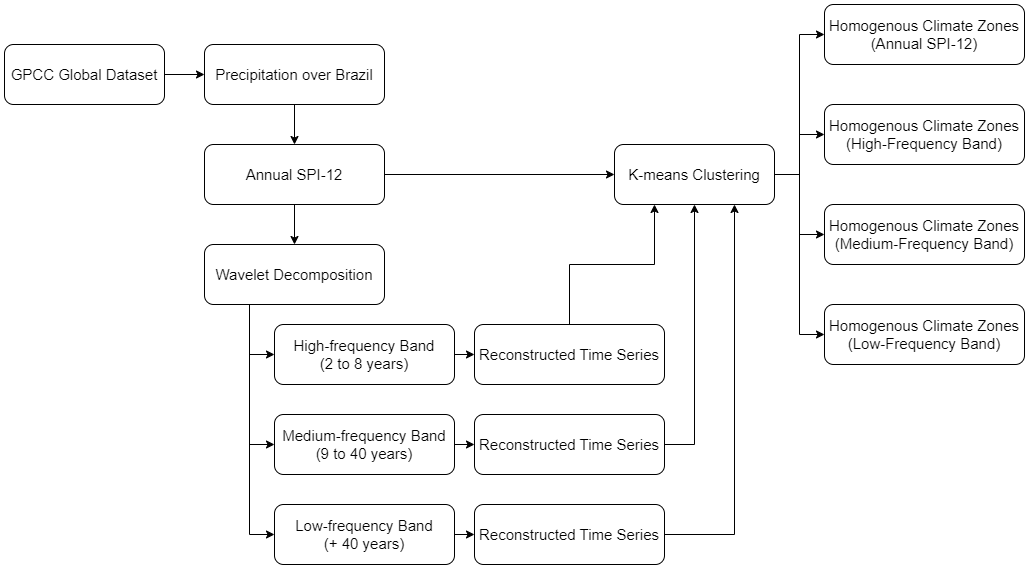
# introdução

# metodologia

A metodologia utilizada para determinação das zonas homogêneas de risco climático no Brasil considera:i) técnicas de padronização das séries temporais de precipitação, através do *Standard Precipitation Index* (SPI)*,* ii) Transformada em Ondeletas (T.O) para decomposição das bandas de alta, média e baixa frequência que compõem a série temporal original, iii) reconstrução dessas bandas de frequência no domínio do tempo e iv) a determinação dos agrupamentos para as séries de precipitação padronizada e para cada uma das séries temporais reconstruídas utilizando k-means. A Figura 1 apresenta um fluxograma esquemático que sintetiza os procedimentos supracitados.

Figure 1 –



## Área de Estudo

Falar um pouco sobre a variabilidade climática do Brasil. A área de estudo do presente trabalho é apresentada na Figura 1.

## Dados Utilizados

As informações de precipitação ao longo de todo o território brasileiro foram obtidas da versão 2022 do *grid* global de precipitações mensais sob a superfície terrestre fornecido pelo *Global Precipitation Climatology Center* (GPCC). Nessa versão, são disponibilizadas séries históricas de precipitação que abrangem o período entre os anos de 1891 e 2020 para diferentes resoluções de grade: 0.25°, 0.5°, 1.0° e 2.5°. Adotou-se uma resolução de grade de 1° no presente trabalho. (Schneider *et al.,* 2022).

Na versão de 2022, o grid de precipitações sobre a superfície terrestre é estimado com base em, aproximadamente, 86.000 estações nas mais diversas regiões do planeta, as quais apresentam pelo menos 10 anos de registros (Schneider *et al.,* 2022).

Para períodos mais antigos, onde não há registro de precipitação em algumas regiões, o GPCC atribui a precipitação climatológica da região em caso de ausência de dados, garantindo que não haja anomalias entre esse período preenchido e o restante da série temporal (Schneider *et al.*, 2018). Dessa forma, para evitar os períodos preenchidos com a climatologia da precipitação, optou-se por adotar as séries de precipitação a partir de 1951 até 2020.

