



AGRUPAMENTO DE INSTALAÇÕES DE ENERGIA SOLAR EM TERRENOS ABANDONADOS NORTE-AMERICANOS

David Gabriel de Barros Franco

Pontifícia Universidade Católica do Paraná
R. Imaculada Conceição, 1155 - Prado Velho, Curitiba - PR
david.barros@pucpr.br

Maria Teresinha Arns Steiner

Pontifícia Universidade Católica do Paraná
R. Imaculada Conceição, 1155 - Prado Velho, Curitiba - PR
maria.steiner@pucpr.br

RESUMO

O presente artigo valeu-se de um modelo híbrido *fuzzy c-means* para *clusterizar* (agrupar) e definir locais aptos para instalações de captação de energia solar no território continental norte-americano. Inicialmente foi realizado o pré-processamento dos dados, para substituição de dados faltantes, normalização e análise dos componentes principais (*PCA*). Em seguida, foi aplicado o algoritmo híbrido proposto envolvendo *clusterização*. Trata-se de um modelo *fuzzy c-means* inicializado por meta-heurísticas, a saber, algoritmo genético (*GA*), evolução diferencial (*DE*) e *particle swarm optimization* (*PSO*). O número de *clusters* foi validado por três métricas: *Calinski-Harabasz Index*, *Davies-Bouldin Index* e *Silhouette Coefficient*. Como resultado da abordagem proposta observou-se que houve um incremento na velocidade de treinamento do algoritmo. Visualmente, pode-se perceber a predominância da alocação das instalações em estados de maior incidência média de radiação solar. Por fim, são considerados os ganhos ambiental-econômico-social com a revitalização de terrenos improdutivos e contaminados para a implantação de usinas solares.

PALAVRAS CHAVE. *Clusterização, Fuzzy C-means, Energia Solar.*

Tópicos: MH – Meta-heurísticas. EN – PO na Área de Energia.

ABSTRACT

The present article used a hybrid fuzzy c-means model to cluster and define suitable locations for solar energy capture facilities in the continental United States. Initially, data preprocessing was performed for substitution of missing data, normalization and principal component analysis (PCA). Then, the proposed hybrid clustering algorithm was applied. It is a fuzzy c-means model initialized by metaheuristics, namely genetic algorithm (GA), differential evolution (DE) and particle swarm optimization (PSO). The number of clusters was validated by three metrics: Calinski-Harabasz Index, Davies-Bouldin Index and Silhouette Coefficient. As a result of the proposed approach it was observed that there was an increase in the training speed of the algorithm. Visually, we can see the predominance of the allocation of the facilities in states with a higher average incidence of solar radiation. Finally, the environmental-economic-social gains are considered with the revitalization of unproductive and contaminated land for the implantation of solar plants.

KEYWORDS. Clustering. Fuzzy C-means. Solar Energy.

Paper topics: MH – Metaheuristics. EN – OR in Energy.



1. Introdução

Áreas abandonadas, que possuem substâncias potencialmente perigosas ao meio ambiente ou à saúde humana, estão se tornando o centro de uma preocupação mundial [Bergius e Öberg 2007; Greenberg e Lewis 2000; Li et al. 2017; van Straalen 2002]. Entre estas áreas de risco, no contexto norte-americano, podemos citar: minas abandonadas, que incluem *spoil banks* (pilhas de rejeitos) e plantas de processamento de metais, geralmente contaminadas por metais pesados [Kovacs e Szemmelveisz 2017]; *brownfields* (campos marrons), que podem ser definidos como instalações industriais ou comerciais que apresentam dificuldades para reutilização, devido à presença de substâncias perigosas, poluentes ou contaminantes [U.S. Government Publishing Office 2002]; locais que se enquadram no *Superfund*, programa do governo federal norte-americano voltado para a localização e limpeza de áreas contaminadas com substâncias perigosas ou poluentes [U.S. Government Publishing Office 2015]; aterros sanitários, que nos países desenvolvidos incluem, basicamente, restos de alimentos e embalagens [Rong et al. 2017]; e áreas abrangidas pela legislação do *Resource Conservation and Recovery Act (RCRA)*, de destinação de rejeitos sólidos [U.S. Government Publishing Office 2011].

Ao mesmo tempo, a crescente ocupação do espaço urbano e rural têm-se tornado um problema no mundo moderno [Lambin e Meyfroidt 2011; Morio, Schädler, e Finkel 2013], o que demanda maior eficiência na ocupação territorial, principalmente a reutilização de áreas abandonadas, que apresentam maior desafio [Morio, Schädler, e Finkel 2013; Nuissl e Schroeter-Schlaack 2009; U.S. Government Publishing Office 2002]. Ainda mais se tais áreas possuem considerável tamanho e estão contaminadas, gerando, além dos riscos ambiental e de saúde, riscos econômicos [Apostolidis e Hutton 2006; Cao e Guan 2007; de Sousa 2003; Kaufman, Rogers, e Murray 2005; Morio, Schädler, e Finkel 2013]. A partir do uso de informações públicas sobre as principais áreas em desuso com algum tipo de contaminação no território continental norte-americano, elaborou-se uma metodologia híbrida de *clusterização* (agrupamento) visando classificar tais áreas em apropriadas ou não para a implantação de usinas de captação de energia solar.

