

Análise de Similaridade Climática entre Marte e Diferentes Regiões do Planeta Terra por Meio dos algoritmos K-Means e Fuzzy C-Means

Fábio José Justo dos Santos^{1,2}

¹Instituto Hardware BR
Campinas – SP – Brasil

²Instituto Federal de Educação, Ciência e Tecnologia (IFSP)
Araraquara – SP – Brasil
fabiojjs@ifsp.edu.br

Abstract. *This work investigates the climatic similarity between Mars and 19 regions on Earth using the K-means and Fuzzy C-means algorithms. The variables were defined based on the measurements available for Mars. The results show that, with lower granularity, some stations in Antarctica tend to cluster with Mars and, with more clusters, Mars tends to isolate into a single group. Regions of the Atacama Desert also demonstrated some similarity to Mars. The fuzzy analysis indicated that the Antarctic HOEM weather station is the most similar to Martian data. Finally, the application of K-means and Fuzzy C-means is satisfactory for identifying climatic analogies such as described in this study.*

Resumo. *Este trabalho investiga a similaridade climática entre Marte e 19 regiões da Terra por meio dos algoritmos K-means e Fuzzy C-means. As variáveis foram definidas a partir das medições disponíveis em Marte. Os resultados mostram que, com menor granularidade, algumas estações na Antártida tendem a agrupar-se com Marte e, com mais clusters, Marte tende a isolar-se em um grupo único. Regiões do Atacama também demonstraram proximidade com Marte. A análise fuzzy indicou a estação climática Antártica HOEM como a mais semelhante aos dados de Marte. Ao final, conclui-se que a aplicação de K-means e do Fuzzy C-means é satisfatória para identificar analogias climáticas como a descrita neste trabalho.*

1. Introdução

A Inteligência Artificial (IA) tem sido utilizada como instrumento de avanço tecnológico em diversos domínios do conhecimento, como engenharia [Joskowicz e Slomovitz 2024], saúde [Shaw e Gohel 2025], logística [Kumar e Chidambara 2025], esportes [Li et. al. 2025], trânsito [Lee e Shin 2019] e, também, na área aeroespacial [Scharf et al. 2004], [Gaurav et al. 2024] e [Ramírez et al. 2025].

Com o crescimento da exploração interplanetária, o interesse na exploração do planeta Marte também aumentou também. Somente em Biswal (2020), são mencionadas 44 missões à Marte, das quais 18 obtiveram êxito e 26 tiveram algum contratempo. A publicação online "*The Mars Report*", mantida pela NASA, informa que atualmente

existem 5 missões ativas e que o início da exploração do planeta vermelho está completando 60 anos. No entanto, outras agências espaciais também possuem projetos de exploração, não só do planeta Marte, mas também de outros planetas. Dentre algumas das principais missões de investigação do planeta Marte, podemos citar a *Phoenix Mars*, com lançamento realizado pela NASA em 2007, e liderada pelo cientista brasileiro Nilton de Oliveira Rennó, a *Mars 2020*, que levou o veículo espacial *Perseverance* até Marte em busca de vida microbiana e a MSL (*Mars Science Laboratory*), também conhecida como *Spirit*, que levou a estação de monitoramento climático REMS (*Rover Environmental Monitoring Station*) até Marte para coletar informações como temperatura, umidade relativa e vento.

Diversos trabalhos científicos têm focado na avaliação do clima em Marte. Fisher et al (2019) apresentam um estudo onde após uma recalibragem do sensor PHX TECP (*Phoenix Thermal and Electrical Conductivity Probe*) foi possível identificar que há mais umidade relativa do ar na atmosfera de Marte do que os índices relatados anteriormente em Zent et al. (2010) e Zent et al. (2016). Azua-Bustos et al. (2022) apontam o deserto do Atacama como o local no planeta Terra com clima mais próximo ao existente em Marte, principalmente em virtude do clima extremamente seco e, em virtude desta semelhança, justificam a razão de diversos testes de equipamentos e experimentos serem realizados neste ambiente. No entanto, o deserto do Atacama não possui temperaturas tão baixas quanto as encontradas em Marte. Fang et al. (2025) apresentam um estudo sobre a análise global da hipsometria (formação do relevo e das bacias) em Marte, na Terra, na Lua em fractais artificiais e em formas de relevo simuladas. Os resultados indicam que a formação do relevo marciano ocorreu sobre condições mais áridas que os demais.

Diante do exposto, este trabalho tem como propósito avaliar diferentes regiões do planeta Terra para comparar com o clima em Marte. Para esta análise, foram realizados agrupamentos por meio dos algoritmos K-Means e Fuzzy C-Means. Ao final foi realizada uma análise dos resultados, que indicaram quais regiões do planeta Terra possuem maior similaridade climática com Marte. O restante deste artigo está organizado como segue: na Seção 2 os algoritmos Fuzzy C-Means e K-Means são apresentados, bem como as métricas de avaliação dos resultados. Na Seção 3 são apresentados os dados utilizados, o pré-processamento realizado, os resultados obtidos e sua análise. Na seção 4 as considerações finais e as possibilidades de trabalhos futuros.

