

Clusterização Temporal da Ingestão de Microplásticos: Estabilidade e Transição entre Grupos Regionais

Gagini P. A. M.¹, Salmaze P. H.¹, Joioso A. V. B.¹

¹Instituto De Ciências Matemáticas e de Computação – Universidade de São Paulo (USP)
São Carlos – SP – Brazil

Resumo. Diante da crescente preocupação com a ingestão humana de microplásticos, este estudo investigou padrões globais de exposição alimentar a esses contaminantes entre 1990 e 2018, com base em dados organizados por país, categoria alimentar e região geopolítica. A proposta consistiu em aplicar algoritmos de clusterização (Hierarchical, K-Means e DBSCAN) para identificar agrupamentos de países com perfis semelhantes de ingestão, analisar sua estabilidade temporal e revelar desigualdades regionais. A metodologia incluiu pré-processamento e normalização dos dados, otimização dos parâmetros de cada algoritmo e avaliação dos agrupamentos por métricas como Silhouette Score e Índice de Davies–Bouldin. Os resultados revelaram um crescimento global de 12,7% na ingestão de microplásticos, com regiões desenvolvidas apresentando níveis sistematicamente mais altos. A clusterização expôs dois grandes grupos estáveis ao longo do tempo: países desenvolvidos com altos níveis de ingestão e países em desenvolvimento com níveis mais baixos, porém crescentes. Conclui-se que a exposição alimentar a microplásticos segue padrões estruturais relacionados ao nível de desenvolvimento e organização dos sistemas alimentares, sendo a análise temporal por clusterização uma ferramenta robusta para monitoramento global e formulação de políticas públicas preventivas.

1. Objetivos

A crescente presença de microplásticos no meio ambiente tem despertado preocupações significativas no âmbito da saúde pública, segurança alimentar e sustentabilidade ambiental. Estudos recentes demonstram que fragmentos de plásticos com dimensões inferiores a 5 mm — oriundos da degradação de resíduos plásticos, de embalagens e de processos industriais — têm sido detectados em diversos alimentos de origem vegetal e animal. Como consequência, a ingestão humana desses contaminantes tornou-se praticamente inevitável, especialmente considerando cadeias alimentares globalizadas e altamente industrializadas.

Embora os efeitos toxicológicos da ingestão crônica de microplásticos ainda estejam sendo investigados, evidências preliminares apontam para potenciais riscos inflamatórios, imunológicos e metabólicos, além da capacidade desses contaminantes agirem como vetores de substâncias químicas nocivas. Frente a esse cenário, compreender padrões de exposição alimentar a microplásticos em diferentes contextos geográficos e ao longo do tempo torna-se uma necessidade urgente para subsidiar políticas públicas, orientar a regulação alimentar e fomentar mudanças nos sistemas de produção e consumo.

Neste contexto, este estudo se propõe a explorar um conjunto de dados internacionais sobre ingestão estimada de microplásticos provenientes de alimentos, organizados

por país, categoria alimentar e ao longo de quase três décadas (1990 a 2018). É esperado revelar agrupamentos latentes entre países com padrões similares de contaminação alimentar, identificar tendências temporais e evidenciar contrastes regionais.

1.1. Proposta Analítica

A proposta é aplicar algoritmos de clusterização (como modelos hierárquicos, K-Means, DBSCAN), em diferentes janelas temporais, de forma a:

- Identificar agrupamentos coesos de países com padrões semelhantes de ingestão de microplásticos, considerando múltiplas fontes alimentares;
- Acompanhar a evolução dos *clusters* ao longo do tempo, observando persistência, transições e mudanças estruturais nas exposições;
- Analisar a variabilidade geográfica, observando se determinadas regiões tendem a se concentrar em *clusters* específicos.