Uma vez que o aumento do consumo de energia entra em choque com as implicações do consumo de combustíveis fósseis e a consequente emissão de gases tóxicos e de efeito estufa, fazem-se necessários investimentos em pesquisa e desenvolvimento de novas fontes limpas e renováveis de energia (Almeida et al., 2017; Baños et al., 2011; Cadez e Czerny, 2016; Manzano-Agugliaro et al., 2012; Perea-Moreno et al., 2017). Desse modo, as energias renováveis, como a solar, por exemplo, têm-se colocado como fortes candidatas na nova corrida por produtividade e bem-estar ambiental e social [González, Gonçalves, e Vasconcelos 2017; Lima, Ferreira, e Vieira 2013], além de estarem em voga nos discursos políticos, empresariais e sociais em geral [Onat, Kucukvar, e Tatari 2014; Simas e Pacca 2013]. Neste cenário, a energia solar é considerada uma fonte abundante, gratuita e limpa [Fernández-García et al. 2015].

O objetivo do presente artigo é definir locais adequados para instalações de captação de energia solar no território continental norte-americano. Para tanto, fez-se uso de um modelo híbrido *fuzzy c-means (FCM)* para *clusterizar* estes locais de acordo com suas características. Trata-se de um modelo *fuzzy c-means* inicializado por meta-heurísticas, a saber, algoritmo genético (*GA — Genetic Algorithm*), evolução diferencial (*DE — Differential Evolution*) e *particle swarm optimization (PSO)*. Inicialmente foi realizado o pré-processamento dos dados, para substituição de dados faltantes, normalização e análise dos componentes principais (*PCA — Principal Component Analysis*). Em seguida, foi aplicado o algoritmo híbrido proposto envolvendo *clusterização*. O número de *clusters* foi validado por três métricas: *Calinski-Harabasz Index*, *Davies-Bouldin Index* e *Silhouette Coefficient*.

O presente trabalho está organizado da seguinte maneira: após esta seção introdutória, a Seção 2 apresenta a metodologia empregada, incluindo uma breve revisão de literatura, apresentação dos dados utilizados e descrição do modelo híbrido de *clusterização* proposto. Na Seção 3 são apresentados e discutidos os resultados. Por fim, a Seção 4 sumariza as conclusões da pesquisa.



2. Metodologia

É apresentada a seguir a revisão de literatura referente ao processo de descoberta de conhecimento em base de dados (*KDD — Knowledge Discovery in Databases*), tendo em vista a *clusterização*. Logo na sequência é apresentada a base de dados utilizada e o pré-processamento empregado. E, finalmente, o modelo proposto de *clusterização* híbrida.

2.1 Revisão de Literatura

KDD é uma área do conhecimento dedicada à identificação e extração de padrões significativos de informação a partir de bases de dados [Fayyad, Piatetsky-Shapiro, e Smyth 1996]. A aplicação do *KDD* se dá em múltiplos estágios, começando com a seleção, pré-processamento e transformação dos dados, seguida da mineração (*data mining*) e terminando com a interpretação dos resultados e a geração do conhecimento [Fayyad, Piatetsky-Shapiro, e Smyth 1996; Orriols-Puig et al. 2013; Gamarra, Guerrero, e Montero 2016].

A etapa de mineração de dados pode utilizar um ou vários algoritmos em busca de padrões, tendências e estruturas na base de dados, as quais podem assumir variadas formas, como equações, redes, grafos, conjuntos de regras, entre outros [Roiger 2017; Witten et al. 2017]. Nessa etapa de aprendizado pode-se adotar duas abordagens distintas: na primeira abordagem, de aprendizado supervisionado, consideram-se, *a priori*, estruturas e padrões pré-definidos; na segunda abordagem, de aprendizado não-supervisionado, não se consideram essas possíveis estruturas e padrões, deixando ao algoritmo a tarefa de identifica-las [Orriols-Puig et al. 2013].

A *clusterização* se enquadra na segunda abordagem, de aprendizado não-supervisionado, na qual as instâncias são agrupadas com base em alguma regra inerente à sua estrutura, como a distância entre elas [Roiger 2017; Bramer 2016]. Simplificadamente, pode-se dizer que, após a definição do número de *clusters* e seus respectivos centros, há uma primeira etapa de designação das instâncias para o mais próximo centro, seguida pela otimização da localização do mesmo, minimizando a função objetivo [Aggarwal 2015].

Há um variado número e tipo de algoritmos de *clusterização* [Halkidi, Batistakis, e Vazirgiannis 2001; Xu e Wunsch II 2005]. Alguns deles são apresentados na Tabela 1.

Tabela 1. Algoritmos de *Clusterização*

Tipo	Algoritmo	Referência
	<i>K-Means</i>	[MacQueen 1967]
Particional	<i>PAM (Partitioning Around Medoids)</i> , <i>CLARA (Clustering Large Applications)</i> , <i>CLARANS (CLARA based on Randomized Search)</i>	[Ng e Han 1994]
	<i>K-prototypes</i> , <i>K-mode</i>	[Huang 1998]
Hierárquico	<i>BIRCH (Balanced Iterative Reducing and Clustering using Hierarchies)</i>	[Zhang, Ramakrishnan, e Livny 1996]
	<i>CURE (Clustering Using Representatives)</i>	[Guha, Rastogi, e Shim 1998]
	<i>ROCK (Robust Clustering using linKs)</i>	[Guha, Rastogi, e Shim 2000]
Density-based	<i>DBSCAN (Density Based Spatial Clustering of Applications with Noise)</i>	[Ester et al. 1996]
	<i>DENCLUE (Density-based Clustering)</i>	[Hinneburg e Keim 1998; Hinneburg e Keim 2003]
Grid-based	<i>STING (Statistical Information Grid-based method)</i>	[Wang, Yang, e Muntz 1997]
	<i>WaveCluster</i>	[Sheikholeslami, Chatterjee, e Zhang 1998]
Fuzzy (soft)	<i>Fuzzy c-means</i>	[James C. Bezdek, Ehrlich, e Full 1984]
Neural	<i>EM (Expectation Maximization)</i>	[Dempster, Laird, e Rubin 1977]
Network-based	<i>GLVQ (Generalized Learning Vector Quantization)</i> , <i>SOFM (Self-Organizing Feature Maps)</i>	[Pal, Bezdek, e Tsao 1993]
Kernel-based	<i>SVC (Support Vector Clustering)</i>	[Ben-Hur et al. 2001]