2. Clusterização

Os algoritmos de agrupamento de dados são considerados, em sua essência, uma técnica de aprendizado de máquina não supervisionado, pois não requerem dados rotulados ou qualquer outro tipo de atividade supervisionada para descobrir, automaticamente, agrupamentos no conjunto de dados disponibilizados.

Um dos principais algoritmos de agrupamento, o K-Means, foi empregado pela primeira vez em por MacQueen (1967). No entanto, Lloyd (1957) desenvolveu e aplicou o K-means padrão nos laboratórios Bell inicialmente. Porém, a publicação do seu trabalho ocorreu somente em 1982 [Lloyd 1982]. Outro algoritmo de agrupamento amplamente utilizado é o Fuzzy C-means [Bezdek et al 1981], uma extensão fuzzy do K-means, ou seja, os dados, ou instâncias, podem possuir graus de pertinência em diferentes clusters, o que pode ser uma vantagem do Fuzzy C-means na análise de dados

com representações linguísticas (quente, alto, perto, muito frio, ...) e em situações em que a fronteira entre os clusters não é bem definida. Na sequência são apresentados os conceitos dos algoritmos de agrupamento K-means e Fuzzy C-means.

2.1. K-Means

O objetivo do K-means [MacQueen 1967] é agrupar dados semelhantes em um mesmo cluster por meio da menor soma das distâncias quadradas entre os dados e o centroide, ou seja, criar grupos de dados homogêneos. A escolha do número de clusters, identificado pela variável K , impacta diretamente nos resultados obtidos, de modo a influenciar a granularidade e a generalização dos clusters.

O primeiro passo do algoritmo é selecionar aleatoriamente dentre os dados disponíveis, os K centroides que serão os centros iniciais dos grupos. Na sequência, cada amostra de dado é associada ao centroide mais próximo, sendo que ao final desta etapa teremos o agrupamento inicial.

Definido o agrupamento inicial, os centroides devem ser atualizados por meio da média e, assim, teremos novos centroides a cada nova etapa. O algoritmo encerra a sua execução após a diferença do erro médio quadrático entre uma iteração e outra ficar abaixo de um limite pré-definido ou após o algoritmo atingir um número máximo de iterações.

2.2. Fuzzy C-Means

O algoritmo Fuzzy C-means (FCM), proposto por Bezdek et al (1981), é uma extensão do agrupamento clássico realizado pelo K-means [MacQueen 1967] e baseia-se na teoria dos conjuntos fuzzy [Zadeh 1965]. Enquanto no K-means cada dado pertence exclusivamente a um único cluster, no Fuzzy C-means cada dado pode pertencer simultaneamente a múltiplos clusters, com diferentes graus de pertinência. A soma dos graus de pertinência de cada dado em todos os clusters deve ser igual a 1. Uma das principais vantagens do FCM é lidar satisfatoriamente com situações em que os limites entre os clusters não são muito bem definidos. Para isso, o Fuzzy C-means busca encontrar as partições fuzzy que minimizem sua função objetivo, apresentada na Equação 1.

Equação 1 - Função Objetivo do Fuzzy C-means

$$J(U, c_1, \dots, c_c) = \sum_{i=1}^c J_i = \sum_{i=1}^c \sum_j^n u_{ij}^m d_{ij}^2$$

onde c_i é o centroide do cluster i , C é o número de clusters; n a quantidade de dados; u_{ij} é um valor entre 0 e 1 que representa o grau de pertinência do dado j no cluster i , definido pela Equação 2; d_{ij} é a distância euclidiana entre i -ésimo cluster e o j -ésimo dado; e m $[1, \infty]$ é o fator de fuzzificação, que usualmente é definido entre 1,25 e 2. Quanto mais próximo de 1 for m , mais rígidas serão as partições, ou seja, se $m=1$, então teremos partições crisp resultantes do processo de agrupamento. Na sequência é apresentado o passo a passo do algoritmo Fuzzy C-means:

1. Defina o número C de clusters, o fator de fuzzificação m e as condições de parada (número máximo de iterações e a convergência da matriz de pertiência U);
2. Inicialize, aleatoriamente, a matriz de pertinência U com o grau de pertinência de cada dado j em cada um dos clusters i (u_{ij});

3. Iteração:

- 3.1. Atualizar os centroides por meio da Equação 3;
 - 3.2. Calcular os novos graus de pertinência dos dados conforme Equação 2;
 - 3.3. Verificar se atingiu número máximo de iterações ou se houve a convergência da matriz de pertinência; Se não atingiu condições de parada, retornar ao passo 3.1.
4. Concluída a etapa anterior, o resultado é obtido por meio da matriz de pertinência U_{ij} onde o elemento ij indica o grau de pertinência do dado j no cluster i .

Equação 2 – Cálculo do grau de pertinência do dado j no cluster i .

$$u_{ij} = \frac{1}{\sum_{k=1}^c \left(\frac{d_{ij}}{d_{kj}} \right)^{2/(m-1)}}$$

Os centroides para cálculo da matriz de pertinência U , são atualizados conforme cálculo apresentado na Equação 3.