1.2. Perguntas de Pesquisa

Para orientar a investigação, formulam-se as seguintes perguntas principais:

- **A ingestão de microplásticos tem aumentado globalmente ao longo do tempo?**
Verifica-se se há uma tendência ascendente na contaminação alimentar por microplásticos.
- **Há diferenças significativas entre regiões quanto à tendência de ingestão?**
Explora-se se determinadas regiões geográficas (por exemplo, África Subsaariana, América Latina e Caribe, Sudeste Asiático e Leste Asiático) apresentam trajetórias distintas, refletindo desigualdades estruturais no sistema alimentar ou diferenças em políticas ambientais.
- **Como os *clusters* regionais variam ao longo do tempo?**
Analisa-se a estabilidade e transformação dos *clusters* ao longo das décadas, com foco na rotatividade dos países entre grupos, emergências de novos perfis alimentares contaminados e convergência/divergência regional.

2. Fundamentação Teórica

Diferente de métodos supervisionados que requerem uma variável-alvo conhecida, a clusterização é apropriada quando se busca reconhecer estruturas ocultas nos dados, especialmente em contextos exploratórios como o presente, onde não há uma classificação prévia de países por grau de contaminação. Dentre os diversos métodos disponíveis na literatura, este estudo faz uso de três abordagens principais (Aglomerativo, K-Means e DBSCAN). Essa triangulação metodológica confere maior validade e confiabilidade à análise exploratória realizada.

2.1. Hierarchical Clustering (Aglomerativo)

O Aglomerativo é um método que constrói uma hierarquia de agrupamentos por meio de fusões sucessivas entre os pares de clusters mais próximos, com base em uma métrica de distância (como single linkage, complete linkage, average linkage ou o método de Ward). Esse algoritmo gera uma estrutura em forma de dendrograma, o que permite uma análise visual e interpretativa dos níveis de similaridade entre os dados. Uma vantagem fundamental da clusterização hierárquica é a ausência da necessidade de definir previamente o número de clusters. Entretanto, seu custo computacional é mais elevado em grandes conjuntos de dados, e decisões iniciais tomadas durante a fusão dos clusters não podem ser revertidas. Veja mais em [1].

2.2. K-Means

Por outro lado, o K-Means é um algoritmo particional que busca dividir os dados em k grupos mutuamente exclusivos, minimizando a soma das distâncias quadráticas entre os pontos e os centróides dos clusters. A sua implementação parte da escolha inicial de k centróides, seguida por etapas iterativas de atribuição de pontos ao centróide mais próximo e de recalculação dos centróides com base nas novas atribuições. A simplicidade e a eficiência computacional tornam o K-Means especialmente adequado para conjuntos de dados grandes e bem distribuídos. Contudo, o algoritmo assume que os clusters são aproximadamente esféricos e de tamanhos similares, e sua performance depende fortemente da escolha apropriada do número de clusters, que pode ser guiada por métricas como o método do cotovelo (Elbow Method) ou o índice de Silhueta. Veja mais em [1].

2.3. DBSCAN (Density-Based Spatial Clustering of Applications with Noise)

Complementando essas abordagens, foi utilizado o DBSCAN, baseado em densidade, identificando clusters como regiões densamente povoadas separadas por regiões de baixa densidade. O algoritmo define um cluster como o conjunto de pontos que estão dentro de uma vizinhança de raio ϵ e que contêm pelo menos um número mínimo de pontos (minPts). Ele distingue entre pontos centrais (densamente conectados), pontos de borda (vizinhos de centrais) e ruídos (outliers). A principal vantagem do DBSCAN está na sua capacidade de descobrir agrupamentos com formas arbitrárias e de tratar ruídos de forma natural, sem necessidade de especificar o número de clusters a priori. No entanto, ele é sensível à definição dos parâmetros e pode apresentar dificuldades com dados de densidade variável. Veja mais em [1].