A principal motivação para adoção dos dados fornecidos pelo GPCC para determinação da precipitação ao longo do território brasileiro reside na uniformidade, tanto espacial quanto temporal, da grade em que esses dados são disponibilizados. Além disso, vale destacar que as estações pluviométricas disponibilizadas no Hidroweb da Agência Nacional de Águas (ANA), em geral, apresentam descontinuidades no período observado e diferentes durações e períodos de observação.

De maneira adicional às justificativas supracitadas para utilização das informações de precipitação fornecidas pelo GPCC, vale ressaltar os estudos que avaliaram a performance dessas informações na representação da variabilidade climática, na estimativa de extremos climáticos, na modelagem hidrológica, dentre outras aplicações, tais como: Song *et al.* (2022), Tostes *et al.* (2017), Gomes *et al.* (2022), Kouakou *et al.* (2023), Dhungana *et al.* (2023), Santos *et al.* (2023). Em geral, esses estudos classificam essa informação como, no mínimo, satisfatória.

## Standard Precipitation Index

O índice *Standard Precipitation Index* (SPI) foi proposto inicialmente por McKee *et al.* (1993), objetivando uma ferramenta versátil para a análise de secas meteorológicas, visto que utiliza apenas dados de precipitação para sua determinação.

Como apresentado por McKee *et al.* (1993), esse índice pode ser determinado para uma variedade de escalas temporais na qual a precipitação mensal é agregada. Isso se apresenta como um ponto positivo, pois permite uma flexibilidade na avaliação dos efeitos dos déficits de precipitação em diferentes escalas nos recursos hídricos, por exemplo, escalas temporais curtas refletem impactos na umidade do solo enquanto a longas refletem impactos na disponibilidade hídrica superficial.

O procedimento para determinação do SPI consiste na aproximação da série temporal de precipitação agregada em uma escala de n-meses a uma distribuição de probabilidade acumulada, geralmente do tipo Gamma, possibilitando a determinação da probabilidade de não excedência das precipitações agregadas que compõem a série temporal supracitada. Essas probabilidades, então, dão origem ao SPI através da função inversa da distribuição normal padrão (Equação 1) (ROCHA *et al.*, 2019; GUENANG e KAMGA, 2014; MCKEE *et al.,* 1993).

Em que: é a precipitação agregada na escala adotada para o SPI; é a Função de Distribuição Acumulada (FDA) do tipo Gamma, que retornar à probabilidade de não excedência do evento ; é a função inversa da distribuição normal padrão , a qual possui média igual a zero e desvio padrão unitário.

Por definição, o SPI representa a quantidade de desvios padrões que um evento específico de precipitação mensal agregada em uma janela de n-meses possui em relação à média de uma distribuição normal padronizada oriunda da série temporal de longo prazo de precipitações mensais agregadas nessa mesma janela. De acordo com o valor do SPI, pode-se classificar o evento anômalo de precipitação em 7 categorias, como apresentado na Tabela 1.

Tabela 1 – Classificação dos valores do SPI e os limites de probabilidade correspondentes (Adaptado de BAZRAFSHAN et al., 2014).

|  |  |  |
| --- | --- | --- |
| Valor do SPI | Categoria | Limites de probabilidade |
|  | Extremamente úmido |  |
|  | Severamente úmido |  |
|  | Moderadamente úmido |  |
|  | Normal |  |
|  | Moderadamente seco |  |
|  | Severamente seco |  |
|  | Extremamente seco |  |

No presente trabalho, adotou-se o SPI para uma escala de 12 meses (SPI-12), calculando-o para cada um dos pontos da grade do GPCC contidos no território brasileiro.