Equação 3 – Atualização do centroide do cluster i .

$$c_i = \frac{\sum_{j=1}^n u_{ij}^m x_j}{\sum_{j=1}^n u_{ij}^m}$$

2.3. Métricas de Avaliação

Como mencionado anteriormente, o número de clusters K influencia diretamente no resultado obtido. Tanto para o algoritmo K-means quanto para o Fuzzy C-means, podem ser utilizadas abordagens para determinar, matematicamente, o melhor número de clusters de acordo com os dados disponíveis. Os métodos cotovelo [Thorndike 1953] coeficiente de silhueta [Rousseeuw, 1986] e coeficiente de silhueta fuzzy [Campello e Hruschka 2006] são os mais comumente utilizados.

À medida que aumenta o número de clusters k , a variação dentro do grupo diminui. Após um certo número de clusters, essa variação dentro do grupo torna-se menos significativa. No método cotovelo, essa redução na variação forma um ponto inflexão no gráfico, o que indica, de acordo com o método, o número ideal de clusters. Em outras palavras, com o método cotovelo buscamos o ponto em que a inclusão de clusters adicionais resulta em ganhos marginais na explicação da variância, definido pela WCSS (*Within-Cluster Sum of Squares*), conforme apresentado na Equação 4, onde k é o número de clusters, x é um dado dentro do cluster C_i , c_i é o centroide do cluster C_i .

Equação 4 – Cálculo WCSS para explicação da variância

$$WCSS = \sum_{i=1}^k \sum_{x \in C_i} \|x - c_i\|^2$$

Em diversas situações o método do cotovelo pode não ser suficiente para determinar o número ideal de clusters. Nestes casos o coeficiente de silhueta pode ser útil para determinar o melhor número k de clusters. O valor do coeficiente de silhueta varia entre $[-1, 1]$, sendo que quanto mais próximo de 1 estiver o índice, melhor é o k . A Equação 5 apresenta o modelo matemático utilizado para o cálculo do coeficiente de silhueta.

Equação 5 – Cálculo do coeficiente de silhueta

$$S_j = \frac{b(j) - a(j)}{\max \{a(j), b(j)\}},$$

onde S_j representa o coeficiente de silhueta para o objeto j , $a(j)$ é a distância média entre o objeto j e todos os demais dados do seu cluster (distância intra-cluster) e $b(j)$ é a menor distância média entre o objeto j e todos os demais objetos de cada um dos demais clusters, ou seja, a distância média entre o objeto j e os objetos do cluster mais próximo.

Embora o valor S_j seja um dos principais parâmetros de referência na análise, também é preciso observar se os coeficientes individuais dos clusters não estão, por exemplo, abaixo da média e se também não há grandes flutuações no tamanho dos clusters. Pontuações de silhueta dos clusters próximas e acima da média e uma uniformidade no tamanho dos clusters tendem a indicar um melhor número k .

O coeficiente de silhueta fuzzy é uma métrica de avaliação de agrupamento que estende o conceito crisp do coeficiente de silhueta para algoritmos de agrupamento fuzzy. Este índice considera os dois maiores graus de pertinência (μ_{pj} e μ_{qj}) de cada objeto j para o cálculo do índice, além de introduzir um fator $\alpha \geq 0$ de ponderação. A Equação 6 para apresenta o modelo matemático para o cálculo do coeficiente de silhueta fuzzy.

Equação 6 – Cálculo do coeficiente de silhueta fuzzy (FS)

$$FS = \frac{\sum_{j=1}^N (\mu_{pj} - \mu_{qj})^\alpha s_j}{\sum_{j=1}^N (\mu_{pj} - \mu_{qj})^\alpha}$$

3. Experimentos e Resultados

Os experimentos realizados consistem na coleta dos dados climáticos de 19 diferentes regiões do planeta Terra para uma análise de similaridade com o clima existente no planeta Marte por meio dos algoritmos de agrupamento K-means e Fuzzy C-means. Ao final é apresentada uma comparação dos resultados obtidos.

Os dados referentes ao clima em Marte foram coletados por meio da plataforma Kaggle [Rojas 2022]. Os dados climáticos das diferentes regiões do planeta Terra foram obtidos por meio da plataforma Open-Meteo (<https://open-meteo.com/en/docs/historical-weather-api>). A escolha dos locais foi realizada após um estudo que para identificar diferentes regiões climáticas do planeta Terra, inclusive com a identificação de regiões com climas extremos. A partir deste estudo, foram definidas 19 regiões do planeta Terra com diferentes dados climáticos coletados por suas estações climáticas. São elas: Alaska BRW, Antártida CAAM, Antártida FLMM, Antártida FRSM, Antártida HOEM, Antártida VAAM, Atacama-APEX, Atacama Meteo, Deserto da Arábia, Deserto da Austrália, Deserto de Gobi, Atacama Central, Deserto de Neguev,

Deserto do Saara, Groenlândia *Summit Station*, Índia Mawsynram, Sibéria Golomyanny, São José dos Cordeiros e São Paulo.

Com exceção de São Paulo, que foi incluída para que tenhamos um parâmetro conhecido de comparação, a escolha destes locais levou em consideração a baixa umidade relativa e/ou a baixa temperatura. Apesar do deserto do Atacama ser mundialmente conhecido com o deserto mais seco do planeta Terra, durante as pesquisas foi identificado que em algumas épocas do ano em determinadas regiões da Antártida, como os Vales Secos de McMurdo, o clima torna-se mais seco que no próprio deserto do Atacama.

