



Dayana Telles Bertuzzi

**Comportamento endêmico e epidêmico da
dengue no Paraná e sua associação com
variáveis climáticas**

Maringá,
22 de abril de 2026

Dayana Telles Bertuzzi

Comportamento endêmico e epidêmico da dengue no Paraná e sua associação com variáveis climáticas

Dissertação apresentada ao Programa de Pós-graduação em Bioestatística do Centro de Ciências Exatas da Universidade Estadual de Maringá, como requisito parcial para a obtenção do título de Mestre em Bioestatística.

Universidade Estadual de Maringá - UEM

Orientador: Prof^a. Dr^a. Eniuce Menezes

Maringá,
22 de abril de 2026

Dados Internacionais de Catalogação-na-Publicação (CIP)
(Biblioteca Central - UEM, Maringá - PR, Brasil)

B552c

Bertuzzi, Dayana Telles

Comportamento endêmico e epidêmico da Dengue no Paraná e sua associação com variáveis climáticas / Dayana Telles Bertuzzi. -- Maringá, PR, 2026.
101 f. : il. color., figs., tabs., mapas

Orientadora: Profa. Dra. Eniuce Menezes.

Dissertação (mestrado) - Universidade Estadual de Maringá, Centro de Ciências Exatas, Departamento de Estatística, Programa de Pós-Graduação em Bioestatística, 2026.

1. Bioestatística - Dengue. 2. Epidemiologia - Dengue. 3. Análise climatológica - Paraná - Dengue. I. Menezes, Eniuce, orient. II. Universidade Estadual de Maringá. Centro de Ciências Exatas. Departamento de Estatística. Programa de Pós-Graduação em Bioestatística. III. Título.

CDD 23.ed. 610.72

Rosana de Souza Costa de Oliveira - 9/1366

DAYANA TELLES BERTUZZI

**Comportamento endêmico e epidêmico da dengue no Paraná e sua
associação com variáveis climáticas**

Dissertação apresentada ao Programa de Pós-Graduação em Bioestatística do Centro de Ciências Exatas da Universidade Estadual de Maringá, como requisito parcial para a obtenção do título de Mestre em Bioestatística.

BANCA EXAMINADORA



Prof.^a. Dra. Eniuce Menezes
Universidade Estadual de Maringá – PBE/UEM

Firmado por DURBAN REGUERA
MARIA LUZ - DNI ***2568** el
día 02/03/2026 con un
certificado emitido por AC

Prof.^a. PhD. Maria Durban
Universidade Carlos III de Madrid - Espanha

gov.br

Documento assinado digitalmente
JUCELIA STADINICKI DOS SANTOS
Data: 02/03/2026 09:58:16-0300
Verifique em <https://validar.itu.gov.br>

Prof.^a. Dra. Jucelia Stadinicki dos Santos
Universidade Federal do Paraná – UFPR

Maringá, 27 de fevereiro de 2026.

AGRADECIMENTOS

As vezes, as palavras parecem insuficientes para expressar a imensidão da gratidão que carrego em meu coração. Ao longo da minha vida, houve uma presença constante, fonte inesgotável de amor, apoio e sabedoria: minha mãe, Rosa. Se há alguém a quem devo não só esta conquista, mas grande parte do que sou, é a ela. Mãe, você foi extraordinária e única. Seu apoio incansável e seu sacrifício são a essência por trás dos meus passos firmes e desta jornada.

À minha irmã, por sempre acreditar no meu potencial e torcer pelas minhas conquistas.

À minha sobrinha, cuja existência e pureza trazem uma perspectiva de leveza ao futuro.

A você, Felipi, meu companheiro de vida. Obrigada por segurar a minha mão com tanta firmeza nos dias de exaustão, por celebrar cada pequena vitória comigo e por transformar nossos planos em realidade.

Expresso minha admiração por minha orientadora, Eniuce Menezes. Agradeço por sua dedicação e paciência ao longo de todo este processo e por ter acreditado no potencial desta pesquisa desde o início. Ter contado com seu apoio foi, sem dúvida, um grande privilégio em minha formação.

A todos os meus amigos, expresso minha profunda gratidão a cada um de vocês que caminhou ao meu lado durante esta jornada. Agradeço pelas trocas de conhecimento, pelas discussões e pela parceria em cada etapa desse desafio.

A todos os professores do Programa de Pós-Graduação em Bioestatística (PBE) que participaram da minha formação e à CAPES pelo apoio financeiro.

RESUMO

A dengue representa um desafio crítico de saúde pública no Brasil, com as condições climáticas exercendo uma forte influência nas dinâmicas de transmissão. **Objetivo:** investigação da associação entre variáveis climáticas e a incidência de dengue nos municípios do Paraná entre 2016 e 2024, visando o desenvolvimento de um sistema de alerta precoce. **Metodologia:** baseou-se na distinção de padrões endêmico e epidêmico, onde foram avaliados diferentes algoritmos de agrupamento e métricas de distância para distinção de comportamentos entre cidades. Para capturar o comportamento não linear entre as variáveis, a relação foi modelada através de *Generalized Additive Mixed Models* (GAMM) com suavização utilizando P-splines. Foi utilizado os dados da SESA-PR para a taxa da incidência e o conjunto meteorológico ERA5-Land para as variáveis climáticas. A validação do modelo foi conduzida por meio de uma base de teste externa referente ao ano de 2025. **Resultados:** Com a distinção dos padrões endêmico e surto/epidêmico, foi possível identificar 5 *clusters* com semelhanças temporal e espacial da dengue no estado, em que o ano de 2024 atingiu o ápice da série histórica com ($n = 89$) municípios em surto/epidemia. Enquanto as variáveis temperatura e precipitação apresentaram, individualmente, tendência de influência positiva sobre o desfecho, a umidade média exibiu uma associação linear negativa ($p = 0,004$). A interação das variáveis climáticas mostrou-se estatisticamente significativa ($p < 0,001$), indicando que umidades mais baixas (45% e 50%) aumentam o risco de surto/epidemia em condições de temperaturas amenas e chuvas moderadas, enquanto que níveis superiores funcionam como fator redutor. Na fase de validação externa, foi identificado o melhor limiar de decisão, atingindo uma acurácia de 66,45% (58,43% - 73,83%). O modelo demonstrou especificidade de 77,92% e um valor preditivo positivo de 71,67%. A arquitetura proposta oferece suporte à vigilância epidemiológica, permitindo a antecipação de surto/endemia e a otimização de recursos assistenciais pela Secretaria da Saúde do Paraná em cenários de mudanças climáticas.

Palavras-chaves: dengue; *clustering*; variáveis climatológicas; endemia; epidemia.

ABSTRACT

Dengue fever represents a critical public health challenge in Brazil, with climatic conditions exerting a strong influence on transmission dynamics. **Objective:** To investigate the association between climatic variables and dengue incidence in the municipalities of Paraná between 2016 and 2024, aiming to develop an early warning system. **Methodology:** The study was based on the distinction between endemic and epidemic patterns, where different clustering algorithms and distance metrics were evaluated to distinguish behavioral patterns between cities. To capture non-linear behavior among variables, the relationship was modeled using Generalized Additive Mixed Models (GAMM) with smoothing using P-splines. Data from SESA-PR were used for incidence rates, and the ERA5-Land meteorological dataset was used for climatic variables. Model validation was conducted using an external test set from the year 2025. **Resultados:** By distinguishing endemic and epidemic patterns, it was possible to identify 5 clusters with temporal and spatial similarities of dengue in the state, in which the year 2024 reached the peak of the historical series with ($n = 89$) municipalities in outbreak. While temperature and precipitation variables individually showed a positive influence trend on the outcome, average humidity exhibited a negative linear association ($p = 0.004$). The interaction of climatic variables proved to be statistically significant ($p < 0.001$), indicating that lower humidity levels (45% and 50%) increase the risk of an outbreak under conditions of mild temperatures and moderate rainfall, while higher levels act as a reducing factor. In the external validation phase, the optimal decision threshold was identified, reaching an accuracy of 66.45% (58.43% - 73.83%). The model demonstrated a specificity of 77.92% and a positive predictive value of 71.67%. The proposed architecture provides support for epidemiological surveillance, allowing for the anticipation of outbreaks and the optimization of healthcare resources by the Paraná State Health Department in climate change scenarios.

Key-words: dengue; clustering; climatological variables; endemicity; epidemic.

LISTA DE ILUSTRAÇÕES

Figura 1 – Fluxograma metodológico: etapas seguidas para clusterização de municípios do estado do Paraná por situação endêmica e epidêmica de acordo com casos de dengue.	21
Figura 2 – Dinâmica espaço-temporal baseado no canal endêmico (diagrama de controle) da dengue nos municípios do Paraná entre 2016 e 2024.	32
Figura 3 – Avaliação da qualidade e separação dos agrupamentos via Índice de Davies-Bouldin nos cenários: (a) endêmico e (b) surto/epidêmico.	33
Figura 4 – Agrupamentos (<i>clusters</i>) identificados pelo algoritmo PAM ($k = 5$) para o cenário endêmico da dengue no Paraná. A cor vermelha representa o medoide da incidência do <i>cluster</i> e a cor preta identifica as cidades pertencentes ao agrupamento	34
Figura 5 – Agrupamentos (<i>clusters</i>) identificados pelo algoritmo PAM ($k = 5$) para o cenário de surto/epidemia da dengue no Paraná. A cor vermelha representa o medoide da incidência do <i>cluster</i> e a cor preta identifica as cidades pertencentes ao agrupamento	34
Figura 6 – Efeito suavizado da Taxa de Incidência e componente sazonal sobre a probabilidade de surto/epidemia de dengue.	46
Figura 7 – Distribuição semanal da proporção de acertos e erros do modelo preditivo GAMM no conjunto de validação externa (2025)	47
Figura 8 – Mapas de contorno do risco de surto/epidemia de dengue em relação à temperatura e precipitação, considerando níveis de umidade fixa (45%, 50%, 55%, 60%, 65% e 70%). As cores refletem a contribuição do termo de interação para o risco: verde indica áreas de menor risco com linhas de contorno negativa, enquanto que a cor laranja indica maior risco de surto/epidemia com linhas de contorno positiva. Áreas em branco indicam que não há dados suficientes para fazer observações.	49

Figura 9 – Comparação da Função de Autocorrelação (FAC) dos resíduos para diferentes estruturas de erro: (a) modelo sem tratamento de dependência temporal; (b) estrutura autorregressiva de ordem 1; e (c) estrutura autorregressiva de ordem 3.	50
Figura 10 – Distribuição temporal da taxa média de incidência de dengue por semana epidemiológica e cluster: (a) cenário endêmico e (b) cenário de surto/epidemia.	52
Figura 11 – Incidência média semanal de casos de dengue confirmados para o contexto endêmico (a) e epidêmico (b)	52
Figura 12 – Distribuição espacial da classificação climática do estado do Paraná. Fonte: FUNDAÇÃO ABC (2025).	53
Figura 13 – Agrupamentos (<i>clusters</i>) identificados pelo algoritmo K-means ($k = 5$) para a dengue no Paraná: (a) cenário endêmico; (b) cenário epidêmico.	65
Figura 14 – Agrupamentos (<i>clusters</i>) identificados pelo algoritmo ACF ($k = 5$) para a dengue no Paraná: (a) cenário endêmico; (b) cenário epidêmico.	66
Figura 15 – Agrupamentos (<i>clusters</i>) identificados pelo algoritmo CORT ($k = 5$) para a dengue no Paraná: (a) cenário endêmico; (b) cenário epidêmico.	67
Figura 16 – Agrupamentos (<i>clusters</i>) identificados pelo algoritmo PDC ($k = 5$) para a dengue no Paraná: (a) cenário endêmico; (b) cenário epidêmico.	68
Figura 17 – Agrupamentos (<i>clusters</i>) identificados pelo algoritmo CID ($k = 5$) para a dengue no Paraná: (a) cenário endêmico; (b) cenário epidêmico.	69
Figura 18 – Agrupamentos (<i>clusters</i>) identificados pelo algoritmo PER ($k = 5$) para a dengue no Paraná: (a) cenário endêmico; (b) cenário epidêmico.	70
Figura 19 – Agrupamentos (<i>clusters</i>) identificados pelo algoritmo PIC ($k = 5$) para a dengue no Paraná: (a) cenário endêmico; (b) cenário epidêmico.	71
Figura 20 – Agrupamentos (<i>clusters</i>) identificados pelo algoritmo PAM ($k = 6$) para a dengue no Paraná: (a) cenário endêmico; (b) cenário epidêmico.	72
Figura 21 – Agrupamentos (<i>clusters</i>) identificados pelo algoritmo K-means ($k = 6$) para a dengue no Paraná: (a) cenário endêmico; (b) cenário epidêmico.	73
Figura 22 – Agrupamentos (<i>clusters</i>) identificados pelo algoritmo ACF ($k = 6$) para a dengue no Paraná: (a) cenário endêmico; (b) cenário epidêmico.	74
Figura 23 – Agrupamentos (<i>clusters</i>) identificados pelo algoritmo CORT ($k = 6$) para a dengue no Paraná: (a) cenário endêmico; (b) cenário epidêmico.	75
Figura 24 – Agrupamentos (<i>clusters</i>) identificados pelo algoritmo PDC ($k = 6$) para a dengue no Paraná: (a) cenário endêmico; (b) cenário epidêmico.	76
Figura 25 – Agrupamentos (<i>clusters</i>) identificados pelo algoritmo CID ($k = 6$) para a dengue no Paraná: (a) cenário endêmico; (b) cenário epidêmico.	77
Figura 26 – Agrupamentos (<i>clusters</i>) identificados pelo algoritmo PER ($k = 6$) para a dengue no Paraná: (a) cenário endêmico; (b) cenário epidêmico.	78

Figura 27 – Agrupamentos (<i>clusters</i>) identificados pelo algoritmo PIC ($k = 6$) para a dengue no Paraná: (a) cenário endêmico; (b) cenário epidêmico.	79
Figura 28 – Agrupamentos (<i>clusters</i>) identificados pelo algoritmo PAM ($k = 7$) para a dengue no Paraná: (a) cenário endêmico; (b) cenário epidêmico.	80
Figura 29 – Agrupamentos (<i>clusters</i>) identificados pelo algoritmo K-means ($k = 7$) para a dengue no Paraná: (a) cenário endêmico; (b) cenário epidêmico.	81
Figura 30 – Agrupamentos (<i>clusters</i>) identificados pelo algoritmo ACF ($k = 7$) para a dengue no Paraná: (a) cenário endêmico; (b) cenário epidêmico.	82
Figura 31 – Agrupamentos (<i>clusters</i>) identificados pelo algoritmo CORT ($k = 7$) para a dengue no Paraná: (a) cenário endêmico; (b) cenário epidêmico.	83
Figura 32 – Agrupamentos (<i>clusters</i>) identificados pelo algoritmo PDC ($k = 7$) para a dengue no Paraná: (a) cenário endêmico; (b) cenário epidêmico.	84
Figura 33 – Agrupamentos (<i>clusters</i>) identificados pelo algoritmo CID ($k = 7$) para a dengue no Paraná: (a) cenário endêmico; (b) cenário epidêmico.	85
Figura 34 – Agrupamentos (<i>clusters</i>) identificados pelo algoritmo PER ($k = 7$) para a dengue no Paraná: (a) cenário endêmico; (b) cenário epidêmico.	86
Figura 35 – Agrupamentos (<i>clusters</i>) identificados pelo algoritmo PIC ($k = 7$) para a dengue no Paraná: (a) cenário endêmico; (b) cenário epidêmico.	87
Figura 36 – Agrupamentos (<i>clusters</i>) identificados pelo algoritmo PAM ($k = 8$) para a dengue no Paraná: (a) cenário endêmico; (b) cenário epidêmico.	88
Figura 37 – Agrupamentos (<i>clusters</i>) identificados pelo algoritmo k-means ($k = 8$) para a dengue no Paraná: (a) cenário endêmico; (b) cenário epidêmico.	89
Figura 38 – Agrupamentos (<i>clusters</i>) identificados pelo algoritmo ACF ($k = 8$) para a dengue no Paraná: (a) cenário endêmico; (b) cenário epidêmico.	90
Figura 39 – Agrupamentos (<i>clusters</i>) identificados pelo algoritmo CORT ($k = 8$) para a dengue no Paraná: (a) cenário endêmico; (b) cenário epidêmico.	91
Figura 40 – Agrupamentos (<i>clusters</i>) identificados pelo algoritmo EUCL ($k = 8$) para a dengue no Paraná: (a) cenário endêmico; (b) cenário epidêmico.	92
Figura 41 – Agrupamentos (<i>clusters</i>) identificados pelo algoritmo PDC ($k = 8$) para a dengue no Paraná: (a) cenário endêmico; (b) cenário epidêmico.	93
Figura 42 – Agrupamentos (<i>clusters</i>) identificados pelo algoritmo CID ($k = 8$) para a dengue no Paraná: (a) cenário endêmico; (b) cenário epidêmico.	94
Figura 43 – Agrupamentos (<i>clusters</i>) identificados pelo algoritmo PER ($k = 8$) para a dengue no Paraná: (a) cenário endêmico; (b) cenário epidêmico.	95
Figura 44 – Agrupamentos (<i>clusters</i>) identificados pelo algoritmo PIC ($k = 8$) para a dengue no Paraná: (a) cenário endêmico; (b) cenário epidêmico.	96
Figura 45 – Distribuição espacial dos clusters por classificação epidemiológica: a) endemia e b) surto/epidemia. Observação: as cores atribuídas são independentes e aleatória entre os mapas.	100

LISTA DE TABELAS

Tabela 1 – Algoritmos e métricas de dissimilaridade do pacote TSclust utilizados na análise de concordância.	25
Tabela 2 – Comparação da similaridade (<i>Sim</i>) entre o método basal PAM e demais métricas para $k = 5$	37
Tabela 3 – Distribuição dos municípios por cluster nos cenários endêmico e de surto/epidemia.	38
Tabela 4 – Métricas de desempenho das variáveis climáticas por cenário epidemiológico intra-cluster.	40
Tabela 5 – Resumo das especificações dos modelos GAMM selecionados.	41
Tabela 6 – Métricas de desempenho dos modelos para as distâncias PAM, HCA-EUCL e CORT, na etapa do treino	42
Tabela 7 – Métricas detalhadas para o modelo de teste (2025) utilizando suavização P-splines.	43
Tabela 8 – Estimativas dos coeficientes paramétricos, Razão de Chances (OR) e termos suavizados do modelo final.	44
Tabela 9 – Comparação de modelos GAMM com diferentes estruturas de autocorrelação residual para incidência de dengue.	51
Tabela 10 – Comparação da similaridade (<i>Sim</i>) entre o método PAM e demais algoritmos/métricas para $k = 5$ a $k = 8$	97
Tabela 11 – Distribuição dos municípios por classificação epidemiológica (Endemia e Epidemia) e agrupamentos (Clusters)	97

LISTA DE ABREVIATURAS E SIGLAS

ACF	<i>Autocorrelation Function</i> (Função de Autocorrelação)
AIC	<i>Akaike Information Criterion</i> (Critério de Informação de Akaike)
<i>Sim</i>	<i>Clustering Evaluation Index Based on Known Ground Truth</i> (Índice de Avaliação de Agrupamento Baseado em Verdade Terrestre Conhecida)
AUC	<i>Area Under the Curve</i> (Área Sob a Curva)
BIC	<i>Bayesian Information Criterion</i> (Critério de Informação Bayesiano)
CID	<i>Complexity-Invariant Distance</i> (Distância Invariante à Complexidade)
CORT	<i>Correlation-based distance</i> (Distância baseada em correlação)
DENV	<i>Dengue Virus</i> (Vírus da Dengue)
DG	Dengue Grave
DIR	<i>Dengue Incidence Rate</i> (Taxa de Incidência de Dengue)
DSA	Dengue com Sinais de Alarme
ECMWF	<i>European Centre for Medium-Range Weather Forecasts</i> (Centro Europeu de Previsões Meteorológicas de Médio Prazo)
EPSG	<i>European Petroleum Survey Group</i> (Código de referência de sistemas de coordenadas)
ERA5 - Land	<i>ECMWF Reanalysis 5th Generation</i> (Reanálise de quinta geração do ECMWF)
GAMM	<i>Generalized Additive Mixed Models</i> (Modelos Mistos Aditivos Generalizados)

IC	Intervalo de Confiança
ICTV	<i>International Committee on Taxonomy of Viruses</i> (Comitê Internacional de Taxonomia de Vírus)
LS	Limite Superior
PAM	<i>Partitioning Around Medoids</i> (Partição em Torno de Medoides)
PDC	<i>Permutation Distribution Clustering</i> (Agrupamento por Distribuição de Permutação)
PER	<i>Periodogram-based distance</i> (Distância baseada em Periodograma)
PQL	<i>Penalized Quasi-Likelihood</i> (Quase-Verossimilhança Penalizada)
REML	<i>Restricted Maximum Likelihood</i> (Máxima Verossimilhança Restrita)
SESA-PR	Secretaria de Estado da Saúde do Paraná
SINAN	Sistema de Informação de Agravos de Notificação
VPN	Valor Preditivo Negativo
VPP	Valor Preditivo Positivo
WGS 84	<i>World Geodetic System 1984</i> (Sistema Geodésico Mundial 1984)

SUMÁRIO

1	Introdução	15
2	Material e Métodos	19
2.1	Dados da dengue e diagrama de controle	19
2.2	Dados climáticos	22
2.3	Análise de concordância entre métodos de partição	22
2.4	Coerência climática dos agrupamentos	25
2.5	Modelos Mistos Aditivos Generalizados (GAMM)	26
2.5.1	Estrutura da autocorrelação temporal	27
2.6	Seleção de modelos e métricas de avaliação	28
2.7	Implementação e recursos computacionais	29
3	Resultados e Discussão	31
3.1	Índice Davies-Bouldin	32
3.1.1	Níveis de correlação entre os métodos de partição	35
3.1.2	Níveis de correlação intracluster	38
3.2	Modelos Mistos Aditivos Generalizados (GAMM)	41
3.2.1	Análise do desempenho dos modelos no treinamento	41
3.2.2	Resultados da validação externa (conjunto de teste)	42
3.2.3	Análise das estimativas de risco do modelo final	43
3.2.4	Análise da interação climática sobre o risco de transmissão	47
3.2.5	Análise da dependência temporal para o modelo PAM quantitativo	49
3.2.6	Classificação de risco e dinâmica de transmissão e análise do modelo final	51
4	Conclusão	54
	Referências	56
	Apêndices	63
	APÊNDICE A Dinâmica temporal dos agrupamentos de dengue	64
A.1	Agrupamentos nos cenários endêmico e epidêmico	64
A.1.1	Configuração da partição ($k = 5$)	65
A.1.2	Configuração da partição ($k = 6$)	72

A.1.3	Configuração da partição ($k = 7$)	80
A.1.4	Configuração da partição ($k = 8$)	88
A.1.5	Avaliação da similaridade entre algoritmos de partição e métricas hierárquicas para $k = 5$ a $k = 8$	97
A.2	Distribuição geográfica nos cenários de endemia e surto/epidemia para o algoritmo PAM	97
A.3	Distribuição espacial nos cenários de endemia e surto/epidemia para o algoritmo PAM	100

INTRODUÇÃO

A dengue faz parte do grupo das arboviroses, doenças que dependem da interação entre vírus e artrópodes para se espalhar. No Brasil, essa transmissão ocorre principalmente através da fêmea do mosquito *Aedes aegypti*. O agente etiológico pertence à família Flaviviridae e ao gênero *Orthoflavivirus* (International Committee on Taxonomy of Viruses, 2024), que se apresenta em quatro sorotipos distintos, DENV-1, DENV-2, DENV-3 e DENV-4, cada um com suas próprias variações genéticas (Ministério da Saúde, 2025). A infecção pode variar de quadros assintomáticos a formas graves. As manifestações iniciais incluem febre alta, dores no corpo e náuseas, podendo evoluir para Dengue com Sinais de Alarme (DSA), caracterizada por dor abdominal intensa e vômitos persistentes, ou para Dengue Grave (DG), manifestada por sangramento de mucosas e comprometimento grave de órgãos, condições que exigem atenção médica imediata (Secretaria de Estado da Saúde do Paraná, 2025).

As epidemias de dengue representam um desafio para os sistemas de saúde pública, tanto pelo aumento da demanda por atendimentos quanto pelos riscos associados às formas graves da doença, incluindo hospitalizações e óbitos. De acordo com as novas diretrizes nacionais, o foco recai sobre a vigilância entomológica e o controle vetorial. O objetivo é viabilizar intervenções mais ágeis e, sensíveis às particularidades de cada território. Nesse sentido, as diretrizes reforçam que a eficácia no combate aos surtos regionais depende de estratégias de controle capazes de responder com a urgência que cada realidade local exige (BRASIL, 2025).

Nos últimos anos, a dengue tem sido a principal arbovirose em circulação nas Américas, com exceção do Canadá, e tem apresentado crescimento alarmante, culminando em um aumento histórico de casos registrados em 2024 (Organização Pan-Americana da Saúde, 2025). Ao longo de todo o ano de 2024, a situação atingiu níveis alarmantes, com a Região das Américas contabilizando 13.027.747 casos notificados por 50 países. Desse total, 6.906.396 foram confirmados laboratorialmente, 22.684 foram caracterizados como dengue grave e 8.186 foram casos fatais (Organização Pan-Americana da Saúde, 2025). Este cenário, no Brasil, foi acompanhado pela circulação simultânea dos quatro sorotipos (Organização Pan-Americana da Saúde, 2025; World Health Organization, 2024). Países do hemisfério norte registraram um

crescimento expressivo de notificações antes do início do período sazonal de alta transmissão, previsto para o segundo semestre de 2024 (World Health Organization, 2024).

No Brasil, a primeira ocorrência documentada de dengue foi registrada entre 1981 e 1982. Desde então, a doença tem se mantido de forma contínua (endêmica), intercalada por episódios epidêmicos, geralmente associados à introdução ou substituição de sorotipos e à expansão do vetor (Ministério da Saúde, 2025). O volume de notificações alcançou 10.232.872 casos, desses, 5.973.340 confirmados e 6.321 óbitos, consolidando 2024 como o ano de maior magnitude epidemiológica da série histórica nacional (Pan American Health Organization, 2026).

A expansão territorial da dengue no Brasil tem sido associada às mudanças climáticas. O estudo de Barcellos et al. (2024) identificou que anomalias térmicas, especialmente em áreas de maior altitude nas regiões Sul e Centro-Oeste, estão ligadas ao avanço da doença. Microrregiões com aumento persistente de temperatura, combinadas à urbanização e ao histórico prévio de incidência, apresentaram maior risco de transmissão sustentada. Esses fatores vêm tornando o padrão de transmissão mais estável e duradouro em áreas anteriormente menos afetadas.

De acordo com a World Health Organization (2024), esse agravamento tem sido atribuído a fatores como o início precoce e a maior duração das temporadas de transmissão, além das mudanças climáticas e eventos meteorológicos periódicos, como El Niño e La Niña, que alteram padrões de temperatura, precipitação e umidade, favorecendo a proliferação do vetor. Nesse contexto, Morin, Comrie e Ernst (2013) ressaltam que esses impactos não são imediatos, mas apresentam efeitos defasados (lags) que refletem o tempo biológico de desenvolvimento do vetor e da replicação do vírus. Compreender essas associações e de suas não-linearidades é fundamental para compreender a sazonalidade da doença e fundamentar estratégias preditivas.

Entre os estados da região Sul do Brasil, o Paraná destaca-se pela crescente vulnerabilidade ao avanço da dengue impulsionado pelas mudanças climáticas. A quebra dessas barreiras climáticas, evidenciada pelas anomalias de temperatura registradas nos últimos anos, permite que o vetor colonize novos territórios e sustente a transmissão viral em regiões anteriormente sem classificação de endemia (BARCELLOS et al., 2024; BRASIL, 2025).

No Paraná, o período epidemiológico 2023/2024 atingiu 624.995 casos confirmados e 741 óbitos decorrentes da dengue (Secretaria de Estado da Saúde do Paraná, 2024), evidenciando a expansão da transmissão regional. Esse volume de notificações destaca a consolidação de áreas de alta incidência, especialmente nas regiões norte e noroeste do estado. Estudos apontam que a formação de agrupamentos espaciais de municípios com elevadas taxas de casos sugere um processo de interiorização da doença e a manutenção de focos epidêmicos regionais (FREIRE et al., 2023).

Entre 2014 e 2020, Barcellos et al. (2024) identificaram um número expressivo de

microrregiões paranaenses com taxas de incidência da dengue (Dengue Incidence Rate – DIR) superiores a 300 casos por 100 mil habitantes, sobretudo no oeste e noroeste do estado, em áreas coincidentes com a bacia do rio Paraná. Esses territórios, anteriormente classificados como de baixa transmissão, passaram a apresentar condições favoráveis à disseminação do vírus, confirmando a tendência de interiorização e expansão territorial da dengue no Paraná (BARCELLOS et al., 2024).

Essa heterogeneidade na distribuição da doença também é abordada por Melo et al. (2014), que destacam a influência de fatores regionais, como o clima e a circulação viral entre áreas vizinhas, na dinâmica da transmissão. Os autores enfatizam que compreender esses padrões espaciais é fundamental para subsidiar estratégias de prevenção mais eficazes. Desse modo, fatores climáticos, como temperatura, precipitação e umidade, influenciam diretamente o ciclo de vida e a dinâmica populacional do vetor *Aedes aegypti*, afetando desde a alimentação e o voo até o desenvolvimento e a densidade populacional do vetor, devido à maior disponibilidade de criadouros em condições favoráveis (REINHOLD; LAZZARI; LAHONDÈRE, 2018). Nesse sentido, investigar variáveis meteorológicas no estado do Paraná torna-se essencial para identificar condições ambientais que possam proceder eventos de maior incidência da dengue e, assim, antecipar surtos por meio de métodos que considerem defasagens temporais.

Embora a dengue seja endêmica em diversos municípios paranaenses e o comportamento de ocorrências siga um padrão sazonal, sua dinâmica apresenta duas características importantes que norteiam as hipóteses investigadas neste trabalho. A primeira refere-se à instabilidade temporal, em que localidades com histórico endêmico podem apresentar alterações bruscas no comportamento da doença, como a antecipação do pico em caso de epidemias. A segunda diz respeito à heterogeneidade espacial, onde a intensidade da transmissão, seja em períodos endêmicos ou de surto/epidemia, varia entre os municípios, embora existam agrupamentos que compartilham comportamentos epidemiológicos semelhantes.

Churakov et al. (2019), ao analisarem a circulação sazonal da dengue no Brasil entre 2001 e 2016, identificaram ondas epidêmicas recorrentes que se iniciam nos estados do oeste e se propagam em direção ao leste e ao nordeste, com variações regionais no início e pico das epidemias. A região Sul, por sua vez, não segue esse padrão de forma uniforme: enquanto Santa Catarina e Rio Grande do Sul apresentaram baixa incidência e ausência de sazonalidade, o Paraná demonstrou maior alinhamento com a onda nacional, embora não tenha sido analisado de forma individualizada. Segundo os autores, a análise baseou-se em séries temporais contínuas e na coerência anual dos sinais, sem distinção entre anos endêmicos e epidêmicos.

Almeida, Lana e Codeço (2022) destacam que a classificação de perfis temporais de transmissão pode subsidiar estratégias mais específicas de vigilância e controle, especialmente diante de fatores como clima e densidade populacional, que moldam diferentes padrões epidemiológicos no território nacional. No estado do Paraná, observaram uma separação espacial nítida entre os perfis, com transmissão persistente inclusive em municípios pequenos, sugerindo

dinâmicas locais complexas como a conectividade regional. Essa estratégia segue as recomendações de estudos recentes como os de Chen e Moraga (2025), que apontam a incorporação de dados temporais, climáticos e de mobilidade como essenciais para melhorar a previsão de surtos e a sensibilidade de sistemas de alerta precoce.

Considerando tais lacunas na abordagem tradicional, observa-se que um mesmo município pode alternar entre períodos endêmicos, ou seja, de transmissão persistente para períodos de surto/epidemia ao longo do tempo, o que dificulta muito investigar e modelar as séries temporais tradicionalmente construídas de modo contínuo ao longo dos anos. Villela (2024) reforçam essa limitação ao demonstrarem que os surtos de dengue não seguem um padrão fixo de ocorrência e que ciclos sazonais anuais, aqueles que presumem repetições regulares ano a ano, são insuficientes para previsões precisas. A variabilidade temporal entre diferentes anos exige, portanto, abordagens que considerem essa complexidade.

Este trabalho caracteriza-se como um estudo ecológico que propõe uma nova abordagem baseada na construção de séries temporais agregadas por semanas epidemiológicas que representem separadamente os comportamentos endêmicos e comportamentos de surto/epidemia em cada município. Para a separação entre as duas séries, os anos epidêmicos (surto) de cada município são identificados de acordo com o comportamento temporal histórico ao longo dos anos. Para identificar municípios paranaenses com comportamento similar, propõe-se o uso da técnica de *clustering* para agrupar municípios segundo padrões temporais de ocorrência da dengue, distinguindo contextos endêmicos e epidêmicos ao longo do período de 2016 a 2024.

Essa abordagem visa identificar grupos com comportamentos semelhantes entre as semanas epidemiológicas e analisar suas associações com variáveis climáticas, contribuindo para o entendimento das relações entre o clima e a transmissão da doença. A predição da incidência média semanal por cluster foi realizada por meio de modelos Modelos Aditivos Generalizados Mistos (GAMM), empregando bases de suavização do tipo *P-spline* para o ajuste das variáveis climáticas. Essa abordagem visa antecipar períodos de surtos/epidemias por meio da identificação de padrões associados às condições climáticas, contribuindo para estratégias mais eficazes de vigilância e controle no contexto regional do estado do Paraná.

MATERIAL E MÉTODOS

2.1 Dados da dengue e diagrama de controle

Neste estudo, foram utilizados dados de casos de dengue registrados no estado do Paraná entre 2016 e 2025. As informações foram obtidas por meio do Departamento de Informática do Sistema Único de Saúde (DATASUS), via Sistema de Informação de Agravos de Notificação (SINAN). Foram considerados todos os casos confirmados em pacientes residentes nos 399 municípios paranaenses.

A clusterização dos municípios foi realizada a partir da série temporal de casos confirmados. Foram incluídos apenas os municípios com registros em pelo menos sete anos do período analisado, totalizando 218 municípios dentre os 399 existentes no estado, com séries temporais semanais completas. Essa exigência se deve ao fato de que séries muito curtas não permitem o cálculo adequado das estatísticas necessárias para a construção do diagrama de controle. Assim, municípios com menos de sete anos de dados disponíveis foram excluídos da análise, por não apresentarem comportamento endêmico da doença. A Figura 1 apresenta o fluxograma que resume as etapas da metodologia adotada para a clusterização.

A metodologia buscou identificar padrões temporais da dengue no Paraná, distinguindo duas dimensões distintas: o comportamento endêmico (ocorrência regular) do surto/epidêmico (definido pela incidência que supera o limiar endêmico estabelecido) (Ministério da Saúde, 2009). Para reduzir o impacto de surtos isolados e caracterizar a endemicidade de forma estável, os registros foram agrupados por semana epidemiológica (SE), adotando-se a data de início dos sintomas, gerando uma distribuição que reflete a frequência esperada da doença ao longo do período analisado. A organização dos dados para a SE seguiu o calendário oficial do Ministério da Saúde, no qual as unidades temporais são contabilizadas de domingo a sábado.

O processamento foi realizado no software R (R Core Team, 2024), com auxílio do pacote *dengueControl* (MENEZES, 2025). No Brasil, o cálculo da tendência central do canal endêmico tem sido realizado pela mediana. No entanto, para essa representação, pode ser também calculado a média removendo os anos epidêmicos. Nesse trabalho, na abordagem

endêmica, foi utilizada a média para representar o padrão esperado via Diagrama de Controle. Já na análise epidêmica, o perfil foi delineado pelo limite superior (LS) da série (terceiro quartil), capturando os picos recorrentes. Em ambas as abordagens, foram utilizadas as taxas de incidência por 100 mil habitantes. Após a padronização das taxas de incidência pelo método Z -score, procedeu-se à identificação do número ideal de grupos para a análise.