Os procedimentos supracitados para determinação do SPI retornam uma série mensal contínua desse índice, a qual apresenta uma dependência entre seus valores. Como apontado por Estácio *et al.* (2021), as precipitações mensais agregadas para determinação do SPI-12 para dois meses consecutivos se diferenciam apenas em 1 dos 12 valores agregados. Dessa forma, objetivando a remoção dessa dependência, optou-se por utilizar apenas o último mês de cada um dos anos que compunham a série temporal mensal do SPI-12, dando origem a uma série anual de SPI-12.

## Transformada em Ondeletas

A Transformada em Ondeletas (TO) foi utilizada para analisar os modos de variabilidade das séries anuais de SPI-12 para cada um dos pontos de grid do GPCC contidos no território brasileiro.

A TO permite a análise dos padrões de multifrequências que compõem a série temporal original, sendo uma metodologia bastante utilizada para determinação dos modos de variabilidade dominantes e da variação desses modos ao longo de uma série temporal não estacionária. A decomposição do sinal original no domínio do tempo-frequência é feita através da translação e dilatação de uma função geradora base (ondeleta mãe) que gera um conjunto de funções (ondeletas filhas) de frequência variada, devido a dilatação, e de posicionamento variado ao longo do sinal, devido a translação (Rocha e Souza Filho, 2020; Lima *et al.* 2021; Torrence e Compo, 1998).

Como apontado por Alves *et al.* (2013), a função geradora, ou ondeleta mãe, pode ser do tipo discreta ou contínua. Para cada um desses tipos de ondeleta, há uma variedade de funções base, cada qual com suas especificidades e mais adequadas para determinadas aplicações.

No presente trabalho, adotou-se a ondeleta contínua de Morlet como função base geradora das ondeletas filhas. A escolha da ondeleta de Morlet se baseia em sua recorrente aplicação na análise de variabilidade de séries hidrometeorológicas, tais como: Rocha e Souza Filho (2020), Lima *et al.* (2021), Alves *et al.* (2013) e Kömüşcü e Aksoy (2023). Além disso, Torrence e Compo (1998) recomendam a sua utilização para análise de variáveis geofísicas. A ondeleta contínua de Morlet é dada pela Equação 2.

DESCREVER VARIÁVEIS.

Após a decomposição do sinal original em um conjunto de ondeletas filhas originadas da ondeleta de Morlet, em outras palavras, a obtenção dos coeficientes das ondeletas no domínio da frequência, filtrou-se o sinal decomposto em duas diferentes bandas de frequência: Alta Frequência (2 a 8 anos) e Média Frequência (9 até 40 anos), para uma posterior reconstrução dessas bandas no domínio do tempo.

No processo de reconstrução sinal, considerou-se uma terceira banda, a de Baixa Frequência (+40 anos), interpretada como resíduo. Como apontado por Lima *et al.* (2021), tendo em vista a ortogonalidade das ondeletas e de suas bandas, logo uma correlação insignificante, pode-se determinar o sinal da banda de baixa frequência reconstruído no domínio do tempo através da Equação 3.

Em que: , respectivamente, o i-ésimo valor reconstruído no domínio do tempo das bandas de alta, média e baixa frequência; o i-ésimo valor do sinal original, no caso, da série temporal anual do SPI-12 de determinado ponto de grade selecionado.

A escolha dessas bandas de frequência se baseou nos trabalhos de Lima *et al.* (2021), Alves *et al.* (2013). Uma justificativa para a escolha dessas bandas pode ser dada à luz dos mecanismos de variabilidade climática. Nesse contexto, pode-se associar os mecanismos de variabilidade interanual à banda de alta frequência, os de variabilidade decadal à banda de média frequência e os de variabilidade multidecadal às bandas de média e baixa frequência. Ademais, a banda de baixa frequência também pode ser associada às tendências de crescimento ou decrescimento da série temporal, relacionadas, por exemplo, aos impactos das mudanças climáticas.