Os resultados da *clusterização* devem ser validados por alguma métrica, que pode ser externa, como rótulo de classes; interna, quando a avaliação se dá por característica inerentes aos dados, como variância e separação dos pontos em diferentes *clusters*; ou relativa, onde as métricas visam comparar o uso de diferentes parâmetros durante a execução do algoritmo. A Tabela 2 sumariza algumas dessas métricas apontando, em negrito, as utilizadas neste trabalho [Zaki e Meira Jr. 2014; Halkidi, Batistakis, e Vazirgiannis 2001; Amigó et al. 2009; Meilă 2007].

Tabela 2. Métricas de validação da *clusterização*

Externa	Interna	Relativa
<i>Purity</i>	<i>BetaCV Measure</i>	<i>Calinski-Harabasz Index</i>
<i>Maximum Matching</i>	<i>Normalized Cut Measure</i>	<i>Gap Statistic</i>
<i>F-Measure</i>	<i>Modularity</i>	
<i>Conditional Entropy</i>	<i>Dunn Index</i>	
<i>Normalized Mutual Information</i>	<i>Davies-Bouldin Index</i>	
<i>Variation of Information</i>	<i>Silhouette Coefficient</i>	
<i>Jaccard Coefficient</i>	<i>Hubert Statistic</i>	
<i>Rand Statistic</i>		
<i>Fowlkes-Mallows Measure</i>		
<i>Discretized Hubert Statistic</i>		
<i>Normalized Discretized Hubert Statistic</i>		

2.2 Coleta e Pré-processamento dos Dados

Os dados utilizados são oriundos da *National Solar Radiation Database (NSRDB)*, coleção de medidas horárias de radiação solar e dados meteorológicos, e do projeto *RE-Powering America's Land*, da *United States Environmental Protection Agency (EPA)*, cujo propósito é identificar áreas abandonadas e contaminadas que são ideais para projetos de energia renovável. A Fig. 1 mostra a união entre os dados de radiação solar direta (*DNI — Direct Normal Irradiance*) média para o período 1998-2014, medida em kWh/m²/dia, e áreas contaminadas, maiores que 100 acres (~400 mil m²), com potencial para instalação de usinas solares, num total de 5.063 pontos (minas abandonadas, a maioria).

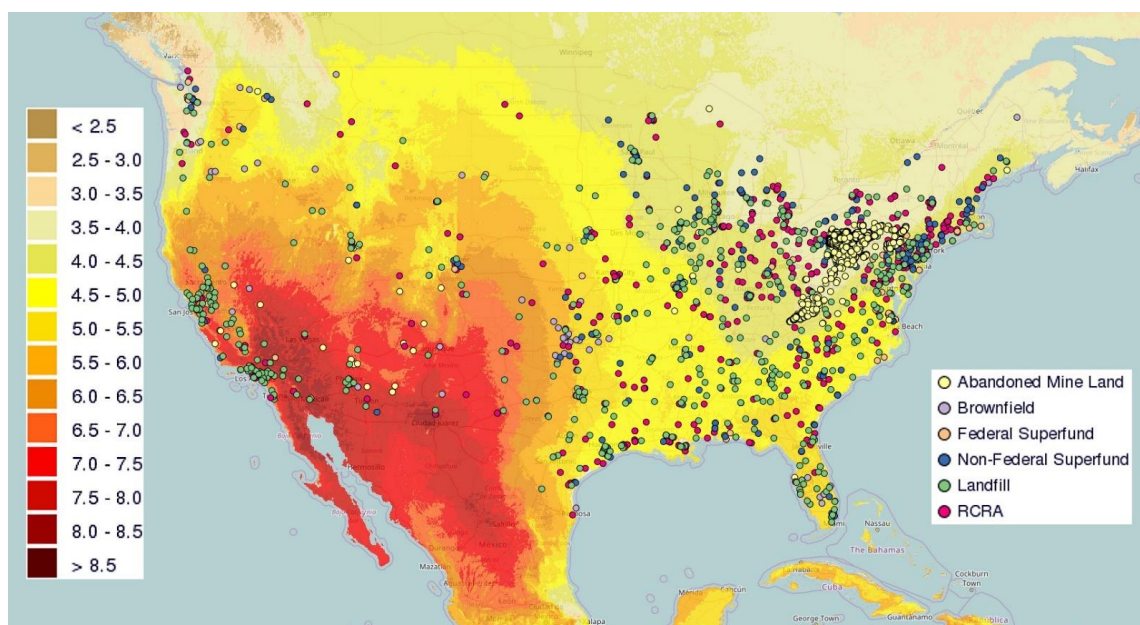


Fig. 1. Radiação solar direta e áreas contaminadas no território continental norte-americano.



As variáveis contidas na base de dados são: área mapeada, medida em acres; distância até linhas de transmissão, medida em milhas; e radiação solar direta, medida em kWh/m²/dia, para escala de utilidade (geração na casa dos megawatts e possibilidade de exportação para a rede) e pequena escala (na qual é inviável a exportação para a rede).

