentos ( $K$ )	2	3	4	5	6
Xie-Beni	0.252	0.118	0.131	<b><u>0.082</u></b>	0.353

Na sequência, o algoritmo foi aplicado à base de dados da mina de cobre de San Manuel no estado do Arizona nos Estados Unidos, a qual foi gerada por Shanley e Mahtab (1976). Esta base é de grande relevância nos estudos dos algoritmos de agrupamento pois é muito utilizada por diversos autores. Dentre os algoritmos testados através dessa base de dados destacam-se : Hammah & Curran (1998), Jimenez & Sitar (2006), Li *et al.* (2015), Liu *et al.* (2017); Cui & Yan (2020) entre outros.

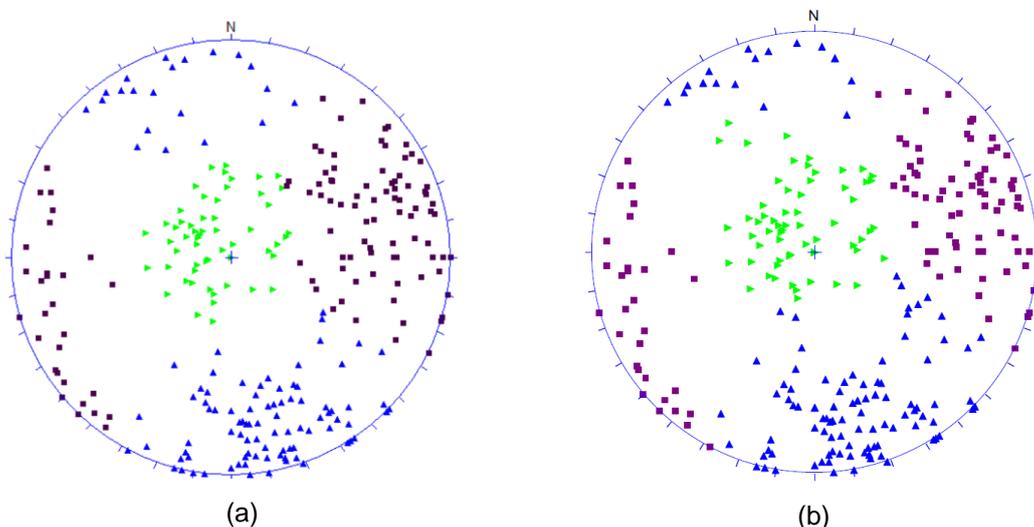
Nesse primeiro estudo de caso o novo método é aplicado para particionar a base de dados da mina de San Manuel e seus resultados são comparados com os resultados alcançados pelos algoritmos de Shanley & Mahtab (1976), Jimenez & Sitar (2006) e Li *et al.* (2015).

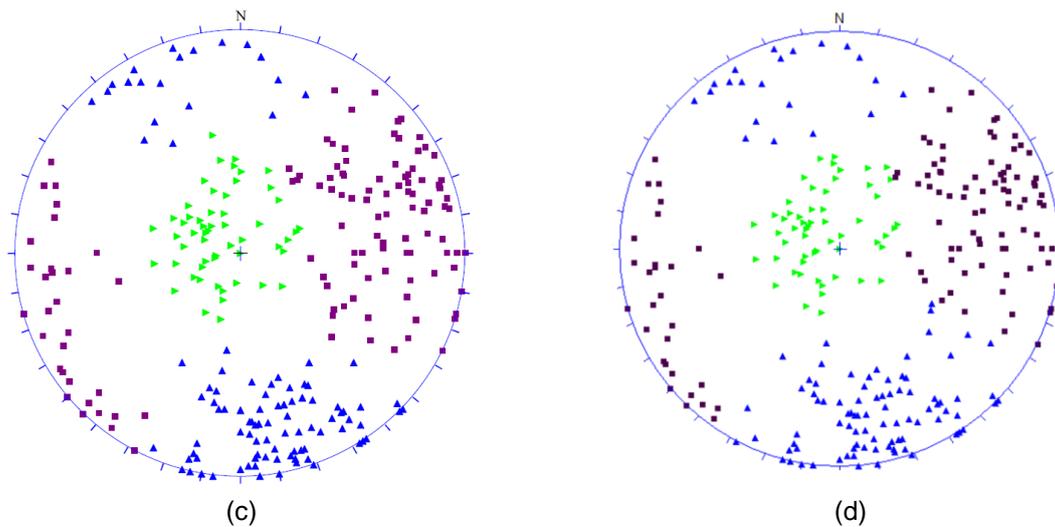
Esses algoritmos são utilizados como referências, pois são frequentemente testados e comparados por diversos autores como Song *et al.* (2017) e Cui & Yan (2020). O algoritmo de Shanley & Mahtab (1976) é o primeiro método a apresentar critérios matemáticos para dividir as descontinuidades em agrupamentos. Já o método de Jimenez e Sitar (2006), *Spectral method*, baseia-se na representação das descontinuidades em um espaço amostral  $R^K$ . Nesse novo espaço os  $K$  agrupamentos são bem separados e de fácil identificação, assim o algoritmo  $K$ -means é aplicado para identificar os conjuntos.