As variáveis utilizadas na análise foram definidas a partir dos dados climáticos disponíveis de Marte. No entanto, devido a um problema ocorrido em seus sensores, a base de dados que contém informações coletadas a partir da REMS, não possui informações sobre o vento e a umidade relativa do ar. Por outro lado, Fischer et al (2019) apresentam um estudo onde é possível concluir que existe mais umidade relativa em Marte do que o estimado nos estudos iniciais. A Figura 1 apresenta os índices da umidade relativa do ar em Marte coletada durante um período de 150 sóis em diferentes regiões. Cumpre destacar que em Marte um dia é referenciado como um sol, pois o dia em Marte possui 24 horas, 39 minutos e 35 segundos.

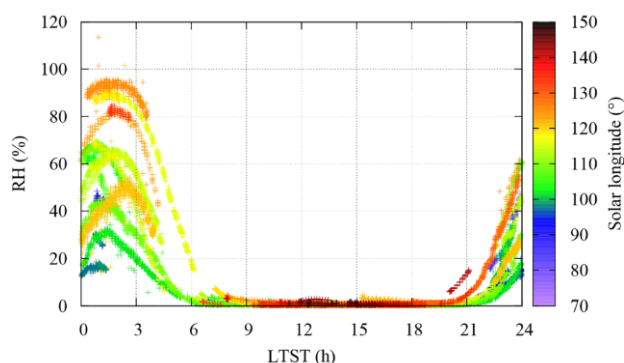


Figura 1. Umidade relativa do ar em Marte

As variáveis disponíveis do conjunto de dados [Rojas 2022] são: *earth_date_time*, *mars_date_time*, *sol_number*, *max_ground_temp(°C)*, *min_ground_temp(°C)*, *max_air_temp(°C)*, *min_air_temp(°C)*, *mean_pressure(Pa)*, *wind_speed(m/s)* e *humidity(%)*.

Das 3197 amostras disponíveis no *data set Mars Enviroment*, 29 foram removidas por apresentar dados incompletos (exceto vento e umidade relativa). Após a remoção das amostras incompletas, foi realizada uma análise dos dados e, em virtude de não haver dados disponíveis, a variável *wind_speed(m/s)* foi removida. No caso da variável *humidity(%)*, dada a sua extrema importância na análise comparativa e a indisponibilidade destes dados para teste, foi solicitado ao ChatGPT que realizasse uma análise do gráfico apresentado na Figura 1 e, a partir desta análise, gerasse 3168 amostras de dados a respeito da umidade relativa do ar, considerando uma estimativa dos índices máximo e mínimo por sol a partir dos dados apresentados no gráfico. Após algumas iterações, os dados obtidos representaram satisfatoriamente o gráfico da Figura 1. Outra variável importante para análise comparatória com o clima na Terra, é a precipitação, pois não é possível afirmar que uma determinada região no planeta Terra

possui clima semelhante a Marte se houver índices expressivos de precipitação nela. Como não havia esta variável nos dados coletados sobre Marte, a variável Precipitação foi adicionada ao conjunto de dados, dado este que foi coletado a partir das 19 estações climáticas mencionadas e, para Marte, preenchida com 0 (zero). Assim, as variáveis utilizadas para a análise, já renomeadas, são: TempMax, TempMin, Precipitação, PaMedia, UmidadeMin, UmidadeMax.

Para o pré-processamento dos dados, foram utilizadas as bibliotecas *Pandas*, para manipulação tabular, e *Scikit-Learn* para transformações matemáticas. Todos os dados foram concatenados em uma única matriz e em seguida foi realizada a normalização por meio da classe *StandardScaler*. Para fins de visualização exploratória bidimensional dos agrupamentos, empregou-se a Análise de Componentes Principais (PCA), reduzindo a dimensionalidade do conjunto de dados padronizado para dois componentes principais. Na sequência são apresentados os resultados obtidos.

3.1. Resultados Obtidos

Nesta subseção são apresentados os resultados obtidos por meio da aplicação dos algoritmos de agrupamento K-means e Fuzzy C-means. Para fins de análise dos dados, antes da execução dos algoritmos foi aplicado o método do cotovelo, a fim de obter um indicador do melhor k para execução dos algoritmos. O resultado é apresentado na Figura 2 e o melhor k indicado pelo modelo matemático é $k=9$. Embora visualmente não seja tão evidente a formação do cotovelo, é possível identificar uma atenuação na queda da inércia/soma dos quadrados na região onde o $k \geq 9$. No conteúdo compartilhado, também são apresentados todos os resultados para os índices coeficiente de silhueta e coeficiente de silhueta fuzzy para $k=2$ até $k=11$.