Os valores faltantes, num total de 12 instâncias para a variável distância até linhas de transmissão, foram substituídos pela média do conjunto. Em seguida foi realizada a normalização dos dados, que passaram a ter média “0” e desvio-padrão “1”, para dar sequência à análise dos componentes principais, que tem por objetivo diminuir o número de atributos, retirando os que são correlacionados entre si e que pouco contribuem para a variância do conjunto [Theodoridis e Koutroumbas 2009].

Para as variáveis testadas, apenas a radiação solar direta em pequena escala foi removida, uma vez que não contribuía para a variância do conjunto de dados (apenas 0,5829%, contra 50,5270% para área mapeada, 25,9960% para distância até linhas de transmissão e 22,8941% para radiação solar direta em escala de utilidade), ficando o conjunto final, utilizado nos testes, com três variáveis.

2.3 Modelo Híbrido Evolucionário-Fuzzy C-means

Na *clusterização fuzzy*, ou *soft*, cada instância possui grau de participação, no intervalo [0;1], em cada um dos *clusters*, de modo que a soma total das participações, para cada instância, é igual a “1”. Após a execução do algoritmo, adotou-se o arredondamento simples da matriz de partição *fuzzy*, que armazena os graus de participação para cada instância em cada *cluster*, de modo que os valores maiores ou iguais a “0,5” foram arredondados para “1” e os valores menores que “0,5” foram arredondados para “0” (uma vez que são apenas dois *clusters*). Poder-se-ia adotar que os valores intermediários de graus de participação (compreendidos entre 0.45 e 0.55, por exemplo) seriam representantes de um terceiro *cluster*, intermediário. Como, porém, o objetivo do trabalho (respaldado por três métricas distintas) era obter apenas dois *clusters*, tal procedimento não foi realizado.

A função objetivo empregada foi a *fuzzy c-means functional*, base para uma ampla família de algoritmos de *clusterização fuzzy* [James C. Bezdek 1981; Dunn 1974; Babuška 1998]. O valor desta função pode ser entendido como uma medida da variância total entre cada instância e o centro de seu respectivo *cluster* [Babuška 1998]. A minimização desta função representa um problema de otimização não-linear, que pode ser solucionado de inúmeras maneiras, entre elas: algoritmos genéticos [Babu e Murty 1994], *simulated annealing* [Desarbo 1982] e *grouped coordinate minimization* [J. C. Bezdek et al. 1987; Hathaway e Bezdek 1991]. A mais popular, entretanto, é a *fuzzy c-means* [Babuška 1998].

Uma vez que o algoritmo *fuzzy c-means* padrão pressupõe a inicialização aleatória da matriz de partição *fuzzy*, existe a necessidade de um maior número de iterações para alcançar a solução final. Pensando nessa limitação, nós propomos a inicialização da matriz de partição *fuzzy* através de três meta-heurísticas distintas: algoritmo genético [Goldberg 1989] e evolução diferencial [Storn 1996; Storn e Price 1997], estratégias evolutivas e *particle swarm optimization* [Kennedy e Eberhart 1995; Kennedy, Eberhart, e Shi 2001; Poli 2008], estratégia de enxames.

3. Resultados

Como esperado, a proposta sugerida conseguiu reduzir o número de iterações. A versão clássica do algoritmo *fuzzy c-means* demandou 28 iterações, enquanto que a inicializada por algoritmo genético e *PSO* demandou 23 iterações, e apresentaram a mesma convergência. A versão inicializada por evolução diferencial precisou de 22 iterações. Todas alcançaram o mesmo valor final para a função objetivo (8.737,8), como pode-se observar na Fig. 2 (a).

Percebe-se que, quando o algoritmo é inicializado pelas meta-heurísticas, o valor da função objetivo é 18,7% mais baixo (10.076,0) do que quando inicializado aleatoriamente (12.394,0). Nota-se também, pelo comportamento da curva de treinamento, que o algoritmo *fuzzy*



c-means realiza um refinamento da solução a partir de certo nível da função objetivo (aproximadamente 10.000).

Quanto ao resultado final da *clusterização*, houve divergência em apenas 1 instância entre as 5.063. Para o algoritmo clássico, 835 instâncias foram classificadas como pertencentes ao *cluster 1* e 4.228 como pertencentes ao *cluster 2*. Para o método proposto, 836 instâncias foram alocadas para o *cluster 1* e 4.227 para o *cluster 2*, conforme ilustrado pela Fig. 2 (b).

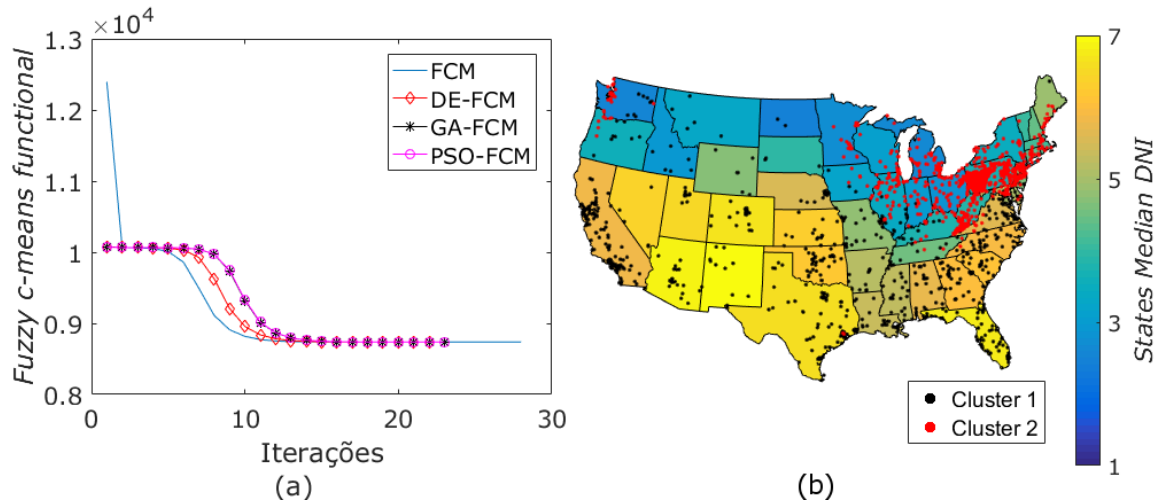


Fig. 2. (a) Resultados da estratégia proposta (*DE-FCM*, inicializado por evolução diferencial; *GA-FCM*, inicializado por algoritmo genético; e *PSO-FCM*, inicializado por *PSO*) e da versão clássica (*FCM*). (b) Resultado final da *clusterização* e incidência solar mediana para o território continental americano.