3. Metodologia

A presente análise foi estruturada em três etapas principais: pré-processamento dos dados, escolha e otimização dos parâmetros dos algoritmos de clusterização e avaliação da qualidade dos agrupamentos. Cada etapa foi desenhada de modo a garantir coerência metodológica, reprodutibilidade e validade estatística na identificação de padrões regionais de exposição alimentar a microplásticos. O código implementado está disponível em [2].

3.1. Pré-processamento dos dados

Inicialmente, os dados foram organizados de modo a favorecer uma análise mais interpretável e generalizável. Para reduzir a granularidade da análise e possibilitar

comparações geográficas mais amplas, os 109 países presentes no conjunto de dados foram segmentados em 9 grandes regiões geopolíticas. As regiões adotadas foram:

1. África Subsaariana;
2. América Latina e Caribe;
3. Europa Ocidental;
4. Sudeste Asiático e Leste Asiático;
5. Ásia Ocidental e Sul;
6. Europa Oriental e Norte;
7. Norte da África;
8. América do Norte;
9. Oceania.

Após a categorização geográfica, os dados foram agrupados por região e ano, com a ingestão média de microplásticos sendo calculada para cada categoria alimentar dentro de cada combinação “região-ano”. Essa agregação regional por média permite observar tendências macroestruturais ao longo do tempo, evitando vieses decorrentes de variações internas entre países da mesma região.

Com os dados organizados, foi aplicada a normalização Z-score (via StandardScaler), garantindo que todas as variáveis — independentemente de suas magnitudes — contribuíssem igualmente para as métricas de distância utilizadas pelos algoritmos de clusterização.

3.2. Escolha dos parâmetros e otimização dos hiperparâmetros

Para garantir a qualidade e a validade estatística dos agrupamentos gerados, foram utilizadas estratégias específicas de otimização de parâmetros para cada algoritmo de clusterização empregado.

- Clusterização Hierárquica (Aglomerativa): Foi utilizada a estratégia de ligação do tipo ward, que minimiza a variância intra-grupo a cada etapa de fusão. O número ideal de clusters foi escolhido por três métricas (Silhouette Score, Índice de Davies–Bouldin e Índice de Calinski–Harabasz)
- K-Means: Para este algoritmo, o número ótimo de clusters foi igualmente definido com base nas três métricas mencionadas. A escolha do valor de k buscou maximizar a coesão intra-cluster e a separação inter-cluster.
- DBSCAN: Por ser um método baseado em densidade, a definição dos parâmetros ε e minPts foi orientada pelo gráfico de distância dos k-vizinhos (k-distance plot), utilizado para estimar o ponto de inflexão (joelho) na curva de distâncias — identificado com a ferramenta KneeLocator.

3.3. Avaliação dos agrupamentos

A qualidade dos agrupamentos foi avaliada com base em métricas quantitativas padronizadas e representações gráficas, de forma a permitir a comparação entre algoritmos e entre períodos de tempo.

As métricas de avaliação aplicadas foram:

- Silhouette Score: indicativo da qualidade de separação entre clusters;
- Índice de Davies–Bouldin: penaliza clusters sobrepostos ou mal definidos;
- Índice de Calinski–Harabasz: favorece agrupamentos com maior variabilidade entre clusters e menor variabilidade interna.

Para o DBSCAN, foram visualizados com projeções em duas dimensões via PCA, permitindo avaliar graficamente a coerência espacial dos clusters e facilitando a identificação de padrões e sobreposições.

Resultados de cada métrica foram organizados em tabelas e gráficos, comparando os algoritmos aplicados em cada ano, com destaque para a evolução temporal das estruturas de agrupamento regionais.

4. Análise de Resultados

A análise dos resultados obtidos a partir da aplicação dos algoritmos de clusterização permitiu identificar padrões latentes na ingestão alimentar de microplásticos entre países, considerando sua evolução temporal e segmentação por região geopolítica.

4.1. Análise direta dos dados

Para responder a duas primeiras perguntas de pesquisa não precisamos de algoritmos rebuscados e complexos, uma análise direta dos dados já é o suficiente.