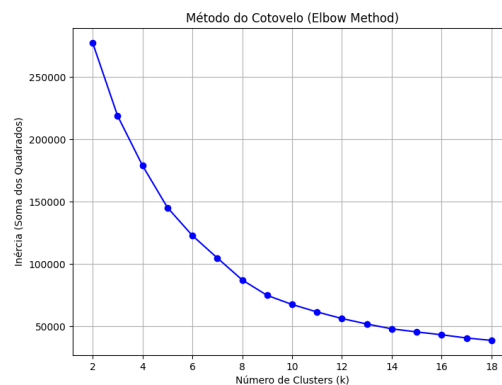


Figura 2. Resultados obtidos com o método do cotovelo para $2 \leq k \leq 18$

3.1.1. Agrupamento pelo K-means

Foram realizados experimentos com $2 \leq k \leq 11$ e os resultados indicam que o melhor coeficiente de silhueta foi obtido com $k=9$. O gráfico do coeficiente de silhueta para $k=9$ é apresentado na Figura 3.

Com 9 clusters não há indicação de semelhança do clima de Marte com nenhuma das regiões da Terra estudadas, o que é compreensível. Como é possível observar na Figura 4, com $k=9$ os dados de Marte formam um cluster único.

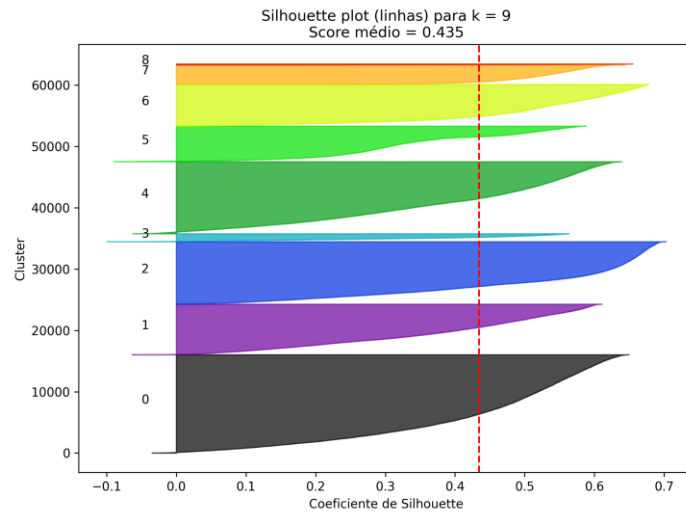


Figura 3. Gráfico do coeficiente de silhueta para $k=9$

Outra característica importante na análise é quais clusters estão próximos ao de Marte. Tal proximidade pode ser um indicativo de similaridade e os clusters obtidos para todos os k indicam que as regiões da Antártida e do Deserto do Atacama são as que mais se aproximam de Marte.

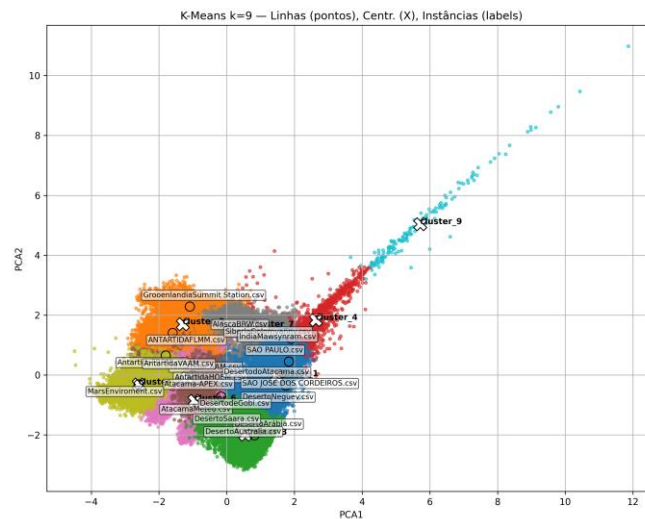


Figura 4. Agrupamento K-means com $k=9$

Na Tabela 1 são apresentados os clusters predominantes de cada região para $2 \leq k \leq 11$. Nas situações em que $k \geq 8$, podemos observar que nenhuma das 19 regiões está no mesmo cluster que Marte. Uma característica importante na análise é que para $2 \leq k \leq 7$ as estações meteorológicas Antártida CAAM, Antártida FRSM, Antártida HOEM e Antártida VAAM, todas elas estão no mesmo cluster de Marte. Esse comportamento deve-se, principalmente, pela semelhança climática existente em virtude das temperaturas extremamente baixas e da baixíssima umidade relativa do ar, características predominantes nos Vales Secos de McMurdo, local onde estão localizadas todas estas estações climáticas na Antártida.

Tabela 1. Resultado do agrupamento no K-means para $2 \leq k \leq 11$

	k=2	k=3	k=4	k=5	k=6	k=7	k=8	k=9	k=10	k=11
AntartidaFLMM	1	0	1	2	3	2	6	1	8	1
AlascaBRW	0	1	3	3	1	4	7	6	0	2
AntartidaCAAM	1	0	1	0	0	5	1	4	3	8
AntartidaFRSM	1	0	1	0	0	5	1	4	8	1
AntartidaHOEM	1	0	1	0	0	5	1	4	3	8
AntartidaVAAM	1	0	1	0	0	5	1	4	3	8
Atacama-APEX	1	0	1	2	3	0	2	5	1	4
AtacamaMeteo	1	2	0	1	5	0	2	5	1	4
DesertoArábia	0	2	0	1	5	6	4	2	6	0
DesertoAustralia	0	2	0	1	5	6	4	2	6	0
DesertoNeguev	0	1	3	3	2	3	0	0	2	10
DesertoSaara	0	2	0	1	5	6	4	2	6	0
DesertodeGobi	0	2	0	1	5	6	4	2	6	0
DesertodoAtacama	0	1	3	3	2	3	0	0	2	10
GroenlandiaSummitStation	1	0	1	2	3	2	6	1	5	7
MarsEnviroment	1	0	1	0	0	5	5	7	7	5
SãoJoséDosCordeiros	0	1	3	3	2	3	0	0	2	10
SãoPaulo	0	1	3	3	2	3	0	0	2	3
SiberiaGolomyanny	0	1	3	3	1	4	7	6	0	2
ÍndiaMawsynram	0	1	3	3	2	3	0	0	2	3