O algoritmo de Li *et al.* (2015), aplica o método estocástico de otimização *Particle Swarm Optimization* para minimizar a função objetivo do  $K$ -means e definir os grupos de descontinuidades. Essa técnica busca pelo ótimo global da função, eliminando, na maioria dos casos, a sensibilidade do FKM à escolha dos centros iniciais.

Essa questão da sensibilidade está presente em todos os algoritmos baseados no  $K$ -means e sua variante *Fuzzy K-means*. Devido a isso, vários métodos foram desenvolvidos para reparar esse problema como é o caso de Jimenez & Sitar (2006) e Klen & Lana (2014).

A Figura 4 mostra os diagramas de polos para três agrupamentos, resultantes do particionamento da base de dados pelos algoritmos.





**Fig. 4:** Comparação entre os resultados dos algoritmos para o base de dados da mina de San Manual: **(a)** Shanley & Mahtab (1976); **(b)** Li *et al.* (2015); **(c)** Novo método; **(d)** Jimenez & Sitar (2006).

A tabela 4 compara a atitude média e o número de descontinuidades dos agrupamentos definidos pelos métodos: Novo método, Jimenez & Sitar (2006), Li *et al.*(2015) e Shanley & Mahtab (1976). Já a tabela 5 aponta os resultados da medida de validação Xie-Beni e mostra que o algoritmo identifica corretamente o número de agrupamentos, pois o menor valor de XB é para  $K=3$ .

**Tab. 4:** Comparação entre os resultados do novo método e dos outros algoritmos

Agrupamentos	Novo método		Jimenez e Sitar (2006) <sup>1</sup>		Li <i>et al.</i> (2015) <sup>1</sup>		Shanley e Mahtab (1976) <sup>1</sup>	
	Dip / Dip direction	Número	Dip / Dip direction	Número	Dip / Dip direction	Número	Dip / Dip Direction	Número
1 – (■)	79/251	115	77/252	124	80/251	108	75/251	127
2 – (▲)	80/347	115	81/347	107	77/344	116	75/343	112
3 – (▶)	17/129	56	12/135	55	16/151	62	8/123	47

<sup>1</sup>Dados extraídos de Liu *et al.* (2017).

**Tab. 5:** Índice Xie-Beni para os agrupamentos gerados pelo Novo Método.

Agrupamentos( $K$ )	2	<u>3</u>	4	5	6
Xie-Beni	0.1902	<u>0.1074</u>	0.2615	0.2102	0.2207

Analisando os resultados, percebe-se que os algoritmos apresentam performances muito parecidas e que a diferença entre o número de descontinuidades e a atitude média dos agrupamentos tem sua origem nos elementos que estão na região de sobreposição entre os conjuntos. A forma como cada método trata a sobreposição altera a associação das descontinuidades aos agrupamentos.

O método de Li *et al.* (2015), na maioria dos casos, atinge o ótimo global da função objetivo, diferentemente dos demais métodos comparados, que geralmente param em um ótimo local. Porém, o ótimo global não garante que o particionamento do banco de dados seja adequado. Além disso, é importante a análise do que é encontrado em campo e o que é definido pelos algoritmos (HAMMAH & CURRAN, 1998).

Conforme Gao *et al.* (2019) e Liu *et al.* (2017), os métodos empregados por Jimenez & Sitar (2006) e Li *et al.* (2015) são complicados em teoria e precisam de parâmetros que são difíceis de estabelecer. Nesse ponto onde o Nébula se destaca, pois por ser baseado no FKM, ele não requer ajuste de parâmetros ou inserção de nenhum outro método de otimização, como é o caso do algoritmo de Li *et al.* (2015). Isso garante, facilidade de implementação, processamento ágil e consistência aos resultados.