Visualmente, pode-se perceber a predominância da alocação das instalações em estados de maior incidência média de radiação solar, sendo este, portanto, o fator predominante na convergência do algoritmo, o que está de acordo com o esperado (embora este atributo contribua menos com a variância total do conjunto).

O número de *clusters* foi validado por três métricas: *Calinski-Harabasz Index*, *Davies-Bouldin Index* e *Silhouette Coefficient*. Os três testes foram unânimes, indicando dois *clusters* como o número ideal, ou seja, um *cluster* para locais com potencial para alocação de instalações de captação de energia solar (principalmente nos estados ensolarados) e outro para locais sem potencial (estados com menor incidência solar).

Para análise da diferença entre os *clusters*, a Tabela 3 apresenta a estatística para cada um deles. Nota-se que as instâncias alocadas para o *cluster 1* possuem maior média para as variáveis área mapeada (Var. 1) e incidência solar, em escala de utilidade (Var. 3) e em pequena escala (Var. 4). Para a distância até linhas de transmissão (Var. 2), o *cluster 2* obteve a menor média, uma diferença de 11,6% em relação ao *cluster 1*. É interessante notar que o *cluster 2*, mesmo possuindo mais de cinco vezes o número de instâncias do *cluster 1*, possui menores desvio padrão e variância, para todas as variáveis, um dos indicativos de qualidade (compacidade) de uma *clusterização* [Babuška 1998].

Tabela 3. Estatística para os *clusters* encontrados

	<i>Cluster 1 (836 instâncias)</i>				<i>Cluster 2 (4227 instâncias)</i>			
	Var. 1	Var. 2	Var. 3	Var. 4	Var. 1	Var. 2	Var. 3	Var. 4
Mínimo	100,80	0,00	3,47	4,17	100,04	0,00	2,17	2,86
Média	347,74	2,50	4,76	5,20	255,50	2,21	3,30	4,21
Máximo	996,00	63,84	7,75	6,68	999,56	49,88	4,36	4,79
Desvio padrão	236,30	5,54	0,86	0,48	147,32	4,07	0,17	0,12
Variância	55.839,95	30,66	0,73	0,23	21.702,83	16,54	0,03	0,01



4. Conclusão

Como resultado da abordagem híbrida proposta, houve um incremento na velocidade de treinamento do algoritmo *fuzzy c-means*, que necessitou de um menor número de iterações para atingir o mesmo valor da função objetivo. Os dois *clusters* resultantes apresentaram características estatísticas que validam a inicialização pelas meta-heurísticas testadas, uma vez que o *cluster 1* apresentou maiores médias para três das quatro variáveis analisadas, sendo o *cluster* das instâncias com maior potencial para instalação de usinas solares. O *cluster 2* apresentou melhor compacidade, tanto em termos das variáveis analisadas quanto em relação à localização geográfica de suas instâncias.

É importante ressaltar a importância do estudo em relação a uma melhor alocação dos parques recursos disponíveis para energias renováveis (em relação às outras energias convencionais, como os combustíveis fósseis). Com a *clusterização* pode-se determinar *a priori* os locais com melhor potencial e, respectivamente, retorno sobre o investimento feito. Entre os principais benefícios decorrentes da instalação de usinas solares, podemos citar a própria economia com custos de energia, seguido dos ganhos ambientais, com redução de emissões diretas e indiretas de gases do efeito estufa, além da criação de empregos. Outro importante benefício é a limpeza de áreas antes contaminadas e improdutivas, que geravam riscos para a saúde e o meio ambiente. Especificamente para o caso norte-americano podemos citar, ainda, os subsídios estatais, como o *Solar Renewable Energy Certificate (SREC)*, uma *commodity* energética (específica para geração solar) não tangível que é emitida ao se gerar 1 MWh de energia de fonte solar e que pode ser vendida no mercado [Bird, Heeter, e Kreyck 2011].

A título de ilustração, com a implantação de 190 unidades de geração de energia solar (em seu potencial máximo), pelo programa *RE-Powering America's Land*, seria possível reduzir as emissões de gases do efeito estufa em 1,7 milhão de toneladas de CO₂ por ano, o equivalente (em termos de dióxido de carbono equivalente) a 500 mil toneladas de lixo ou 193 milhões de galões de gasolina. Com relação a empregos, o *2015 Solar Jobs Census* estimou em 209 mil o número de empregados na industrial solar norte-americana. Tudo isso ressalta o potencial de ganhos ambiental-econômico-social das energias renováveis e, em especial, a solar [EPA 2016].