Depois da região dos Vales Secos de McMurdo, os dados com maior semelhança foram o da estação meteorológica Atacama-APEX, localizada na região leste do deserto, há 5100 metros de altitude, considerada uma das regiões mais secas do deserto do Atacama.

3.1.1. Agrupamento pelo Fuzzy C-means

Assim como no K-means, foram realizados experimentos com $2 \leq k \leq 11$. No entanto, no Fuzzy C-means para cada valor de k foram utilizados dois fatores de fuzzificação m diferentes, sendo 1.5 e 2.0. No caso do fator de fuzzificação $m=1.5$, o melhor coeficiente de silhueta fuzzy foi para $k=8$, com $FS=0.5234$. Ao aumentarmos o fator de fuzzificação para $m=2$, o melhor coeficiente de silhueta fuzzy foi para $k=6$, com $FS=0.4367$. A Figura 5 apresenta os gráficos para ambos os resultados.

Embora o coeficiente de silhueta fuzzy para $m=1.5$ e $k=8$ pareça melhor, é possível identificar no gráfico que o Cluster 3 possui apenas 249 amostras, quantidade muito inferior ao total de amostras de cada região, que é de 3168. Para $m=2$ e $k=6$, os clusters estão mais bem distribuídos e a explicação pelo coeficiente menor deve-se ao fato de termos mais pontos em cada cluster com FS abaixo de 0 (zero), o que indica que tais pontos estão mais próximo de outro cluster. Isso deve-se, principalmente, ao fator de fuzzificação m ser maior.

Na Tabela 2 são apresentados os graus de pertinência de cada local para $2 \leq k \leq 11$, com o fator de fuzzificação $m=2$. Cada célula contém o cluster de maior pertinência com seu respectivo grau de pertinência, ou seja, a estação meteorológica Atacama-APEX, com $k=5$, possui seu maior grau de pertinência no cluster 1, com grau de pertinência 0.414 (1:414).

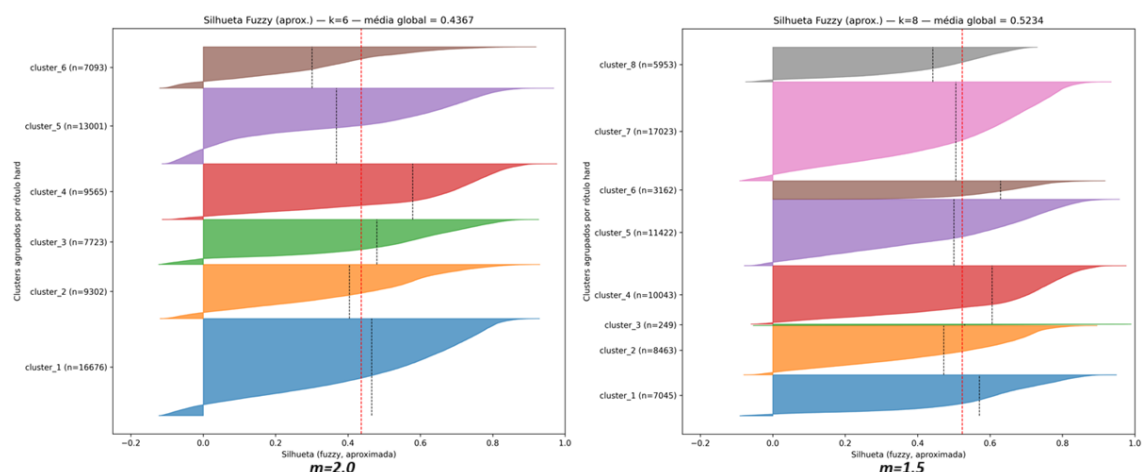


Figura 5. Comparação dos melhores coeficientes de silhueta fuzzy para $m=2.0$ e $m=1.5$