As análises foram conduzidas no ambiente estatístico R (R Core Team, 2024), com o suporte do pacote *dengueControl* (MENEZES, 2025). No Brasil, o cálculo da tendência central do canal endêmico tem sido realizado pela mediana. No entanto, para essa representação, pode ser também calculado a média removendo os anos epidêmicos. Para a caracterização da abordagem endêmica nesse trabalho, foi utilizado a média das séries históricas após a exclusão dos anos classificados como epidêmicos. Esse procedimento visou isolar o comportamento endêmico da doença e estabelecer o padrão esperado por meio do Diagrama de Controle. O critério para identificação e retirada desses anos de surto/epidemia baseou-se no Limite Superior (LS) da série, definido pelo terceiro quartil (Q_3). Em contrapartida, para a abordagem epidêmica, manteve-se a série histórica integral, visando capturar a magnitude real e a variabilidade dos casos durante os períodos de surto/epidemia. Em ambas as abordagens, as análises basearam-se na taxa de incidência por 100 mil habitantes, posteriormente foram padronizadas pelo método Z -score para a identificação do número ideal de agrupamentos na análise de clusterização.

Para determinar o número ideal de agrupamentos (k) utilizou-se o índice de validação interna Davies-Bouldin (DB), que fornece uma medida quantitativa da qualidade do particionamento que equilibra a homogeneidade intra-cluster e a heterogeneidade inter-cluster (DAVIES; BOULDIN, 1979). Ao contrário de métricas que consideram apenas a variância interna, tal como a técnica visual e heurística do conhecido método Elbow (Cotovelo) (SYAKUR et al., 2018), o DB identifica o ponto de configuração onde os clusters estão melhor separados, o que é essencial para a distinção de padrões dinâmicos em séries temporais.

O índice é calculado a partir da razão entre a dispersão intra-cluster e a separação inter-cluster. Para cada par de clusters, define-se uma medida de similaridade R_{ij} :

$$R_{ij} = \frac{S_i + S_j}{M_{ij}} \quad (2.1)$$

em que S_i e S_j representam a dispersão (distância média dos pontos em relação ao medoide) dentro dos clusters i e j , enquanto M_{ij} denota a distância entre os respectivos medoides.

A partir dessas medidas, identifica-se a maior similaridade para cada cluster i (R_i), representando ponto de menor distinção frente aos demais agrupamentos:

$$R_i = \max_{j \neq i} (R_{ij}) \quad (2.2)$$

Assim, o índice DB final é obtido pela média dessas similaridades máximas, conforme a Equação 2.3. O valor mínimo do índice aponta a configuração ideal, pois garante a máxima distinção entre os agrupamentos identificados (DAVIES; BOULDIN, 1979).

$$DB = \frac{1}{k} \sum_{i=1}^k R_i \quad (2.3)$$

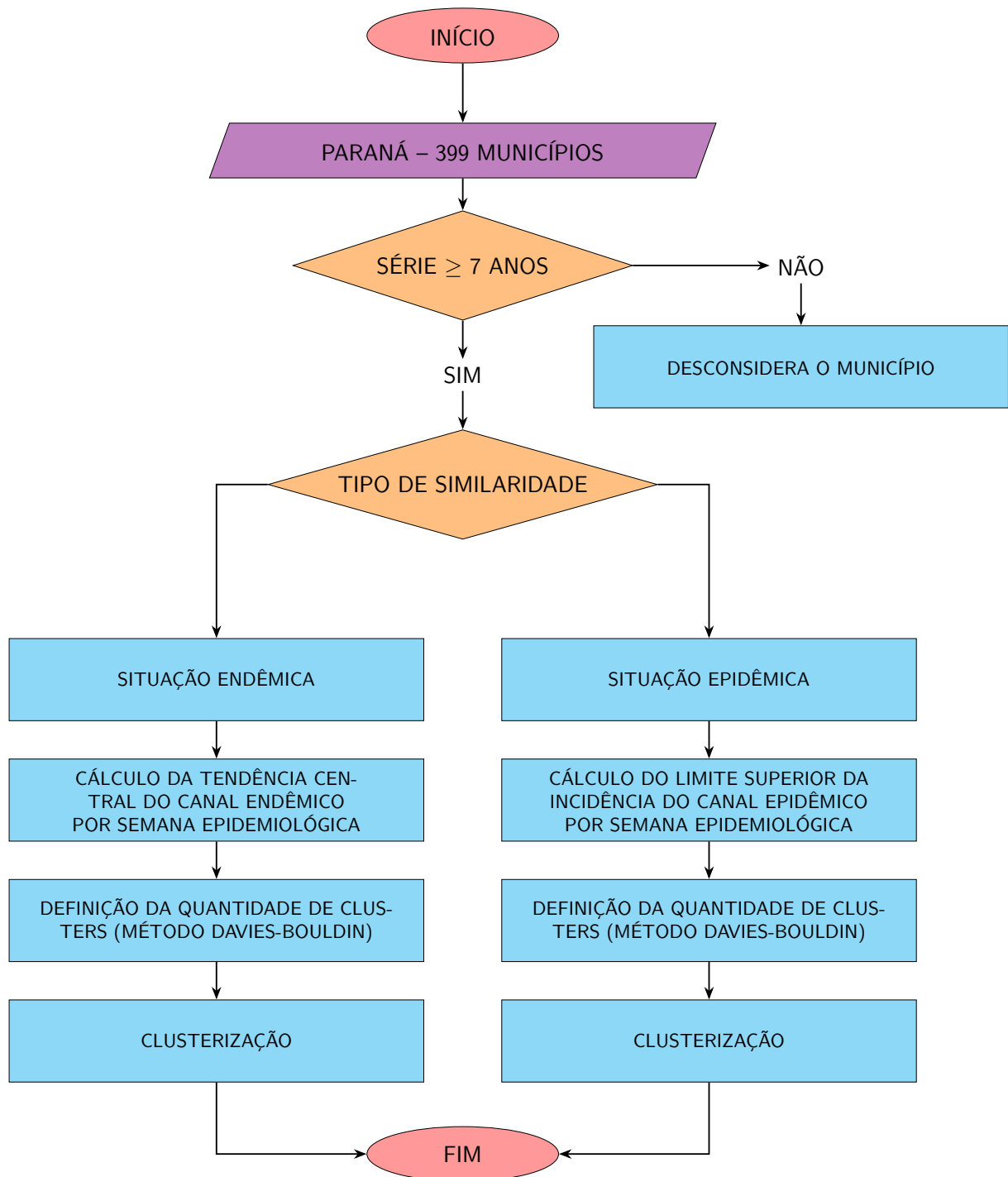


Figura 1 – Fluxograma metodológico: etapas seguidas para clusterização de municípios do estado do Paraná por situação endêmica e epidêmica de acordo com casos de dengue.

2.2 Dados climáticos

A partir da clusterização dos municípios por padrão temporal de ocorrência da dengue, investigou-se a relação entre esses agrupamentos e variáveis climáticas. Foram utilizados dados diários de temperatura, precipitação e ponto de orvalho do conjunto ERA5-Land (resolução $0.1^\circ \times 0.1^\circ$), disponibilizados pelo European Centre for Medium-Range Weather Forecasts (ECMWF, 2024), processados em Python (Python Software Foundation, 2021) e agregados semanalmente por município, calculou-se a média dos totais acumulados semanais para a precipitação e a média das temperaturas semanais. Devido à ausência de dados diretos de umidade relativa na base ERA5-Land, esta variável foi estimada por meio da equação de Magnus-Tetens (ALDUCHOV; ESKRIDGE, 1996).

Para associar os dados climáticos aos municípios do Paraná, utilizou-se a malha vetorial dos limites municipais em formato shapefile, convertida para o sistema WGS 84 (EPSG:4326). Em seguida, foi realizado um recorte espacial com base na grade climática, mantendo apenas os municípios cobertos pelos dados. Os valores diários das variáveis climáticas foram organizados em uma grade de coordenadas e vinculados aos municípios por meio de junção espacial. Após, os dados foram interpolados por aproximação linear para cobrir lacunas e padronizados com exclusão da semana 53.

2.3 Análise de concordância entre métodos de partição

O presente estudo investigou algoritmos de clusterização para séries temporais com diferentes abordagens. Considerou-se o K-means e o *Partitioning Around Medoids* (PAM) como algoritmos de particionamento, além do algoritmo Hierárquico Aglomerativo (Hclust).

Qualquer método de clusterização utiliza alguma medida de dissimilaridade. Para os algoritmos de particionamento (K-means e PAM), adotou-se exclusivamente a distância euclidiana. Esta escolha justifica-se por dois motivos principais:

1. Coerência algorítmica: O algoritmo K-means padrão assume que o centróide é a média aritmética, uma propriedade que minimiza a variância apenas no espaço euclidiano. A utilização de métricas elásticas ou baseadas em forma exigiria adaptações algorítmicas (como DTW - Dynamic Time Warping ou Barycenter Averaging) que descaracterizariam o algoritmo padrão e fugiria ao escopo de comparação dos métodos clássicos tal como proposto por MacQueen (1967)
2. Definição de *Baseline Lock-step*: A utilização da distância euclidiana nestes algoritmos estabelece uma *baseline* para comparações ponto-a-ponto rígidas (*lock-step measures*) (Ratanamahatana; Keogh, 2004), que está de acordo com o interesse nesta pesquisa, permitindo identificar no tempo a ocorrência de semelhanças no padrão sazonal.

Embora o PAM aceite qualquer matriz de distância (ao contrário do K-means), mantê-lo com a euclidiana serve para comparar diretamente a performance de medóides (PAM) versus centróides sob a mesma régua geométrica. Isso isola a "robustez a outliers" (vantagem do PAM) sem a interferência da mudança de métrica.

Em contrapartida, o algoritmo Hierárquico Aglomerativo (Hclust) foi utilizado como plataforma exploratória para medidas de dissimilaridade de maior complexidade e abstração, conforme taxonomia de Esling e Agon (2012), Aghabozorgi, Shirkhorshidi e Wah (2015). A estrutura do método hierárquico permite a pré-computação da matriz de dissimilaridade, viabilizando o uso de métricas com maior custo computacional (baseadas em modelos e compressão) que seriam inviáveis em abordagens iterativas de particionamento.

O objetivo principal do PAM é minimizar a soma das dissimilaridades entre cada objeto e o medoide mais próximo. A função S do modelo de referência é definida pela Equação 2.4:

$$S = \sum_{i=1}^N \min_{j=1}^k d(x_i, m_j) \quad (2.4)$$

em que N representa o número total de municípios, k denota o número de grupos pré-definidos, m_j é o medoide representante do *cluster* j , e $d(x_i, m_j)$ é a medida de dissimilaridade entre o município i e o respectivo medoide.

Para a mensuração dessas dissimilaridades, empregou-se a distância euclidiana baseada na representação de séries temporais. O pacote TSrepr (LAURINEC, 2018) foi utilizado para a extração dos perfis sazonais para o PAM, enquanto o pacote TScluster (MONTERO; VILAR, 2014) foi aplicado na etapa de validação via algoritmo hierárquico. Esse pacote é abordado por (DOVE et al., 2023) como referência para o tratamento de séries biológicas complexas. A métrica adotada para mensurar a semelhança entre as observações e os centros dos agrupamentos (centroides ou medoides) foi a distância euclidiana (TAN et al., 2019), expressa pela Equação 2.5:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2.5)$$

em que $i = 1, 2, 3, \dots, n$, na qual n denota o número de atributos (52 semanas epidemiológicas), x_i representa o valor observado na i -ésima semana e y_i a coordenada correspondente do centro do agrupamento (SYAKUR et al., 2018). Esta métrica de distância é compartilhada por todos os algoritmos que utilizam a base euclidiana, garantindo que a concordância medida posteriormente reflita a consistência da estrutura sazonal entre os diferentes métodos.

Para fins de validação, os agrupamentos do PAM foram confrontados com o algoritmo K-means (MACQUEEN, 1967). Embora compartilhe a métrica definida na Equação 2.5, o K-means busca minimizar a variância total intragrupo através de centroides (μ_j), que representam

médias aritméticas, agrupando dados com comportamento semelhante (NIENNATTRAKUL; RATANAMAHATANA, 2007). Sua fórmula é definida por:

$$J = \sum_{k=1}^K \sum_{x_i \in c_k} \|x_i - \mu_k\|^2 \quad (2.6)$$

O processo de otimização ocorre de forma iterativa: após cada iteração, os centroides são recalculados levando em conta as novas composições dos grupos, e o processo se repete até a convergência ou estabilização dos agrupamentos (TAN et al., 2019; NIENNATTRAKUL; RATANAMAHATANA, 2007).

Adicionalmente, utilizou-se o algoritmo Hierárquico Aglomerativo (*hclust*) (MURTAGH; CONTRERAS, 2012). Diferente das técnicas de partição, este método constrói uma hierarquia de grupos de forma ascendente (*bottom-up*), fundindo municípios sucessivamente com base em um critério de ligação (*linkage*). A lógica de fusão entre dois grupos A e B é regida pela distância $D(A, B)$:

$$D(A, B) = \min_{x \in A, y \in B} d(x, y) \quad (2.7)$$

As métricas de dissimilaridade para séries temporais são classificadas em quatro eixos principais: baseadas em forma (*shape-based*), características (*feature-based*), modelos (*model-based*) e compressão (*compression-based*) (ESLING; AGON, 2012; AGHABOZORGI; SHIRKHORSHIDI; WAH, 2015). Essa taxonomia permite avaliar desde a proximidade entre os valores (ponto a ponto) até a identificação de padrões que compõem a complexidade intrínseca dos dados.

Nesta etapa, o algoritmo hierárquico serviu como base para comparar a distância euclidiana (medida do tipo *shape-based/lock-step*) com métricas que capturam outras propriedades das séries, abrangendo as diferentes taxonomias mencionadas. Esta diversidade metodológica, detalhada na Tabela 1, garante que a concordância medida reflita a consistência da estrutura sazonal sob múltiplas perspectivas.

Para quantificar a concordância entre a métrica PAM e os demais métodos, utilizou-se o índice de avaliação de agrupamento baseado na partição de referência (*Clustering Evaluation Index Based on Known Ground Truth*). Baseada inicialmente nos estudos de Larsen e Aone (1999) e Gavrilov et al. (2000), a medida de similaridade externa avalia o grau de sobreposição entre os grupos. O índice de similaridade proposto por Gavrilov et al. (2000) é representado por Sim , em que varia entre 0 (completamente dissimilar) a 1 (idênticos), sendo definido por:

$$Sim(G, A) = \frac{1}{K} \sum_{i=1}^K \max_{1 \leq j \leq K} Sim(G_i, A_j) \quad (2.8)$$

em que $G = \{G_1, \dots, G_K\}$ representa o *ground-truth* dos K agrupamentos e $A = \{A_1, \dots, A_K\}$ é o agrupamento a ser avaliado. A similaridade entre dois clusters específicos, $Sim(G_i, A_j)$, é calculada por:

$$Sim(G_i, A_j) = \frac{2|G_i \cap A_j|}{|G_i| + |A_j|} \quad (2.9)$$

Conforme observado por Gavrilov et al. (2000), foca na capacidade do algoritmo testado em recuperar os padrões estabelecidos pela referência. No contexto de séries temporais, esta abordagem foi consolidada por Kalpakis, Gada e Puttagunta (2001) e validada em estudos de simulação por Pértiga e Vilar (2010).

Tabela 1 – Algoritmos e métricas de dissimilaridade do pacote TSclust utilizados na análise de concordância.

Tipo	Método / Métrica	Base Matemática	Descrição
Shape	K-medoides - Euclidiana (PAM)	Equação 2.4	Método de referência. Particionamento via Medoides. Utiliza observações reais do banco de dados como centros.
Shape	K-means - Euclidiana (KMEANS)	Equação 2.6	Particionamento via Centroides. Utiliza médias aritméticas como centros dos grupos.
Shape	Hierárquico - Euclidiana (HCA-EUCL)	Equação 2.7	Algoritmo aglomerativo para testar a estabilidade via lógica de conectividade ponto a ponto.
Feature	Hierárquico - Autocorrelação (ACF)	$d_{ACF} = \frac{\sqrt{(\hat{\rho}_x - \hat{\rho}_y)^T \Omega (\hat{\rho}_x - \hat{\rho}_y)}}{\sqrt{(\hat{\rho}_x - \hat{\rho}_y)^T \Omega (\hat{\rho}_x - \hat{\rho}_y)}}$	Baseada nos coeficientes de autocorrelação estimada das séries temporais.
Feature	Hierárquico - Correlação (CORT)	$d_{CORT} = \Phi[\text{cor}(x, y)] \cdot d_E$	Mede a proximidade temporal e a similaridade de tendência das curvas.
Feature	Hierárquico - Complexidade (CID)	$d_{CID} = \frac{\max(CEx, CEy)}{\min(CEx, CEy)} \cdot d_E$	Distância euclidiana corrigida por uma estimativa de complexidade da série.
Feature	Hierárquico - Periodograma (PER)	$d_{PER} = \frac{1}{n} \sum I_x(\lambda_j) - I_y(\lambda_j) $	Distância entre as estimativas de densidade espectral (domínio da frequência).
Model	Hierárquico - Modelagem AR (PIC)	$d_{PIC} = \sqrt{\sum (\hat{\phi}_x, i - \hat{\phi}_y, i)^2}$	Dissimilaridade entre coeficientes de modelos autorregressivos ajustados.
Compres.	Hierárquico - Permutação (PDC)	$d_{PDC} = \frac{1}{2} \sum P(x) - P(y) $	Distância entre distribuições de permutação baseada em dinâmica simbólica.

2.4 Coerência climática dos agrupamentos

A avaliação da consistência climática foi conduzida de forma independente para cada cenário (endêmico e surto/epidêmico). Dentro de cada agrupamento (*cluster*) identificado em

ambos os cenários, a homogeneidade das variáveis ambientais foi quantificada por meio do Índice de Correlação Média Absoluta ($\bar{\rho}_{abs}$). Esta abordagem permite quantificar a sincronia climática regional dentro de cada grupo, validando se os municípios agregados compartilham, efetivamente, comportamentos atmosféricos coerentes entre si (BJØRNSTAD; IMS; LAMBIN, 1999).

O índice foi calculado a partir da magnitude média das associações presentes na triangular superior da matriz de correlação de Spearman (ρ) de cada *cluster*, conforme a Equação 2.10:

$$\bar{\rho}_{abs} = \frac{1}{N} \sum_{i < j} |\rho_{ij}| \quad (2.10)$$

em que N representa o número de pares únicos de municípios dentro do agrupamento e ρ_{ij} a correlação entre as séries dos municípios i e j . A utilização de Spearman justifica-se pela natureza não linear das relações climáticas. Para a comparação final entre os métodos de partição, utilizou-se ainda, o desvio padrão, e métricas de erro, como o Erro Quadrático Médio (EQM) e a sua respectiva raiz (RMSE) do método para seu respectivo cenário (endemia ou epidemia), permitindo identificar a técnica que gerou grupos com maior interdependência climática.

2.5 Modelos Mistos Aditivos Generalizados (GAMM)

Para modelar e permitir a interpretação da relação não-linear do clima e da dengue, utilizou-se modelos GAMM. A construção deste modelo pode ser feita de modo a prever a situação epidemiológica binária dos municípios (Endêmica ou não) ou prever a taxa de ocorrência de dengue de modo que a variável resposta seja quantitativa. No caso da variável resposta binária, o modelo pode ser representado da seguinte forma, com função de ligação *logit*:

$$\text{logit}(P(Y_{ij} = 1)) = \beta_0 + \sum_j f_j(X_{ij}) + b_i \quad (2.11)$$

em que β_0 representa o intercepto e f_j denota as funções de suavização. A variável Y_{ij} assume valor 1 no cenário epidêmico e 0 no endêmico. O vetor de covariáveis X compreende as variáveis climáticas, temperatura média, umidade relativa média e precipitação acumulada média. Além destas, o modelo inclui a informação categórica dos *clusters* identificados, e a semana epidemiológica (SE). A inclusão dos *clusters* como efeito fixo permite que o modelo capture as diferentes linhas de base de transmissão e os padrões sazonais específicos de cada agrupamento municipal.

A incorporação de efeitos aleatórios por município ($b_i \sim N(0, \sigma_b^2)$) é fundamental, pois permite tratar a dependência longitudinal e a heterogeneidade espacial não observada (PINHEIRO et al., 2019). Esta estrutura corrige a correlação residual entre observações repetidas na mesma unidade geográfica e captura trajetórias epidemiológicas individuais que não seriam explicadas apenas pelos preditores fixos (CURRIE; DURBAN, 2002).

A flexibilidade do ajuste f_j baseou-se em P-splines, que equilibram a aderência aos dados e a suavidade da curva através de uma penalidade de diferenças (Δ^d) sobre os coeficientes (EILERS; MARX, 1996), em que d é a penalidade de segunda ordem, o que implica que o modelo penaliza mudanças bruscas na inclinação da curva (análogo à segunda derivada). Foi utilizado também a base *Cyclic Cubic Regression Splines* para capturar a sazonalidade dos dados epidemiológicos. Esta especificação garante a continuidade biológica interanual, impondo que os valores e derivadas na semana 1 coincidam com os da semana 52, representando o ciclo ininterrupto de transmissão.

2.5.1 Estrutura da autocorrelação temporal

O tratamento da dependência temporal apresenta desafios metodológicos que variam de acordo com a natureza da variável resposta. Na abordagem binomial, a modelagem da autocorrelação apresenta limitações teóricas. Devido ao contexto de medidas repetidas, em que múltiplas observações (0 e 1) ocorrem sob a mesma combinação de tempo e espaço (semanas e municípios), a estrutura dos dados não apresenta uma formulação probabilística binomial tradicionalmente esperada. Consequentemente, a avaliação dos resíduos autocorrelacionados para modelos binários torna-se estatisticamente complexa para capturar os erros correlacionados em GAMM.

Assim, uma alternativa para incorporar a dependência temporal no modelo, ao invés de estimar diretamente a situação epidemiológica do município, é considerar a taxa de incidência de dengue por semana epidemiológica com variável resposta quantitativa. A situação endêmica pode ser obtida indiretamente avaliando a taxa predita pelo modelo em comparação ao canal endêmico histórico de cada município.

A fim de reduzir a forte assimetria da variável dependente, para essa análise, esta foi definida como o logaritmo da taxa de incidência semanal acrescido de um ($\log(\text{Taxa} + 1)$). A adoção dessa transformação permite que a variável resposta se aproxime de uma distribuição Gaussiana. Consequentemente, ao especificar a família *gaussian* com função de ligação identidade (*identity link*), o ajuste do modelo GAMM é realizado via estimação direta, evitando a necessidade do algoritmo *Penalized Quasi-Likelihood*, comumente utilizado em famílias não-gaussianas, resultando em um processo de convergência mais estável. Dessa maneira, para avaliar a autocorrelação foi considerado a seguinte estrutura abaixo:

$$Y_{ijt} = \beta_0 + \beta_1 \text{Tipo}_j + \beta_2 \text{Cluster}_i + f_1(\text{Semana}_t, \text{Tipo}_j) + f_2(\text{Temp}) + f_3(\text{Umid}) + f_4(\text{Prec}) + b_i + \epsilon_{ijt} \quad (2.12)$$

em que β_0 é o intercepto e β_1 e β_2 são os coeficientes dos efeitos fixos paramétricos associados às variáveis *Tipo* e *Cluster*. As funções de suavização não lineares são denotadas por f , sendo f_1 estimada via *splines* de regressão cúbica cíclica para capturar a sazonalidade epidemiológica específica para cada contexto epidemiológico j ($by = \text{Tipo}$). As funções f_2 , f_3 e f_4 são modeladas via *P-splines* para as covariáveis climáticas de temperatura média, umidade relativa média e precipitação acumulada média, respectivamente. O modelo também incorpora b_i , intercepto aleatório do município i , assumindo $b_i \sim \mathcal{N}(0, \sigma_b^2)$ e ϵ_{ijt} que é o termo de erro residual associado à observação do município i , no regime j e na semana t .

A construção do termo de erro residual ϵ_{ijt} baseou-se na premissa de que a dependência temporal ocorre de forma independente dentro de cada município. Para tanto, utilizou-se a interação entre o identificador do município (*ID*) e o tipo epidemiológico (*Tipo*), criando-se estruturas autorregressivas isoladas para cada série histórica.

Para verificar a ordem da dependência temporal, os resíduos do modelo foram avaliados por meio da Função de Autocorrelação (FAC). Desta forma, procedeu-se à comparação de três especificações para o termo de erro residual: (i) pressuposto de independência (sem estrutura de correlação temporal); (ii) estrutura autorregressiva de ordem 1, denotada por ARMA(1,0); e (iii) estrutura autorregressiva de ordem 3, denotada por ARMA(3,0). A seleção baseou-se na capacidade de cada estrutura em remover a autocorrelação residual.

2.6 Seleção de modelos e métricas de avaliação

Devido às limitações conhecidas dos diagnósticos residuais (*gam.check*) para desfechos binários e algoritmos Penalized Quasi-Likelihood (PQL - Quasi-Verossimilhança Penalizada), que tendem a apresentar inconsistências por conta da natureza discreta dos dados e da linearização do processo (WOOD, 2017), a validação definitiva do modelo fundamentou-se na sua capacidade de generalização.

A escolha da estrutura de preditores baseou-se no equilíbrio entre ajuste e parcimônia via critérios o Critério de Informação de Akaike (AIC) e o Critério de Informação Bayesiano (BIC). Adicionalmente, aplicou-se a Análise de Variância (ANOVA) para testar a significância estatística de cada termo (paramétrico e suavizado) incluído nos modelos, assegurando a relevância das variáveis selecionadas. Os componentes paramétricos assumem uma relação linear e constante entre o preditor e a resposta, representados por um coeficiente fixo (β) que indica o efeito global das covariáveis climáticas e dos clusters. Em contrapartida, os termos suavizados utilizam funções de base (*splines*) para modelar relações não-lineares complexas, permitindo que o modelo capture dinâmicas sazonais e efeitos de limiar que seriam omitidos

em uma abordagem estritamente linear. A performance foi avaliada no conjunto de validação externa (dados de teste referentes ao ano de 2025) através de métricas derivadas da matriz de confusão: acurácia, sensibilidade, especificidade, valores preditivos positivo (VPP) e negativo (VPN), a estatística Kappa de Cohen (κ) e a área sob a curva ROC (AUC), utilizando abordagens presentes na literatura (HOSMER; LEMESHOW; STURDIVANT, 2013).

Exclusivamente na etapa de teste, para abordagem binomial, foram avaliados diferentes pontos de corte nos valores de probabilidade estimados (0,3; 0,4; 0,5 e 0,6) para a definição do limiar de alerta. A escolha final visou o equilíbrio entre sensibilidade e a redução de falsos positivos no conjunto de dados de 2025, priorizando a utilidade prática do modelo em dados não observados.

Em uma avaliação de modelos preditivos quantitativos, a performance pode ser mensurada através do Coeficiente de Determinação (R^2) e da proporção da variância explicada (*Deviance Explained*), que indicam a capacidade do modelo em capturar a variabilidade da incidência. Complementarmente, a precisão das estimativas no conjunto de teste pode ser verificada por meio do Erro Quadrático Médio (RMSE) e do Erro Médio Absoluto (MAE), métricas que quantificam o desvio entre os valores observados e preditos.

2.7 Implementação e recursos computacionais

A implementação estatística foi realizada no software R (versão 4.4.2). Os padrões de agrupamento foram verificados mediante técnicas de validação externa, implementadas pela função `cluster.evaluation` do pacote `TSclust` (MONTERO; VILAR, 2014). Esta etapa possibilitou a comparação entre os diferentes algoritmos de partição e de dissimilaridade temporal.

Para avaliar o grau de interdependência das variáveis climáticas entre os municípios do Paraná, foi implementada uma função no ambiente R onde utilizou-se o Índice de Correlação Média Absoluta. A métrica fundamenta-se na matriz de correlação de Spearman, escolhida por sua distribuição não-linear.

A modelagem GAMM foi executada por meio do pacote `mgcv` (WOOD, 2017). Para o ajuste dos termos de efeitos aleatórios e a estimação por Máxima Verossimilhança Restrita (REML), a função `gamm` operou em interface com a biblioteca `nlme` (PINHEIRO et al., 2019). A flexibilidade do ajuste foi feita pela especificação de bases de suavização do tipo *P-splines* para os preditores contínuos, enquanto a base *Cyclic Cubic Regression Splines* foi adotada para capturar a sazonalidade intrínseca dos dados.

Na presença de autocorrelação temporal para o desfecho quantitativo, a implementação pode ser realizada por meio da mesma estrutura de modelos mistos com a função `gamm`. A dependência entre as observações temporais é tratada mediante o argumento `correlation`,

utilizando classes de correlação da biblioteca `nlme`. Especificamente, implementou-se a estrutura autorregressiva (`corARMA`), com os resíduos estratificados pelo identificador do município e pelo tipo epidemiológico (`form = ~Semana | ID/Tipo`). Essa abordagem permite que as estimativas considerem a correlação temporal.

RESULTADOS E DISCUSSÃO

A transição dos municípios paranaenses para o regime de surto/epidemia apresentou frequências anuais heterogêneas, conforme ilustrado na Figura 2 elaborada a partir da metodologia do Diagrama de Controle. O volume de cidades classificadas em situação de surto/epidemia variou entre os anos de 2016 ($n = 2$), 2019 ($n = 1$), 2020 ($n = 29$), 2022 ($n = 22$), 2023 ($n = 16$) e atingiu o ápice da série histórica em 2024 ($n = 89$). O ano de 2024 diferenciou-se dos ciclos epidêmicos de 2020 e 2022 por apresentar uma disseminação geográfica generalizada.

O aumento do número de municípios em epidemia em 2024 reflete a dependência biológica do *Aedes aegypti* em relação às variações térmicas. Este fenômeno tem sido amplamente enfatizado em diversas contribuições científicas. Como destacado por Lowe et al. (2021), o aumento da temperatura reduz o período de incubação extrínseca, acelerando o ciclo de vírus dentro do mosquito e antecipando sua capacidade de transmissão. Esse mecanismo biológico ocorreu de forma mais acentuada em 2024, ano que se consolidou como o mais quente da história brasileira (Instituto Nacional de Meteorologia, 2025a). Impulsionado por anomalias térmicas extremas e pelos efeitos do fenômeno El Niño, o cenário no Paraná combinou o calor intenso a volumes significativos de chuva na porção leste (Instituto Nacional de Meteorologia, 2025b). Essa conjuntura foi determinante para afetar a barreira climática que, historicamente, restringiam a expansão geográfica da doença no estado.

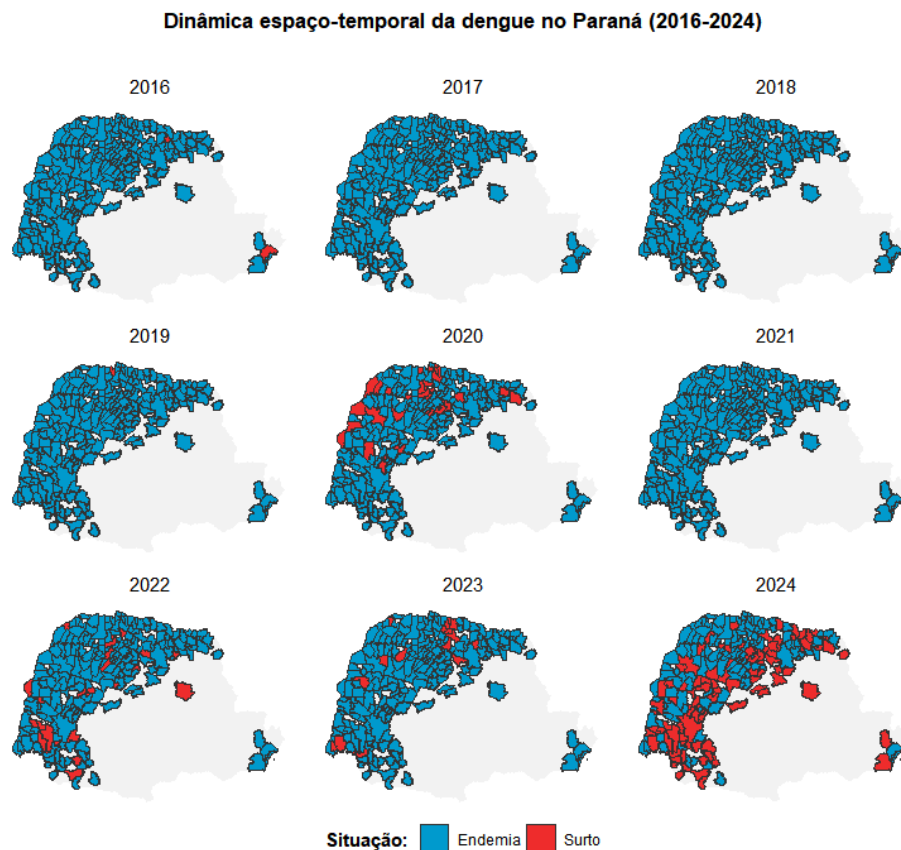


Figura 2 – Dinâmica espaço-temporal baseado no canal endêmico (diagrama de controle) da dengue nos municípios do Paraná entre 2016 e 2024.

Vale ressaltar que a caracterização apresentada na Figura 2 refere-se apenas a identificação de comportamento atípico em relação ao histórico no Diagrama de Controle e não às decisões tomadas efetivamente pelos municípios sobre suas respectivas situações de surto/epidemia, as quais são autônomas e incluem diversos fatores tais como, georreferenciamento, índice de infestação predial, positividade laboratorial, inserção de novo sorotipo, etc.

3.1 Índice Davies-Bouldin

A definição do número de agrupamentos foi fundamentada na análise do índice DB (Figura 3), buscando equilibrar parcimônia estatística e sensibilidade epidemiológica. No cenário endêmico, embora o mínimo global do índice tenha ocorrido em $k = 2$, essa configuração foi descartada por simplificar excessivamente a heterogeneidade espacial do Paraná, ocultando nuances sobre a dinâmica de transmissão. Optou-se, portanto, pelo mínimo local em $k = 5$ para identificar padrões de maior relevância prática (KAUFMAN; ROUSSEUW, 1990).

Ao observar a trajetória do índice após $k = 2$, nota-se que o valor para $k = 5$ é inferior ao observado em $k = 4$. Essa redução no índice indica que a transição para cinco grupos promoveu um ganho de compactação interna e separação entre os agrupamentos. Essa adoção

foi mantida de forma padronizada para ambos os cenários (endêmico e de surto/epidemia), a fim da necessidade de garantir a comparação direta entre os contextos ao fixar o número de agrupamentos na análise.

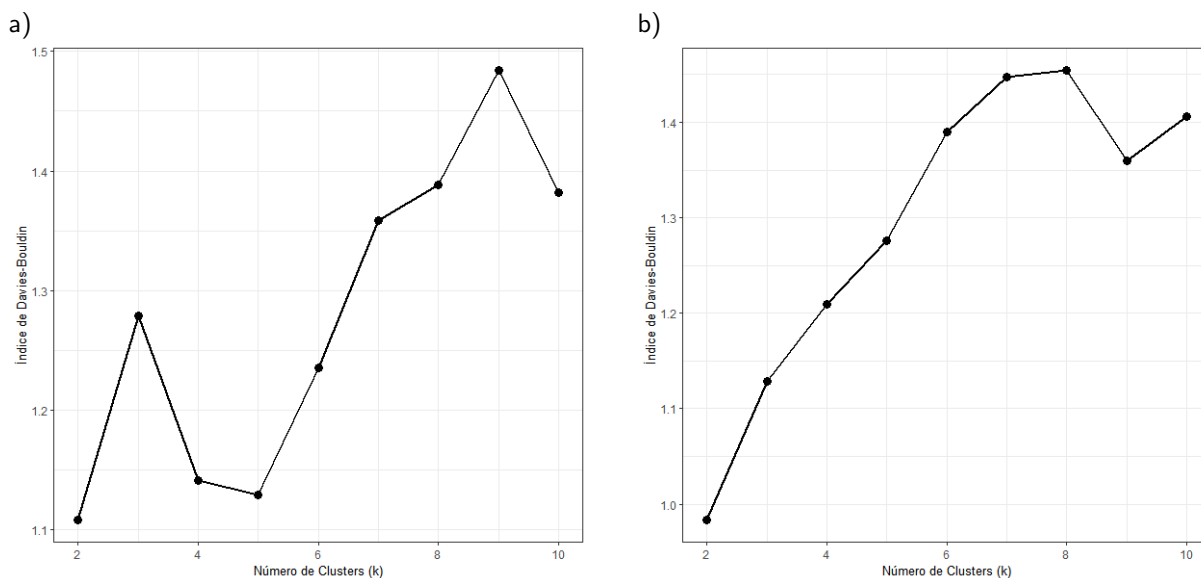


Figura 3 – Avaliação da qualidade e separação dos agrupamentos via Índice de Davies-Bouldin nos cenários: (a) endêmico e (b) surto/epidêmico.