Referências

- Aggarwal, Charu C. 2015. *Data Mining: The Textbook*. Springer International Publishing. doi:10.1007/978-3-319-14142-8.
- Almeida, C. M V B, F. Agostinho, D. Huisingh, e B. F. Giannetti. 2017. "Cleaner Production towards a sustainable transition". *Journal of Cleaner Production* 142. Elsevier Ltd: 1–7. doi:10.1016/j.jclepro.2016.10.094.
- Amigó, Enrique, Julio Gonzalo, Javier Artilles, e Felisa Verdejo. 2009. "A comparison of extrinsic clustering evaluation metrics based on formal constraints". *Information Retrieval* 12 (4): 461–86. doi:10.1007/s10791-008-9066-8.
- Apostolidis, N., e N. Hutton. 2006. "Integrated water management in brownfield sites: More opportunities than you think". *Desalination* 188 (1–3): 169–75. doi:10.1016/j.desal.2005.04.114.
- Babu, G. Phanendra, e M. Narasimha Murty. 1994. "Clustering with evolution strategies". *Pattern Recognition* 27 (2): 321–29. doi:10.1016/0031-3203(94)90063-9.
- Babuška, Robert. 1998. *Fuzzy Modeling for Control*. New York: Springer International Publishing. doi:10.1007/978-94-011-4868-9.



- Baños, R., F. Manzano-Agugliaro, F. G. Montoya, C. Gil, A. Alcayde, e J. Gómez. 2011. "Optimization methods applied to renewable and sustainable energy: A review". *Renewable and Sustainable Energy Reviews* 15 (4): 1753–66. doi:10.1016/j.rser.2010.12.008.
- Ben-Hur, Asa, David Horn, Hava T. Siegelmann, e Vladimir Vapnik. 2001. "Support Vector Clustering". *Journal of Machine Learning Research* 2: 125–37. www.jmlr.org/papers/volume2/horn01a/rev1/horn01ar1.pdf.
- Bergius, Kristine, e Tomas Öberg. 2007. "Initial screening of contaminated land: A comparison of US and Swedish methods". *Environmental Management* 39 (2): 226–34. doi:10.1007/s00267-006-0005-4.
- Bezdek, J. C., R. J. Hathaway, R. E. Howard, C. A. Wilson, e M. P. Windham. 1987. "Local convergence analysis of a grouped variable version of coordinate descent". *Journal of Optimization Theory and Applications* 54 (3): 471–77. doi:10.1007/BF00940196.
- Bezdek, James C. 1981. *Pattern Recognition with Fuzzy Objective Function Algorithms*. SIAM Review. Vol. 25. Springer US. doi:10.1007/978-1-4757-0450-1.
- Bezdek, James C., Robert Ehrlich, e William Full. 1984. "FCM: The fuzzy c-means clustering algorithm". *Computers & Geosciences* 10 (2–3): 191–203. doi:10.1016/0098-3004(84)90020-7.
- Bird, Lori, Jenny Heeter, e Claire Kreycik. 2011. "Solar Renewable Energy Certificate (SREC) Markets: Status and Trends". <http://www.nrel.gov/docs/fy12osti/52868.pdf>.
- Bramer, Max. 2016. *Principles of Data Mining*. 3^{ed}. Undergraduate Topics in Computer Science. London: Springer-Verlag London. doi:10.1007/978-1-4471-7307-6.
- Cadez, Simon, e Albert Czerny. 2016. "Climate change mitigation strategies in carbon-intensive firms". *Journal of Cleaner Production* 112. Elsevier Ltd: 4132–43. doi:10.1016/j.jclepro.2015.07.099.
- Cao, Kang, e Hua Guan. 2007. "Brownfield redevelopment toward sustainable urban land use in China". *Chinese Geographical Science* 17 (2): 127–34. doi:10.1007/s11769-007-0127-5.
- de Sousa, Christopher A. 2003. "Turning brownfields into green space in the City of Toronto". *Landscape and Urban Planning* 62 (4): 181–98. doi:10.1016/S0169-2046(02)00149-4.
- Dempster, A. P., N. M. Laird, e D. B. Rubin. 1977. "Maximum Likelihood from Incomplete Data via the EM Algorithm". *Journal of the Royal Statistical Society. Series B - Methodological* 39 (1): 1–38. <https://www.jstor.org/stable/2984875>.
- Desarbo, Wayne S. 1982. "Gennclus: New models for general nonhierarchical clustering analysis". *Psychometrika* 47 (4): 449–75. doi:10.1007/BF02293709.
- Dunn, J. C. 1974. "A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters". *Journal of Cybernetics* 3 (3): 32–57. doi:10.1080/01969727308546046.
- EPA. 2016. "RE-Powering America's Land Initiative: Benefits Matrix". EPA. <https://goo.gl/XMov1T>.
- Ester, Martin, Hans-Peter Kriegel, Jörg Sander, e Xiaowei Xu. 1996. "A Density-Based