4.1.1. A ingestão de microplásticos tem aumentado globalmente ao longo do tempo?

A Figura 1 mostra uma tendência geral de aumento no consumo (em $\mu\text{g/kg}$) de microplásticos entre 1990 e 2018, passando de aproximadamente 1493,6 para 1683,3, o que representa um crescimento de cerca de 12,7% ao longo do período. Embora tenha ocorrido uma leve queda entre 2005 e 2010, esse recuo foi temporário, com os valores voltando a subir posteriormente e alcançando o pico em 2018. Essa variação indica uma tendência contínua de crescimento, possivelmente refletindo mudanças nos fatores que influenciam essa métrica ao longo do tempo.

4.1.2. Há diferenças entre regiões quanto à tendência de ingestão?

A Figura 2, por sua vez, mostra a ingestão média anual de microplásticos (em $\mu\text{g/kg}$) por diferentes regiões do mundo entre 1990 e 2018. Observa-se que América do Norte, Europa Ocidental e Europa Oriental e Norte apresentam os maiores valores, com picos próximos a 2500 $\mu\text{g/kg}$, embora tenham uma leve queda após 2005. Oceania também mantém níveis elevados, mas com variações ao longo dos anos. Regiões como Norte da África e América Latina e Caribe apresentam valores intermediários, com tendência geral de crescimento. Já Sudeste Asiático e Leste Asiático, África Subsaariana e Ásia Ocidental e Sul apresentam os níveis mais baixos, embora o Sudeste Asiático e Leste Asiático mostre um aumento contínuo ao longo do período. Isso sugere disparidades regionais significativas na exposição a microplásticos, possivelmente refletindo diferenças em hábitos alimentares, industrialização e políticas ambientais.

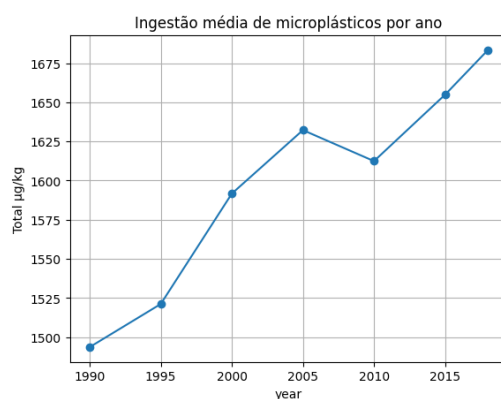


Figura 1. Ingestão média de microplásticos por ano

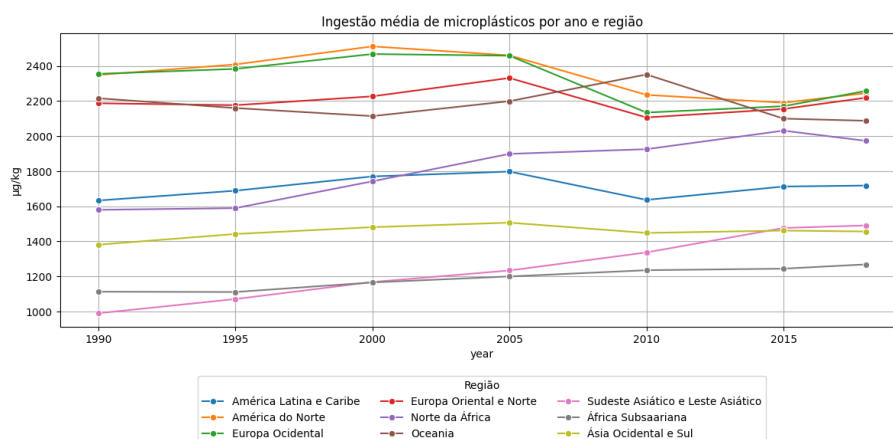


Figura 2. Ingestão média de microplásticos por ano e região

4.2. Algoritmos de agrupamento

A análise temporal dos clusters revelou padrões consistentes de exposição a microplásticos entre regiões com características socioeconômicas semelhantes, independentemente da proximidade geográfica. Desde 1990, observou-se a formação de dois agrupamentos principais: um composto por países desenvolvidos (América do Norte, Europa Ocidental e Oceania), com níveis elevados de ingestão, e outro por regiões em desenvolvimento (América Latina, África, Sudeste Asiático), com níveis mais baixos ou intermediários. Essa divisão se manteve relativamente estável ao longo do tempo.