Uma vez estabelecida a estrutura de cinco grupos, a aplicação do algoritmo de Particionamento em torno de Medoides (PAM) foi adotada como referência basal para a definição dos agrupamentos. A escolha reflete a capacidade do algoritmo em lidar com valores atípicos, frequentes em surtos isolados e pela utilização de observações reais do banco de dados (medoides) como representantes de cada *cluster*, evitando as distorções inerentes às médias aritméticas dos centroides (KAUFMAN; ROUSSEUW, 1990). Essa abordagem tem se mostrado eficaz em investigações sobre a dinâmica da dengue. Cindy et al. (2022) reforçam que a utilização do algoritmo permite a estabilidade, sendo determinante para identificar correlações entre a variáveis climáticas e a ocorrência surtos/epidemia.

A análise visual das trajetórias nas Figuras 4 e 5 revela que os agrupamentos capturam diferentes formas na dinâmica da dengue no Paraná. Em ambos os cenários, a linha vermelha destaca o medoide (perfil de referência do grupo), enquanto as linhas pretas sobrepostas exibem o comportamento individual dos municípios, permitindo avaliar a aderência das cidades ao padrão identificado. Nestas Figuras 4 e 5, ilustra-se o particionamento realizado pelo algoritmo PAM. Para os demais algoritmos, pode-se consultar o Apêndice A. Nas duas subseções seguintes, apresenta-se as análises dos diferentes métodos de agrupamentos, assim como o destaque ao desempenho do algoritmo PAM.

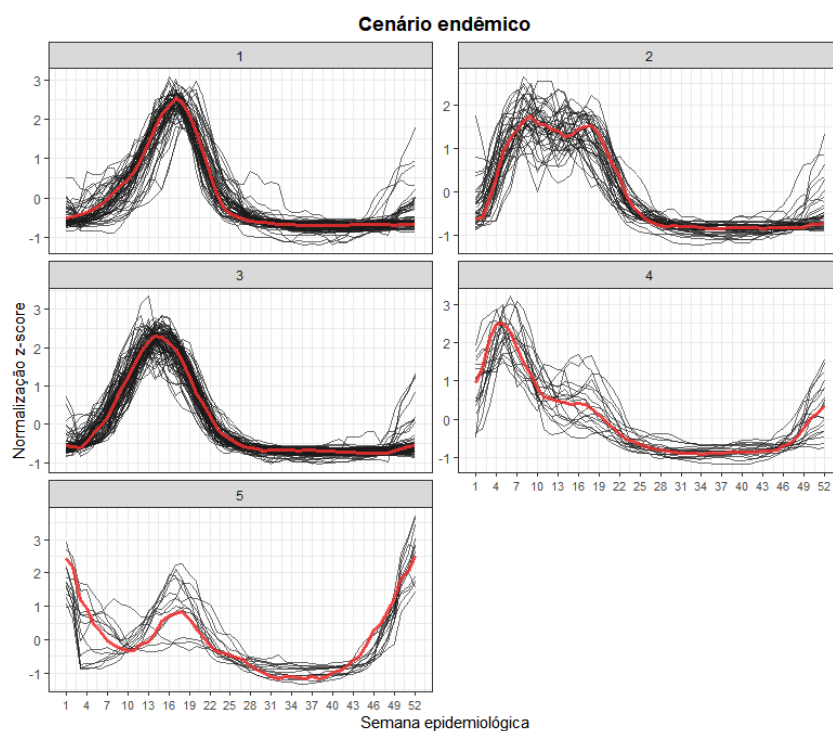


Figura 4 – Agrupamentos (*clusters*) identificados pelo algoritmo PAM ($k = 5$) para o cenário endêmico da dengue no Paraná. A cor vermelha representa o medoide da incidência do *cluster* e a cor preta identifica as cidades pertencentes ao agrupamento

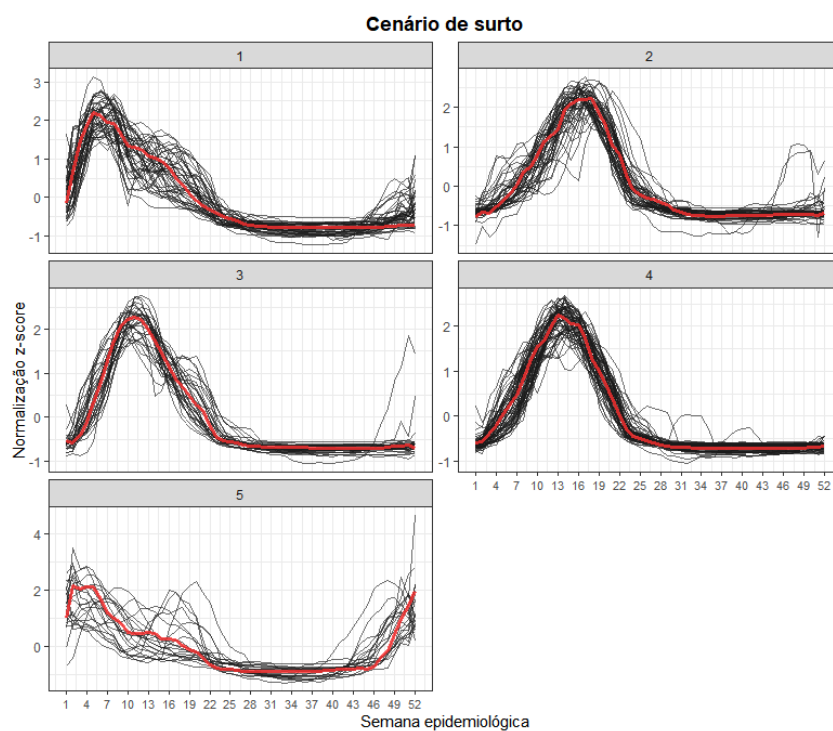


Figura 5 – Agrupamentos (*clusters*) identificados pelo algoritmo PAM ($k = 5$) para o cenário de surto/epidemia da dengue no Paraná. A cor vermelha representa o medoide da incidência do *cluster* e a cor preta identifica as cidades pertencentes ao agrupamento

No cenário endêmico (Figura 4), observa-se uma maior dispersão das trajetórias muni-

cipais em torno do medoide, refletindo dinâmicas locais mais heterogêneas. Em contrapartida, no cenário de surto/epidemia (5), embora os padrões estejam normalizados, os picos tendem a ser mais agudos.

Embora o índice DB tenha indicado 5 clusters, o leitor poderia questionar se a utilização de mais clusters poderia ser benéfica para redução da variabilidade intra-cluster. Apesar da redução ser natural com o aumento de clusters, o Apêndice A traz a ilustração dos particionamentos para k variando de 5 a 8, nos quais pode-se observar agrupamentos contendo apenas um município e ainda sem melhoria expressiva na redução da variabilidade intra-cluster.

3.1.1 Níveis de correlação entre os métodos de partição

Ao analisar os métodos que compartilham a distância euclidiana como métrica de dissimilaridade, especificamente o PAM, o K-means e o algoritmo hierárquico (HCA-Eucl), observa-se uma forte convergência de resultados na Tabela 10, no cenário endêmico.

A elevada similaridade entre o K-means e o PAM ($Sim = 0,973$) indica que, em períodos de estabilidade epidemiológica, a estrutura dos dados é bem definida. Ambos os algoritmos buscam minimizar a variância interna e organizar os municípios em agrupamentos compactos. Como o cenário endêmico apresenta séries temporais mais estáveis, a diferença entre usar um centroide (K-means) ou um medoide (PAM) torna-se mínima, resultando em classificações quase idênticas (KAUFMAN; ROUSSEUW, 1990; OSEMWENKHAE; EKHATOR; IDUSERI, 2016).

Contudo, a literatura destaca que o PAM oferece uma vantagem maior em relação ao K-means devido à sua robustez. Enquanto o K-means é sensível a ruídos e pode ter seus centroides deslocados por picos isolados de incidência, o PAM, por estar ancorado em observações reais, mantém a estabilidade do grupo mesmo diante de flutuações bruscas (VELMURUGAN; SANTHANAM, 2010). Apesar da maior complexidade computacional, o PAM demonstra superioridade na resistência a valores atípicos, sendo mais eficaz para a identificação de padrões em bases de dados complexas (NIRMAL, 2019). Isso é observado pelo recuo generalizado do índice no cenário de surto, onde o K-means declinou para 0,749.

Diferentemente dos métodos de partição (PAM e K-means), o algoritmo hierárquico (HCA) apresentou uma concordância mais reduzida ($Sim = 0,804$ na endemia e 0,603 no surto/epidemia). Essa distância nos resultados reflete a lógica interna do algoritmo: enquanto os métodos de partição realizam uma otimização global para encontrar o melhor k , o método constrói os pares mais similares de forma sucessiva e permanente. A cada iteração, os pares mais próximos são mesclados para formar grupos maiores (JAIN; DUBES, 1988).

Conforme observado por Everitt et al. (2011), uma vez que dois pontos são unidos em um degrau da hierarquia, essa decisão não pode ser desfeita, o que pode propagar erros de classificação se houver ruído nas séries temporais. Na prática, pequenas diferenças na semana

de início de transmissão entre cidades vizinhas podem atrapalhar o algoritmo hierárquico, criando ramificações isoladas que ele não pode mais desfazer que se distanciam da estrutura do métodos de particionamento, PAM e K-means, que demonstram melhor flexibilidade de ajuste.

No cenário de surto/epidemia, a métrica de Correlação Temporal (CORT) apresentou a maior estabilidade entre os métodos hierárquicos ($Sim = 0,632$), superando a distância euclidiana ($Sim = 0,603$). Essa performance decorre da capacidade da CORT em priorizar o comportamento de crescimento e declínio das curvas em detrimento da magnitude absoluta. Enquanto a abordagem euclidiana ignora a interdependência temporal, a CORT foca na sincronização das taxas de incidência e na tendência do movimento das séries (CHOUAKRIA; NAGABHUSHAN, 2007). Nesse sentido, Montero e Vilar (2014) argumentam que tal característica é superior para capturar a dinâmica de eventos súbitos, como o início sincronizado. Na prática, a métrica permite agrupar municípios que compartilham o mesmo ritmo de transmissão e semanas de pico, independentemente de aumentos no volume total de notificações.

O desempenho das demais abordagens hierárquicas revela limitações inerentes à estrutura das séries temporais da dengue. A métrica de Distância Invariante à Complexidade (CID) apresentou um $Sim = 0,667$, destacando-se pela sua capacidade de corrigir distorções causadas pela variação nas séries temporais. Conforme observado por Batista et al. (2014), a CID introduz um fator de correção à distância euclidiana para evitar que séries com comportamentos similares, mas com níveis de complexidade (ruído e picos) distintos, sejam indevidamente separadas. No entanto, observou-se uma queda de precisão no cenário de surto/epidemia ($Sim = 0,547$). Este resultado sugere que, em contexto de surto/epidemia, a volatilidade extrema da dengue torna a complexidade das séries instáveis, o que faz com que o fator de correção perca estabilidade e poder discriminativo, tornando a métrica menos eficaz do que a CORT para capturar a sincronização dos picos.

Os menores níveis de similaridade foram registrados pela métrica Agrupamento por Distribuição de Permutação (PDC), com valores de Sim inferiores a 0,35. Embora Brandmaier (2015) descreva o PDC como uma abordagem resistente a valores atípicos e independente da inclinação da curva, a sua aplicação em dados epidemiológicos de curto prazo enfrenta limitações estruturais. O método fundamenta-se na análise da distribuição de permutação das séries para identificar a dinâmica do sistema. Contudo, a curta extensão das séries anuais (52 semanas) e o excesso de semanas com zero casos em municípios de menor porte possivelmente impedem que o algoritmo extraia uma estatística confiável.

As métricas baseadas em Autocorrelação (ACF) e Periodogramas (PER) mantiveram valores intermediários de concordância na Tabela 10. É possível notar que PER apresentou um aumento de similaridade durante a epidemia ($Sim = 0,606$), o que se explica pela imposição de uma frequência anual muito forte e sincronizada pelo surto/epidemia, característica facilmente detectada por métodos espectrais. Conforme mencionado por Caiado, Maharaj e

D'Urso (2006), embora as métricas baseadas no domínio da frequência sejam eficazes para identificar o processo e seus padrões, elas tendem a ser insensíveis a deslocamentos temporais e à magnitude absoluta. No contexto da dengue, essa falha em capturar o pico da dengue torna a utilidade dessas métricas limitada para o presente estudo.

Adicionalmente, observa-se uma tendência de declínio nos índices de similaridade à medida que o número de agrupamentos (k) é ampliado para 6, 7 e 8, conforme detalhado na Tabela 10. Essa redução nos valores de *Sim* indica que a partição em 5 clusters representa o ponto de equilíbrio ideal, corroborando com os resultados apresentados na Seção 3.1. O aumento da fragmentação em níveis superiores de k resultou em grupos excessivamente pequenos e com baixa coesão, perdendo a capacidade de refletir padrões climáticos-epidemiológicos robustos.

Tabela 2 – Comparação da similaridade (*Sim*) entre o método basal PAM e demais métricas para $k = 5$.

Algoritmo / Métrica	<i>Sim</i> (Endemia)	<i>Sim</i> (surto/epidemia)
<i>Algoritmo de Partição</i>		
K-means (Euclidiana)	0,973	0,749
<i>Algoritmos Hierárquicos</i>		
ACF	0,581	0,502
CID	0,667	0,547
CORT	0,660	0,632
HCA-EUCL	0,804	0,603
PDC	0,294	0,318
PER	0,533	0,606
PIC	0,294	0,339

Nota: *Sim*: Clustering Evaluation Index Based on Known Ground Truth baseado em Gavrilov et al. (2000). A análise completa, abrangendo a variação de $k = 5$ a $k = 8$, encontra-se detalhada na Tabela 10 do Apêndice A.1.5.

A Tabela 3 detalha a distribuição dos municípios por agrupamento. Observa-se que, no cenário endêmico, as métricas CID, CORT, HCA-EUCL e PER tendem à formação de mega-agrupamentos, concentrando a grande maioria das cidades em um único grupo. Já no cenário de surto/epidemia, as métricas CID, PER e PIC resultam em clusters unitários ($n = 1$), o que impede a identificação de padrões coletivos. Observa-se ainda nas Figuras 17 e 18 (CID e PER), por exemplo, uma baixa coesão interna nos clusters identificados. Devido a essas limitações, avançaram para a etapa seguinte apenas os métodos que apresentaram mais de um município por cluster.

Tabela 3 – Distribuição dos municípios por cluster nos cenários endêmico e de surto/epidemia.

Método	C1	C2	C3	C4	C5
<i>Cenário endêmico</i>					
PAM	61	40	84	19	14
ACF	96	42	60	13	7
CID	144	43	6	13	12
CORT	148	23	15	26	6
HCA-EUCL	132	39	18	16	13
PDC	94	39	13	57	15
PER	149	42	14	12	1
PIC	74	108	26	9	1
<i>Cenário de surto/epidemia</i>					
PAM	46	50	33	67	22
ACF	74	92	41	4	7
CID	34	32	149	2	1
CORT	63	5	60	80	10
HCA-EUCL	41	19	17	129	12
PDC	17	45	55	62	39
PER	62	87	63	5	1
PIC	8	163	42	4	1

3.1.2 Níveis de correlação intracluster

A coesão interna dos clusters foi avaliada quanto a homogeneidade das características climáticas validando se os municípios integrados ao mesmo cluster compartilham, de fato, perfis climáticos semelhantes (temperatura, precipitação e umidade). De acordo com Hastie, Tibshirani e Friedman (2009), o Erro Quadrático Médio (MSE) pode ser decomposto em componentes de viés ao quadrado, variância e um erro irreduzível. Consequentemente, a sua raiz quadrada, o RMSE, agrega em uma única métrica a informação sobre a acurácia (baixo viés) e a precisão (baixa variância) de um modelo, permitindo identificar o método que melhor minimiza as discrepâncias intracluster (Tabela 4).

Os resultados indicam alta consistência interna em todos os cenários analisados, com a temperatura destacando-se como a variável de maior estabilidade tanto na endemia quanto no surto/epidemia. No cenário de endemia, os métodos ACF e PAM apresentaram as maiores correlações médias ($\rho = 0,9749$ e $\rho = 0,9745$, respectivamente), com o ACF registrando o RMSE mais baixo (1,0596), seguido do HCA-EUCL (1,0831) e PAM (1,0874). Enquanto no cenário de surto/epidemia observou-se um ganho de sincronia em todos os métodos, com a métrica CORT atingindo o melhor ajuste ($\rho = 0,9843$ e $RMSE = 1,0123$). Esse comportamento indica que as transições térmicas tornam-se ainda mais sincronizadas durante anos de surto/epidemia, facilitando a identificação de padrões climáticos comuns entre os agrupamentos.

Quando se trata da variável de precipitação, é importante notar que ela apresenta maior variabilidade em relação à temperatura. É possível notar que a correlação média tam-

bém aumentou em todos os métodos de partição, acompanhada por uma redução do RMSE no cenário de surto/epidemia em relação à endemia. Isso sugere que, nos anos em que ocorrem surto/epidemia da doença, as chuvas tendem a ser mais regulares e consistentes em determinadas áreas, o que reduz as diferenças locais entre municípios do mesmo cluster. Em especial, o método PAM obteve o menor RMSE (6,4631) na endemia, apresentando-se como o mais equilibrado para esse período; já a métrica CORT destacou-se no surto/epidemia, alcançando a maior correlação média ($\rho = 0,9099$) e o menor RMSE (5,5594).

Além disso, a umidade também manteve a consistência e estabilidade, com os métodos apresentando correlações médias superiores a 0,9. No período da endemia, o método ACF registrou o melhor RMSE, seguido pelo PAM (3,3418 e 3,4909, respectivamente). No cenário de surto/epidemia a CORT sobressaiu-se novamente ao alcançar a maior correlação ($\rho = 0,9361$) e o menor RMSE (3,2260) entre os métodos de partição, capturando com maior eficácia a dinâmica de umidade dentro dos agrupamentos durante a instabilidade.

Tabela 4 – Métricas de desempenho das variáveis climáticas por cenário epidemiológico intra-cluster.

Variável	Cenário	Método	Correlação Média (ρ)	DP	EQM	RMSE
Temperatura	Endemia	PAM	0,9745	0,0069	1,1824	1,0874
		HCA-EUCL	0,9726	0,0099	1,1732	1,0831
		CORT	0,9695	0,0119	1,3130	1,1459
		PDC	0,9714	0,0026	1,4354	1,1981
		ACF	0,9749	0,0047	1,1228	1,0596
	Surto/epidemia	PAM	0,9806	0,0057	1,2534	1,1196
		HCA-EUCL	0,9811	0,0044	1,2891	1,1354
		CORT	0,9843	0,0063	1,0248	1,0123
		PDC	0,9796	0,0028	1,3081	1,1437
		ACF	0,9827	0,0044	1,1619	1,0779
Precipitação	Endemia	PAM	0,8746	0,0335	41,7713	6,4631
		HCA-EUCL	0,8748	0,0335	46,2643	6,8018
		CORT	0,8739	0,0349	52,6181	7,2538
		PDC	0,8576	0,0076	46,7945	6,8406
		ACF	0,8664	0,0250	43,9235	6,6275
	Surto/epidemia	PAM	0,8844	0,0242	41,2977	6,4263
		HCA-EUCL	0,8996	0,0282	38,0542	6,1688
		CORT	0,9099	0,0471	30,9074	5,5594
		PDC	0,8741	0,0123	43,7484	6,6143
		ACF	0,9035	0,0425	32,4396	5,6956
Umidade	Endemia	PAM	0,9191	0,0497	12,1866	3,4909
		HCA-EUCL	0,9070	0,0704	14,5569	3,8153
		CORT	0,8995	0,0728	18,3851	4,2878
		PDC	0,9078	0,0371	13,5074	3,6752
		ACF	0,9184	0,0463	11,1676	3,3418
	Surto/epidemia	PAM	0,9118	0,0331	13,9617	3,7365
		HCA-EUCL	0,9253	0,0265	14,4873	3,8062
		CORT	0,9361	0,0460	10,4068	3,2260
		PDC	0,9070	0,0294	14,3914	3,7936
		ACF	0,9342	0,0433	10,8944	3,3007

Nota: Coeficiente de correlação de Spearman (ρ); DP: Desvio Padrão; EQM: Erro Quadrático Médio; RMSE: Raiz do Erro Quadrático Médio.

A avaliação qualitativa das Figuras 14 e 16, apresentada nos Apêndice A, revelou uma baixa coesão interna nos clusters identificados a partir dos algoritmos baseados em auto-correlação (ACF) e distribuição de permutação (PDC). Nestas abordagens, observou-se uma fragmentação excessiva das séries temporais, demonstrando que tais métricas foram incapazes de isolar padrões meteorológicos distintos de forma clara em comparação com as demais métricas. Assim, os métodos ACF e PDC foram desconsiderados para seguir para a próxima etapa de modelagem.

3.2 Modelos Mistos Aditivos Generalizados (GAMM)

3.2.1 Análise do desempenho dos modelos no treinamento

A modelagem preditiva de surto/epidemia de dengue construída através do GAMM, permitiu a integração de componentes fixos, como os clusters e as variáveis climáticas, com efeitos aleatórios para os municípios a fim de absorver as características específicas de cada município. As especificações detalhadas dos modelos selecionados para cada métrica de dissimilaridade, focando na flexibilidade conferida pelas bases P-splines, estão apresentadas na Tabela 5.

Tabela 5 – Resumo das especificações dos modelos GAMM selecionados.

Distância Base	Ajuste da Taxa	Sazonalidade	Variáveis Climáticas
PAM	$f_{ps}(\text{Taxa}_{j-1})$	$f_{cc}(\text{Semana})$	Componentes Lineares
CORT	$f_{ps}(\text{Taxa}_{j-1})$	$f_{cc}(\text{Semana})$	Componentes Lineares
HCA-EUCL	$f_{ps}(\text{Taxa}_{j-1})$	$f_{cc}(\text{Semana})$	Umidade: f_{ps} ; Outras: Lineares

Nota: f_{ps} : Suavização por P-splines; f_{cc} : Spline cúbico cíclico. Todos os modelos incluem o intercepto aleatório municipal (b_i).

Na avaliação dos modelos durante a fase de treinamento comparou-se a performance de três diferentes partições de agrupamento sobre o mesmo conjunto de dados. Conforme a Tabela 6, o modelo baseado na distância PAM apresentou os menores valores de AIC (102.672,8) e BIC (102.769,1), indicando um ajuste com maior parcimônia. No entanto, nos demais indicadores de desempenho, a métrica PAM demonstrou limitações, apresentando resultados inferiores aos observados para HCA-EUCL e CORT. A performance inferior de PAM frente às demais métricas sugere que a técnica de K-medoides no treino tendeu a suavização excessiva das variações entre municípios.

De modo geral, o método HCA-EUCL apresentou um desempenho melhor, liderando a maioria dos indicadores de avaliação, como AUC (0,8332), Acurácia (0,7688) e Kappa (0,5377). Esse resultado indica que o agrupamento hierárquico com distância euclidiana permitiu uma classificação mais equilibrada e precisa das semanas epidemiológicas. Essa tendência é pontualmente alterada pela distância CORT, que se mostrou superior nos critérios de Sensibilidade (0,6949) e Valor Preditivo Negativo (0,7219), revelando uma capacidade maior em identificar corretamente a presença dos eventos epidêmicos. Enquanto o HCA-EUCL oferece um modelo mais preciso e com menos alarmes falsos, o CORT prevê uma maior detecção de surtos/epidemias reais, garantindo que o sistema não falhe em identificar o início de uma crise.

Tabela 6 – Métricas de desempenho dos modelos para as distâncias PAM, HCA-EUCL e CORT, na etapa do treino

Métrica	Métodos		
	PAM	HCA-EUCL	CORT
AIC	102.672,8	113.430,1	108.495,1
BIC	102.769,1	113.534,5	108.591,5
AUC (ROC)	0,7159	0,8332	0,7886
Acurácia	0,6549	0,7688	0,7435
Kappa	0,3097	0,5377	0,4870
Sensibilidade	0,5731	0,6417	0,6949
Especificidade	0,7366	0,8960	0,7920
VPP	0,6851	0,8605	0,7697
VPN	0,6331	0,7143	0,7219
F1 Score	0,6241	0,7352	0,7304

3.2.2 Resultados da validação externa (conjunto de teste)

A etapa de validação externa buscou verificar o comportamento dos modelos de melhor desempenho em um conjunto de dados independente, simulando sua utilidade prática com dados de 155 municípios referentes ao ano de 2025, os quais são integrantes do grupo original de 218 municípios analisados entre 2016 e 2024. Esse processo possibilitou avaliar a constância do desempenho dos métodos diante de novos dados, viabilizando o teste do impacto de diferentes limiares de decisão sobre a capacidade preditiva. Os resultados que detalham a resposta do modelo frente a novos dados são apresentados na Tabela 7.

A distância HCA-EUCL confirmou sua consistência com uma AUC de 0,7835, destacando-se pela alta especificidade e no controle de alarmes falsos em todos os limiares testados. Por outro lado, o PAM destacou-se pelo maior efetividade para a tomada de decisão. Esse método, no limiar 0,4, atingiu maiores índices de acurácia (0,664) e concordância ($Kappa = 0,330$) de toda a etapa de teste, consolidando-se como a abordagem de melhor desempenho global para o monitoramento epidemiológico.

Esse comportamento ilustra o desafio de equilibrar a precisão do alertas na epidemiologia, em que a busca pela detecção precoce por meio de limiares mais baixos (0,3) costuma ocorrer em detrimento da precisão dos alertas, que é assegurada apenas por níveis mais conservadores (0,5). No contexto atual do Paraná, onde a dengue avança sobre regiões historicamente menos propensas a endemia em virtude de transformações climáticas e ecológicas (BAVIA et al., 2020), o método PAM com distância euclidiana apresentou-se como uma abordagem mais equilibrada. Ao adotar o limiar moderado de 40%, a modelagem assegura melhor eficiência para alocação de recursos, alinhando-se às necessidades práticas de alerta precoce fundamentados em modelagem probabilística (LOWE et al., 2021). Essa configuração permite que o gestor atue com maior segurança e fundamentada na realidade epidemiológica local.

Tabela 7 – Métricas detalhadas para o modelo de teste (2025) utilizando suavização P-splines.

Métrica	Limiares			
	0.3	0.4	0.5	0.6
K-medoides - Euclidiana (PAM)				
AUC (ROC)	0,6873			
Acurácia	0,580	0,664	0,593	0,535
Kappa	0,161	0,330	0,189	0,074
Sensibilidade	0,589	0,551	0,333	0,205
Especificidade	0,571	0,779	0,857	0,870
VPP	0,582	0,716	0,702	0,615
VPN	0,578	0,631	0,559	0,519
F1 Score	0,586	0,623	0,452	0,307
Hierárquico - Euclidiana (HCA-EUCL)				
AUC (ROC)	0,7835			
Acurácia	0,600	0,567	0,554	0,529
Kappa	0,203	0,139	0,114	0,063
Sensibilidade	0,294	0,192	0,153	0,102
Especificidade	0,909	0,948	0,961	0,961
VPP	0,766	0,789	0,800	0,727
VPN	0,560	0,536	0,528	0,513
F1 Score	0,425	0,309	0,258	0,179
Hierárquico - Correlação (CORT)				
AUC (ROC)	0,6527			
Acurácia	0,567	0,516	0,509	0,503
Kappa	0,138	0,035	0,023	0,010
Sensibilidade	0,346	0,230	0,217	0,205
Especificidade	0,792	0,805	0,805	0,805
VPP	0,627	0,545	0,531	0,516
VPN	0,544	0,508	0,504	0,500
F1 Score	0,446	0,324	0,309	0,293

3.2.3 Análise das estimativas de risco do modelo final

Os resultados do modelo GAMM utilizando a técnica de suavização P-splines são apresentados na Tabela 8. Em relação aos coeficientes paramétricos, todos os agrupamentos espaciais mostraram-se estatisticamente significativos ($p < 0,001$).

O Cluster 2 apresentou um efeito positivo moderado, municípios desse cluster possuem uma chance de surto/epidemia 41% maior ($OR = 1,41$) em comparação aos municípios situados no nível de referência (Cluster 1). As diferenças na distribuição da doença ficam evidentes nos Cluster 4 e 5, que exibiram efeitos de maior vulnerabilidade do estado. Municípios situados nos Cluster 4 tem chance de enfrentar um surto/epidemia 10,6 vezes maior ($OR = 10,6$) do que no grupo de referência, enquanto no Cluster 5, esse risco é de 16,3 vezes maior ($OR = 16,30$). A concentração de risco representadas por esses agrupamentos corrobora com a consolidação da dengue no estado paranaense (FREIRE et al., 2023; BARCELLOS et al., 2024). Enquanto, o Cluster 3 foi o único com perfil redutor do risco ($OR = 0,24$), sugerindo que fatores locais ou microclimas possam agir como barreiras à transmissão nessas áreas, como discutido por Pereda, Valverde e Silva (2014).

Dentre as variáveis meteorológicas, a umidade média apresentou uma relação inversa com o risco de dengue ($p = 0,004$), ou seja, quanto maior a umidade, menor a probabilidade de surto/epidemia. Esse resultado, é salientado por Lowe et al. (2021), que explicam que a umidade extrema muitas vezes está ligada a chuvas intensas que lavam os criadouros, eliminando as larvas, além de dificultarem o voo e a atividade do mosquito no ambiente urbano.

Em relação a temperatura e a precipitação, essas apresentaram coeficientes positivos no limiar da significância ($p = 0,089$ e $p = 0,052$, respectivamente). De acordo com Assunção, Rodrigues et al. (2025), o aumento da temperatura acelera o ciclo de vida do *Aedes aegypti* e a replicação viral, mas esse efeito depende de um período específico do ano. Esse comportamento pode ser explicado pela biologia do vetor, que não responde ao clima de forma linear, mas através de janelas específicas de desenvolvimento ou que sua influência foi absorvida pelos termos espaciais e sazonais, conforme observado em estudos de modelagem espacial no Paraná (LEANDRO; CICHELO et al., 2022; ALMEIDA; RUDKE et al., 2021).

A análise dos termos suavizados indicou que tanto a taxa de incidência ($s(\text{Taxa})$) quanto a sazonalidade ($s(\text{Semana})$) exercem efeitos não lineares significativos sobre a dinâmica da doença ($p < 0,001$). Isso evidencia que o risco não evolui de forma constante, mas responde a ciclos biológicos e temporais complexos, onde a repetição de padrões anuais isolados é insuficiente para capturar a real variabilidade dos surtos/epidemias (VILLELA, 2024).

Tabela 8 – Estimativas dos coeficientes paramétricos, Razão de Chances (OR) e termos suavizados do modelo final.

Coeficientes Paramétricos					
Efeito	Estimativa (β)	Valor t	OR	IC 95% (OR)	Valor- p
Intercepto	-0,0439	-0,045	0,957	[0,14 – 6,41]	0,964
Cluster 2	0,3415	6,409	1,407	[1,26 – 1,56]	< 0,001
Cluster 3	-1,4324	-22,327	0,238	[0,21 – 0,27]	< 0,001
Cluster 4	2,3615	35,193	10,60	[9,30 – 12,10]	< 0,001
Cluster 5	2,7914	26,577	16,30	[13,27 – 20,03]	< 0,001
Temperatura	0,0467	1,701	1,048	[0,99 – 1,10]	0,089
Precipitação	0,0036	1,940	1,003	[0,99 – 1,01]	0,052
Umidade	-0,0207	-2,852	0,979	[0,96 – 0,99]	0,004
Significância dos termos suaves					
Efeito	edf	Ref.gl	Estatística F	Valor- p	
$s(\text{Taxa})$	5,335	5,335	178,2	< 0,001	
$s(\text{Semana})$	7,500	8,000	59,7	< 0,001	
n = 22.568					

Nota: OR: Razão de Chances (*Odds Ratio*); IC 95%: Intervalo de Confiança; edf: Graus de liberdade efetivos; Ref.gl: Graus de liberdade de referência.

A interpretação da dinâmica não linear, é detalhada pelos gráficos de efeitos parciais (Figura 6). O gráfico à esquerda, referente à Taxa de Incidência, demonstra que o risco de transição para um surto/epidemia é sensível em níveis iniciais de transmissão. Observa-se que

a probabilidade de surto/epidemia sobe rápido até atingir 400 casos por 100 mil habitantes, sugerindo um efeito positivo onde o acúmulo de casos acelera a probabilidade de descontrole. Após esse limiar, a curva atinge uma estabilização, indicando que o estado de surto/epidemia já se consolidou no Paraná.

No gráfico à direita, a componente sazonal revela a janela de vulnerabilidade temporal do estado. O ponto mínimo observado entre as semanas 10 e 20 indica um período de estabilidade endêmica. Nestas semanas, que compreendem a transição para meses mais frios, a transmissão se torna previsível e mantida dentro da normalidade, resultando na menor probabilidade de evolução para surto/epidemia.

A alta adaptabilidade do *Aedes aegypti* é um grande desafio para o controle da doença. A capacidade de seus ovos sobreviverem por até 450 dias em ambientes secos, permanecendo viáveis por mais de um ano (FUNDAÇÃO OSWALDO CRUZ, 2016), pode ser observada pela componente sazonal do modelo a medida que a probabilidade de surto/epidemia aumenta a partir da 25 semana e se consolida na 40 semana. Com o retorno do clima quente e chuvoso, pode ocorrer a eclosão desses ovos depositados anteriormente, impactando diretamente na incidência da dengue nos próximos meses (RIBEIRO et al., 2006; STEWART-IBARRA et al., 2013).

Adicionalmente, a transmissão transovariana, onde as larvas já nascem infectadas com o vírus, reforça a necessidade de ações de controle vetorial focadas na remoção manual de criadouros durante os meses de baixa transmissão. Sem a eliminação desses ovos durante o inverno, o sistema de saúde permanece vulnerável, onde a transição ambiental para o verão pode culminar em novos surtos (MOTTA, 2016; ASSUNÇÃO; RODRIGUES et al., 2025).

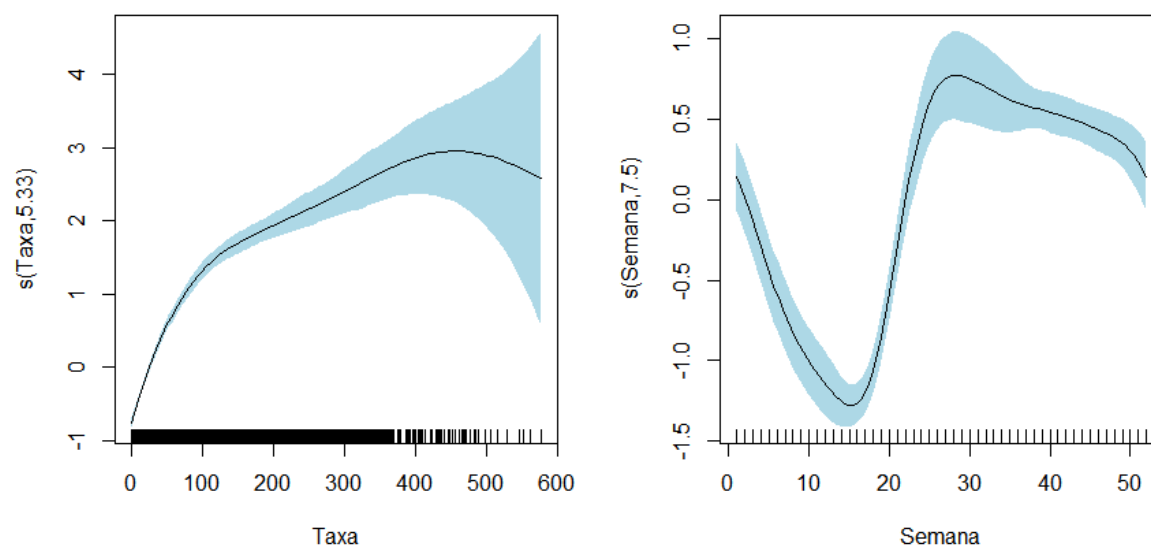


Figura 6 – Efeito suavizado da Taxa de Incidência e componente sazonal sobre a probabilidade de surto/epidemia de dengue.