Tabela 2. Maior grau de pertinência nos clusters para $2 \leq k \leq 11$

	k=2	k=3	k=4	k=5	k=6	k=7	k=8	k=9	k=10	k=11
AntartidaFLMM	2:.826	2:.802	2:.756	2:.709	2:.719	6:.451	6:.452	6:.464	6:.465	6:.436
AlascaBRW	1:.566	3:.583	3:.448	3:.394	3:.693	3:.669	1:.619	3:.595	3:.566	3:.550
AntartidaCAAM	2:.724	2:.523	1:.661	5:.623	5:.604	5:.559	5:.541	8:.440	8:.401	8:.368
AntartidaFRSM	2:.885	2:.849	2:.497	2:.419	2:.448	6:.615	6:.610	6:.547	6:.529	6:.456
AntartidaHOEM	2:.685	2:.479	1:.635	5:.608	5:.582	5:.553	5:.531	8:.429	8:.384	8:.351
AntartidaVAAM	2:.743	2:.551	1:.702	5:.689	5:.652	5:.611	5:.588	8:.456	8:.393	8:.350
Atacama-APEX	2:.604	2:.418	2:.367	1:.414	6:.427	7:.330	8:.319	1:.286	10:.277	10:.232
AtacamaMeteo	1:.508	1:.597	4:.444	1:.567	6:.620	7:.659	8:.638	1:.625	10:.589	10:.581
DesertoArábia	1:.737	1:.708	4:.716	4:.709	4:.684	4:.667	4:.624	7:.608	4:.551	7:.535
DesertoAustralia	1:.759	1:.647	4:.638	4:.633	4:.596	4:.576	4:.520	7:.499	4:.427	7:.409
DesertoNaguev	1:.867	3:.506	3:.496	3:.454	1:.517	1:.504	7:.498	4:.476	7:.347	4:.331
DesertoSaara	1:.705	1:.771	4:.760	4:.705	4:.689	4:.661	4:.634	7:.611	4:.557	7:.541
DesertodeGobi	1:.505	1:.538	4:.405	1:.331	6:.262	7:.269	8:.226	5:.225	5:.189	5:.184
DesertodoAtacama	1:.886	3:.828	3:.836	3:.819	1:.752	1:.735	3:.480	9:.469	7:.419	4:.396
GroenlandiaSummitStation	2:.669	2:.571	2:.676	2:.631	2:.563	2:.719	2:.713	2:.703	2:.693	2:.672
MarsEnviroment	2:.687	2:.512	1:.399	5:.302	5:.279	6:.239	6:.218	8:.182	8:.170	5:.153
SãoJoseDosCordeiros	1:.894	3:.720	3:.743	3:.706	1:.788	1:.780	7:.575	4:.592	7:.615	4:.606
SaoPaulo	1:.834	3:.807	3:.793	3:.758	1:.657	1:.628	3:.564	9:.549	9:.473	9:.471
SiberiaGolomyanny	1:.555	3:.572	3:.431	3:.375	3:.730	3:.704	1:.664	3:.640	3:.610	3:.594
ÍndiaMawsynram	1:.736	3:.668	3:.592	3:.514	1:.413	1:.372	3:.368	9:.340	9:.329	9:.307

Com base nos resultados apresentados na tabela acima, é possível afirmar com mais precisão que os dados das estações meteorológicas Antártida CAAM, Antártida HOEM e Antártida VAAM possuem maior semelhança com os dados climáticos de Marte, uma vez que encontram-se nos mesmos clusters para $k=2$, $k=3$, $k=4$, $k=5$, $k=6$, $k=9$ e $k=10$. Ao analisarmos os graus de pertinência de cada uma destas estações meteorológicas e compararmos seus graus de pertinência com MarsEnviroment, dos três locais com maior semelhança indicados, a estação meteorológica Antártida HOEM é a mais semelhante à MarsEnviroment, pois possui o menor valor absoluto das diferenças dos graus de pertinência nos clusters para $k=2$, $k=3$, $k=4$, $k=5$, $k=6$, $k=9$ e $k=10$. Em outras palavras, os dados da estação meteorológica Antártida HOEM são mais

semelhantes, pois possuem a menor variação dentro dos clusters quando comparados com MarsEnvironment, uma vez que possuem graus de pertinência mais próximos.

4. Conclusões e Trabalhos Futuros

Os resultados obtidos demonstram que Marte apresenta características suficientemente distintas para formar um cluster isolado a medida de aumentamos k . Entretanto, com granularidades até próxima de 8 observou-se que estações localizadas nos Vales Secos de McMurdo na Antártida, tendem a se agrupar com Marte, indicando a existência de similaridade em termos de temperaturas muito baixas e umidade relativa do ar reduzida.

A análise pelo Fuzzy C-means permitiu identificar nuances de semelhança, demonstrando que a estação Antartida HOEM possui o menor desvio nos graus de pertinência em relação aos dados de Marte, sendo, portanto, o local mais semelhante ao ambiente de Marte dentre os avaliados. A estação Atacama-APEX também demonstrou similaridade com Marte em alguns cenários, corroborando estudos que apontam o Atacama como local com clima análogo ao de Marte, embora, em geral, com temperaturas predominantemente mais elevadas.

Por fim, a combinação de K-means e Fuzzy C-means aliada às métricas de avaliação de clusters, mostrou-se satisfatória para analisar e identificar a similaridade climática de Marte com diferentes regiões da Terra. Como trabalhos futuros, propõe-se ampliar o número de variáveis analisadas e obter dados mais precisos de umidade relativa e vento em Marte, além de incluir um conjunto maior de regiões terrestres.