- Algorithm for Discovering Clusters in Large Spatial Databases with Noise”. In *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining - KDD '96*, 226–31. Portland: AAAI Press. <https://www.aaai.org/Papers/KDD/1996/KDD96-037.pdf>.
- Fayyad, Usama, Gregory Piatetsky-Shapiro, e Padhraic Smyth. 1996. “From data mining to knowledge discovery in databases”. *AI magazine*, 37–54. doi:10.1145/240455.240463.
- Fernández-García, Aránzazu, Esther Rojas, Manuel Pérez, Ricardo Silva, Quetzalcoatl Hernández-Escobedo, e Francisco Manzano-Agugliaro. 2015. “A parabolic-trough collector for cleaner industrial process heat”. *Journal of Cleaner Production* 89: 272–85. doi:10.1016/j.jclepro.2014.11.018.
- Gamarra, Carlos, Josep M. Guerrero, e Eduardo Montero. 2016. “A knowledge discovery in databases approach for industrial microgrid planning”. *Renewable and Sustainable Energy Reviews* 60. Elsevier: 615–30. doi:10.1016/j.rser.2016.01.091.
- Goldberg, David E. 1989. *Genetic Algorithms in Search, Optimization & Machine Learning*. Boston: Addison-Wesley Professional.
- González, Mario Orestes Aguirre, Joeberson S. Gonçalves, e Rafael M. Vasconcelos. 2017. “Sustainable development: Case study in the implementation of renewable energy in Brazil”. *Journal of Cleaner Production* 142: 461–75. doi:10.1016/j.jclepro.2016.10.052.
- Greenberg, Michael, e M. Jane Lewis. 2000. “Brownfields Redevelopment, Preferences and Public Involvement: A Case Study of an Ethically Mixed Neighbourhood”. *Urban Studies* 37 (13): 2501–14. doi:10.1080/00420980020005442.
- Guha, Sudipto, Rajeev Rastogi, e Kyuseok Shim. 1998. “CURE: An Efficient Clustering Algorithm for Large Databases”. In *Proceedings of the 1998 ACM SIGMOD International Conference on Management of Data - SIGMOD '98*, 73–84. New York: ACM. doi:10.1145/276304.276312.
- . 2000. “Rock: a robust clustering algorithm for categorical attributes”. *Information Systems* 25 (5): 345–66. doi:10.1016/S0306-4379(00)00022-3.
- Halkidi, Maria, Yannis Batistakis, e Michalis Vazirgiannis. 2001. “On clustering validation techniques”. *Journal of Intelligent Information Systems* 17 (2–3): 107–45. doi:10.1023/A:1012801612483.
- Hathaway, R. J., e J. C. Bezdek. 1991. “Grouped coordinate minimization using Newton’s method for inexact minimization in one vector coordinate”. *Journal of Optimization Theory and Applications* 71 (3): 503–16. doi:10.1007/BF00941400.
- Hinneburg, Alexander, e Daniel A Keim. 2003. “A general approach to clustering in large databases with noise”. *Knowledge and Information Systems* 5 (4): 387–415. doi:10.1007/s10115-003-0086-9.
- Hinneburg, Alexander, e Daniel A. Keim. 1998. “An efficient approach to clustering in large multimedia databases with noise”. In *Proceedings of 4th International Conference on Knowledge Discovery and Data Mining - KDD '98*, 58–65. New York: AAAI Press. <http://www.aaai.org/Papers/KDD/1998/KDD98-009.pdf>.
- Huang, Zhexue. 1998. “Extensions to the k-means algorithm for clustering large data sets with categorical values”. *Data Mining and Knowledge Discovery* 2 (3): 283–304.



doi:10.1023/A:1009769707641.

- Kaufman, Martin M., Daniel T. Rogers, e Kent S. Murray. 2005. “An empirical model for estimating remediation costs at contaminated sites”. *Water, Air, and Soil Pollution* 167 (1–4): 365–86. doi:10.1007/s11270-005-0214-0.
- Kennedy, James, e Russell Eberhart. 1995. “Particle swarm optimization”. In *Proceedings of International Conference on Neural Networks ICNN '95*, 4:1942–48. IEEE. doi:10.1109/ICNN.1995.488968.
- Kennedy, James, Russell C. Eberhart, e Yuhui Shi. 2001. *Swarm intelligence. Science*. Morgan Kaufmann.
- Kovacs, Helga, e Katalin Szemmelveisz. 2017. “Disposal options for polluted plants grown on heavy metal contaminated brownfield lands: A review”. *Chemosphere* 166. Elsevier Ltd: 8–20. doi:10.1016/j.chemosphere.2016.09.076.
- Lambin, Eric F., e Patrick Meyfroidt. 2011. “Global land use change, economic globalization, and the looming land scarcity”. *Proceedings of the National Academy of Sciences of the United States of America* 108 (9): 3465–3472. doi:10.1073/pnas.1100480108.
- Li, Xiaonuo, Wentao Jiao, Rongbo Xiao, Weiping Chen, e Wen Liu. 2017. “Contaminated sites in China: Countermeasures of provincial governments”. *Journal of Cleaner Production* 147. Elsevier Ltd: 485–96. doi:10.1016/j.jclepro.2017.01.107.
- Lima, Fátima, Paula Ferreira, e Filipa Vieira. 2013. “Strategic impact management of wind power projects”. *Renewable and Sustainable Energy Reviews* 25. Elsevier: 277–90. doi:10.1016/j.rser.2013.04.010.
- MacQueen, J. 1967. “Some methods for classification and analysis of multivariate observations”. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1:281–97. Berkeley: University of California Press.
- Manzano-Agugliaro, F., M. J. Sanchez-Muros, F. G. Barroso, A. Martínez-Sánchez, S. Rojo, e C. Pérez-Bañón. 2012. “Insects for biodiesel production”. *Renewable and Sustainable Energy Reviews* 16 (6): 3744–53. doi:10.1016/j.rser.2012.03.017.
- Meilă, Marina. 2007. “Comparing clusterings: an information based distance”. *Journal of Multivariate Analysis* 98 (5): 873–95. doi:10.1016/j.jmva.2006.11.013.
- Morio, Maximilian, Sebastian Schädler, e Michael Finkel. 2013. “Applying a multi-criteria genetic algorithm framework for brownfield reuse optimization: Improving redevelopment options based on stakeholder preferences”. *Journal of Environmental Management* 130. Elsevier Ltd: 331–46. doi:10.1016/j.jenvman.2013.09.002.
- Ng, Raymond T., e Jiawei Han. 1994. “Efficient and effective clustering methods for spatial data mining”. In *Proceedings of the 20th International Conference on Very Large Data Bases - VLDB '94*, 144–55. San Francisco: Morgan Kaufmann.
- Nuissl, Henning, e Christoph Schroeter-Schlaack. 2009. “On the economic approach to the containment of land consumption”. *Environmental Science and Policy* 12 (3): 270–80. doi:10.1016/j.envsci.2009.01.008.
- Onat, Nuri Cihat, Murat Kucukvar, e Omer Tatari. 2014. “Integrating triple bottom line input-



output analysis into life cycle sustainability assessment framework: The case for US buildings”. *International Journal of Life Cycle Assessment* 19 (8): 1488–1505. doi:10.1007/s11367-014-0753-y.