O algoritmo K-Means identificou uma breve reconfiguração em 2010, indicando possíveis mudanças temporárias nos padrões de exposição, possivelmente ligadas a políticas ambientais ou alterações nos sistemas de produção alimentar. Já o DBSCAN evidenciou variações mais sensíveis, com destaque para o comportamento da América Latina e Caribe como outlier em anos como 1990 e 2000, sugerindo padrões únicos de contaminação. De modo geral, a Aglomerativa se manteve homogênea na maior parte do tempo.

O ano 2000 se destacou pela alta dispersão entre regiões, sugerindo um momento de reorganização global. No entanto, a partir de 2015, os agrupamentos retornaram a uma estrutura semelhante à de 1990, sugerindo uma consolidação nos perfis regionais

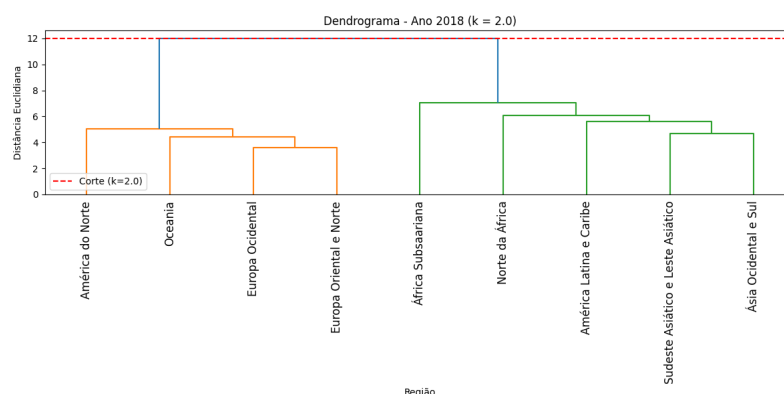


Figura 3. Dendrograma - Ano 2018

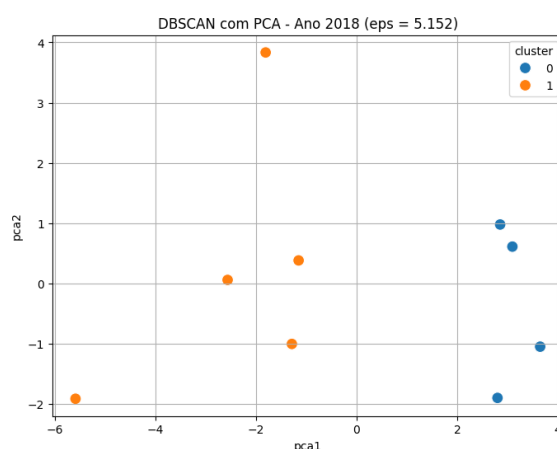


Figura 4. DBSCAN com PCA - Ano 2018

de exposição. Os resultados apontam que fatores como desenvolvimento econômico e estrutura dos sistemas alimentares têm mais influência sobre os padrões de contaminação do que a localização geográfica em si.