## Zonas Homogêneas de risco Climático

As zonas homogêneas de risco climático foram determinadas com a utilização da técnica de K-means para determinação dos agrupamentos homogêneos da série anual de SPI-12 e das séries reconstruídas da banda de alta, média e baixa frequência obtidas com a aplicação da TO nessa série.

O K-means é uma técnica de clusterização amplamente utilizada nas mais diversas áreas pela sua simplicidade, rápida implementação, capacidade de operar grandes bases de dados, dentre outros benefícios (Ghasempour *et al.* 2022).

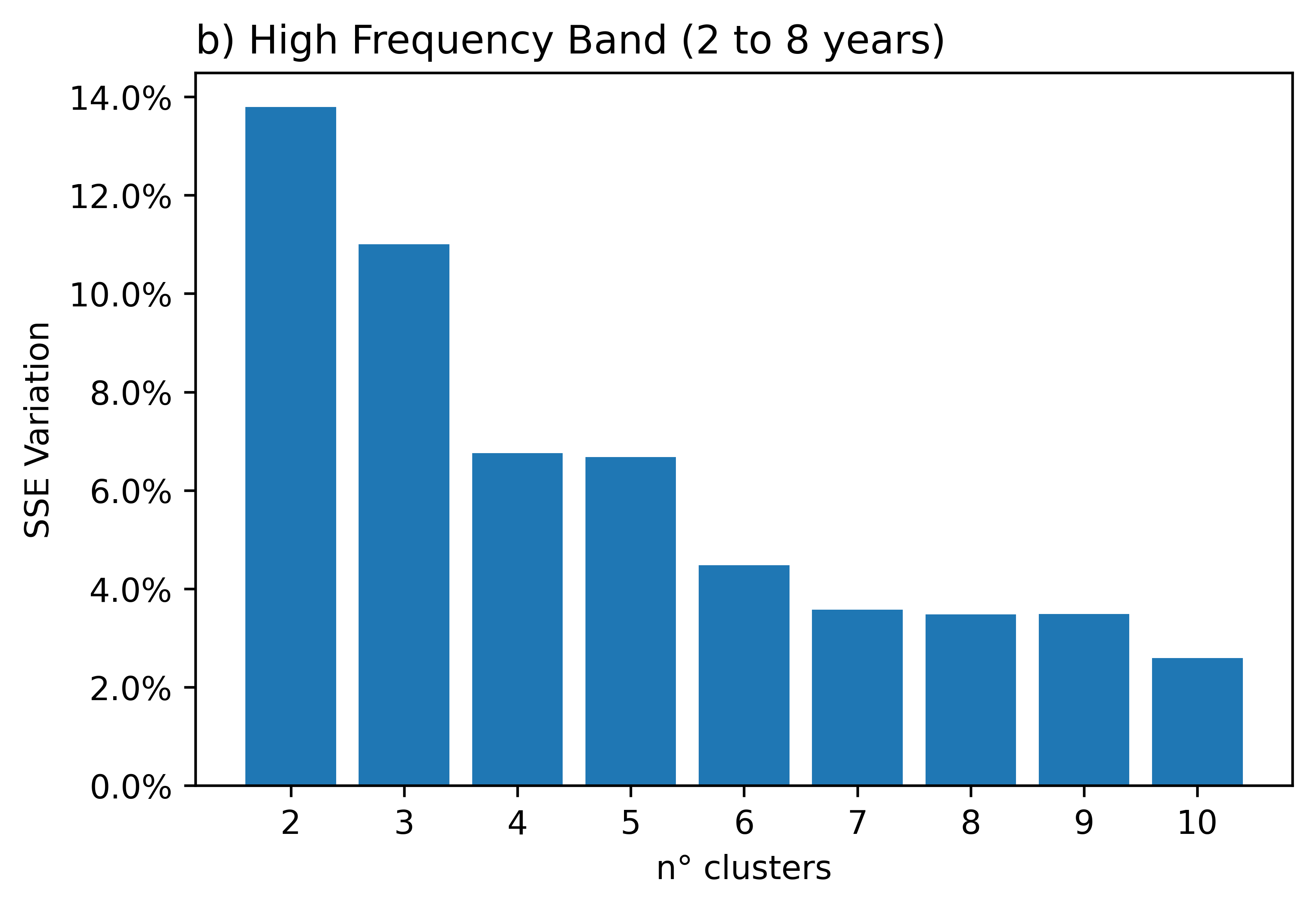