- Orriols-Puig, Albert, Francisco J. Martínez-López, Jorge Casillas, e Nick Lee. 2013. “Unsupervised KDD to creatively support managers’ decision making with fuzzy association rules: A distribution channel application”. *Industrial Marketing Management* 42 (4). Elsevier Inc.: 532–43. doi:10.1016/j.indmarman.2013.03.005.
- Pal, Nikhil R., James C. Bezdek, e Eric C.-K. Tsao. 1993. “Generalized Clustering Networks and Kohonen’s Self-Organizing Scheme”. *IEEE Transactions on Neural Networks* 4 (4): 549–57. doi:10.1109/72.238310.
- Perea-Moreno, Alberto-Jesús, Amós García-Cruz, Nuria Novas, e Francisco Manzano-Agugliaro. 2017. “Rooftop analysis for solar flat plate collector assessment to achieving sustainability energy”. *Journal of Cleaner Production* 148: 545–54. doi:10.1016/j.jclepro.2017.02.019.
- Poli, Riccardo. 2008. “Analysis of the Publications on the Applications of Particle Swarm Optimisation”. *Journal of Artificial Evolution and Applications* 2008 (2): 1–10. doi:10.1155/2008/685175.
- Roiger, Richard J. 2017. *Data Mining: A Tutorial-Based Primer*. 2^oed. CRC Press.
- Rong, Liming, Chengliang Zhang, Dongsheng Jin, e Ziyun Dai. 2017. “Assessment of the potential utilization of municipal solid waste from a closed irregular landfill”. *Journal of Cleaner Production* 142. Elsevier Ltd: 413–19. doi:10.1016/j.jclepro.2015.10.050.
- Sheikholeslami, Gholamhosein, Surojit Chatterjee, e Aidong Zhang. 1998. “Wavecluster: A multi-Resolution Clustering Approach for Very Large Spatial Databases”. In *Proceedings of 24rd International Conference on Very Large Data Bases - VLDB '98*, 428–39. San Francisco: Morgan Kaufmann.
- Simas, Moana, e Sergio Pacca. 2013. “Energia eólica, geração de empregos e desenvolvimento sustentável”. *Estudos Avançados* 27 (77): 99–116. doi:10.1590/S0103-40142013000100008.
- Storn, Rainer. 1996. “On the usage of differential evolution for function optimization”. In *Proceedings of North American Fuzzy Information Processing - NAFIPS '96*, 519–23. IEEE. doi:10.1109/NAFIPS.1996.534789.
- Storn, Rainer, e Kenneth Price. 1997. “Differential Evolution – A Simple and Efficient Heuristic for global Optimization over Continuous Spaces”. *Journal of Global Optimization* 11 (4): 341–59. doi:10.1023/A:1008202821328.
- Theodoridis, Sergios, e Konstantinos Koutroumbas. 2009. *Pattern Recognition*. 4th ed. Academic Press.
- U.S. Government Publishing Office. 2002. “Public Law 107-118 - Small Business Liability Relief and Brownfields Revitalization Act”. *H.R. 2869*. <https://goo.gl/UK19n2>.
- . 2011. “40 C.F.R. 239-282 - Solid Wastes”. *Code of Federal Regulations (annual ed)*. <https://goo.gl/UBCLDF>.
- . 2015. “42 U.S.C. 9601-9628 - Hazardous Substances Releases, Liability,



- Compensation”. *United States Code, 2012 Edition, Supplement 3, Title 42 - The Public Health and Welfare, Subchapter I*. <https://goo.gl/y0ki6N>.
- van Straalen, Nico M. 2002. “Assessment of soil contamination: a functional perspective”. *Biodegradation* 13 (1): 41–52. doi:10.1023/a:1016398018140.
- Wang, Wei, Jiong Yang, e Richard R. Muntz. 1997. “STING: A Statistical Information Grid Approach to Spatial Data Mining”. In *Proceedings of 23rd International Conference on Very Large Data Bases - VLDB '97*, 186–95. San Francisco: Morgan Kaufmann.
- Witten, Ian H., Eibe Frank, Mark A. Hall, e Christopher J. Pal. 2017. *Data Mining: Practical Machine Learning Tools and Techniques*. 4^{ed}. Morgan Kaufmann.
- Xu, Rui, e Donald Wunsch II. 2005. “Survey of Clustering Algorithms”. *IEEE Transactions on Neural Networks* 16 (3): 645–78. doi:10.1109/TNN.2005.845141.
- Zaki, Mohammed J., e Wagner Meira Jr. 2014. *Data Mining and Analysis: Fundamental Concepts and Algorithms*. Cambridge University Press.
<https://books.google.com/books?hl=en&lr=&id=Gh9GAwAAQBAJ&pgis=1>.
- Zhang, Tian, Raghu Ramakrishnan, e Miron Livny. 1996. “BIRCH: An Efficient Data Clustering Databases Method for Very Large”. In *Proceedings of the 1996 ACM SIGMOD International Conference on Management of Data - SIGMOD '96*, 1:103–14. New York: ACM. doi:10.1145/233269.233324.