4.3. Comparação entre modelos

A análise dos dendrogramas hierárquicos de 1990 a 2018 revela uma estrutura global notavelmente estável, com a formação recorrente de dois grandes agrupamentos que refletem a divisão entre regiões desenvolvidas (como Europa Ocidental, América do Norte e Oceania) e regiões em desenvolvimento (como África Subsaariana, América Latina e Ásia Sul e Ocidental), evidenciando a persistência das desigualdades socioeconômicas ao longo do tempo. Essa estabilidade estrutural é comparável à observada no modelo K-means, que também manteve dois clusters principais durante a maior parte do período analisado, destacando a robustez dos padrões macrorregionais. Em contraste, o modelo DBSCAN apresentou maior sensibilidade a variações e anomalias locais, identificando outliers como em 2000, quando seis regiões foram classificadas como atípicas, revelando nuances não captadas pelos métodos anteriores. Dessa forma, enquanto os dendrogramas e o K-means capturam a constância da divisão global, o DBSCAN enriquece a análise ao destacar flutuações regionais e transições pontuais nos padrões de exposição a microplásticos.

4.4. Limitações encontradas

Esta análise apresenta três limitações principais. A primeira refere-se à resolução dos dados, que são agregados por país e categoria alimentar, o que pode ocultar variações internas relevantes, como desigualdades socioeconômicas e diferenças entre áreas urbanas e rurais. Em segundo lugar, há sensibilidade às escalas regionais: regiões com grande diversidade interna, como a África Subsaariana, podem ser representadas de forma simplificada, enquanto áreas mais homogêneas dominam determinados agrupamentos.

Por fim, a análise não incorpora fatores externos — como políticas públicas, acordos comerciais ou práticas culturais — que influenciam a ingestão de microplásticos e poderiam ajudar a explicar melhor as diferenças observadas entre regiões.

5. Conclusão

A análise conduzida confirma que a ingestão global de microplásticos tem aumentado ao longo do período entre 1990 e 2018, com um crescimento geral de cerca de 12,7%, evidenciando uma tendência ascendente consistente. Quanto às diferenças regionais, identificou-se que regiões desenvolvidas, como América do Norte e Europa Ocidental, apresentam níveis mais elevados e relativamente estáveis, enquanto regiões em desenvolvimento, como África Subsaariana e América Latina, exibem níveis mais baixos, porém em ascensão, refletindo desigualdades estruturais nos sistemas alimentares e ambientais.

Por meio da aplicação dos algoritmos de clusterização hierárquica, K-Means e DBSCAN, foi possível identificar agrupamentos coesos de países com padrões semelhantes de ingestão de microplásticos, confirmando que esses clusters tendem a se organizar principalmente segundo critérios socioeconômicos e não apenas geográficos. A análise temporal mostrou que tais agrupamentos mantiveram uma estrutura relativamente estável ao longo do tempo, embora tenham ocorrido algumas transições e reconfigurações pontuais, principalmente evidenciadas pelo DBSCAN, que destacou a rotatividade e emergência de perfis regionais atípicos.

Esses resultados corroboram que a exposição alimentar a microplásticos está fortemente influenciada pela dinâmica dos sistemas produtivos e políticas ambientais regionais, e que a clusterização aplicada em janelas temporais distintas é eficaz para revelar tanto padrões persistentes quanto mudanças estruturais nas exposições. Portanto, a abordagem metodológica adotada cumpriu seu objetivo de mapear e acompanhar a evolução dos perfis de ingestão de microplásticos globalmente, fornecendo subsídios importantes para futuras análises e intervenções políticas.

6. Referências

[1] JAMES, Gareth; WITTEN, Daniela; HASTIE, Trevor; TIBSHIRANI, Robert; TAYLOR, Jonathan. *An Introduction to Statistical Learning: With Applications in Python*. 1. ed. Cham: Springer, 2023.

[2] Gagini, Pedro Augusto Martins; Salmaze, Pedro Henrique; Joioso, Amanda Valukas Breviglieri. *Clusterização Temporal da Ingestão de Microplásticos: Estabilidade e Transição entre Grupos Regionais*. Google Colab, 2025. Disponível em: <https://colab.research.google.com/drive/1spVPrhNXyxvvg96aaLkzSZpgvL6OzfWm?usp=sharing>. Acesso em: 29 jun. 2025.