Referências

- Azua-Bustos, A., González-Silva, C. e Fairén, A. G. (2022) "The Atacama Desert in Northern Chile as an Analog Model of Mars", *Front. Astron. Space Sci.* 8:810426. doi:10.3389/fspas.2021.810426.
- Bezdek, J. C., Ehrlich, R. e Full, W. "FCM: The fuzzy c-means clustering algorithm", *Computers & Geosciences*, vol. 10, Issues 2–3, 1984, p. 191-203, [https://doi.org/10.1016/0098-3004\(84\)90020-7](https://doi.org/10.1016/0098-3004(84)90020-7).
- Biswal, M. K. e Annavarapu, R. N. (2020) "Mars Missions Failure Report Assortment Review and Conspectus", *AIAA Propulsion and Energy 2020 Forum*, p. 1-15.
- Campello, R.J.G.B. e Hruschka, E.R. (2006) "A fuzzy extension of The silhouette width criterion for cluster analysis", *Fuzzy Sets and Systems*, vol. 157, p. 2858 – 2875.
- Fang, J., Luo, W., Howard, A.D., Craddock, R.A., Oliveira, E.A. e Pires, R.S. (2025) "Global comparative basin hypsometric analysis of Earth and Mars: Implications for early Mars climate", *Earth Planet. Sci. Lett.*, vol. 653 (2025), <https://doi.org/10.1016/j.epsl.2025.119226>.
- Fischer, E., Martínez, G. M., Rennó, N. O., Tamppari, L. K., e Zent, A. P. (2019) "Relative Humidity on Mars: New Results From the Phoenix TECP Sensor", *JGR Planets*, vol. 124, Issue 11, p. 2780-2792, <https://doi.org/10.1029/2019JE006080>
- Gaurav, A., Gupta, B. B. e Chuí, K. T. (2024) "A Hybrid Deep Learning Model for Intrusion Detection in Aerospace Vehicles", *2024 IEEE Space, Aerospace and Defence Conference*, p. 1244-1247, doi: 10.1109/SPACE63117.2024.10667826.

- Joskowicz, J. e Slomovitz, D. (2024) "Engineer's Perspectives on the Use of Generative Artificial Intelligence Tools in the Workplace", IEEE Engineering Management Review, vol. 52, no. 1, p. 258-267.
- Kumar, I. e Chidambara (2025) "A Comparative Evaluation of Freight Strategies for Online Business Models in Delhi: A Fuzzy AHP and TOPSIS Approach", AIIT 4th International Conference on Transport Infrastructure and Systems, p. 290 - 297.
- Lee, K. B. e Shin, H. S. (2019) "An Application of a Deep Learning Algorithm for Automatic Detection of Unexpected Accidents Under Bad CCTV Monitoring Conditions in Tunnels", 2019 International Conference on Deep Learning and Machine Learning in Emerging Applications (Deep-ML), p. 7 - 11.
- Li, Z., Wang, L. e Wu, X. (2025) "Artificial intelligence based virtual gaming experience for sports training and simulation of human motion trajectory capture", Entertainment Computing, vol. 52, <https://doi.org/10.1016/j.entcom.2024.100828>.
- MacQueen, J. B. (1967) "Some Methods for Classification and Analysis of Multivariate Observations", Proc. of The 5th Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, vol. 1, p. 281–297.
- Ramírez, N., Sánchez-Lozano, J.M. e Peña-Asensio, E. (2025) "Fuzzy Multi-Criteria Decision Framework for Asteroid Selection in Boulder Capture Missions", Aerospace 2025, 12, 800. <https://doi.org/10.3390/aerospace12090800>
- Rojas, J. (2022) "Weather on mars", disponível em: <https://www.kaggle.com/code/joaquinrojas/weather-on-mars>. Acesso realizado em 29/11/2025.
- Rousseeuw, P. J. (1987) "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis", Journal of Computational and Applied Mathematics, vol. 20, p. 53-65.
- Scharf, D. P., Hadaegh, F. Y. e Ploen, S. R. (2004) "A survey of spacecraft formation flying guidance and control (Part II): control", Proceedings of the 2004 American Control Conference, p. 2976 – 2985, Boston, MA, USA.
- Shaw, L. e Gohel, H. A. (2025) "Role of artificial intelligence in health monitoring using IoT based wearable sensors: A survey", Internet of Things, vol. 34, <https://doi.org/10.1016/j.iot.2025.101761>.
- Thorndike, R. L. (1953) "Who Belongs in the Family?", Psychometrika, vol. 18, Issue 4, p. 267 - 276, doi: <https://doi.org/10.1007/BF02289263>.
- Zadeh, L. A. (1965) "Fuzzy Sets. Information and Control", vol. 8, Issue 3, p. 338-353, [https://doi.org/10.1016/S0019-9958\(65\)90241-X](https://doi.org/10.1016/S0019-9958(65)90241-X).
- Zent, A. P., Michael, H. H., Cobos, D. R., Wood, S. E., Hudson, T. L., Milkovich, S. M., DeFlores, L. P. e Mellon, M. T. (2010) "Initial results from the thermal and electrical conductivity probe (TECP) on Phoenix", JGR Planets, vol. 115, Issue E3. <https://doi.org/10.1029/2009JE003420>.
- Zent, A. P., Hecht, M. H., Hudson, T. L. , Wood, S. E. e Chevrier, V. F. (2016) "A revised calibration function and results for the Phoenix mission TECP relative humidity sensor", PGR Planets, vol. 121, Issue4, p. 626-651, <https://doi.org/10.1002/2015JE004933>.