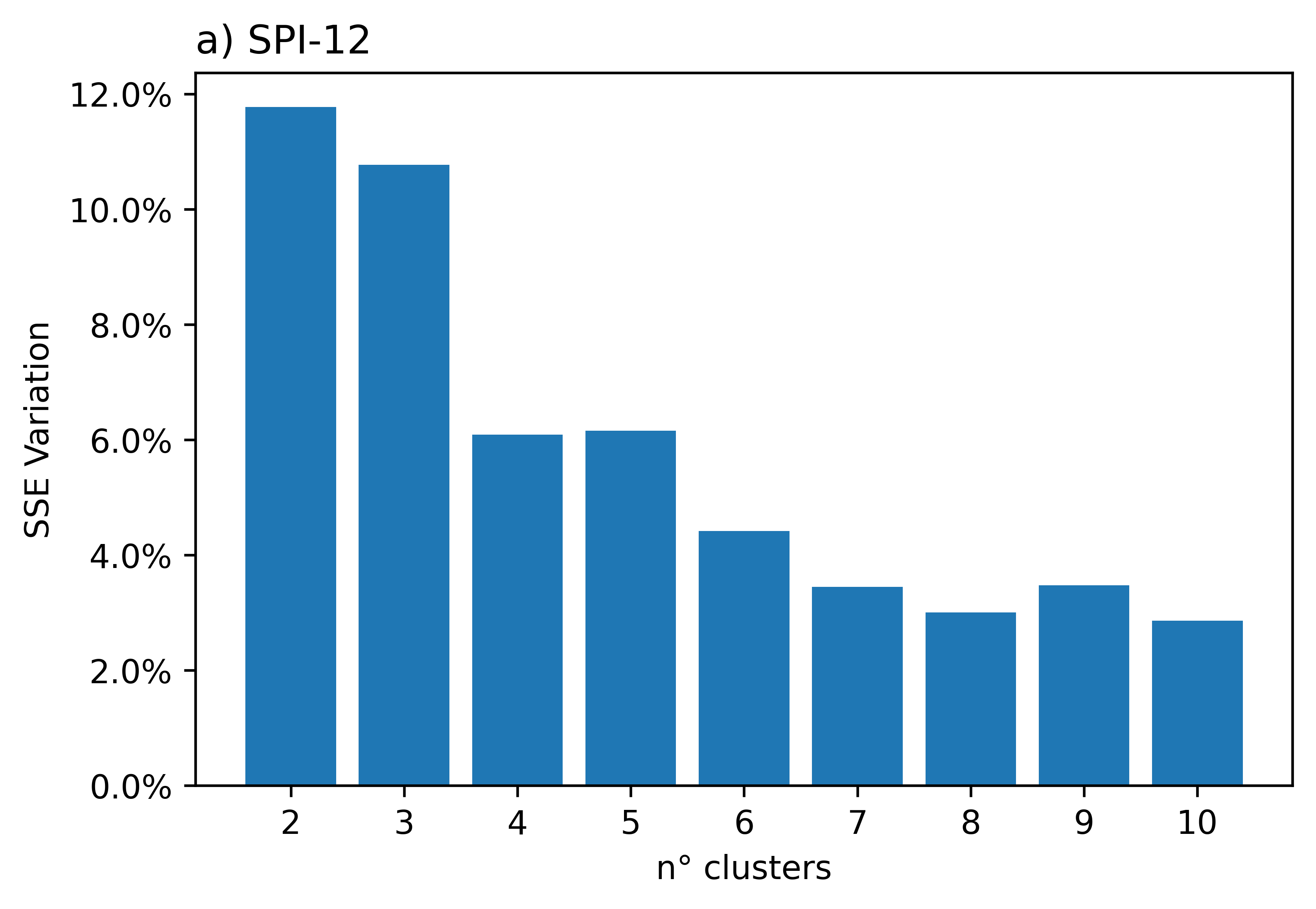
Roushangar e Ghasempour (2022) destacam que o k-means pode ser sintetizado em apenas duas etapas alternantes: i) etapa de assimilação e ii) etapa de atualização. Na etapa de assimilação, cada observação é atribuída ao cluster cujo centroide tem menor distância euclidiana ao quadrado desse ponto. Na etapa de atualização, recalcula-se o centroide com base nos pontos assimilados por cada cluster e repete-se o processo de assimilação supracitada. O loop entre esses dois processos continua até que não seja observada grandes mudanças na atualização dos centroides.