A avaliação final do modelo, estruturada através do algoritmo PAM com limiar de 0,4 e técnica de suavização por P-splines, consolida sua viabilidade como ferramenta de suporte à vigilância epidemiológica no estado do Paraná. Os resultados extraídos da matriz de confusão revelam uma acurácia de 66,45%. O índice Kappa de 0,33 indica uma concordância mínima segundo McHugh (2012), classificação frequente em fenômenos biológicos instáveis e com dados desbalanceados. Conforme discutido por Sim e Wright (2005), a estatística Kappa é sensível à prevalência do evento, tendendo a ser subestimada em séries temporais onde os períodos de endemia predominam significativamente sobre os surtos/epidemias.

O modelo apresenta um equilíbrio relevante entre suas métricas de classificação. A especificidade de 77,92% demonstra eficiência em identificar corretamente períodos de endemia, minimizando alarmes falsos. Ao mesmo tempo, a sensibilidade de 55,13% é compensada por um Valor Preditivo Positivo de 71,67%, garantindo que a maioria dos alertas emitidos corresponda, de fato, a surtos reais. Complementarmente, o Valor Preditivo Negativo de 63,16% reforça a confiabilidade do descarte de riscos em municípios com estabilidade epidemiológica.

Quanto ao desempenho preditivo ao longo das semanas epidemiológicas para o 155 municípios (Figura 7), o modelo manteve estabilidade com acertos acima de 50% em todo o período. O desempenho atingiu seu pico na semana 14 (69,03%), demonstrando eficácia no período de maior pressão epidemiológica. A redução na precisão durante a transição climática, com o mínimo de 52,26% na semana 33, reflete os desafios impostos pela baixa densidade de casos, embora a utilidade preditiva tenha sido preservada em todas as semanas monitoradas, superando consistentemente a taxa de erro.

Apesar dessas limitações métricas, resultados nesta faixa são considerados promissores para o desenvolvimento de sistemas de alerta precoce no Brasil, possuindo alta utilidade prática para a gestão pública ao identificar janelas de risco antes da consolidação do surto/epidemia (LOWE et al., 2016). Portanto, ao alcançar uma concordância superior ao acaso, o modelo demonstra capacidade de capturar padrões reais da dinâmica da doença, superando a aleatoriedade e validando-se como uma base sólida para o monitoramento estadual.

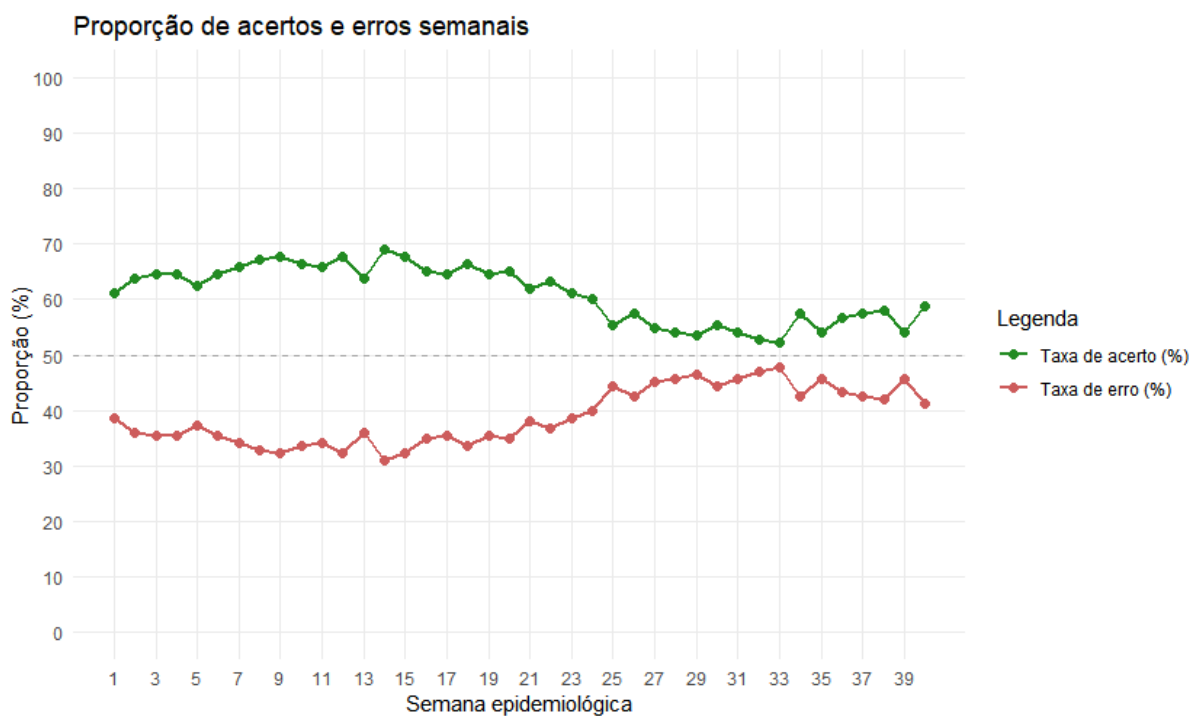


Figura 7 – Distribuição semanal da proporção de acertos e erros do modelo preditivo GAMM no conjunto de validação externa (2025)

3.2.4 Análise da interação climática sobre o risco de transmissão

Para melhorar a visualização das janelas climáticas, foi ajustado um modelo que considera a interação entre as três variáveis climáticas. Esse modelo permitiu capturar a superfície de risco tridimensional por meio do termo de interação pelo produto tensorial $te(Temperatura, Precipitação, Umidade)$ o qual apresentou significância estatística ($p < 0,001$). O termo $te()$ é construído a partir de suavizadores marginais que impõem penalidades diferentes para cada variável que sejam compatíveis com a função $s()$ (WOOD, 2017). Os mapas de contorno que mostram a interação entre temperatura, precipitação e níveis de umidade relativa (45%, 50%, 55%, 60%, 65% e 70%) são exibidos na Figura 8.

Nota-se que o risco é mais elevado em condições de baixa umidade (45% e 50%). Nesse contexto, existe uma área crítica (cor laranja) em que a combinação de temperaturas amenas (14°C a 20°C) e chuvas a cima de 40 mm na semana eleva a chance de surto/epidemia. Conforme a umidade relativa sobe para valores entre 55% e 70%, a superfície passa a ser

negativa, indicando que o aumento da umidade atua como um fator de redução de risco, mantendo as mesmas condições de temperatura. Com relação a precipitação, o risco cresce até cerca de 80 mm, indicando possivelmente que o volume excessivo de chuva possa realizar a lavagem dos criadouros conforme observado no estudo de Lowe et al. (2021).

Os resultados obtidos neste trabalho corroboram com as investigações de Barcellos et al. (2024). O estudo aponta que, historicamente, condições de baixa umidade e temperaturas reduzidas funcionavam como barreiras climáticas naturais que restringiam a dispersão da dengue no território brasileiro. No entanto, anomalias térmicas e o processo de urbanização têm ultrapassado essas barreiras, permitindo a sobrevivência do vetor e a circulação mesmo durante períodos de seca ou em regiões de maior altitude e latitude, áreas conhecidas por baixa umidade relativa, levando à rápida propagação da doença para regiões antes vistas como seguras.

Essa alteração tem gerado preocupações, especialmente frente as anomalias de temperatura (CODEÇO; VILLELA; COELHO, 2018; BARCELLOS et al., 2024). Isso ocorre porque o intervalo de geração, definido como o tempo decorrido entre a infecção de um indivíduo (caso primário) e o surgimento de novos infectados (casos secundários) a partir dele, é altamente sensível ao clima. Dessa forma, o aumento da temperatura diminui o Período de Incubação Extrínseco (PIE), o que acelera o desenvolvimento do vírus no vetor e, como resultado, aumenta a rapidez com que a doença se espalha. Por outro lado, climas frios estendem esse período, diminuindo a eficácia da transmissão e a conectividade entre novos casos.

Além do aspecto biológico, para a interpretação dos resultados é importante ressaltar que os dados epidemiológicos referem-se a casos confirmados da incidência da dengue. Portanto, é necessário considerar o tempo de processamento entre a ocorrência do evento e sua inserção nos sistemas de vigilância. Esse intervalo temporal tende a influenciar a associação entre as variáveis climáticas e os registros oficiais.

O estudo de Mordecai et al. (2017) reforça essa complexidade temporal ao apresentar essa dinâmica por meio da janela térmica diferenciada para cada fase biológica do vetor. Enquanto as taxas de desenvolvimento e a PIE são aceleradas em temperaturas elevadas, atingindo picos acima de 30°C, a longevidade dos adultos e a sobrevivência dos estágios imaturos apresentam declínios após os 25°C.

Complementarmente, Laureano (2025) verificou que os estágios imaturos do vetor são predominantemente influenciados por fatores biológicos relacionados à temperatura e precipitação. Contudo, na fase adulta, observa-se uma maior adaptação às mudanças climáticas, resultando em uma relação menos direta entre os indicadores de captura, como o IPA (Índice de Positividade de Armadilhas) e o IDA (Índice de Densidade de Adultos). Em condições de alta temperatura e baixa umidade, os adultos procuram abrigo em microclimas sombreados, diminuindo sua atividade de voo e a probabilidade de captura em armadilhas, o que pode ocultar a verdadeira densidade populacional. É no estágio adulto que as fêmeas praticam o

hematofagismo para a maturação dos ovos, tornando-se as responsáveis diretas pela transmissão do vírus. Portanto, o risco de epidemia em períodos de baixa umidade apresentado nesse trabalho, reforça a importância de modelos que considerem a complexidade das interações climáticas e o comportamento do mosquito em diferentes estágios.

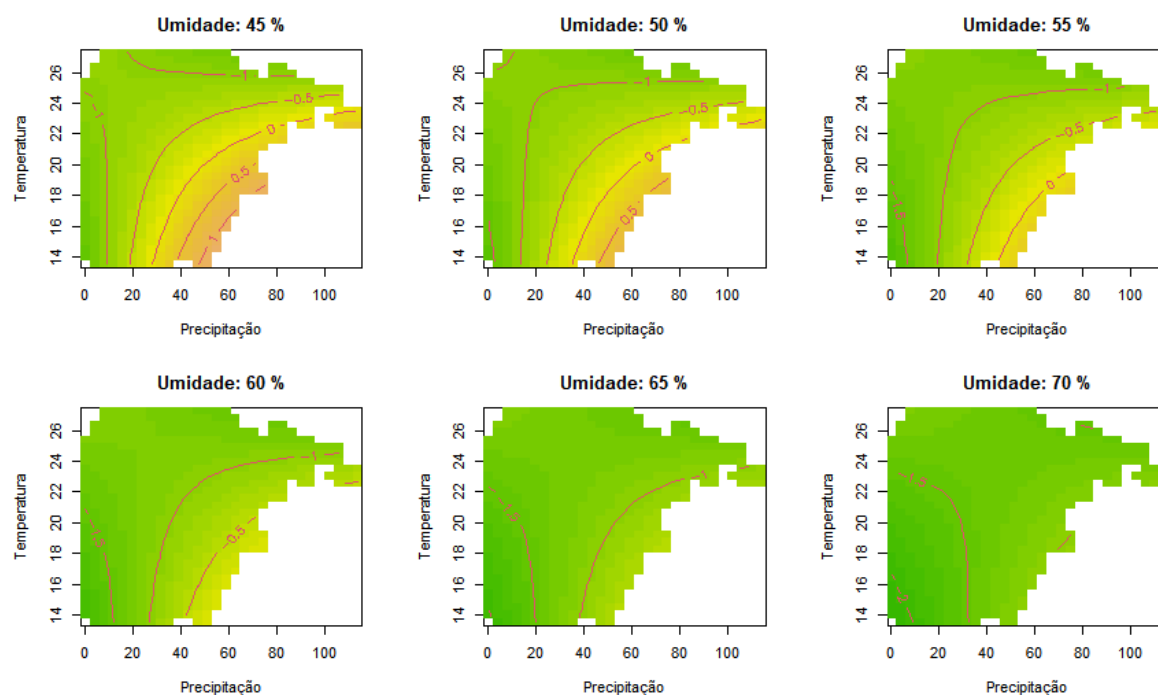


Figura 8 – Mapas de contorno do risco de surto/epidemia de dengue em relação à temperatura e precipitação, considerando níveis de umidade fixa (45%, 50%, 55%, 60%, 65% e 70%). As cores refletem a contribuição do termo de interação para o risco: verde indica áreas de menor risco com linhas de contorno negativa, enquanto que a cor laranja indica maior risco de surto/epidemia com linhas de contorno positiva. Áreas em branco indicam que não há dados suficientes para fazer observações.

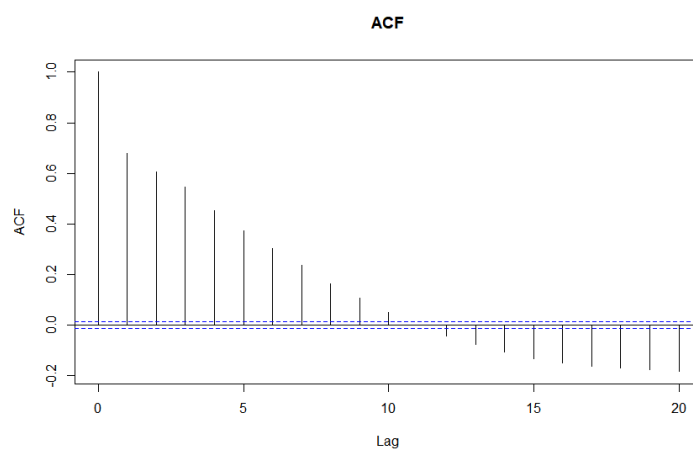
3.2.5 Análise da dependência temporal para o modelo PAM quantitativo

A análise da dependência temporal, iniciou-se pela avaliação dos resíduos do modelo quantitativo sob o pressuposto de independência dos erros. A Figura 9 apresenta a Função de Autocorrelação (FAC) dos resíduos para as três estruturas avaliadas, em que o gráfico (a) representa o pressuposto de independência, sem estrutura de correlação temporal; o gráfico (b) exibe a estrutura autorregressiva de ordem 1; e o gráfico (c) a estrutura autorregressiva de ordem 3.

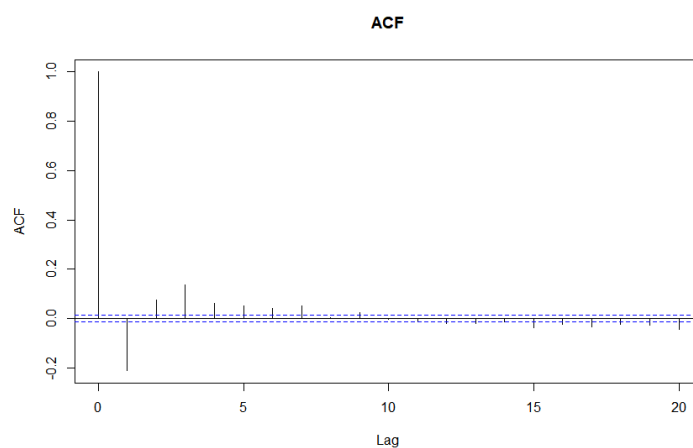
No cenário inicial, sem tratamento de autocorrelação (Figura 9a), observa-se uma violação do pressuposto de independência, com autocorrelações significativas indicando uma forte dependência temporal.

Ao aplicar a estrutura autorregressiva de ordem 1 (Figura 9b), houve uma redução das correlações, porém ainda persistem correlações de curto prazo, demonstrando que a ordem 1 foi insuficiente para capturar a dinâmica completa da série. Somente com a implementação

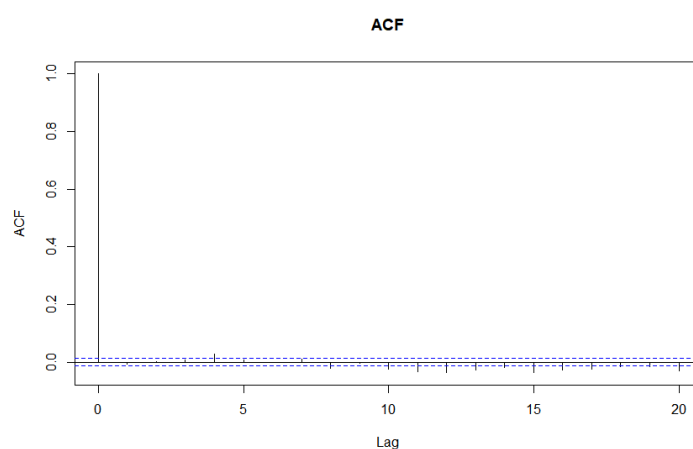
da estrutura ARMA(3,0) (Figura 9c) foi possível capturar adequadamente a autocorrelação temporal.



(a) Modelo Independente



(b) Modelo ARMA(1,0)



(c) Modelo ARMA(3,0)

Figura 9 – Comparação da Função de Autocorrelação (FAC) dos resíduos para diferentes estruturas de erro: (a) modelo sem tratamento de dependência temporal; (b) estrutura autorregressiva de ordem 1; e (c) estrutura autorregressiva de ordem 3.

Para confirmar a melhor estrutura visualmente identificada, realizou-se a comparação dos modelos via Teste de Razão de Verossimilhança, e comparação do AIC E BIC (Tabela 9). Os resultados, corroboram com a análise gráfica, evidenciando que o modelo com estrutura ARMA(3,0) apresenta melhor ajuste aos dados. Importante ressaltar que a modelagem adequada da estrutura de correlação residual possibilita uma estimativa melhor das funções de suavização, evitando o superajuste e garantindo maior precisão na captura das tendências epidemiológicas (WOOD, 2017).

Tabela 9 – Comparação de modelos GAMM com diferentes estruturas de autocorrelação residual para incidência de dengue.

Modelo	df	AIC	BIC	logLik	L.Ratio	p-valor
M1 (Independente)	15	57.855,49	57.975,91	-28.912,74	-	-
M2 (ARMA1)	16	42.724,87	42.853,32	-21.346,43	15.132,62	< 0,0001
M3 (ARMA3)	18	40.146,76	40.291,28	-20.055,38	2.582,11	< 0,0001

3.2.6 Classificação de risco e dinâmica de transmissão e análise do modelo final

A Figura 10 mostra a intensidade da taxa de incidência da dengue por *cluster* utilizando *heatmaps*. Em ambos os cenários, há uma concentração de casos entre as semanas 1 e 20, seguida por uma atividade no final do ano (semanas 50 a 52) em determinados agrupamentos. No cenário epidêmico (Figura 10b), observa-se um aumento na intensidade da taxa, com elevações nos *clusters* 3, 4 e 5, indicando uma maior vulnerabilidade nas cidades que pertencem a esses grupos. A Tabela 11, anexada no Apêndice A, apresenta a distribuição dos municípios de acordo com a classificação epidemiológica e seus respectivos agrupamentos.

O aumento de casos durante o verão é resultado da combinação de altas temperaturas e chuvas intensas, condições que favorecem a proliferação do *Aedes aegypti* (BRASIL. Ministério da Saúde, 2025). De acordo com a Secretaria de Vigilância em Saúde e Ambiente (SVSA), esse ambiente favorável reduz o ciclo de vida do mosquito, fazendo com que ele passe da fase de ovo à fase adulta em apenas 7 a 10 dias (BRASIL. Ministério da Saúde, 2025). Para alguns agrupamentos, a semana 50 pode sinalizar o começo do surto/epidemia, quando o calor e umidade de dezembro induzem a eclosão dos ovos, resultando em uma população de vetores já infectados nas primeiras semanas do ano. Essa dinâmica corrobora com o padrão de sazonalidade apresentado por Assunção, Rodrigues et al. (2025), que indica um aumento da doença nos meses iniciais do ano, seguido por uma diminuição gradual ao longo do segundo semestre.

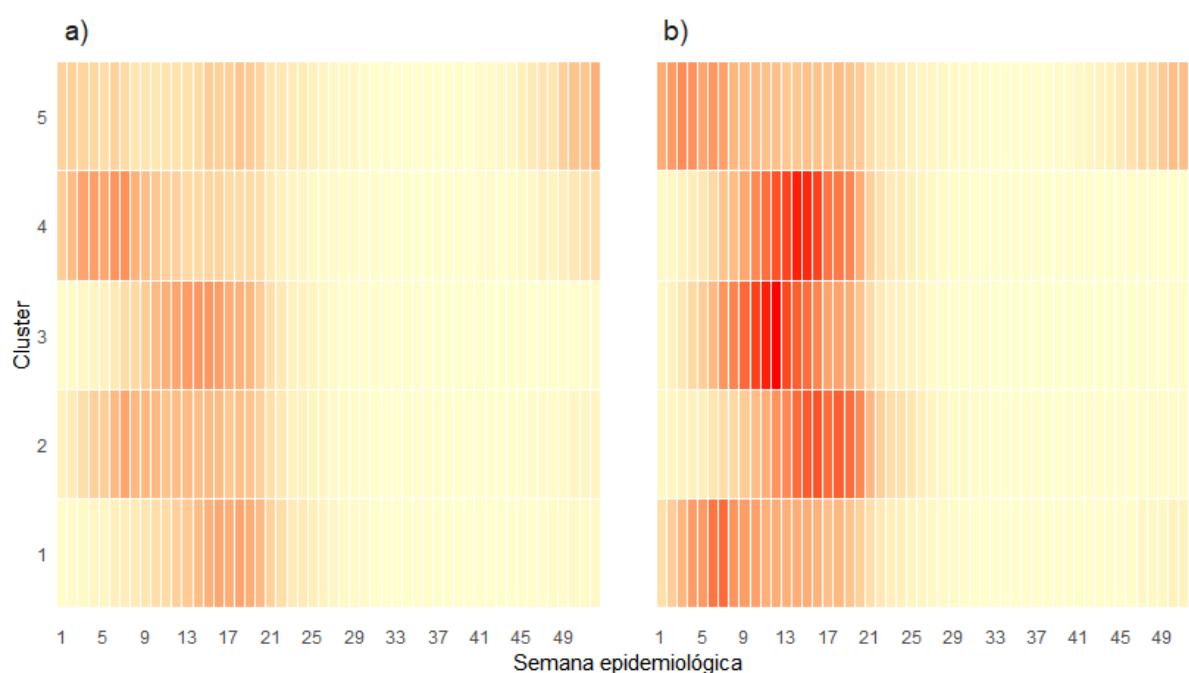


Figura 10 – Distribuição temporal da taxa média de incidência de dengue por semana epidemiológica e cluster: (a) cenário endêmico e (b) cenário de surto/epidemia.

A distribuição espacial das médias de incidência de dengue no Paraná (Figura 11) mostra uma clara heterogeneidade espacial, com um número maior de casos nas regiões norte, noroeste e oeste. Este padrão geográfico está relacionado com a classificação climática do estado (Figura 12), indicando que a intensidade da transmissão é mais intensa em regiões dominadas pelo clima subtropical úmido (Cfa). Nessas áreas marcadas, marcadas por verões quentes e chuvas concentradas, as condições podem influenciar o ciclo reprodutivo do *Aedes aegypti* e a velocidade de transmissão da doença (ABDALLAH; BELENTANI, 2011; MURTA, 2017).

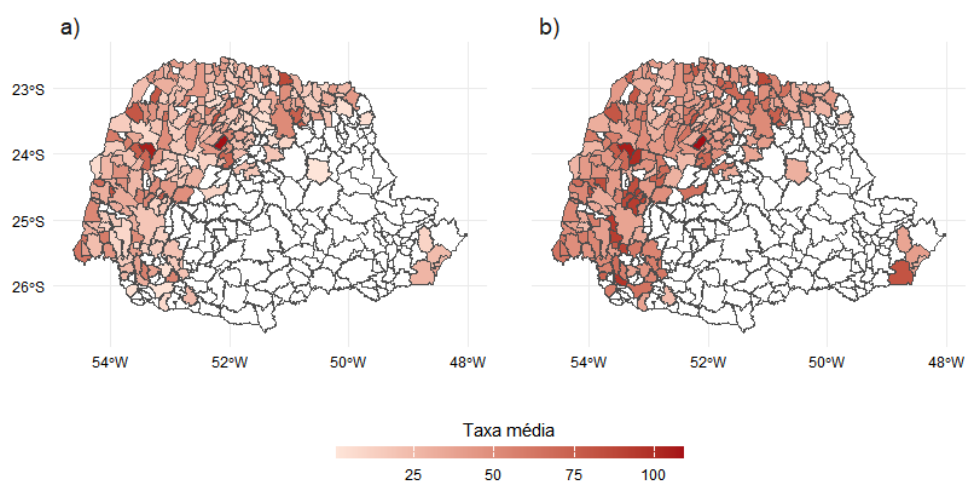


Figura 11 – Incidência média semanal de casos de dengue confirmados para o contexto endêmico (a) e epidêmico (b)

Por outro lado, as áreas afetadas pelo clima Cfb (centro-sul), que possuem verões mais suaves e menor variação de temperatura, não apresentaram casos de endemia confirmada pelo diagrama de controle. Esses achados corroboram com os estudos de Ely, Bertini e Oliveira (2012) sobre a propagação da dengue em direção a regiões que antes eram menos afetadas, em razão do aumento progressivo da temperatura. O vetor se desenvolve melhor em temperaturas entre 17°C e 27°C, combinadas com chuvas intermitentes, enquanto temperaturas mais amenas tendem a restringir sua proliferação. Isso evidencia a influência das condições climáticas na disseminação da doença.

Essa dinâmica ambiental não apenas determina a gravidade das epidemias, como também facilita a alternância e a circulação de sorotipos virais conforme a receptividade de cada área Brigagão e Corrêa (2017). Assim, a sobreposição dos dados epidemiológicos à classificação de Köppen apresenta que o regime climático Cfa é importante para o aumento de casos no verão paranaense, validando a necessidade de estratégias de vigilância diferenciadas para as distintas zonas climáticas do estado.



Figura 12 – Distribuição espacial da classificação climática do estado do Paraná. Fonte: FUNDAÇÃO ABC (2025).

CONCLUSÃO

Neste estudo foi possível investigar a associação entre variáveis climáticas e a incidência de dengue nos municípios do Paraná a partir da proposição de distinção de padrões endêmico e epidêmico, considerando as dinâmicas temporal e espacial da dengue e fatores climáticos, além de modelagem não linear.

Os resultados deste estudo demonstraram que a umidade foi o fator climático mais correlacionado à incidência da doença, quando avaliado individualmente. A interação significativa entre temperatura, precipitação e umidade ($p < 0,001$) observada pela superfície tridimensional sugere que o risco de surto/epidemia resulta da combinação sinérgica desses fatores, tais resultados ressaltam a natureza multivariada da transmissão. É provável que o efeito entre anomalias térmicas e padrões de precipitação possam influenciar a dinâmica do vetor de forma não linear, indicando que o impacto de uma variável pode ser amplificado ou reduzido pela condição da outra. A inovação central desta abordagem reside na diferenciação metodológica entre a estabilidade endêmica e casos de surto/epidemia, permitindo que cada regime de transmissão seja compreendido e antecipado em suas particularidades bioclimáticas.

A separação dos municípios em *clusters* distintos destacou a importância das zonas climáticas como fatores determinantes da endemicidade. Enquanto as regiões de clima Cfa proporcionam condições térmicas estáveis para a manutenção do ciclo viral, as áreas sob influência do clima Cfb tendem a apresentar barreiras ecológicas que limitam a persistência da doença. Essas evidências indicam para uma heterogeneidade espacial, o que requer estratégias preditivas baseadas em indicadores climáticos e adaptadas à dinâmica local.

É importante ressaltar que a eficácia do modelo de alerta proposto adquiri uma dimensão ainda mais estratégica em relação ao novo programa de imunização iniciado em 2026. A introdução da Butantan-DV, primeira vacina de dose única e 100% nacional, com eficácia global de 74% e proteção total contra hospitalizações, representa um marco na saúde pública brasileira (BRASIL. Ministério da Saúde, 2026). Contudo, a oferta de um imunizante não substitui a necessidade de uma vigilância proativa, ao contrário, ela demanda um aprimoramento ainda maior na identificação de áreas prioritárias.

Nesse contexto, o modelo preditivo serve como uma das abordagens necessárias para as medidas de controle vetorial, em que a pressão climática sinaliza um risco iminente de quebra da estabilidade endêmica. Este estudo proporciona ao estado do Paraná uma ferramenta técnica para o auxiliar na tomada de decisões em vigilância epidemiológica, com o objetivo de otimizar a distribuição de recursos e antecipar medidas preventivas em diversas condições meteorológicas.

Embora o modelo GAMM com a estrutura binomial tenha se mostrado satisfatório em capturar as tendências não lineares e correção dos efeitos aleatórios por municípios, a abordagem quantitativa mostrou-se superior por ser capaz de incorporar a correlação temporal residual, estrutura essa não contemplada pela modelagem binomial. Adicionalmente, a investigação de outras arquiteturas de modelos probabilísticos ou mesmo modelos de *machine learning*, podem ser consideradas e comparadas com os modelos construídos neste trabalho.

Além do avanço computacional para prever mudanças nos padrões de incidência, a resolução espacial dos alertas dependem da incorporação de dados mais completos, como disponibilidade de habitats para oviposição, urbanização e indicadores socioeconômicos. Conforme enfatizado por Mordecai et al. (2017), esses elementos interagem de maneira sinérgica com o clima para moldar a variação da incidência, sendo fundamentais para compreender a dinâmica epidemiológica.

REFERÊNCIAS

- ABDALLAH, A. P.; BELENTANI, L. M. Análise dos casos de dengue clássica e febre hemorrágica no Paraná. *Revista Uninga*, v. 28, n. 1, 2011. Disponível em: <<https://revista.uninga.br/uninga/article/view/949>>.
- AGHABOZORGI, S.; SHIRKHORSHIDI, A. S.; WAH, T. Y. Time-series clustering—a decade review. *Information Systems*, Elsevier, v. 53, p. 16–38, 2015.
- ALDUCHOV, O. A.; ESKRIDGE, R. E. Improved magnus form approximation of saturation vapor pressure. *Journal of Applied Meteorology*, v. 35, n. 4, p. 601–609, 1996. DOI: <[https://doi.org/10.1175/1520-0450\(1996\)035<0601:IMFAOS>2.0.CO;2](https://doi.org/10.1175/1520-0450(1996)035<0601:IMFAOS>2.0.CO;2)>.
- ALMEIDA, D. S.; RUDKE, A. P. et al. Estudo da relação entre variáveis meteorológicas e ocorrência de casos de dengue em londrina - pr. *Revista Brasileira de Geografia Física*, v. 14, n. 7, p. 3857–3866, 2021.
- ALMEIDA, I. F. de; LANA, R. M.; CODECO, C. T. How heterogeneous is the dengue transmission profile in brazil? a study in six brazilian states. *PLOS Neglected Tropical Diseases*, v. 16, n. 9, p. e0010746, 2022. Acesso em: 23 jun. 2025. DOI: <<https://doi.org/10.1371/journal.pntd.0010746>>.
- ASSUNÇÃO, L. F. d. A.; RODRIGUES, A. K. G. et al. Sazonalidade na incidência da dengue no brasil nos últimos 10 anos: Uma revisão epidemiológica. *Brazilian Journal of Implantology and Health Sciences*, v. 7, n. 3, p. 139–151, 2025.
- BARCELLOS, C. et al. Climate change, thermal anomalies and the recent progression of dengue in brazil. *Scientific Reports*, v. 14, p. 5948, 2024. Acesso em: 23 jun. 2025. DOI: <<https://doi.org/10.1038/s41598-024-56044-y>>.
- BATISTA, G. E. et al. A complexity-invariant distance measure for time series. *Data Mining and Knowledge Discovery*, v. 28, p. 634–667, 2014.
- BAVIA, L. et al. Epidemiological study on dengue in southern brazil under the perspective of climate and poverty. *Revista da Sociedade Brasileira de Medicina Tropical*, v. 53, p. e20190316, 2020.
- BJØRNSTAD, O. N.; IMS, R. A.; LAMBIN, X. Spatial heterogeneity and regional synchrony in dynamics. *Trends in Ecology & Evolution*, Elsevier, v. 14, n. 11, p. 427–432, 1999.

BRANDMAIER, A. M. pdc: An R package for complexity-based clustering of time series. *Journal of Statistical Software*, v. 67, n. 1, p. 1–23, 2015.

BRASIL. *Diretrizes nacionais para prevenção e controle das arboviroses urbanas: Vigilância entomológica e controle vetorial*. Brasília, 2025. 1ª edição - versão eletrônica.

BRASIL. Ministério da Saúde. *Entenda porque os casos de dengue aumentam no verão*. 2025. Acesso em: 24 jan. 2026. Disponível em: <<https://www.gov.br/saude/pt-br/assuntos/noticias/2025/janeiro/entenda-porque-os-casos-de-dengue-aumentam-no-verao>>.

BRASIL. Ministério da Saúde. *Vacina inédita e 100% brasileira contra a dengue começa a ser aplicada em Botucatu (SP)*. Brasília, DF: Ministério da Saúde, 2026. Disponível em: <<https://www.gov.br/saude/pt-br/assuntos/noticias/2026/janeiro/vacina-inedita-e-100-brasileira-contr-a-dengue-comeca-a-ser-aplicada-em-botucatu-sp>>. Acesso em: 25 jan. 2026.

BRIGAGÃO, G. S.; CORRÊA, N. A. B. Levantamento epidemiológico da dengue no estado do Paraná Brasil nos anos de 2011 a 2015. *Arquivos de Ciências da Saúde da UNIPAR*, v. 21, n. 1, p. 41–45, 2017.

CAIADO, J.; MAHARAJ, E. A.; D'URSO, P. Time series clustering: a comparative study. *De Economist*, Springer, v. 154, p. 331–355, 2006.

CHEN, X.; MORAGA, P. Dengue forecasting and outbreak detection in brazil using lstm: integrating human mobility and climate factors. *medRxiv*, 2025. Acesso em: 23 jun. 2025. DOI: <<https://doi.org/10.1101/2025.03.02.25323168>>.

CHOUAKRIA, A. D.; NAGABHUSHAN, P. N. Adaptive dissimilarity index for measuring time series proximity. *Advances in Data Analysis and Classification*, Springer, v. 1, n. 1, p. 5–21, 2007. Disponível em: <https://www.researchgate.net/publication/225157200_Adaptive_dissimilarity_index_for_measuring_time_series_proximity>.

CHURAKOV, M. et al. Spatio-temporal dynamics of dengue in brazil: Seasonal travelling waves and determinants of regional synchrony. *PLOS Neglected Tropical Diseases*, v. 13, n. 4, p. e0007012, 2019. Acesso em: 23 jun. 2025. DOI: <<https://doi.org/10.1371/journal.pntd.0007012>>.

CINDY et al. Cluster analysis on dengue incidence and weather data using k-medoids and fuzzy c-means clustering algorithms (case study: Spread of dengue in the DKI Jakarta province). *Journal of Mathematical and Fundamental Sciences*, ITB Institute for Research and Community Services, v. 53, n. 3, p. 466–486, 2022. Disponível em: <<https://journals.itb.ac.id/index.php/jmfs/article/view/15445>>.

CODEÇO, C. T.; VILLELA, D. A. M.; COELHO, F. C. Estimating the effective reproduction number of dengue considering temperature-dependent generation intervals. *Epidemics*, Elsevier, v. 25, p. 101–111, dec 2018. Disponível em: <<https://doi.org/10.1016/j.epidem.2018.05.011>>.

CURRIE, I. D.; DURBAN, M. Flexible smoothing with p-splines: a unified approach. *Statistical Modelling*, SAGE Publications, v. 2, n. 4, p. 333–349, 2002.

DAVIES, D. L.; BOULDIN, D. W. A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, IEEE, n. 2, p. 224–227, 1979.

DOVE, S. et al. A user-friendly guide to using distance measures to compare time series in ecology. *Methods in Ecology and Evolution*, Wiley Online Library, v. 13, n. 10, p. 2128–2145, 2023.

ECMWF. *ERA5-Land hourly data on single levels from 1950 to present*. 2024. Copernicus Climate Change Service (C3S) Climate Data Store. Acesso em: 10 mar. 2025. Disponível em: <<https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-land>>.

EILERS, P. H.; MARX, B. D. Flexible smoothing with b-splines and penalties. *Statistical Science*, JSTOR, v. 11, n. 2, p. 89–102, 1996.

ELY, D. F.; BERTINI, I. T.; OLIVEIRA, L. T. Variabilidade climática nas cidades de londrina, maringá (pr) e florianópolis (sc) e a expansão latitudinal da dengue. *Revista Geonorte*, v. 2, n. 5, p. 826–839, 2012. Edição especial 2. Acesso em: 12 jun. 2025. Disponível em: <<https://www.periodicos.ufam.edu.br/index.php/revista-geonorte/article/view/2540>>.

ESLING, P.; AGON, C. Time-series data mining. *ACM Computing Surveys (CSUR)*, ACM New York, NY, USA, v. 45, n. 1, p. 1–34, 2012.