## CONCLUSÕES

O algoritmo proposto foi desenvolvido na linguagem de programação *Scilab* e se baseia no método *Fuzzy K-means* para resolver o problema de agrupamentos de descontinuidades. O algoritmo consegue superar o defeito da inicialização inerente aos métodos *K-means* e apresentou um resultado consistente em relação a outros métodos similares. As principais vantagens desse método são: o fato dele não sofrer influência dos elementos que geram subjetividade no método clássico, ser de fácil implementação, rápida convergência e sua flexibilidade para adaptação a diversos problemas de engenharia. O algoritmo foi testado em dois estudos de casos e apresentou bons resultados. Cabe destacar que, ao ser comparado aos principais algoritmos utilizados como referência na literatura, o novo método mostrou desempenho similar.

Dessa forma, o algoritmo apresentado é uma ferramenta importante no auxílio à definição dos agrupamentos de descontinuidades e pode ser de grande estima para as áreas da construção civil e mineira. A utilização do algoritmo deve ocorrer em consonância com o levantamento geotécnico da área em estudo, e os seus resultados sempre devem ser verificados com as estruturas observadas em campo.

## REFERÊNCIAS BIBLIOGRÁFICAS

CUI, Xuejie; YAN, E.-chuan. A clustering algorithm based on differential evolution for the identification of rock discontinuity sets. **International Journal of Rock Mechanics and Mining Sciences**, v. 126, p. 104181, 2020.

GAO, Feng *et al.* A Fast-Clustering Method for Identifying Rock Discontinuity Sets. **KSCE Journal of Civil Engineering**, v. 23, n. 2, p. 556-566, 2019.

HAMMAH, Reginald E.; CURRAN, John H. Validity measures for the fuzzy cluster analysis of orientations. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 22, n. 12, p. 1467-1472, 2000.

HAMMAH, R. E.; CURRAN, J. H. On distance measures for the fuzzy K-means algorithm for joint data. **Rock Mechanics and Rock Engineering**, v. 32, n. 1, p. 1-27, 1999.

HAMMAH, R. E.; CURRAN, J. H. Fuzzy cluster algorithm for the automatic identification of joint sets. **International Journal of Rock Mechanics and Mining Sciences**, v. 35, n. 7, p. 889-905, 1998.

JIMENEZ-RODRIGUEZ, R.; SITAR, N. A spectral method for clustering of rock discontinuity sets. **International Journal of Rock Mechanics and Mining Sciences**, v. 43, n. 7, p. 1052-1061, 2006.

KLEN, A., M. **Algoritmo para agrupamento de descontinuidades em famílias baseado no método Fuzzy K-means.** 2015, 142 f. Tese (Doutorado Geotecnia aplicada à mineração). Universidade Federal de Ouro Preto. Escola de Minas. Núcleo de Geotecnia da Escola de Minas. Ouro Preto, 2015.

KLEN, A. M.; LANA, M. S. Fuzzy algorithm of discontinuity sets. **REM: R. Esc. Minas**, Ouro Preto, 67(4), p. 439-445, oct. dec. 2014.

LI, Yanyan *et al.* K-means algorithm based on particle swarm optimization for the identification of rock discontinuity sets. **Rock Mechanics and Rock Engineering**, v. 48, n. 1, p. 375-385, 2015.

LIU, Jie; ZHAO, Xing-Dong; XU, Zeng-he. Identification of rock discontinuity sets based on a modified affinity propagation algorithm. **International Journal of Rock Mechanics and Mining Sciences**, v. 94, p. 32-42, 2017.

SHANLEY, R. J.; MAHTAB, M. A. Delineation and analysis of clusters in orientation data. **Journal of the International Association for Mathematical Geology**, v. 8, n. 1, p. 9-23, 1976.

XIE, Xuanli Lisa; BENI, Gerardo. A validity measure for fuzzy clustering. **IEEE Transactions on pattern analysis and machine intelligence**, v. 13, n. 8, p. 841-847, 1991.

ZHANG L. **Engineering properties of rocks.** Amsterdam, Netherlands: Elsevier Science; 2005. p. 304

## Participação em Congressos, publicações e/ou pedidos de proteção intelectual:

O trabalho aqui descrito foi selecionado para integrar o rol de apresentações do XVII Congresso Brasileiro de Geologia de Engenharia e Ambiental, programado para ocorrer em setembro de 2022. URL: (<https://17cbge.abge.org.br/home> )