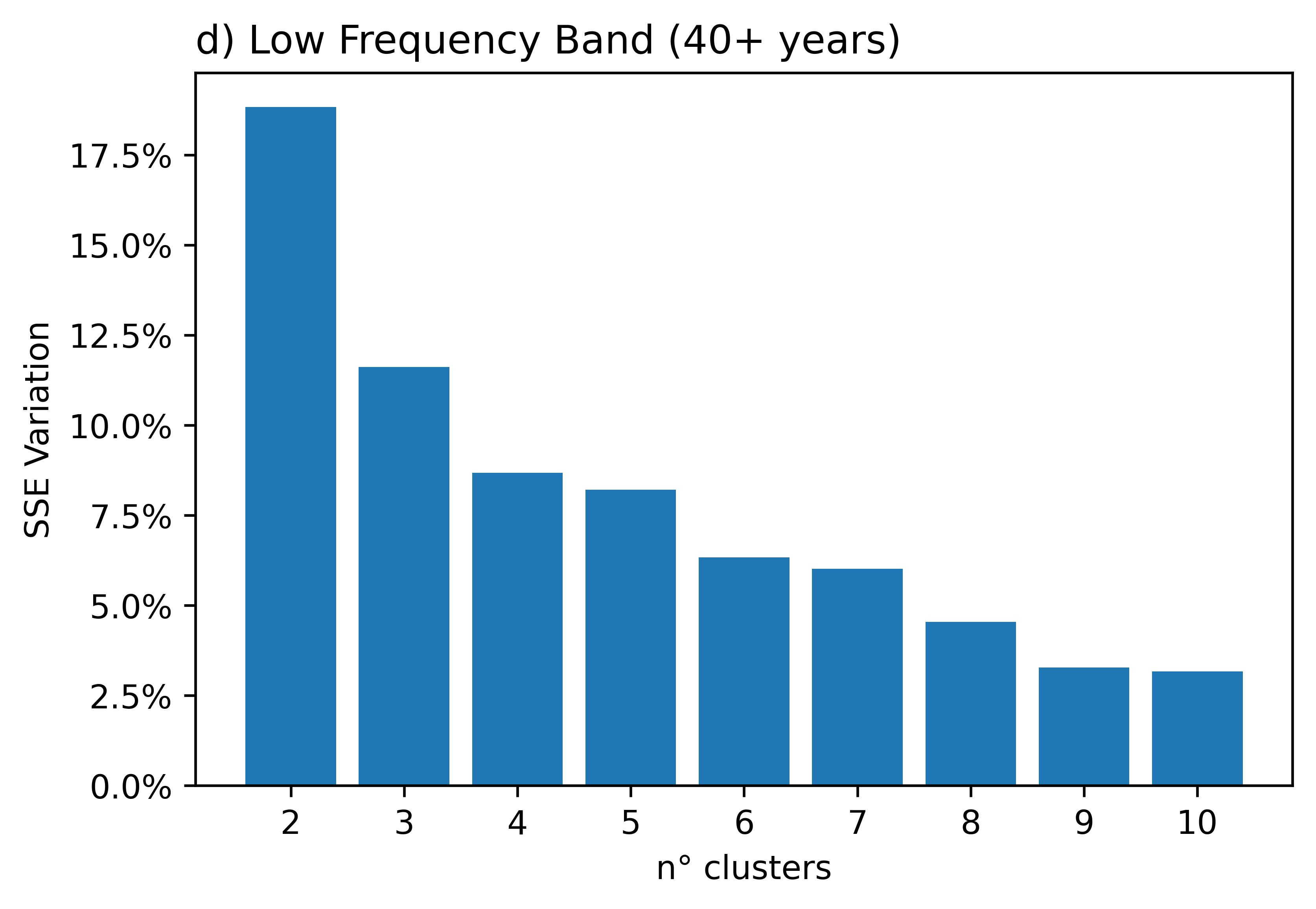
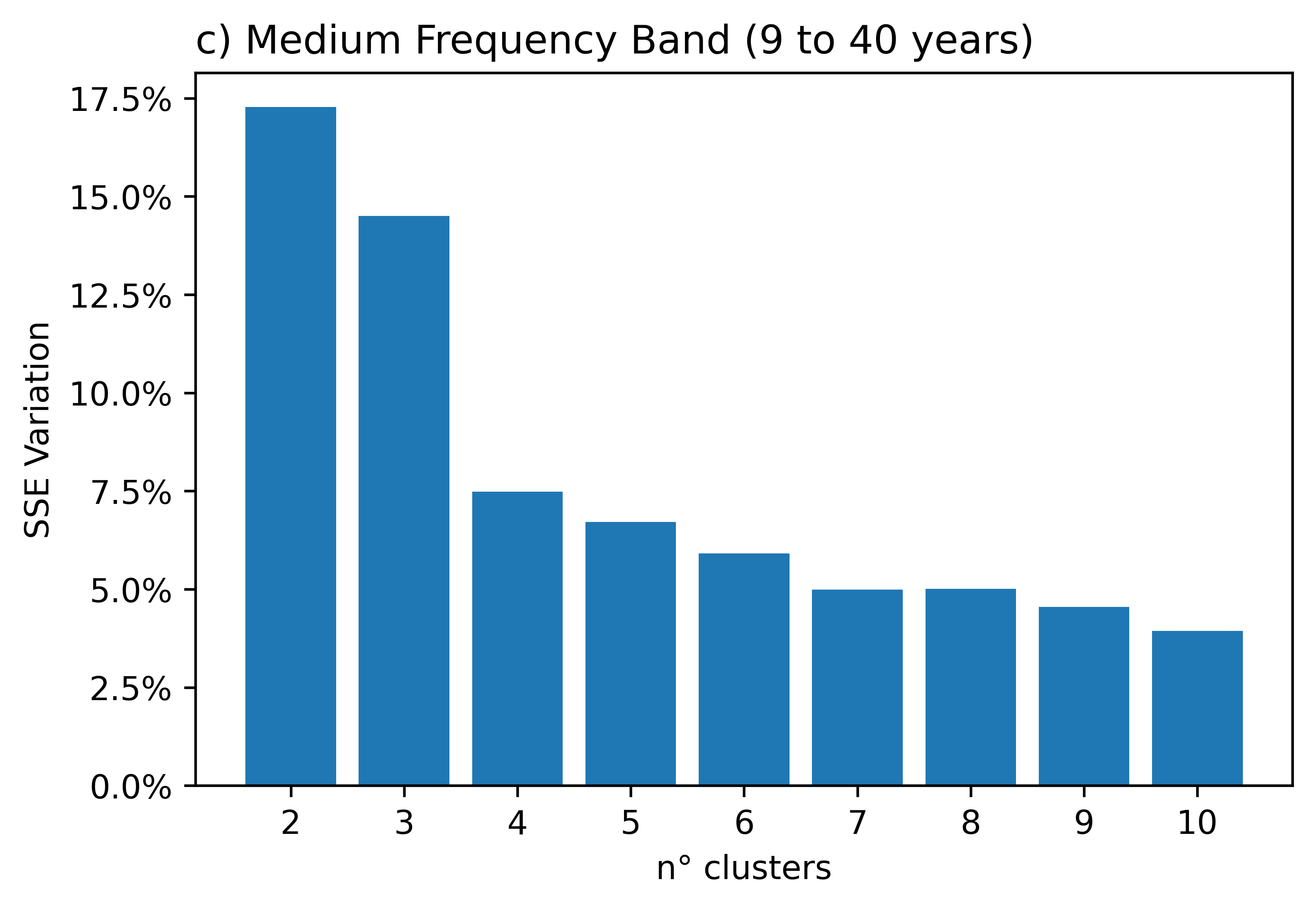
O K-Means otimiza uma função objetivo baseada na distância euclidiana, portanto é bastante sensível à magnitude dos valores. Dessa forma, é comum que se proceda uma padronização da base de dados de maneira prévia à aplicação desse método. No caso da presente análise, o procedimento de padronização não foi necessário, pois o SPI já é um índice padronizado.