EVERITT, B. S. et al. *Cluster Analysis*. 5. ed. Chichester, UK: John Wiley & Sons, 2011. ISBN 978-0-470-74991-3. Disponível em: <https://cicerocq.wordpress.com/wp-content/uploads/2019/05/cluster-analysis_5ed_everitt.pdf>.

FREIRE, V. C. d. S. et al. O espaço da dengue no estado do paran : an lise temporal, sociodemogr fica e ambiental entre 2012 e 2021. *Revista Brasileira de Epidemiologia*, v. 26, p. e230004, 2023. Acesso em: 23 jun. 2025. DOI: <<https://doi.org/10.1590/1980-549720230004>>.

FUNDA  O ABC. *Classifica  o clim tica – Paran *. 2025. Acesso em: 21 abr. 2025. Disponível em: <https://sma.fundacaoabc.org/climatologia/classificacao_climatica/parana>.

FUNDA  O OSWALDO CRUZ. *O mosquito: de longo traje*. 2016. Acesso em: 25 jan. 2026. Disponível em: <<https://www.ioc.fiocruz.br/dengue/textos/longatraje.html>>.

GAVRILOV, M. et al. Mining the stock market: Which measure is best? In: *Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*. [S.l.]: ACM, 2000. p. 487–496.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd. ed. New York: Springer, 2009. Acesso em: 18 jan. 2026. Disponível em: <<https://hastie.su.domains/ElemStatLearn/>>.

HOSMER, D. W.; LEMESHOW, S.; STURDIVANT, R. X. *Applied Logistic Regression*. 3. ed. New York: John Wiley & Sons, 2013.

Instituto Nacional de Meteorologia. *2024   o ano mais quente da s rie hist rica no Brasil*. Portal INMET, 2025. Acesso em: 23 jan. 2026. Disponível em: <<https://portal.inmet.gov.br/noticias/2024-%C3%A9-o-ano-mais-quente-da-s%C3%A9rie-hist%C3%B3rica-no-brasil>>.

Instituto Nacional de Meteorologia. *Ver o 2024/2025 foi o sexto mais quente no Brasil desde 1961*. Portal INMET, 2025. Acesso em: 23 jan. 2026. Disponível em: <<https://portal.inmet.gov.br/noticias/ver%C3%A3o-2024-2025-foi-o-sexto-mais-quente-no-brasil-desde-1961>>.

International Committee on Taxonomy of Viruses. *Virus Taxonomy: 2024 Release*. 2024. Acesso em: 25 jan. 2026. Disponível em: <https://ictv.global/taxonomy/taxondetails?taxnode_id=202403070>.

JAIN, A. K.; DUBES, R. C. *Algorithms for Clustering Data*. [S.l.]: Prentice-Hall, 1988.

KALPAKIS, C.; GADA, D.; PUTTAGUNTA, V. Distance measures for effective clustering of ARIMA time-series. In: IEEE. *Proceedings 2001 IEEE International Conference on Data Mining*. [S.l.], 2001. p. 273–280.

KAUFMAN, L.; ROUSSEEUW, P. J. *Finding groups in data: an introduction to cluster analysis*. Hoboken: John Wiley & Sons, 1990. ISBN 9780471878766.

LARSEN, B.; AONE, C. Fast and effective text mining using linear-time document clustering. In: *Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining*. [S.l.: s.n.], 1999. p. 16–22.

LAUREANO, A. N. A. *Influência de fatores climáticos em Aedes aegypti: relação da temperatura e precipitação com índices entomológicos de monitoramento de dengue*. Dissertação (Trabalho de Conclusão de Curso (Graduação em Ciências Biológicas – Ecologia e Biodiversidade)) — Instituto Latino-Americano de Ciências da Vida e da Natureza, Universidade Federal da Integração Latino-Americana, Foz do Iguaçu, 2025. Disponível em: <<https://dspace.unila.edu.br/items/bbc9f762-7119-4738-9b3d-04a211f49536/full>>.

LAURINEC, P. TSrepr R package: Time Series Representations. *Journal of Open Source Software*, v. 3, n. 23, p. 577, 2018. Disponível em: <<https://doi.org/10.21105/joss.00577>>.

LEANDRO, G. C. W.; CICHELER, L. M. et al. Análise temporal e espacial dos casos municipais de dengue no paran  e indicadores sociais e ambientais, 2012 a 2021: estudo ecol gico. *Revista Brasileira de Epidemiologia*, v. 25, p. e220039, 2022.

LOWE, R.; AL. et. Combined effects of hydrometeorological hazards and urbanization on dengue risk in brazil: a spatiotemporal modelling study. *The Lancet Planetary Health*, Elsevier, v. 5, n. 4, p. e209–e219, 2021.

LOWE, R. et al. The future of dengue in Brazil under climate change: a spatio-temporal modelling study. *The Lancet Planetary Health*, Elsevier, v. 5, n. 3, p. e139–e147, 2021.

LOWE, R. et al. Evaluating probabilistic dengue risk forecasts from a prototype early warning system for brazil. *eLife*, eLife Sciences Publications, Ltd, v. 5, p. e11285, 2016.

MACQUEEN, J. Some methods for classification and analysis of multivariate observations. In: OAKLAND, CA, USA. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*. [S.l.], 1967. v. 1, p. 281–297.

MCHUGH, M. L. Lessons in biostatistics. interrater reliability: the kappa statistic. *Biochemia Medica*, Editorial Office of Biochemia Medica, v. 22, n. 3, p. 276–282, 2012. Acesso em: 25 jan. 2026. Disponível em: <<https://pmc.ncbi.nlm.nih.gov/articles/PMC3900052/>>.

MELO, V. R. d. et al. A expans o da dengue no paran , 2011 e 2012: aumento da distribui o geogr fica e persist ncia da transmiss o. *Revista Paranaense de Sa de P blica*, Secretaria de Estado da Sa de do Paran , v. 15, n. 2, p. 5–12, 2014.

MENEZES, E. *dengueControl: Control Diagrams and Reports of Dengue from Historical Brazilian SINAN data. (Time Series)*. 2025. R package version 0.0.0.9007.

Ministério da Saúde. *Guia de vigilância epidemiológica*. 7. ed. Brasília: Ministério da Saúde, 2009. Acesso em: 22 jun. 2025. Disponível em: <https://bvsms.saude.gov.br/bvs/publicacoes/guia_vigilancia_epidemiologica_7ed.pdf>.

Ministério da Saúde. *Dengue*. 2025. Governo Federal, Portal gov.br. Acesso em: 10 mai. 2025. Disponível em: <<https://www.gov.br/saude/pt-br/assuntos/saude-de-a-a-z/d/dengue>>.

MONTERO, P.; VILAR, J. A. *TSclust: An R Package for Time Series Clustering*. 2014. 1–43 p. Disponível em: <<https://doi.org/10.18637/jss.v062.i01>>.

MORDECAI, E. A. et al. Detecting the impact of temperature on transmission of Zika, dengue, and chikungunya using mechanistic models. *PLOS Neglected Tropical Diseases*, Public Library of Science, v. 11, n. 4, p. e0005568, 2017. Disponível em: <<https://doi.org/10.1371/journal.pntd.0005568>>.

MORIN, C. W.; COMRIE, A. C.; ERNST, K. Climate and dengue transmission: Evidence and implications. *Environmental Health Perspectives*, v. 121, n. 11-12, p. 1264–1272, Nov-Dec 2013. PMID: 24058050; PMCID: PMC3855512. Disponível em: <<https://doi.org/10.1289/ehp.1306556>>.

MOTTA, T. C. A. d. *Apesar de viver apenas 45 dias, mosquito Aedes aegypti pode colocar 450 ovos*. Secretaria de Estado de Saúde de Mato Grosso do Sul, 2016. Acesso em: 25 jan. 2026. Disponível em: <<https://www.saude.ms.gov.br/apesar-de-viver-apenas-45-dias-mosquito-aedes-aegypti-pode-colocar-450-ovos/>>.

MURTA, J. R. M. *Impactos das mudanças climáticas na saúde pública do Distrito Federal em relação à dengue*. Monografia (Monografia) — Faculdade de Planaltina, Distrito Federal, Brasília, 2017.

MURTAGH, F.; CONTRERAS, P. Algorithms for hierarchical clustering: An overview. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, Wiley Online Library, v. 2, n. 1, p. 86–97, 2012.

NIENNATTRAKUL, V.; RATANAMAHATANA, C. A. On clustering multimedia time series data using k-means and dynamic time warping. In: IEEE. *2007 International Conference on Multimedia and Ubiquitous Engineering (MUE'07)*. [S.l.], 2007. p. 733–738. Acesso em: 23 jun. 2025. Disponível em: <https://www.researchgate.net/publication/221282068_On_Clustering_Multimedia_Time_Series_Data_Using_K-Means_and_Dynamic_Time_Warping>.

NIRMAL, S. Comparative study between k-means and k-medoids clustering algorithms. *International Research Journal of Engineering and Technology (IRJET)*, v. 6, n. 3, p. 839–844, March 2019. ISSN 2395-0056. Disponível em: <<https://www.irjet.net/archives/V6/i3/IRJET-V6I3154.pdf>>.

Organização Pan-Americana da Saúde. *Alerta Epidemiológico: Risco de surtos de dengue devido ao aumento da circulação do DENV-3 na Região das Américas*. Washington, D.C., 2025. Retificação - 10 de fevereiro de 2025. Disponível em: <<https://www.paho.org/pt/documentos/alerta-epidemiologico-risco-surtos-dengue-devido-ao-aumento-da-circulacao-do-denv-3-na>>.

Organização Pan-Americana da Saúde. *PLISA Plataforma de Informação em Saúde para as Américas, Portal de Indicadores de Dengue*. 2025. Washington, D.C.: OPAS/OMS. Acesso em: 10 maio 2025. Disponível em: <<https://www3.paho.org/data/index.php/es/temas/indicadores-dengue.html>>.

OSEMWENKHAE, J. E.; EKHATOR, O. F.; IDUSERI, A. A comparative study of k-means and k-medoids clustering methods. *Journal of the Nigerian Association of Mathematical Physics*, v. 36, n. 2, p. 169–176, 2016. Destaca que o K-medoids (PAM) performa melhor que o K-means em dados com estruturas complexas e tamanhos variados.

Pan American Health Organization. *Dengue: analysis by country*. 2026. Accessed: 2026-03-19. Disponível em: <<https://www.paho.org/en/arbo-portal/dengue-data-and-analysis/dengue-analysis-country>>.

PEREDA, P. C.; VALVERDE, M. C. R.; SILVA, L. R. da. *Impacts of Climate Change on Dengue Risk in Brazil*. Washington, DC, 2014. Department of Research and Chief Economist. Acesso em: 23 jun. 2025. Disponível em: <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2533590>.

PÉRTEGA, S.; VILAR, J. A. Comparing several parametric and nonparametric approaches to time series clustering: A simulation study. *Journal of Classification*, Springer, v. 27, n. 3, p. 333–362, 2010.

PINHEIRO, J. et al. *nlme: Linear and Nonlinear Mixed Effects Models*. [S.l.], 2019. R package version 3.1-139. Disponível em: <<https://CRAN.R-project.org/package=nlme>>.

Python Software Foundation. *Python: a programming language*. 2021. Version 3.9.2.

R Core Team. *R: A language and environment for statistical computing*. Vienna: [s.n.], 2024. R Foundation for Statistical Computing.

REINHOLD, J. M.; LAZZARI, C. R.; LAHONDÈRE, C. Effects of the environmental temperature on *aedes aegypti* and *aedes albopictus* mosquitoes: A review. *Insects*, MDPI, v. 9, n. 4, p. 158, 2018. Acesso em: 23 jun. 2025. DOI: <<https://doi.org/10.3390/insects9040158>>.

RIBEIRO, A. D. et al. Associação entre incidência de dengue e variáveis climáticas. *Revista de Saúde Pública*, SciELO Brasil, v. 40, n. 4, p. 671–676, 2006.

Secretaria de Estado da Saúde do Paraná. *Informe Técnico n. 47: Monitoramento das Arboviroses – Período Epidemiológico 2023/2024*. 2024. Acesso em: 3 abr. 2026. Disponível em: <<https://www.dengue.pr.gov.br/Pagina/Boletins-da-Dengue>>.

Secretaria de Estado da Saúde do Paraná. *Dengue: Sintomas e Sinais*. 2025. Portal Dengue Paraná. Acesso em: 15 mai. 2025. Disponível em: <<https://www.dengue.pr.gov.br/Pagina/Dengue-Sintomas-e-Sinais>>.

SIM, J.; WRIGHT, C. C. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Physical Therapy*, Oxford University Press, v. 85, n. 3, p. 257–268, 2005.

STEWART-IBARRA, A. M. et al. Dengue Vector Dynamics (*Aedes aegypti*) Influenced by Climate and Social Factors in Ecuador: Implications for Targeted Control. *PLOS ONE*, v. 8, n. 11, p. e78263, 2013. Disponível em: <<https://doi.org/10.1371/journal.pone.0078263>>.

SYAKUR, M. A. et al. Integration k-means clustering method and elbow method for identification of the best customer profile cluster. *IOP Conference Series: Materials Science and Engineering*, IOP Publishing, v. 336, p. 012017, 2018. Acesso em: 23 jun. 2025. DOI: <<https://doi.org/10.1088/1757-899X/336/1/012017>>.

TAN, P.-N. et al. *Introduction to Data Mining*. 2. ed. Harlow, England: Pearson, 2019. Global Edition.

VELMURUGAN, T.; SANTHANAM, T. Computational complexity between K-means and K-medoids clustering algorithms for normal and uniform distributions of data points. *Journal of Computer Science*, Science Publications, v. 6, n. 3, p. 363–370, 2010.

VILLELA, D. A. M. Predicting high dengue incidence in municipalities of brazil using path signatures. *medRxiv*, 2024. Acesso em: 21 jun. 2025. DOI: <<https://doi.org/10.48550/arXiv.2501.12395>>.

WOOD, S. N. *Generalized Additive Models: An Introduction with R*. 2. ed. Boca Raton: Chapman and Hall/CRC, 2017. ISBN 9781498728331.

World Health Organization. *Dengue in the Region of the Americas – Situation Report*. 2024. Disease Outbreak News, 10 maio 2024. Acesso em: 17 abr. 2025. Disponível em: <<https://www.who.int/emergencies/disease-outbreak-news/item/2024-DON518>>.

Apêndices

DINÂMICA TEMPORAL DOS AGRUPAMENTOS DE DENGUE

A.1 Agrupamentos nos cenários endêmico e epidêmico

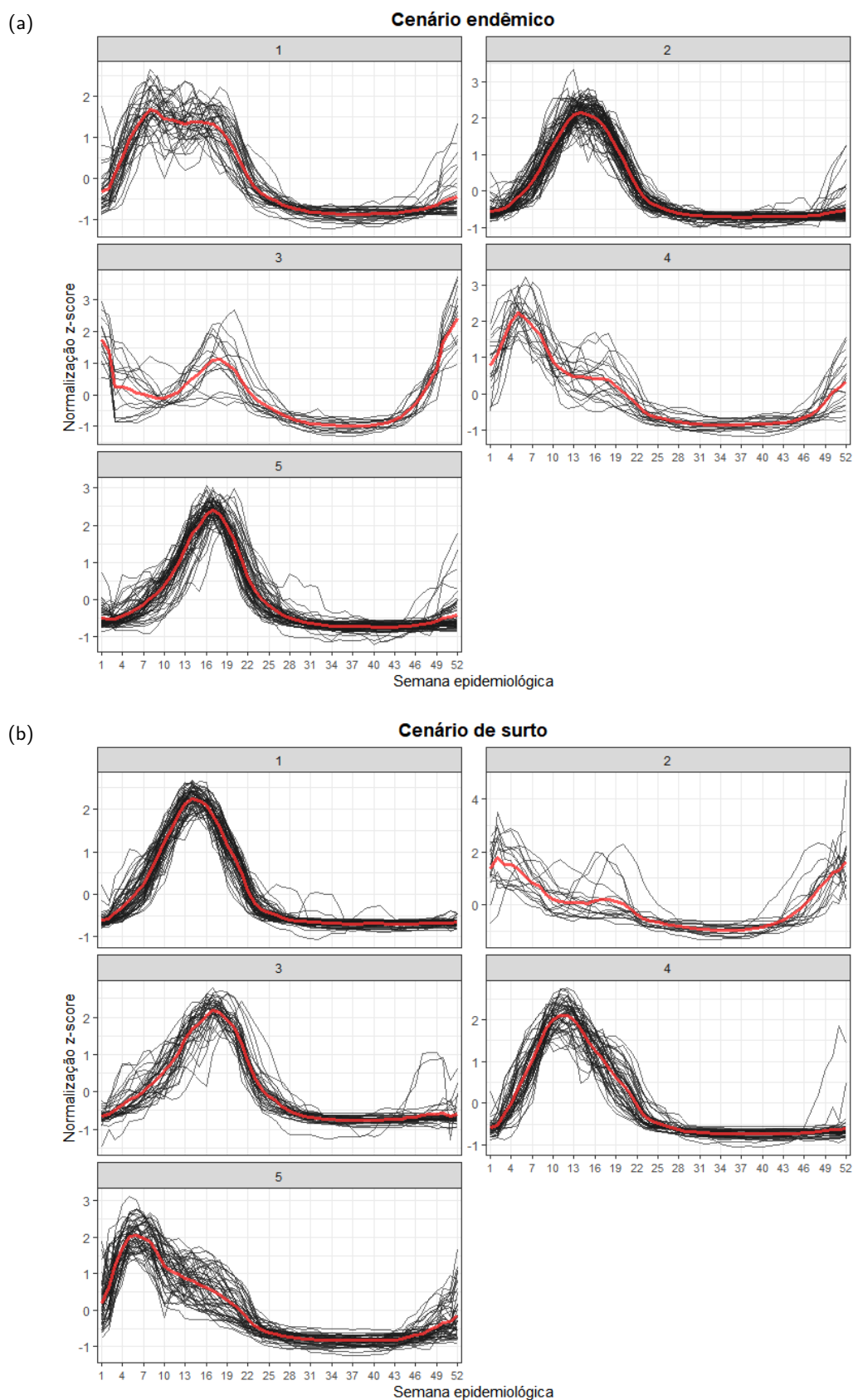
A.1.1 Configuração da partição ($k = 5$)

Figura 13 – Agrupamentos (*clusters*) identificados pelo algoritmo K-means ($k = 5$) para a dengue no Paraná: (a) cenário endêmico; (b) cenário epidêmico.

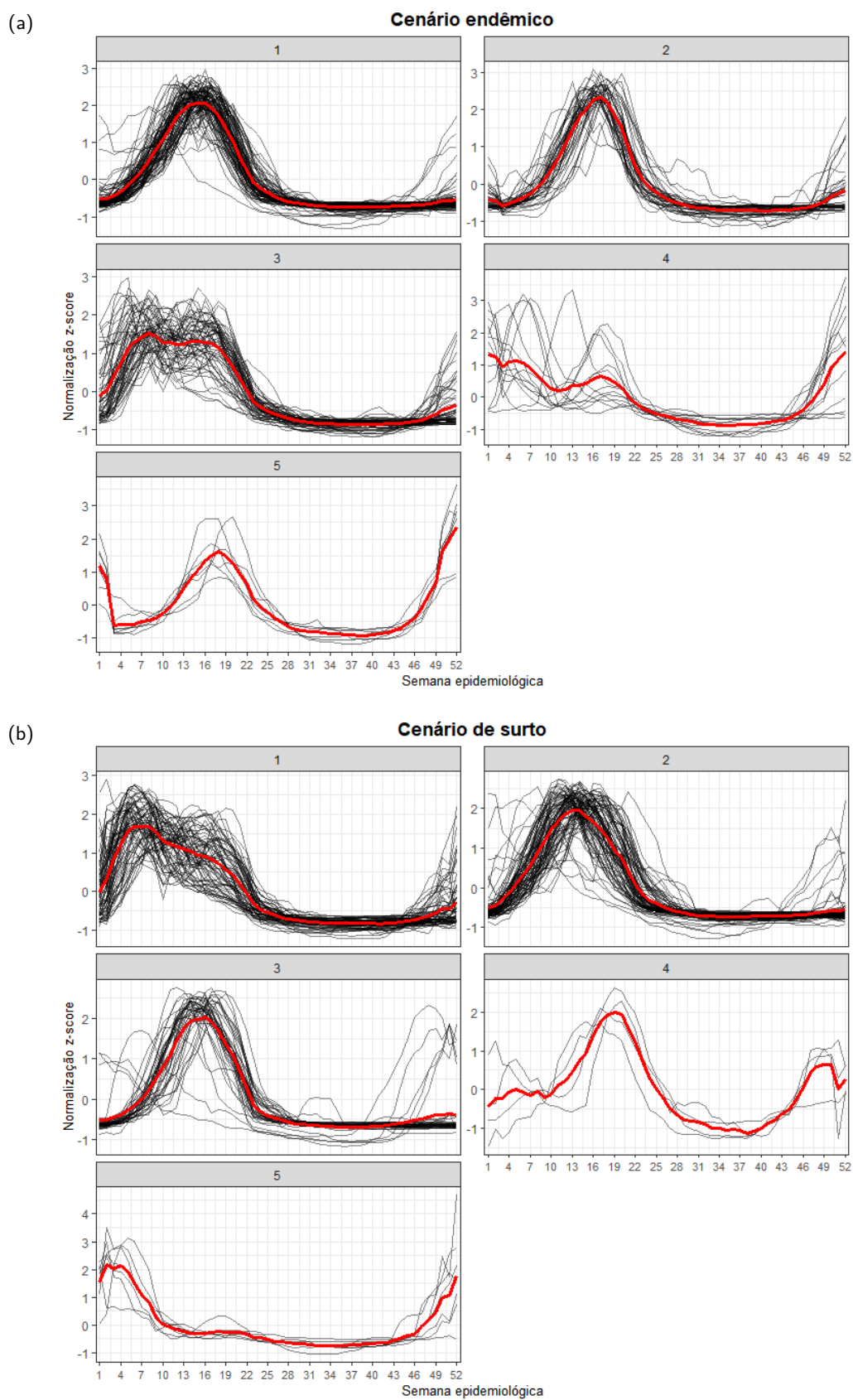


Figura 14 – Agrupamentos (*clusters*) identificados pelo algoritmo ACF ($k = 5$) para a dengue no Paraná: (a) cenário endêmico; (b) cenário epidêmico.

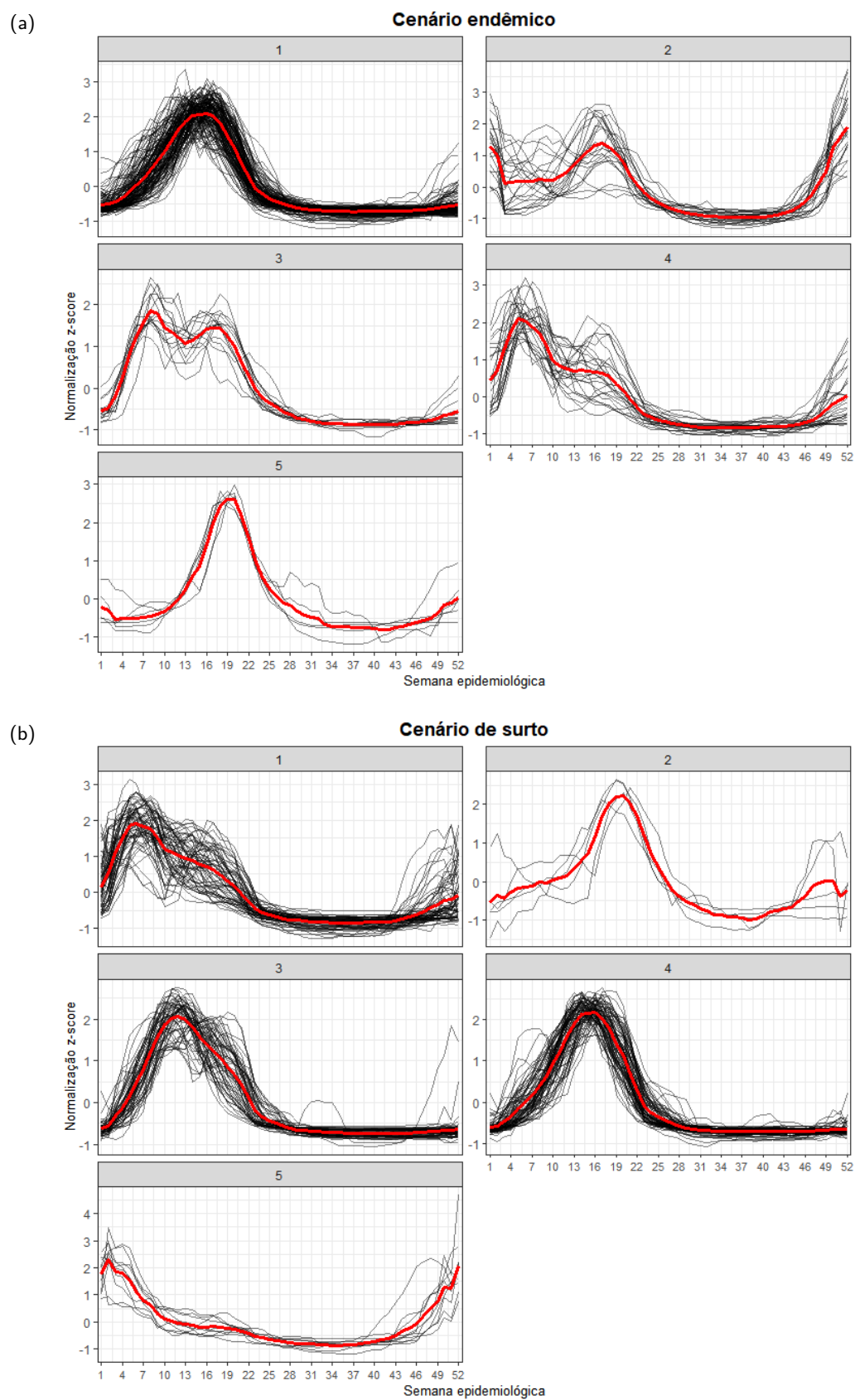


Figura 15 – Agrupamentos (*clusters*) identificados pelo algoritmo CORT ($k = 5$) para a dengue no Paraná: (a) cenário endêmico; (b) cenário epidêmico.

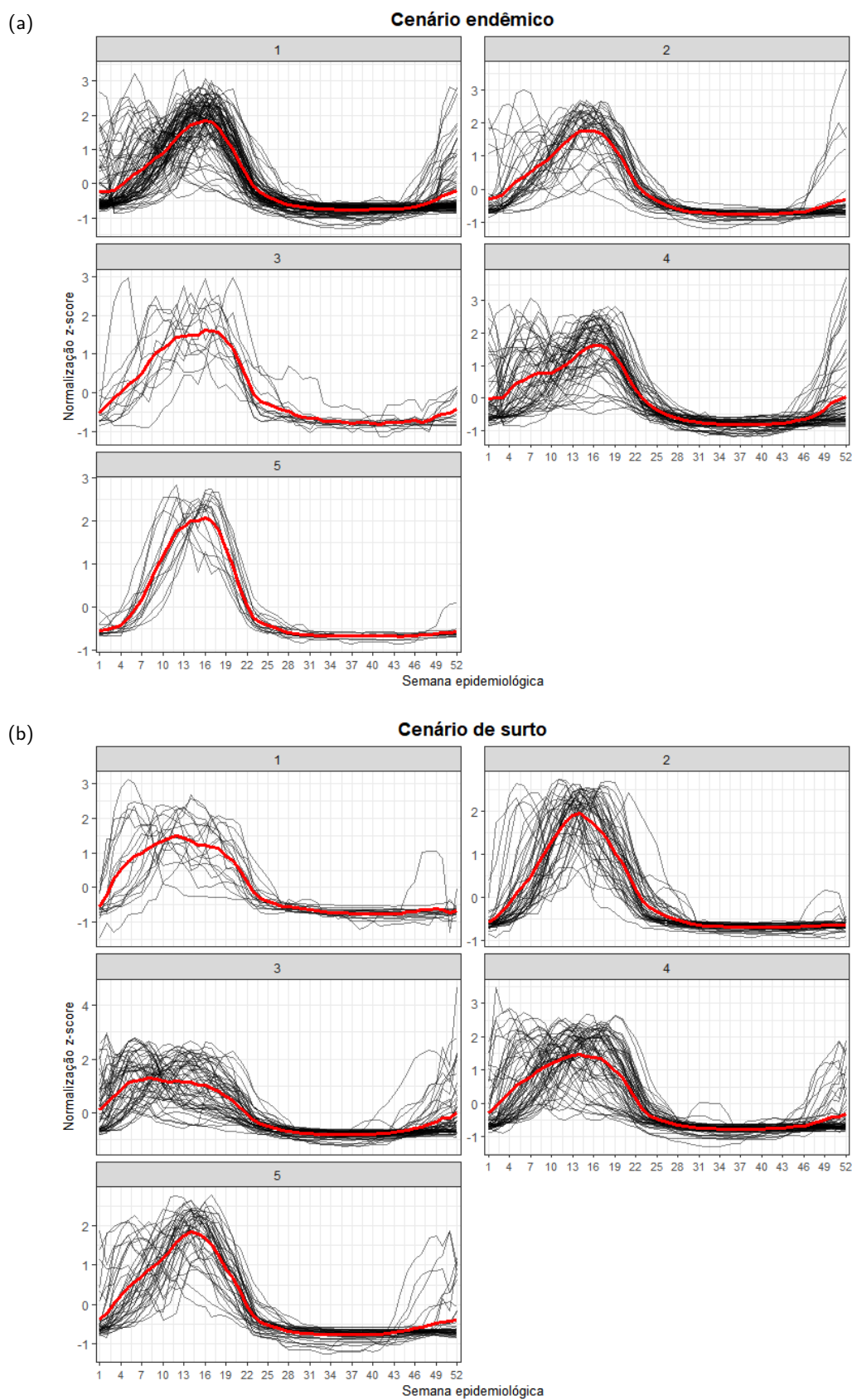


Figura 16 – Agrupamentos (*clusters*) identificados pelo algoritmo PDC ($k = 5$) para a dengue no Paraná: (a) cenário endêmico; (b) cenário epidêmico.

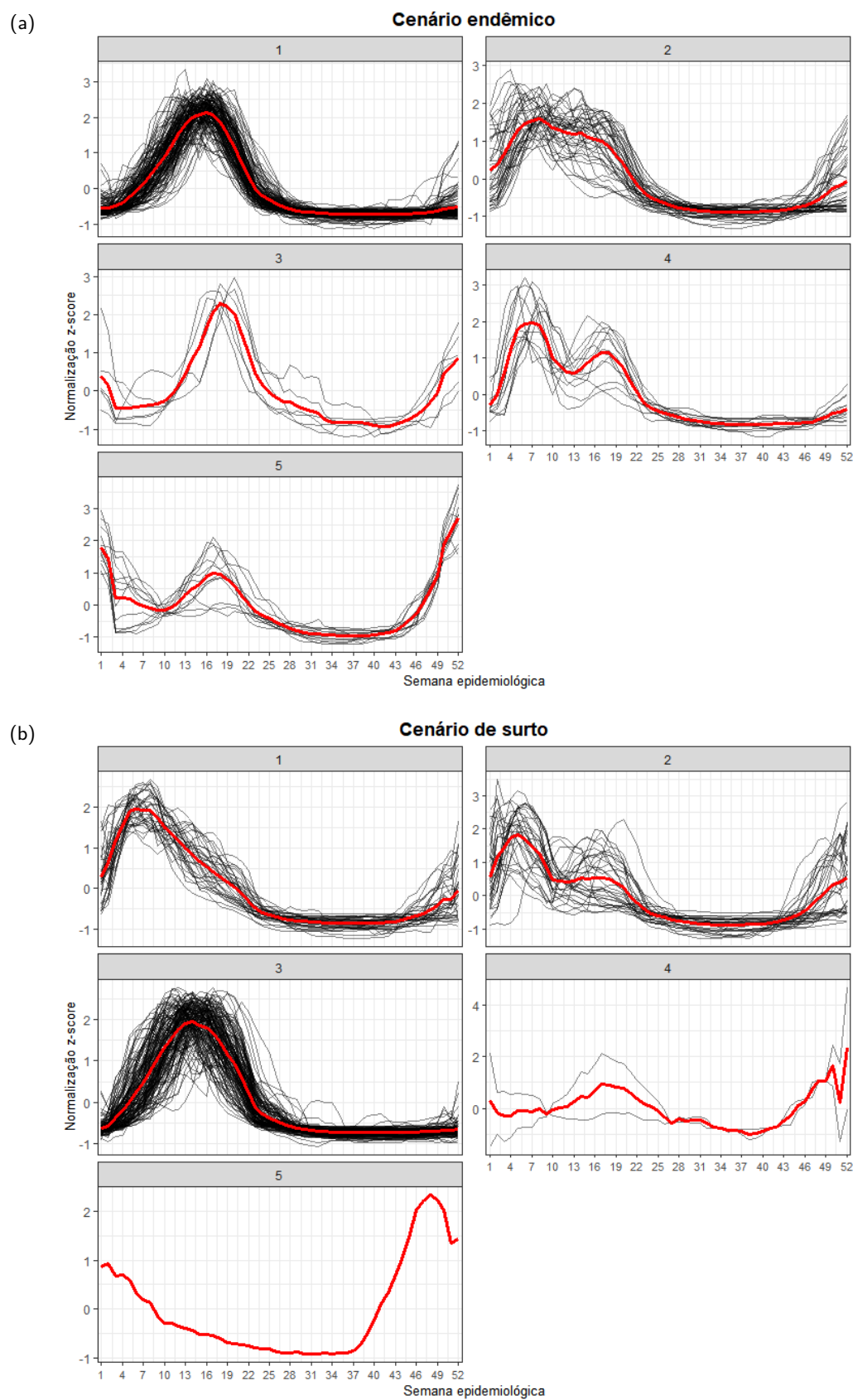


Figura 17 – Agrupamentos (*clusters*) identificados pelo algoritmo CID ($k = 5$) para a dengue no Paraná: (a) cenário endêmico; (b) cenário epidêmico.

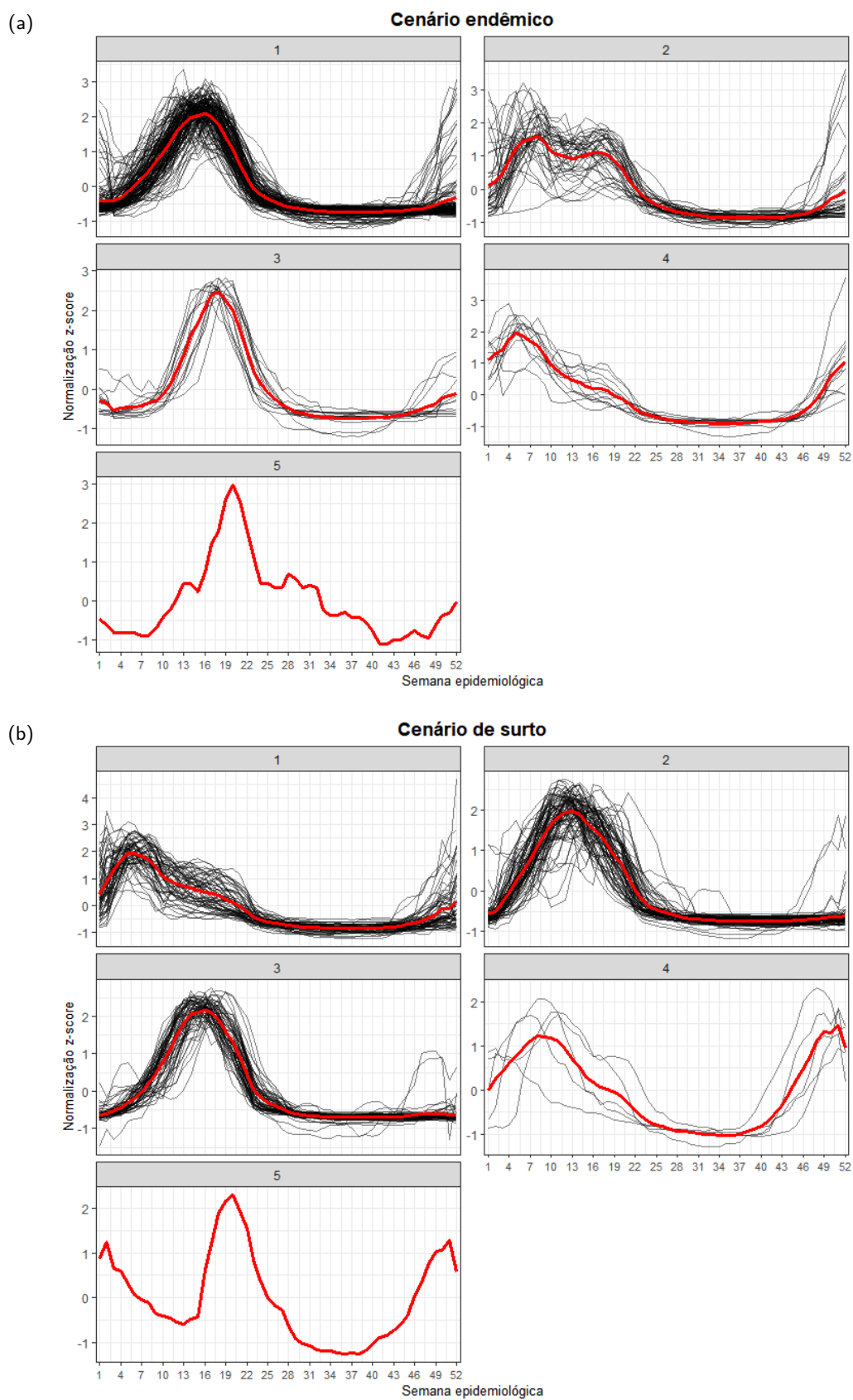


Figura 18 – Agrupamentos (*clusters*) identificados pelo algoritmo PER ($k = 5$) para a dengue no Paraná: (a) cenário endêmico; (b) cenário epidêmico.

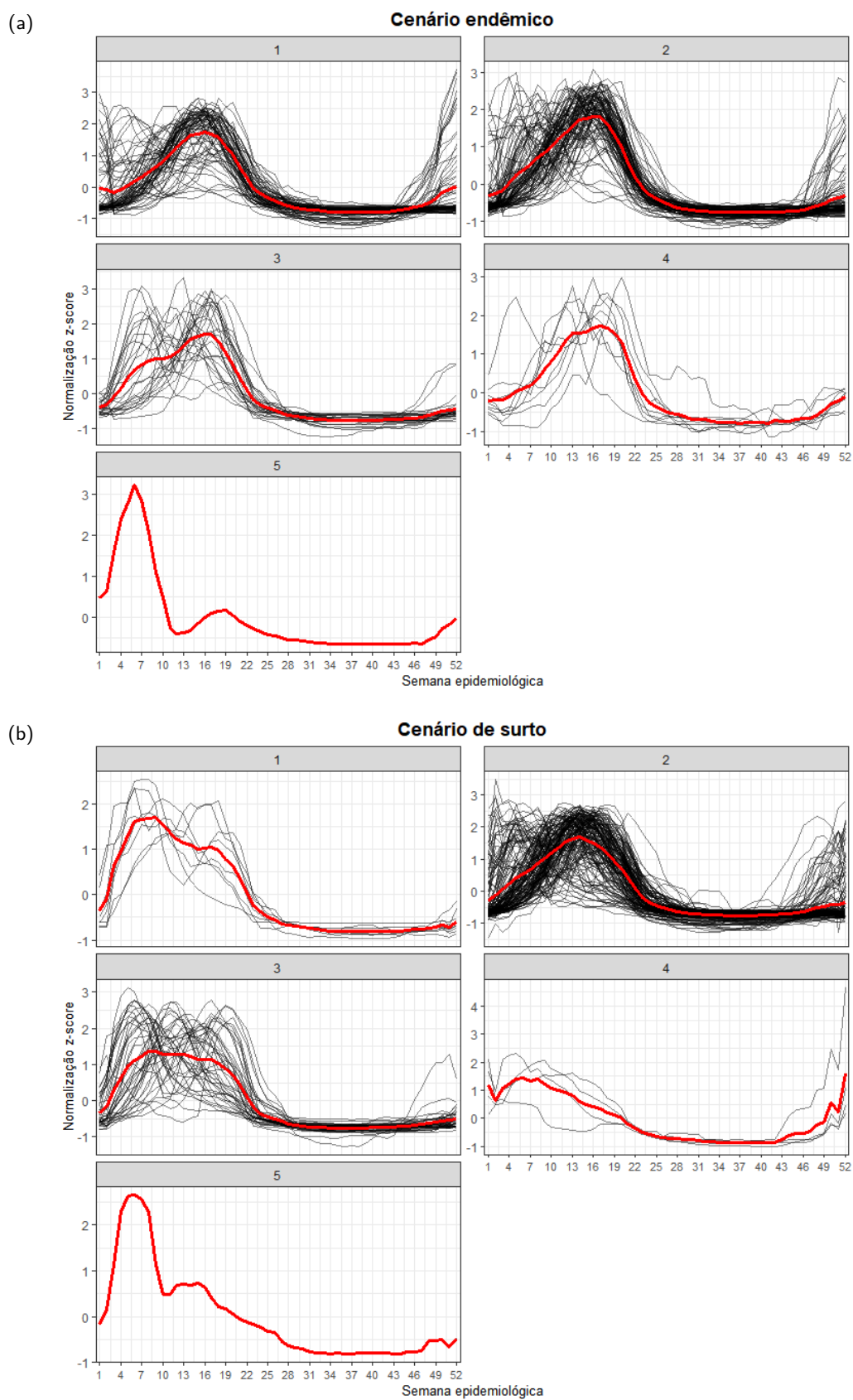


Figura 19 – Agrupamentos (*clusters*) identificados pelo algoritmo PIC ($k = 5$) para a dengue no Paraná: (a) cenário endêmico; (b) cenário epidêmico.

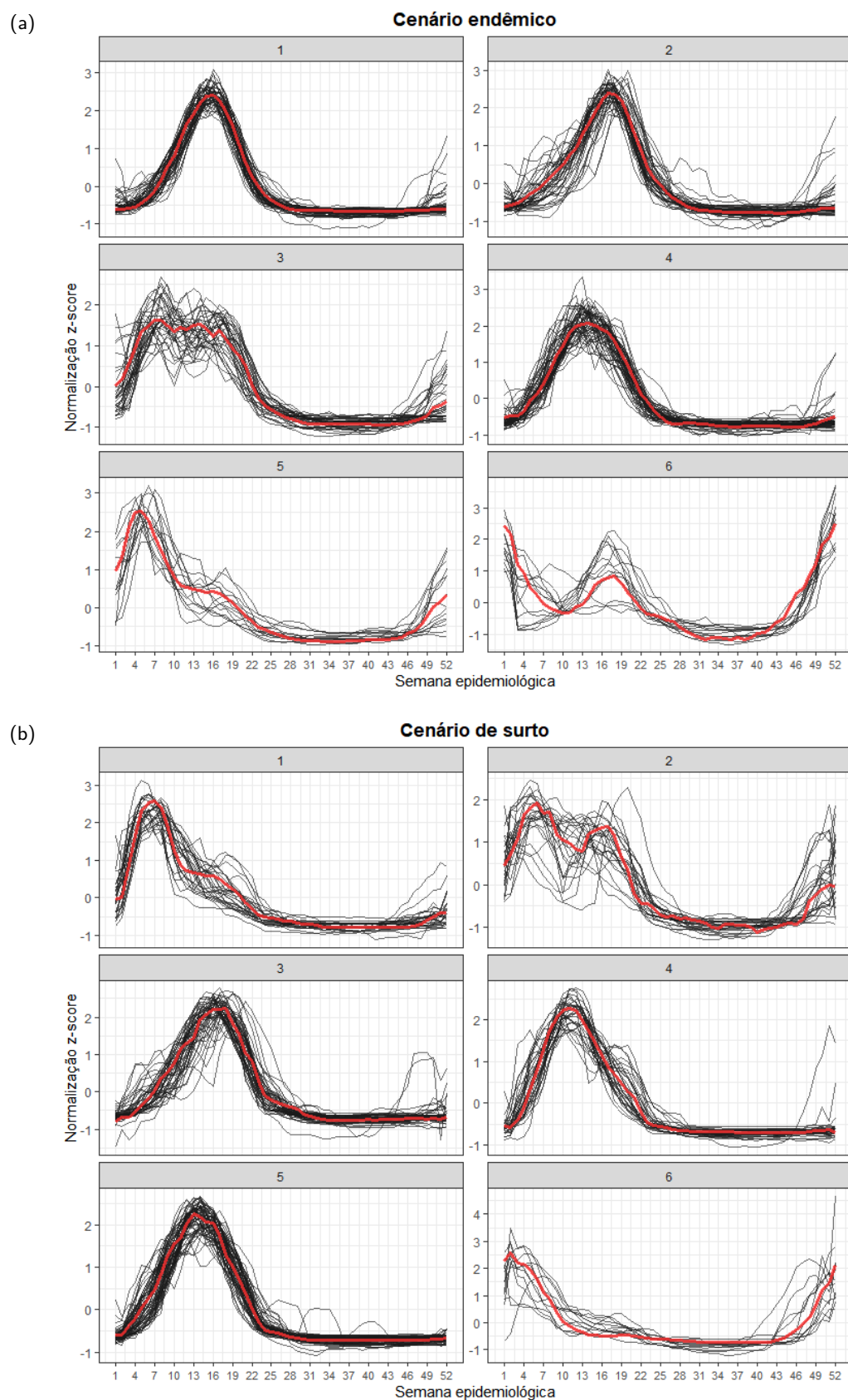
A.1.2 Configuração da partição ($k = 6$)

Figura 20 – Agrupamentos (*clusters*) identificados pelo algoritmo PAM ($k = 6$) para a dengue no Paraná: (a) cenário endêmico; (b) cenário epidêmico.

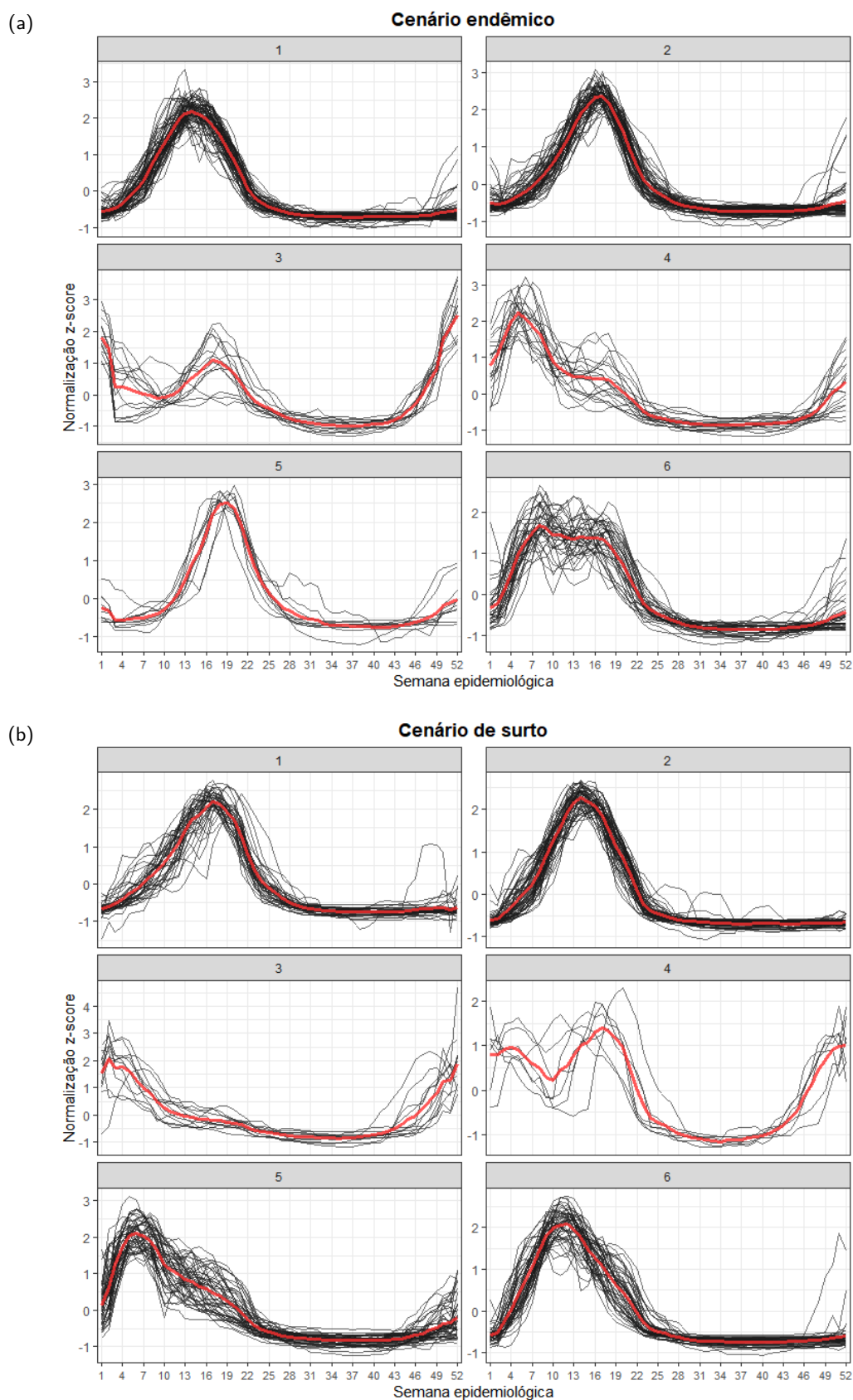


Figura 21 – Agrupamentos (*clusters*) identificados pelo algoritmo K-means ($k = 6$) para a dengue no Paraná: (a) cenário endêmico; (b) cenário epidêmico.

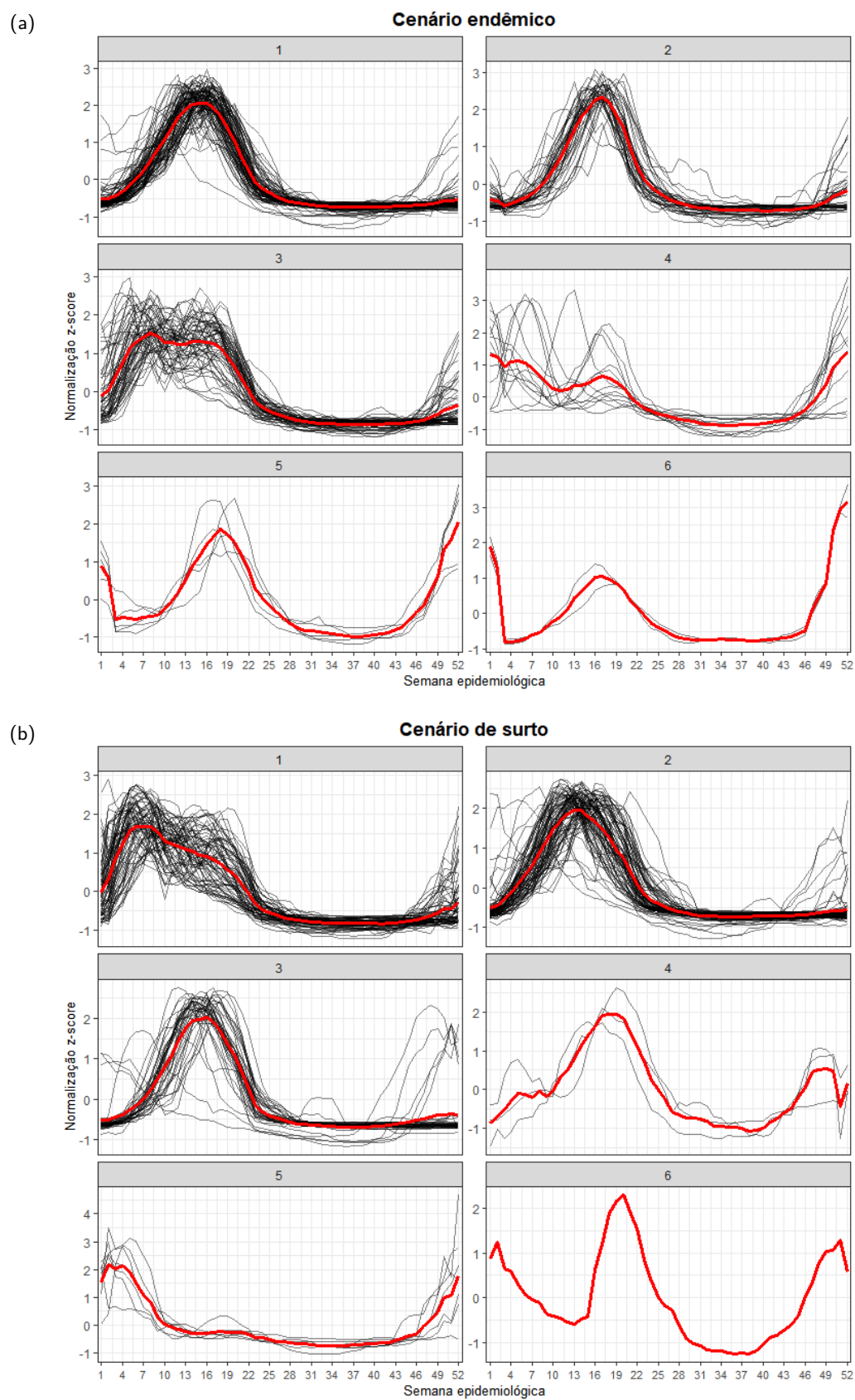


Figura 22 – Agrupamentos (*clusters*) identificados pelo algoritmo ACF ($k = 6$) para a dengue no Paraná: (a) cenário endêmico; (b) cenário epidêmico.

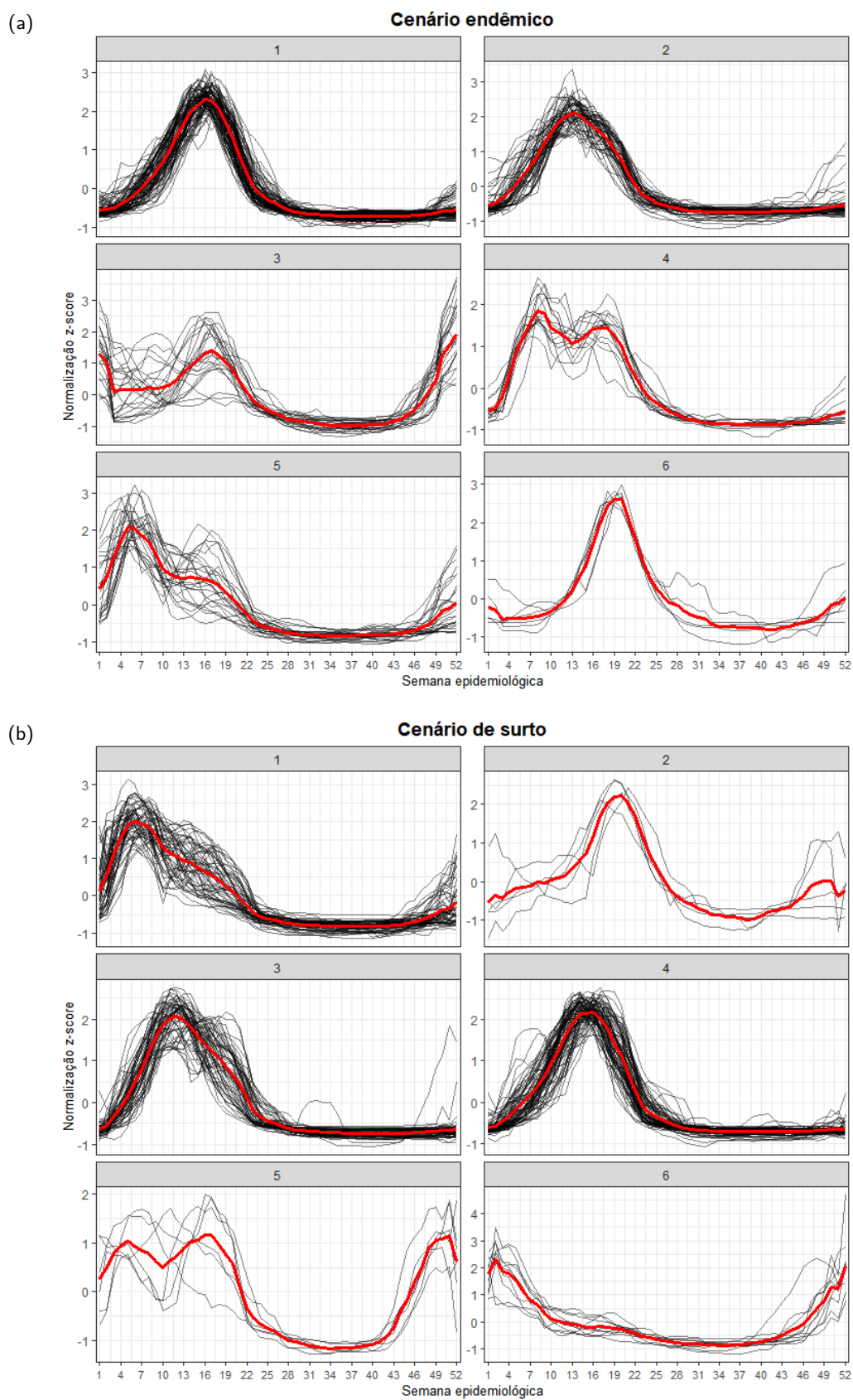


Figura 23 – Agrupamentos (*clusters*) identificados pelo algoritmo CORT ($k = 6$) para a dengue no Paraná: (a) cenário endêmico; (b) cenário epidêmico.

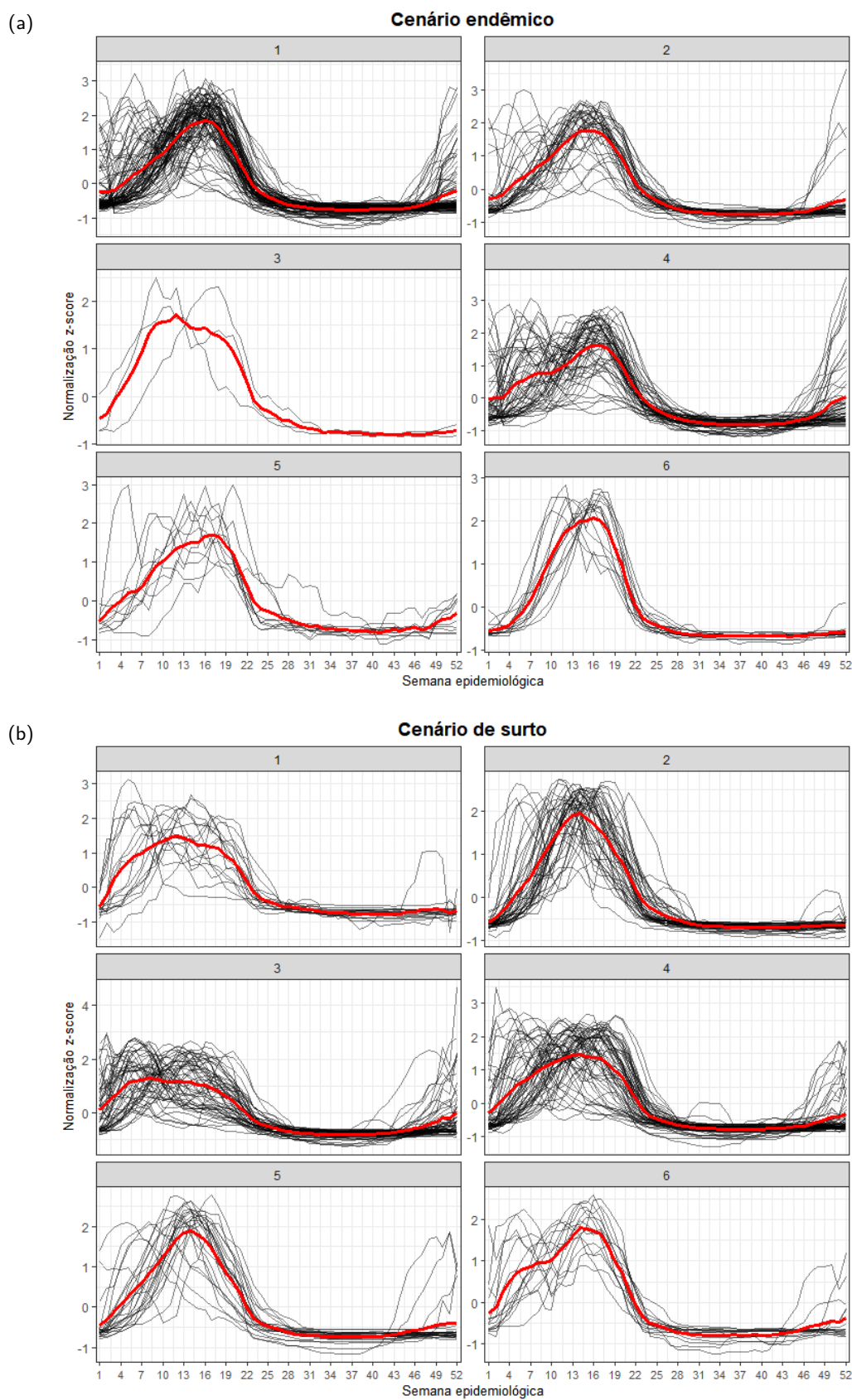


Figura 24 – Agrupamentos (*clusters*) identificados pelo algoritmo PDC ($k = 6$) para a dengue no Paraná: (a) cenário endêmico; (b) cenário epidêmico.

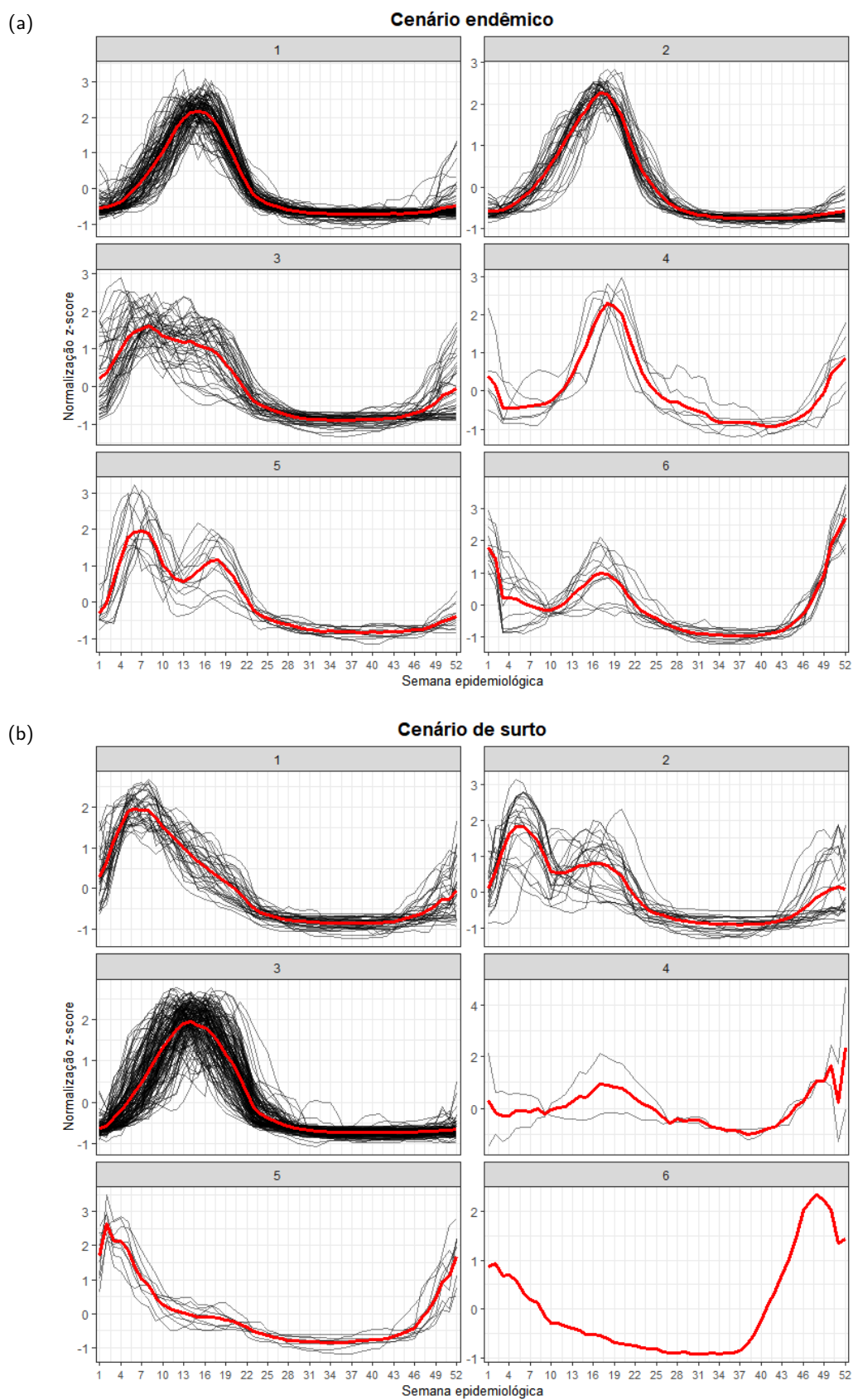


Figura 25 – Agrupamentos (*clusters*) identificados pelo algoritmo CID ($k = 6$) para a dengue no Paraná: (a) cenário endêmico; (b) cenário epidêmico.

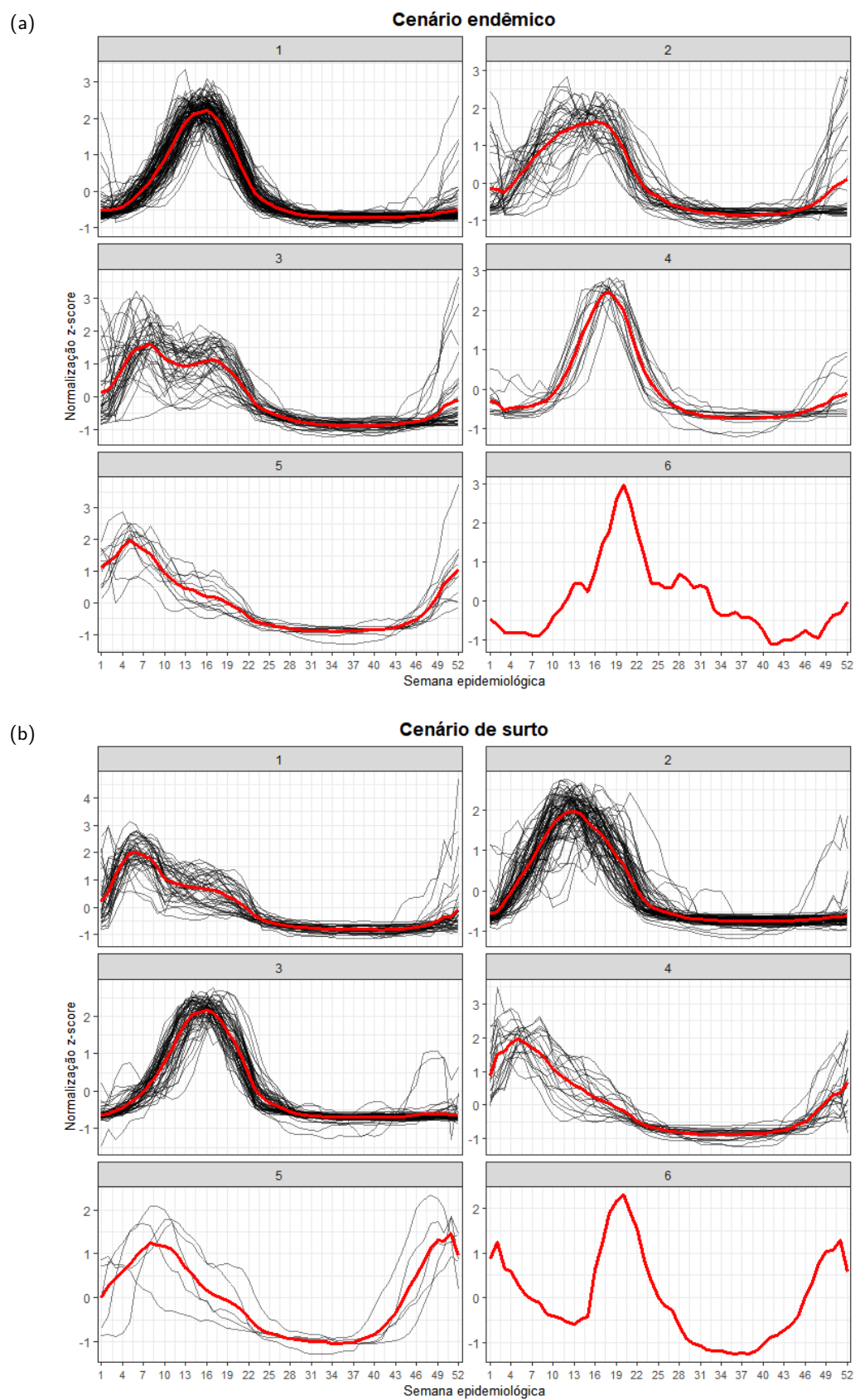


Figura 26 – Agrupamentos (*clusters*) identificados pelo algoritmo PER ($k = 6$) para a dengue no Paraná: (a) cenário endêmico; (b) cenário epidêmico.

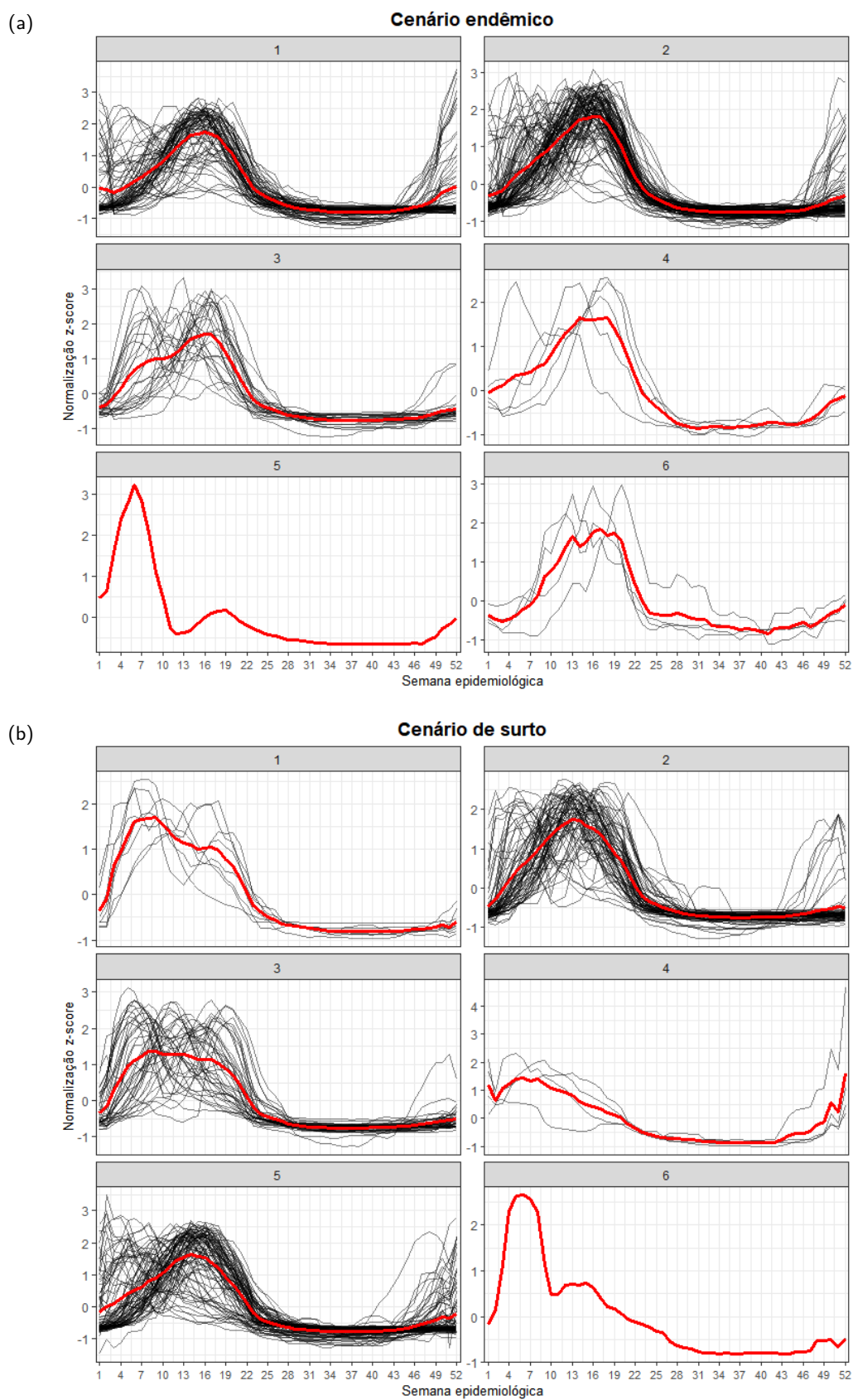


Figura 27 – Agrupamentos (*clusters*) identificados pelo algoritmo PIC ($k = 6$) para a dengue no Paraná: (a) cenário endêmico; (b) cenário epidêmico.