O número de clusters para cada uma das séries temporais consideradas foi definido com base na soma das distâncias das amostras aos seus respectivos centroides aos quais elas estão associadas . O foi determinado para variando de 1 até 10, calculando-se a variação percentual entre dois valores consecutivos, como expresso na Equação 4:

onde: é a variação percentual do entre um número de clusters e ; e valores de para um número de cluster e , respectivamente. A Figura X apresenta essas variações percentuais para as séries de SPI-12 anual e para as séries reconstruídas das bandas de alta, média e baixa frequências.

Figura 1





Essa estratégia para definição do número de clusters visa identificar um ponto de aparente convergência do , onde dois valores de para consecutivos não apresentam grandes variações percentuais. Dessa forma, a partir desse ponto, entende-se como ineficiente o aumento de para redução do somatório das distâncias das amostras aos seus respectivos centroides, adotando‑o como potencial candidato para o número de clusters a ser adotado.

Vale destacar que quando for igual ao número de amostras do conjunto de dados, , pois cada amostra será seu respectivo centroide. Dessa forma, sempre terá um valor decrescente à medida que tende ao tamanho da amostra. Essa estratégia, portanto, não visa minimizar , mas sim encontrar o ponto a partir do qual o incremento do representará pouca variação do seu valor.

Para a série anual de SPI-12 (Figura Xa), por exemplo, observa-se uma considerável redução da variação percentual para , ou seja, de acordo com a Equação 4, uma diminuição da variação percentual do para em relação ao para , indicando uma convergência dos valores para . Assim, é um potencial candidato para o número de clusters adotados para essa série.

Ainda para a série anual de SPI-12, constata-se que a convergência aumenta com as sucessivas reduções das variações percentuais até , atingindo um patamar de variação percentual quase constante para e . Portanto, também é um potencial candidato para número de clusters a ser adotado para essa série. Observa-se que apesar do procedimento supracitado elencar potenciais candidatos à número de clusters adotados, ainda há incertezas quanto ao limiar de variação percentual que deve ser tomado para uma escolha definitiva.

Para definição do número de clusters adotadas na clusterização das séries temporais consideradas no presente trabalho, além do procedimento supracitado, aferiu-se as regiões formadas por cada cluster, buscando sempre otimizar sua homogeneidade espacial ao longo do território brasileiro.

Adotou-se para a série de SPI-12 anual um número de clusters igual a 7 . Para as séries reconstruídas das bandas de alta, média e baixa frequência, adotou-se, respectivamente, , e .

# resultados

# Discussões

# conclusões

# referências