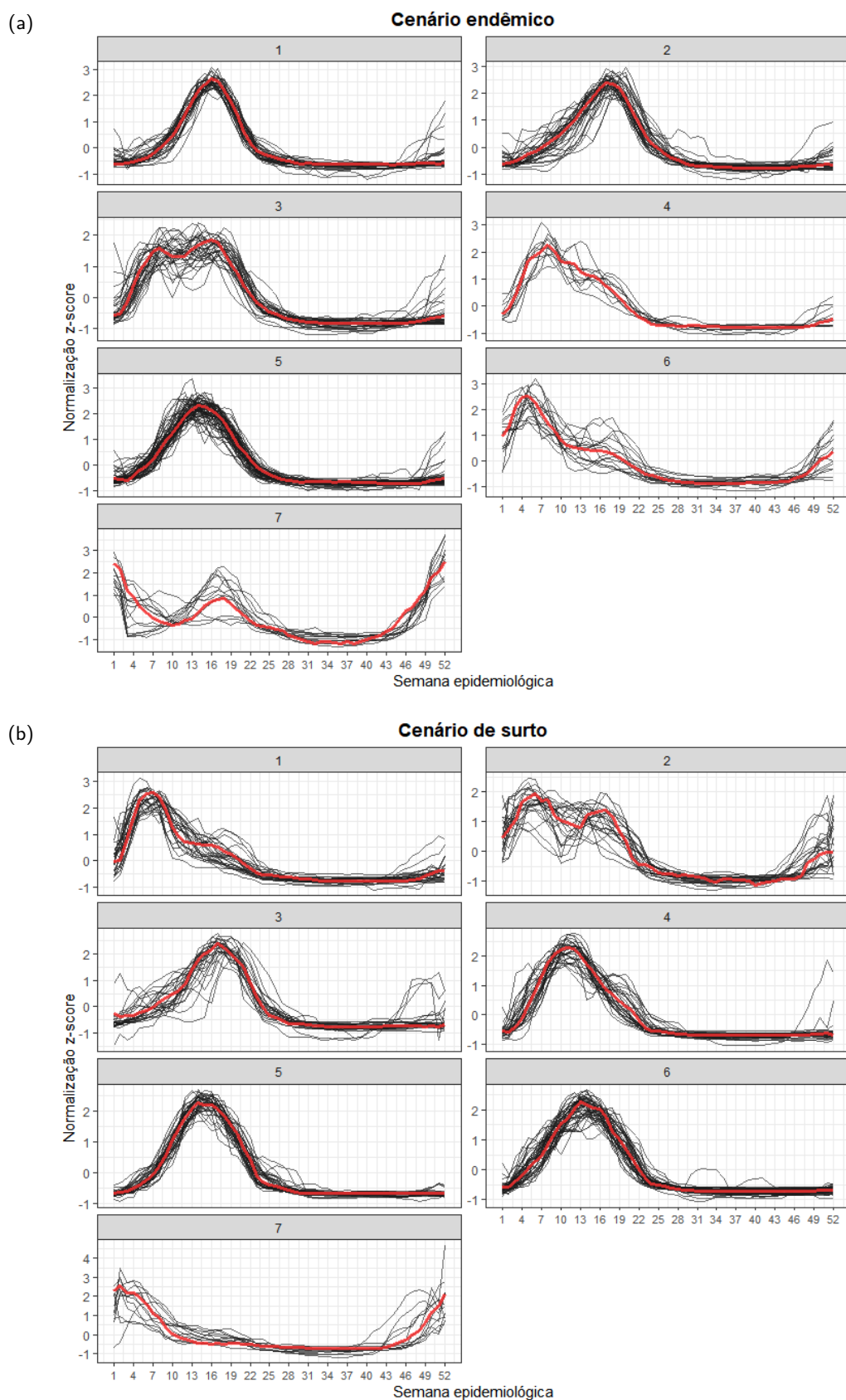
A.1.3 Configuração da partição ($k = 7$)

Figura 28 – Agrupamentos (*clusters*) identificados pelo algoritmo PAM ($k = 7$) para a dengue no Paraná: (a) cenário endêmico; (b) cenário epidêmico.

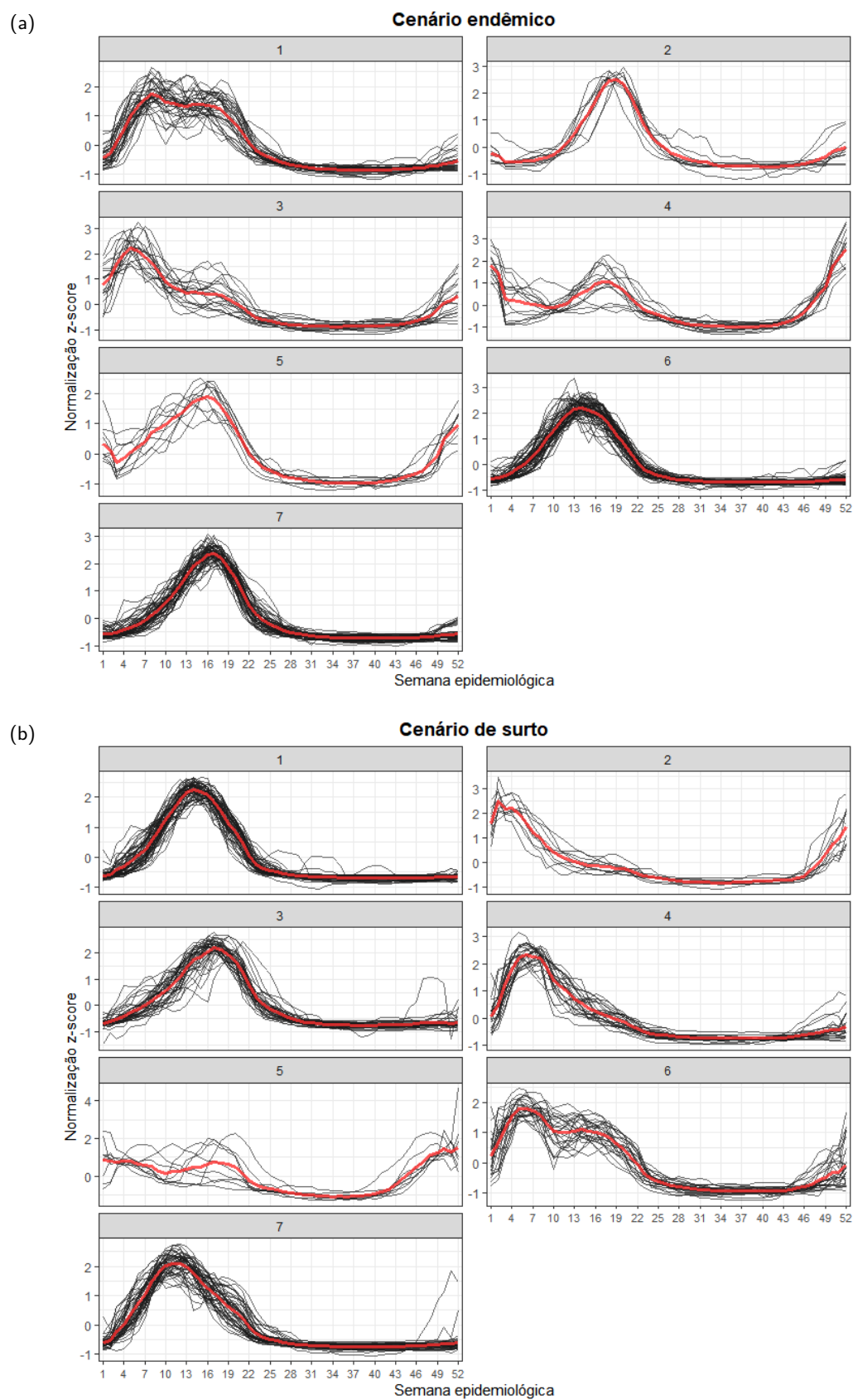


Figura 29 – Agrupamentos (*clusters*) identificados pelo algoritmo K-means ($k = 7$) para a dengue no Paraná: (a) cenário endêmico; (b) cenário epidêmico.

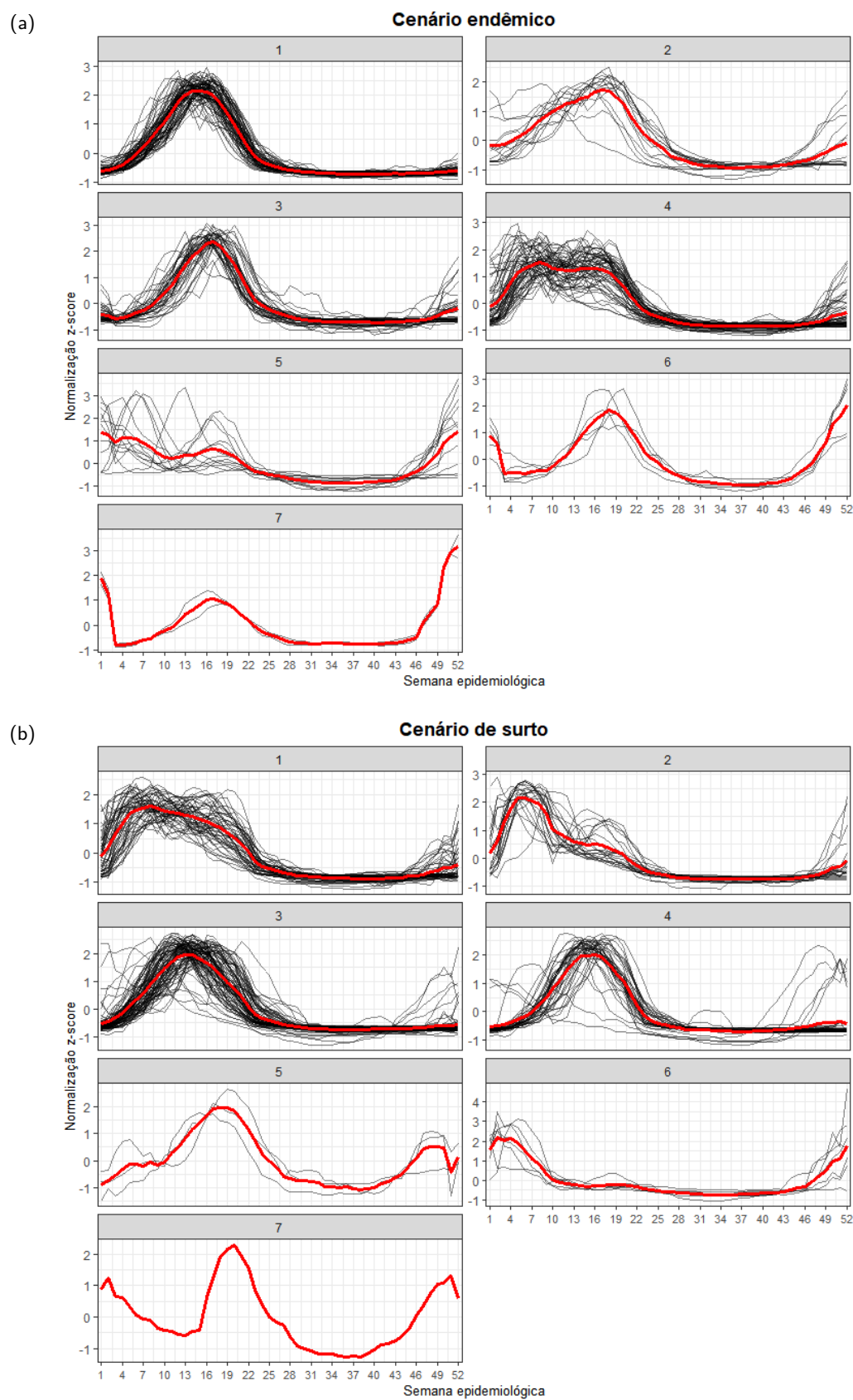


Figura 30 – Agrupamentos (*clusters*) identificados pelo algoritmo ACF ($k = 7$) para a dengue no Paraná: (a) cenário endêmico; (b) cenário epidêmico.

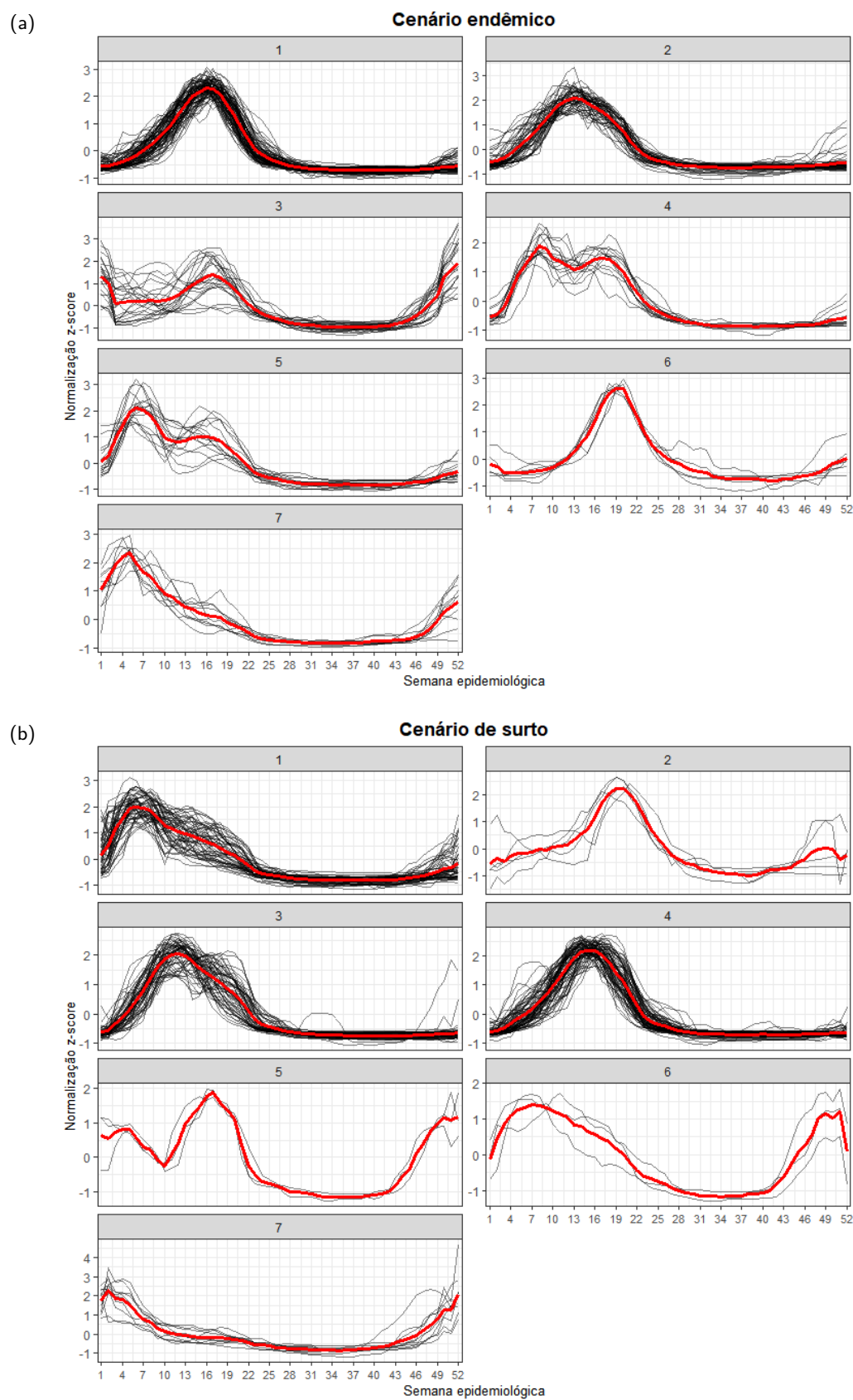


Figura 31 – Agrupamentos (*clusters*) identificados pelo algoritmo CORT ($k = 7$) para a dengue no Paraná: (a) cenário endêmico; (b) cenário epidêmico.

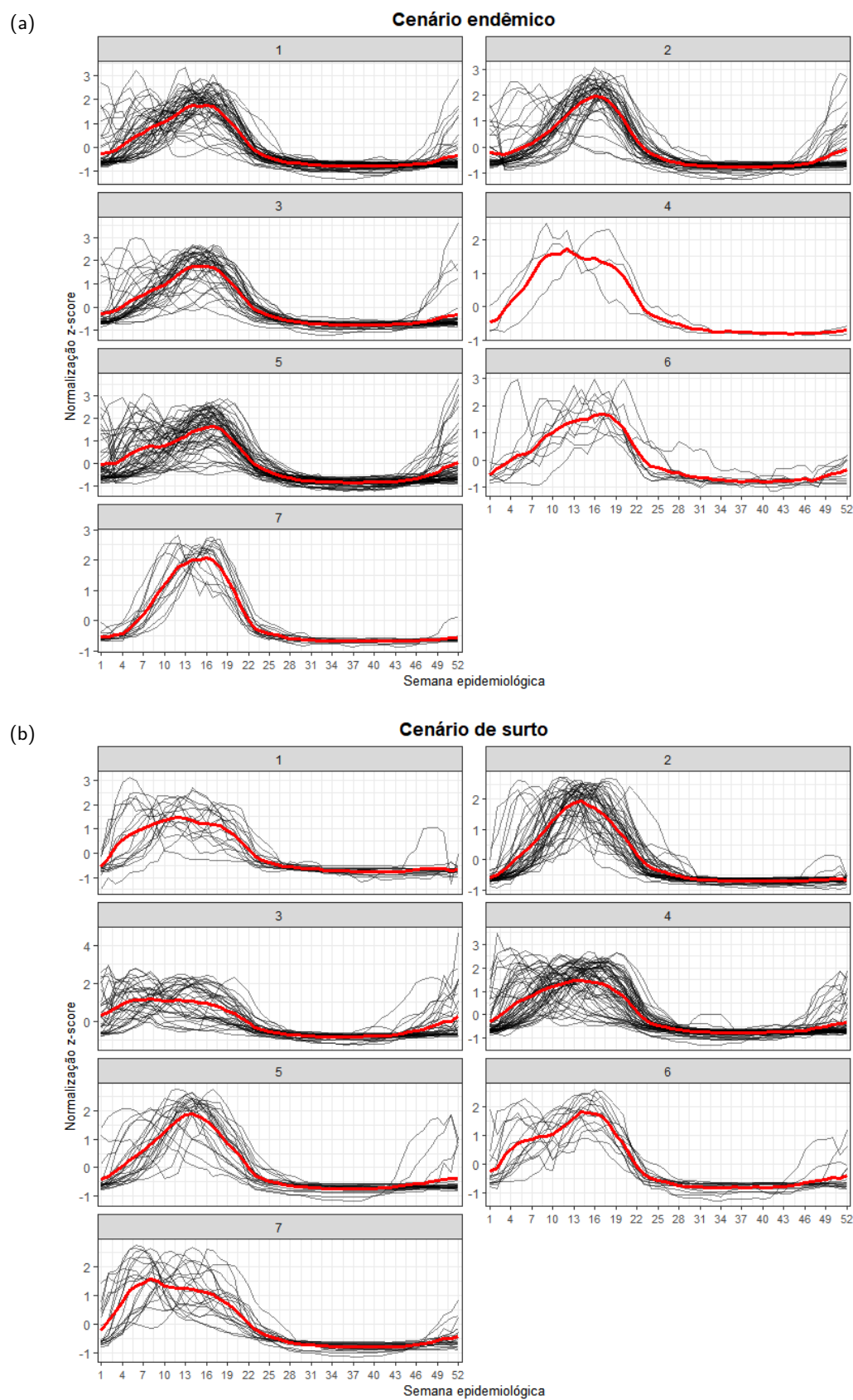


Figura 32 – Agrupamentos (*clusters*) identificados pelo algoritmo PDC ($k = 7$) para a dengue no Paraná: (a) cenário endêmico; (b) cenário epidêmico.

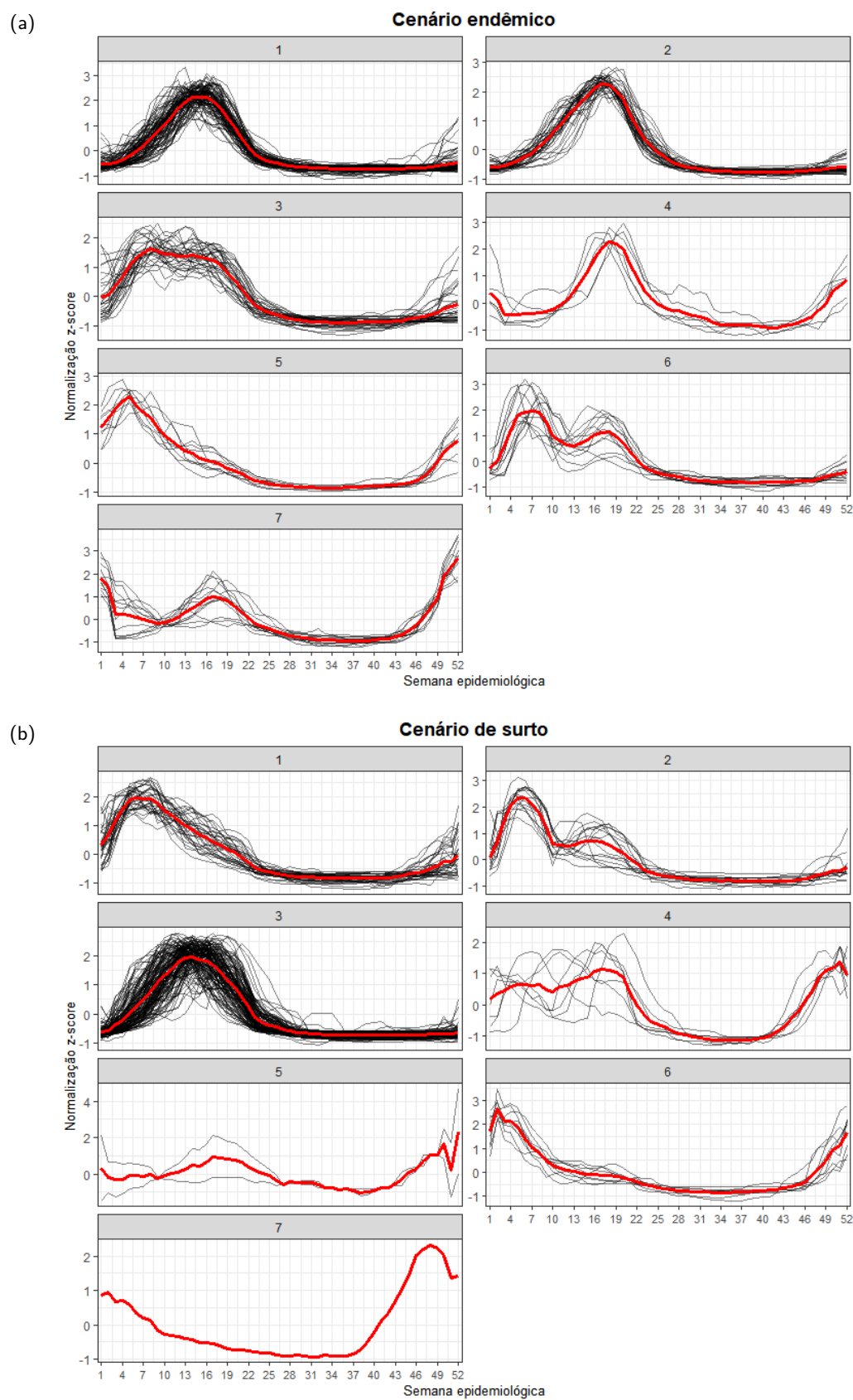


Figura 33 – Agrupamentos (*clusters*) identificados pelo algoritmo CID ($k = 7$) para a dengue no Paraná: (a) cenário endêmico; (b) cenário epidêmico.

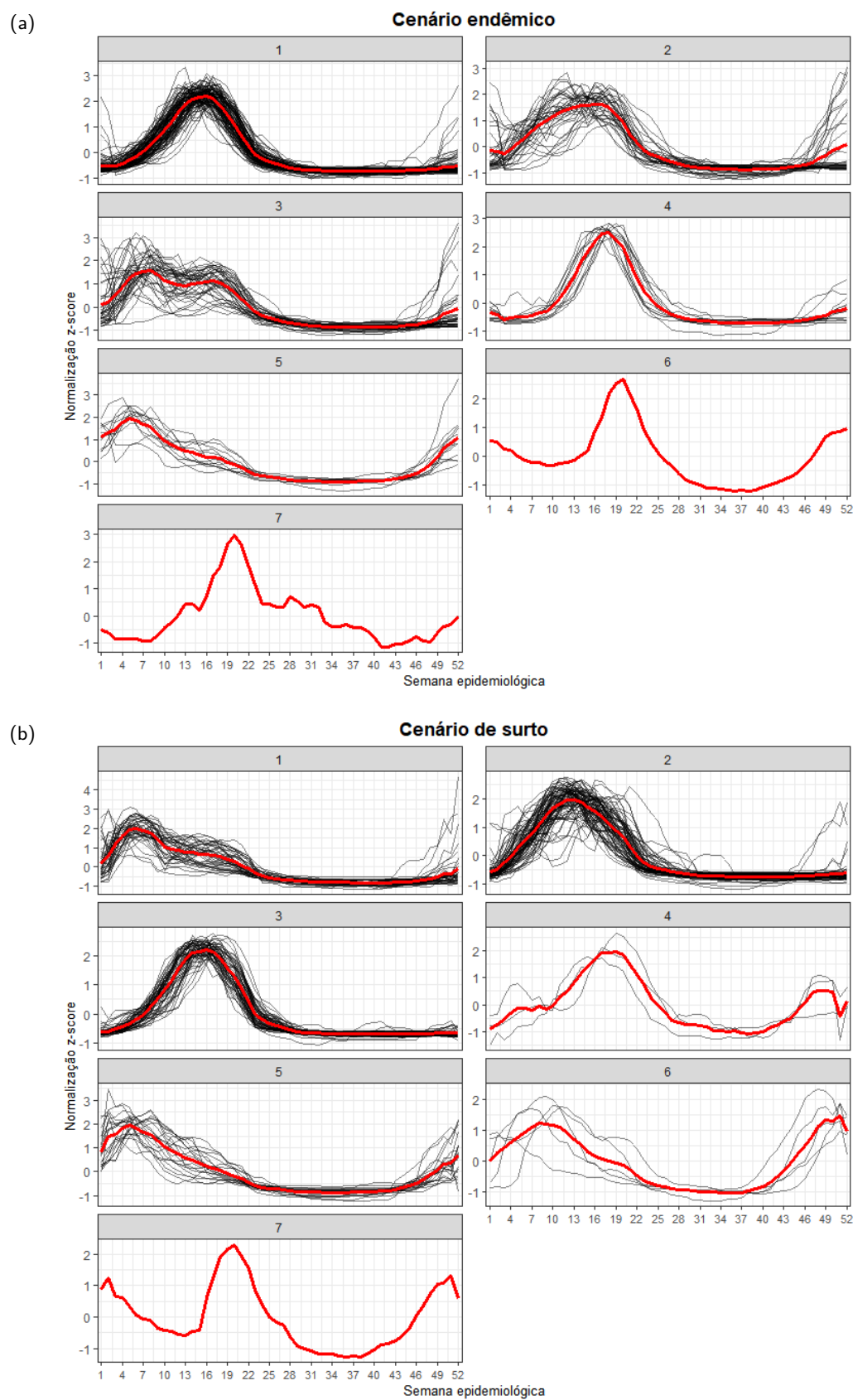


Figura 34 – Agrupamentos (*clusters*) identificados pelo algoritmo PER ($k = 7$) para a dengue no Paraná: (a) cenário endêmico; (b) cenário epidêmico.

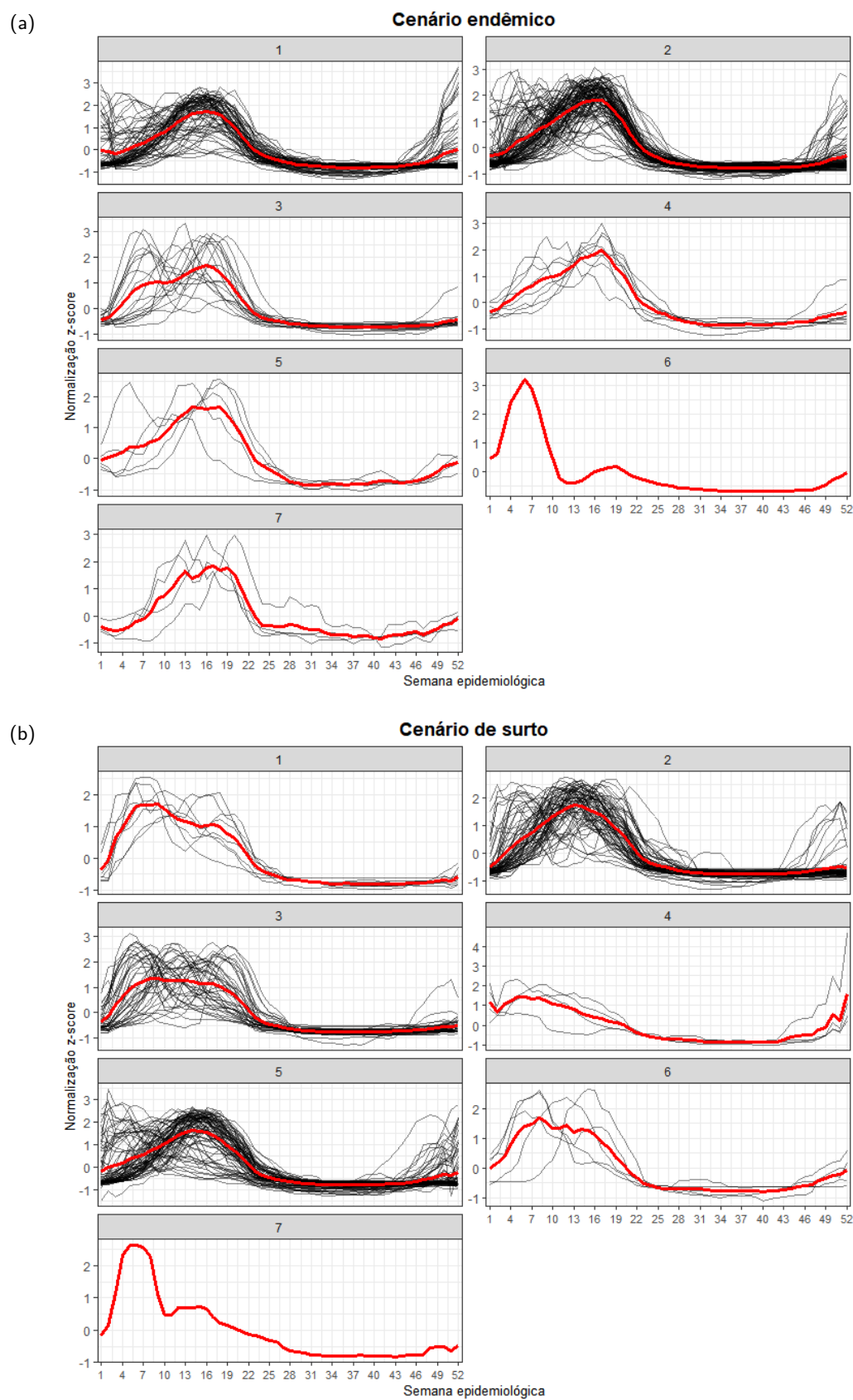


Figura 35 – Agrupamentos (*clusters*) identificados pelo algoritmo PIC ($k = 7$) para a dengue no Paraná: (a) cenário endêmico; (b) cenário epidêmico.

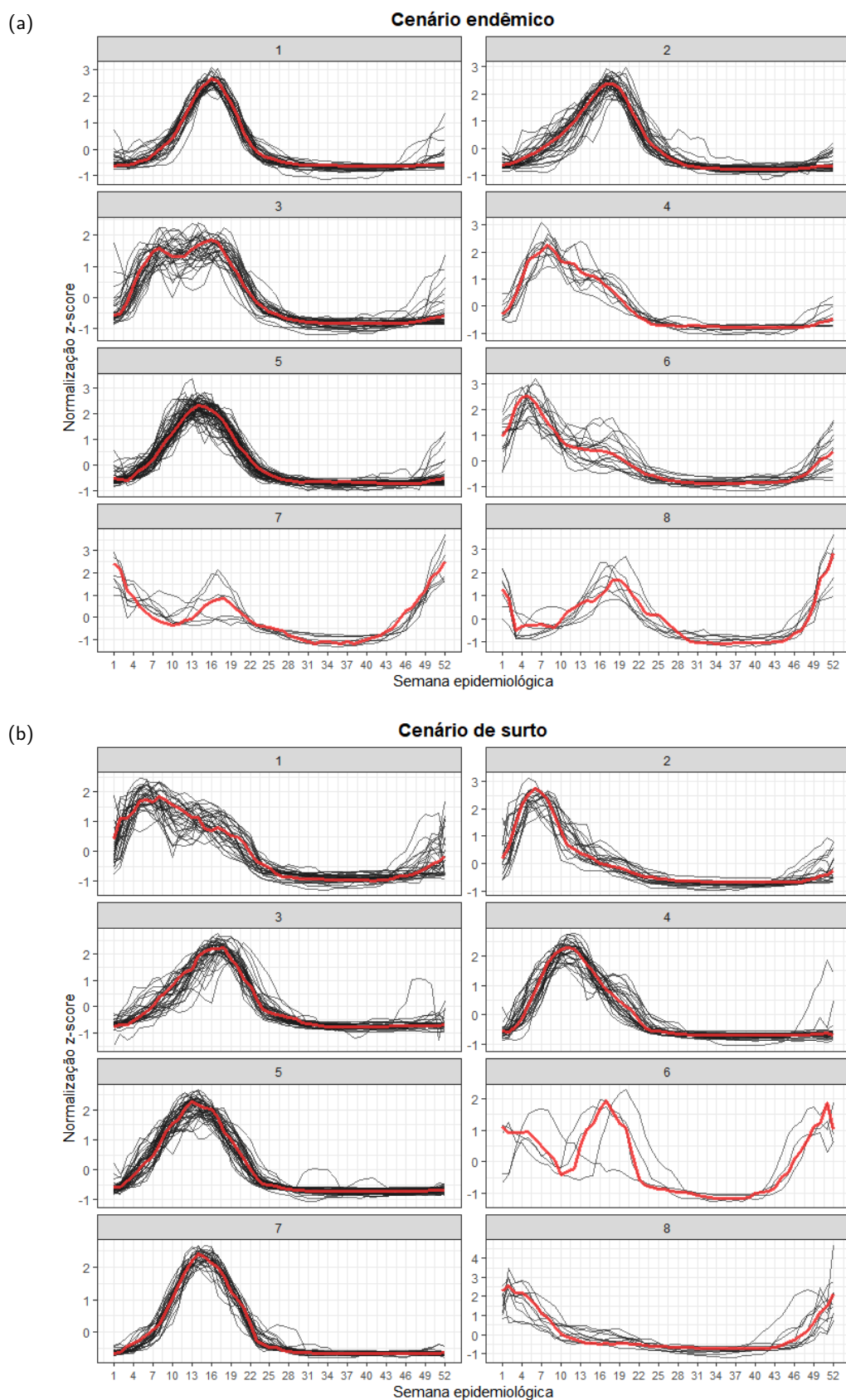
A.1.4 Configuração da partição ($k = 8$)

Figura 36 – Agrupamentos (*clusters*) identificados pelo algoritmo PAM ($k = 8$) para a dengue no Paraná: (a) cenário endêmico; (b) cenário epidêmico.

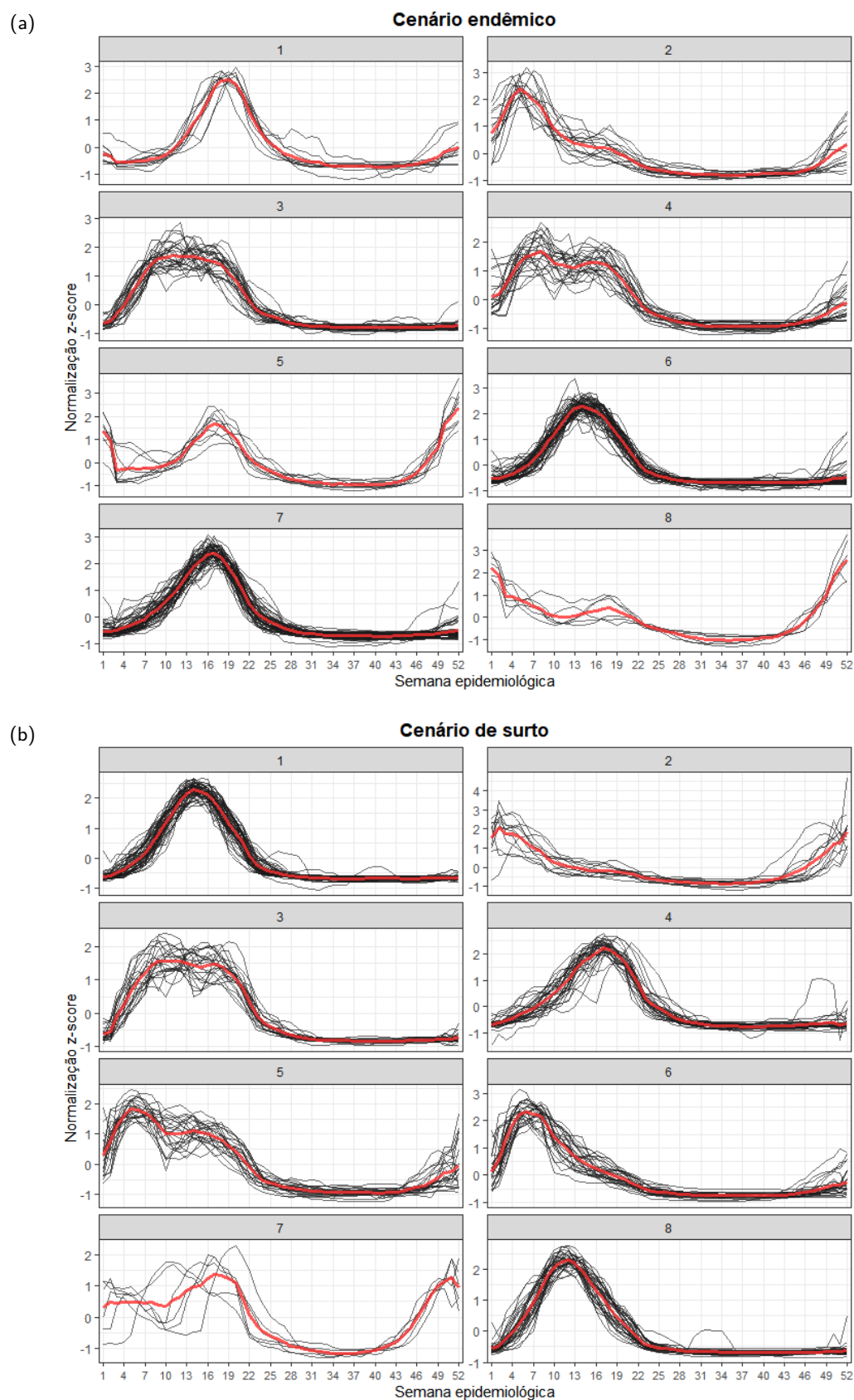


Figura 37 – Agrupamentos (*clusters*) identificados pelo algoritmo k-means ($k = 8$) para a dengue no Paraná: (a) cenário endêmico; (b) cenário epidêmico.

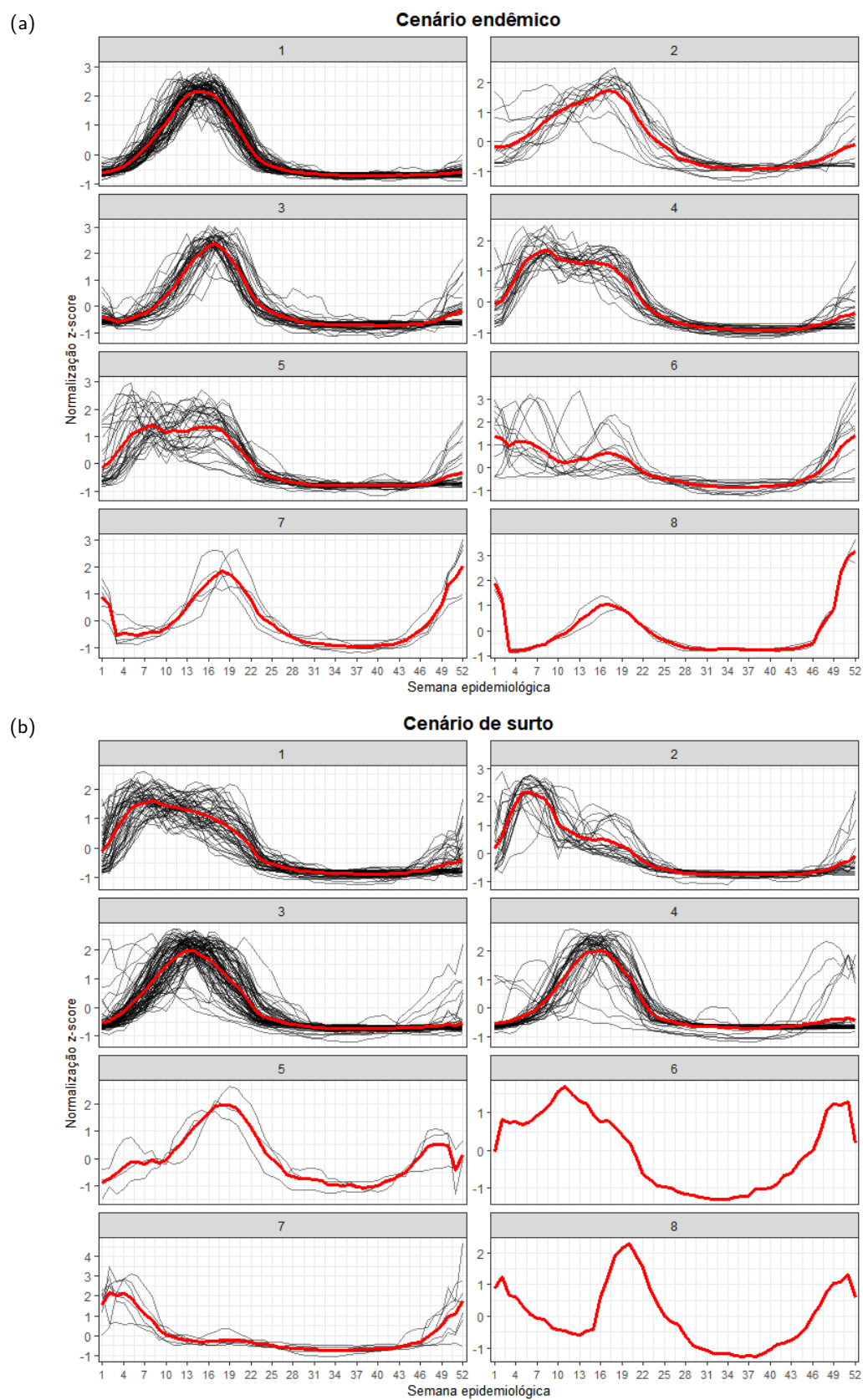


Figura 38 – Agrupamentos (*clusters*) identificados pelo algoritmo ACF ($k = 8$) para a dengue no Paraná: (a) cenário endêmico; (b) cenário epidêmico.

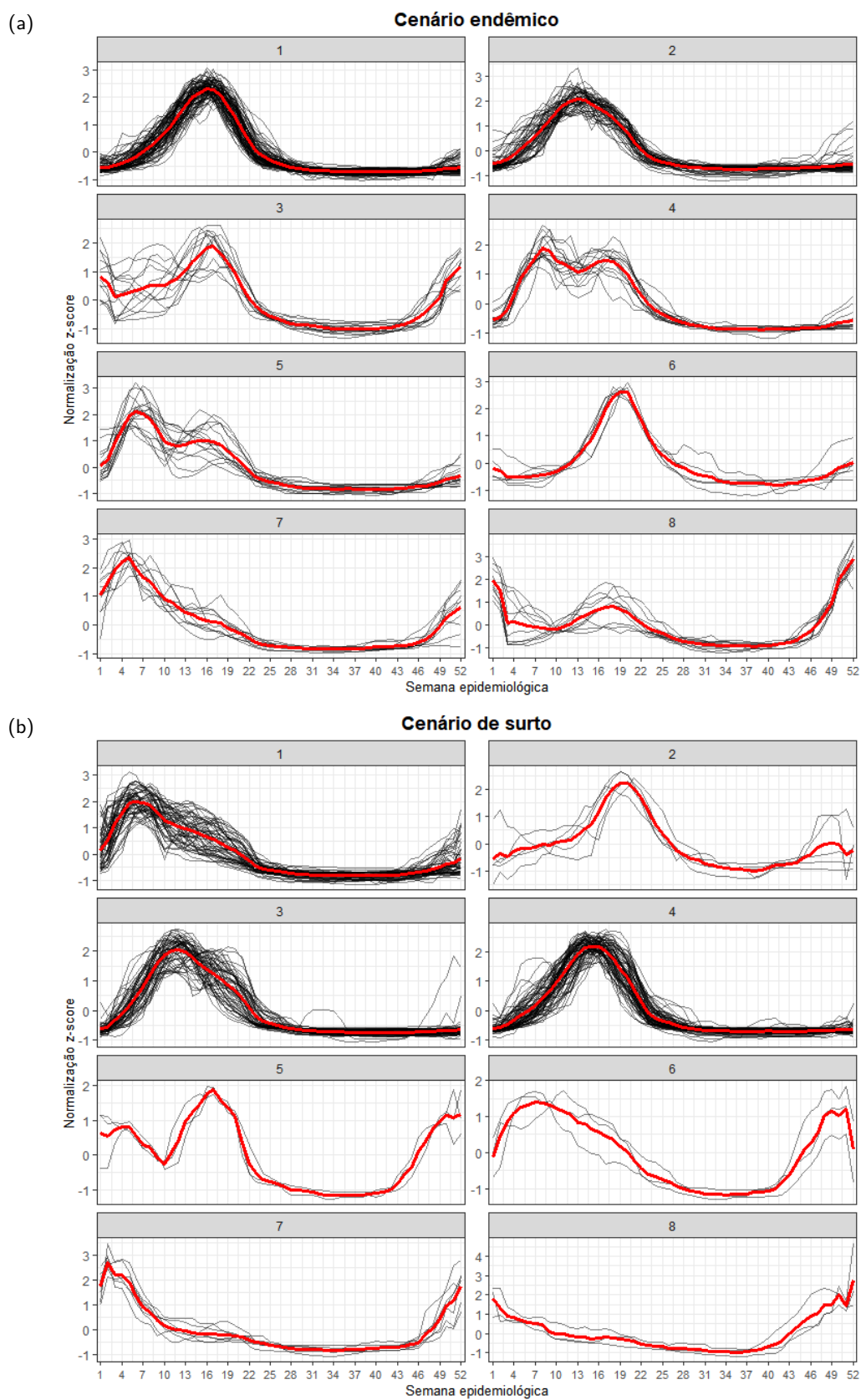


Figura 39 – Agrupamentos (*clusters*) identificados pelo algoritmo CORT ($k = 8$) para a dengue no Paraná: (a) cenário endêmico; (b) cenário epidêmico.

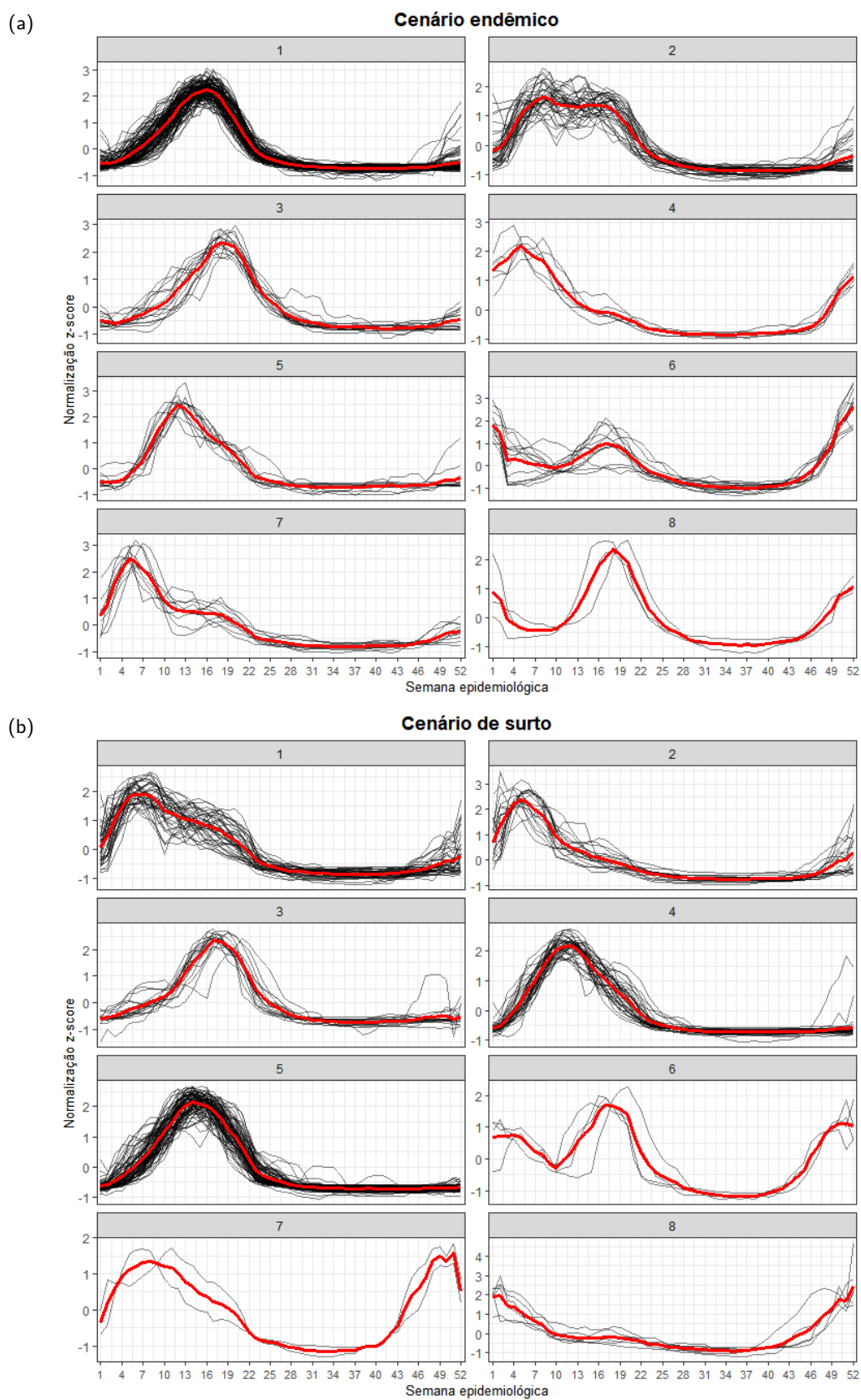


Figura 40 – Agrupamentos (*clusters*) identificados pelo algoritmo EUCL ($k = 8$) para a dengue no Paraná: (a) cenário endêmico; (b) cenário epidêmico.

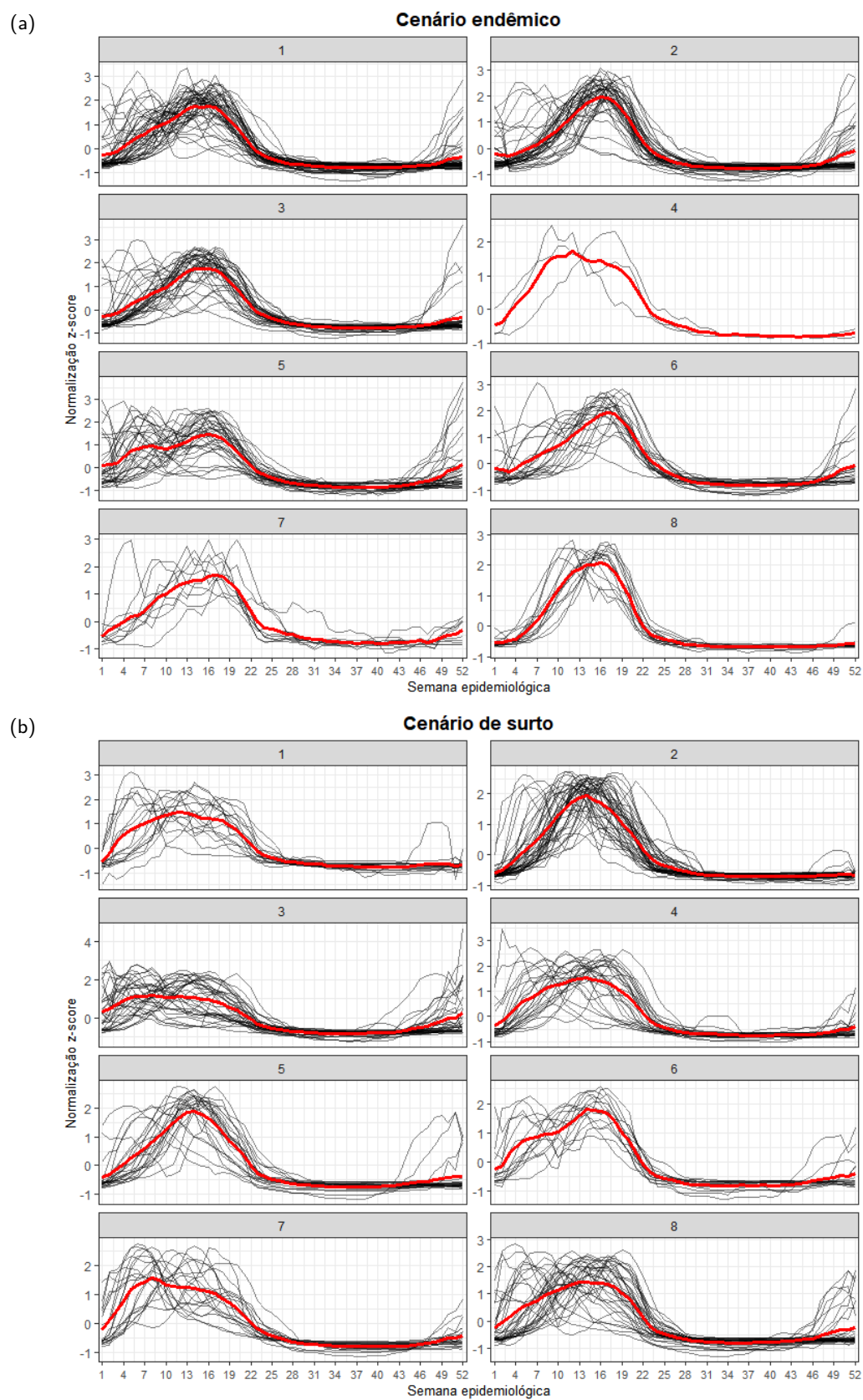


Figura 41 – Agrupamentos (*clusters*) identificados pelo algoritmo PDC ($k = 8$) para a dengue no Paraná: (a) cenário endêmico; (b) cenário epidêmico.

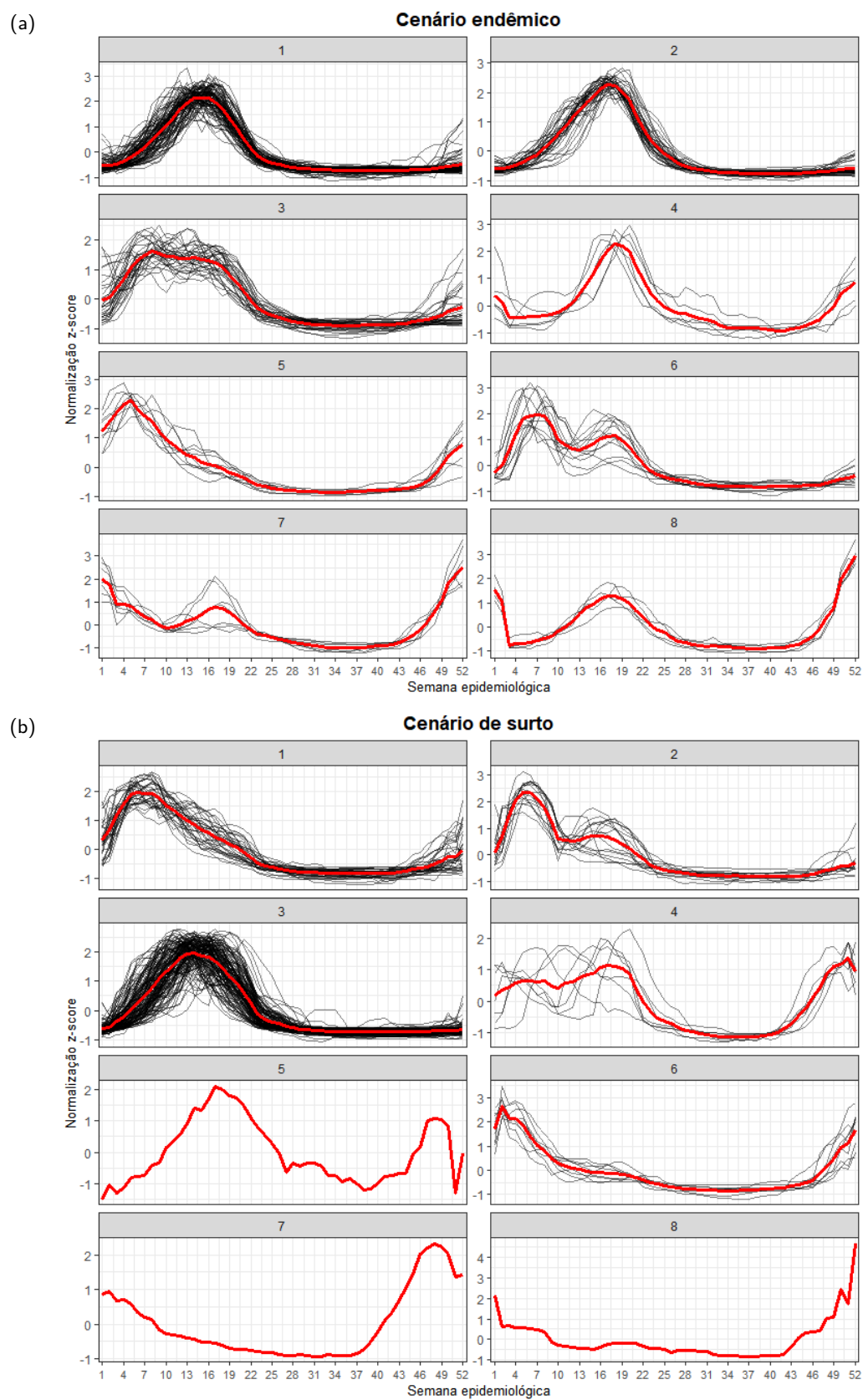


Figura 42 – Agrupamentos (*clusters*) identificados pelo algoritmo CID ($k = 8$) para a dengue no Paraná: (a) cenário endêmico; (b) cenário epidêmico.

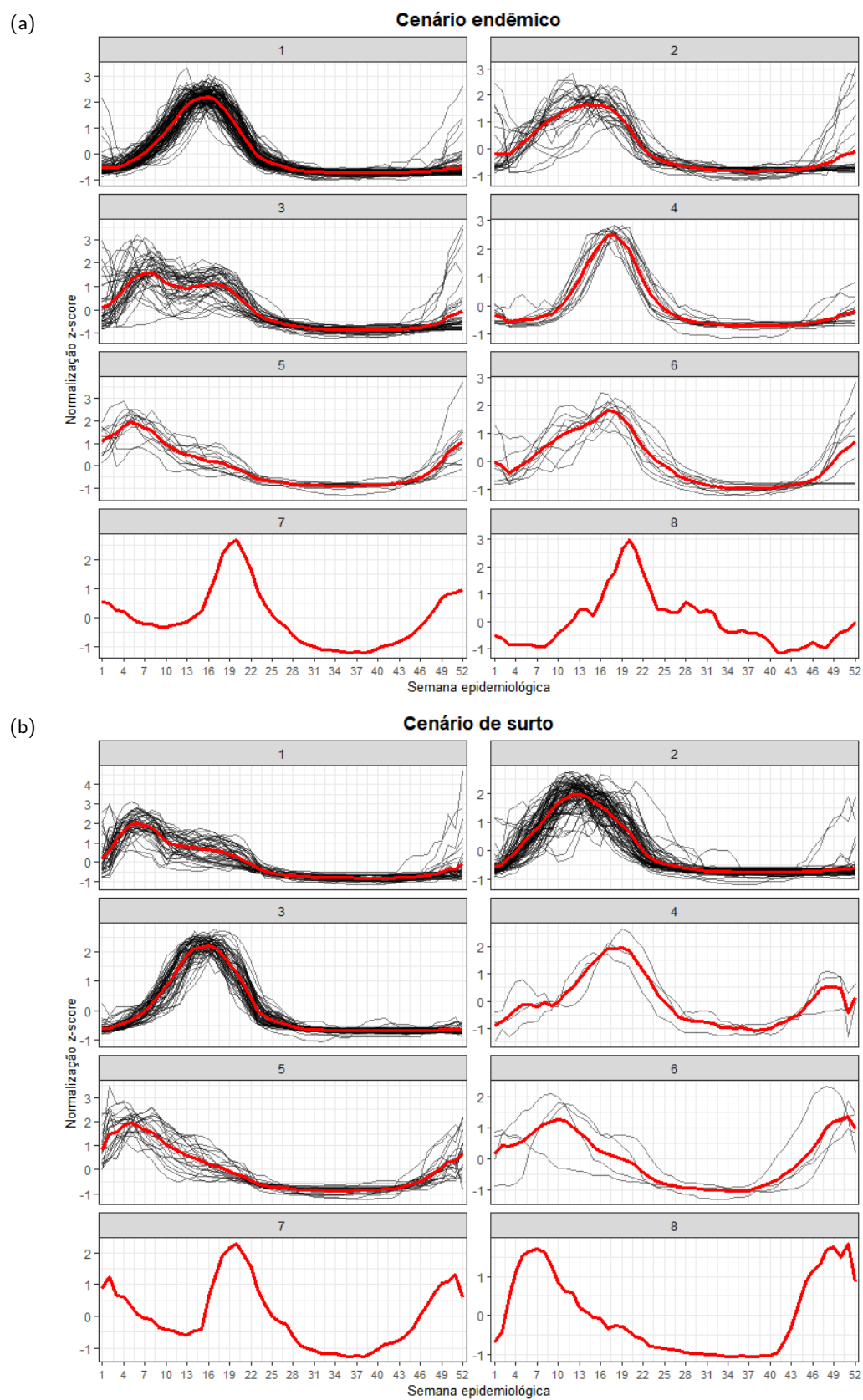


Figura 43 – Agrupamentos (*clusters*) identificados pelo algoritmo PER ($k = 8$) para a dengue no Paraná: (a) cenário endêmico; (b) cenário epidêmico.

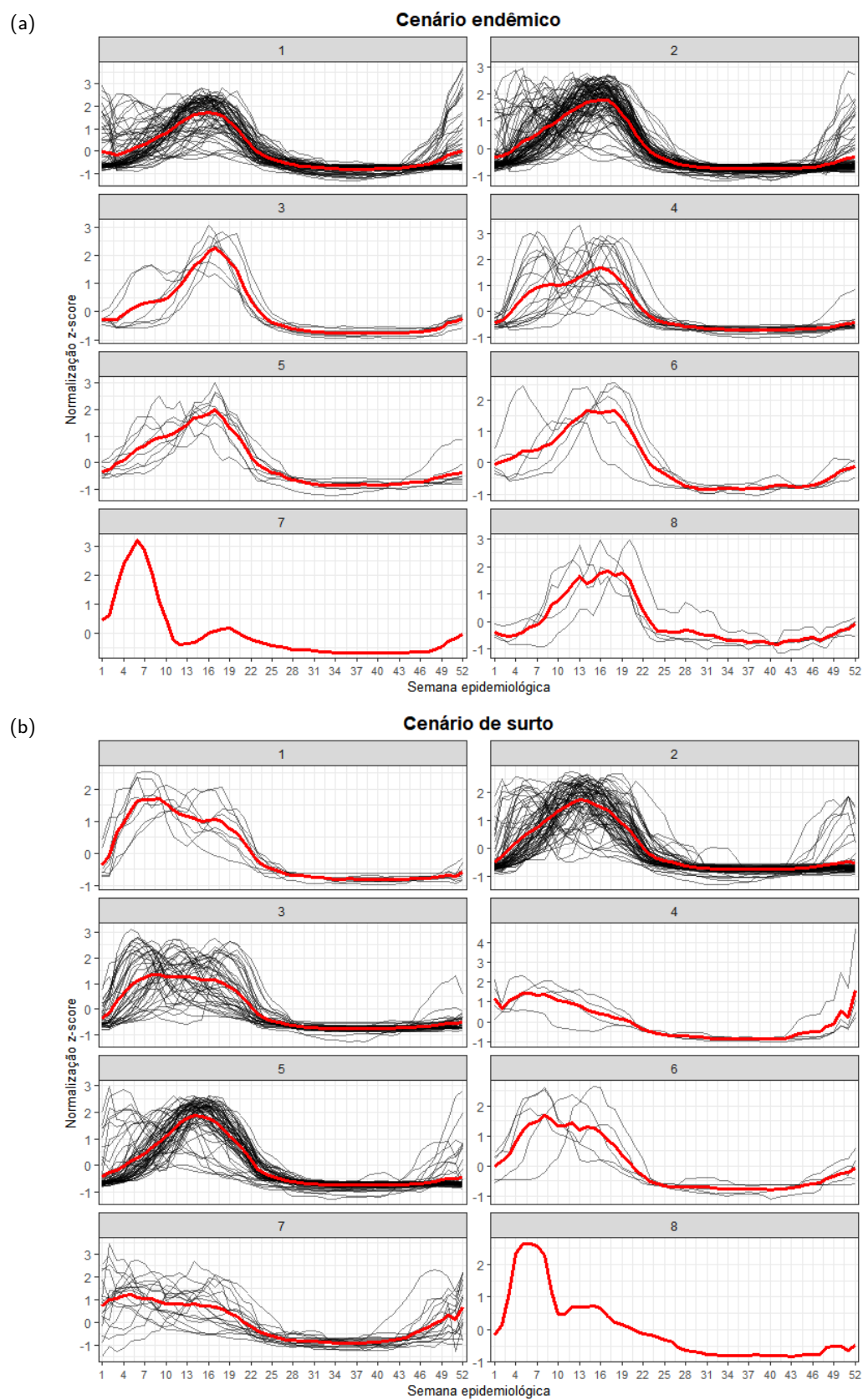


Figura 44 – Agrupamentos (*clusters*) identificados pelo algoritmo PIC ($k = 8$) para a dengue no Paraná: (a) cenário endêmico; (b) cenário epidêmico.

A.1.5 Avaliação da similaridade entre algoritmos de partição e métricas hierárquicas para $k = 5$ a $k = 8$

Tabela 10 – Comparação da similaridade (Sim) entre o método PAM e demais algoritmos/métricas para $k = 5$ a $k = 8$.

Algoritmo / Métrica	Sim (Cenário Endêmico)				Sim (Cenário de Surto)			
	k=5	k=6	k=7	k=8	k=5	k=6	k=7	k=8
<i>Algoritmo de Partição</i>								
K-means (Euclidiana)	0,973	0,748	0,680	0,709	0,790	0,756	0,750	0,659
<i>Algoritmos Hierárquicos</i>								
ACF	0,581	0,501	0,499	0,522	0,502	0,481	0,461	0,487
CID	0,667	0,638	0,634	0,654	0,547	0,555	0,489	0,535
CORT	0,660	0,644	0,580	0,553	0,632	0,618	0,575	0,595
EUCL	0,804	0,747	0,650	0,571	0,603	0,553	0,559	0,709
PDC	0,294	0,263	0,233	0,235	0,318	0,278	0,264	0,244
PER	0,533	0,514	0,442	0,413	0,606	0,563	0,537	0,514
PIC	0,294	0,281	0,270	0,241	0,339	0,305	0,293	0,288

A.2 Distribuição geográfica nos cenários de endemia e surto/epidemia para o algoritmo PAM

Tabela 11 – Distribuição dos municípios por classificação epidemiológica (Endemia e Epidemia) e agrupamentos (Clusters)

Tipo	Cluster	Cidades
Endemia	1	Abatiá, Alto Paraná, Alto Paraíso, Altônia, Arapongas, Assis Chateaubriand, Bandeirantes, Barra do Jacaré, Boa Vista da Aparecida, Cafelândia, Cambará, Carlópolis, Diamante D'Oeste, Dois Vizinhos, Entre Rios do Oeste, Faxinal, Florestópolis, Guaratuba, Iguaçu, Itaipulândia, Ivatuba, Jardim Olinda, Jataizinho, Lindoeste, Loanda, Mandaguari, Mandaguaçu, Marialva, Marilena, Marmeleiro, Marumbi, Nova Esperança, Nova Fátima, Paranguá, Paranapoema, Planalto, Pontal do Paraná, Primeiro de Maio, Pérola D'Oeste, Rancho Alegre D'Oeste, Roncador, Rondon, Sabáudia, Santa Cecília do Pavão, Santa Cruz de Monte Castelo, Santa Helena, Santa Inês, Santa Mônica, Santa Terezinha de Itaipu, Santo Antônio da Platina, São Carlos do Ivaí, São Jorge D'Oeste, São Jorge do Patrocínio, São Manoel do Paraná, São Sebastião da Amoreira, São Tomé, Telêmaco Borba, Terra Roxa, Três Barras do Paraná, Uraí, Ângulo.

Tipo	Cluster	Cidades
	2	Alto Piquiri, Alvorada do Sul, Amaporã, Assaí, Barbosa Ferraz, Brasilândia do Sul, Cafeara, Capitão Leônidas Marques, Centenário do Sul, Cruzeiro do Oeste, Douradina, Ibiporã, Iguatu, Itaguajé, Itambaracá, Jardim Alegre, Jussara, Lobato, Lupionópolis, Maringá, Nova Aurora, Ouro Verde do Oeste, Paiçandu, Planaltina do Paraná, Porecatu, Porto Rico, Quatro Pontes, Ribeirão do Pinhal, Salto do Lontra, Santa Fé, Santa Izabel do Oeste, Santo Antônio do Caiuá, Santo Inácio, Sarandi, Serranópolis do Iguaçu, São Miguel do Iguaçu, Terra Boa, Terra Rica, Tuneiras do Oeste, Xambrê.
	3	Ampére, Anahy, Andirá, Antonina, Apucarana, Araruna, Astorga, Boa Esperança, Bom Sucesso, Cafezal do Sul, Cambé, Campina da Lagoa, Campo Mourão, Capanema, Cascavel, Catanduvas, Cianorte, Cidade Gaúcha, Corbélia, Cornélio Procópio, Corumbataí do Sul, Cruzeiro do Sul, Céu Azul, Formosa do Oeste, Foz do Iguaçu, Francisco Alves, Francisco Beltrão, Fênix, Goioerê, Icaraíma, Indianópolis, Iporã, Ivaté, Jaguapitã, Janiópolis, Japurá, Leópolis, Londrina, Mamborê, Marechal Cândido Rondon, Maria Helena, Marilândia do Sul, Maripá, Matelândia, Matinhos, Medianeira, Mercedes, Mirador, Missal, Moreira Sales, Munhoz de Mello, Nova América da Colina, Nova Londrina, Nova Olímpia, Nova Santa Rosa, Ourizona, Palotina, Pato Bragado, Pato Branco, Perobal, Presidente Castelo Branco, Pérola, Quedas do Iguaçu, Ramilândia, Rancho Alegre, Realeza, Rolândia, Santa Lúcia, Santa Mariana, Santa Tereza do Oeste, Santo Antônio do Sudoeste, Sertanópolis, São Jorge do Ivaí, São João, São João do Ivaí, São Pedro do Ivaí, Tamarana, Tapira, Toledo, Tupãssi, Ubiratã, Umuarama, Vera Cruz do Oeste, Verê.
	4	Atalaia, Braganey, Doutor Camargo, Engenheiro Beltrão, Floresta, Guairaçá, Guaraci, Inajá, Ivaiporã, Jacarezinho, Kaloré, Nossa Senhora das Graças, Paranavaí, Peabiru, Quinta do Sol, Sertaneja, São João do Caiuá, São Pedro do Paraná, Tamboara.
	5	Colorado, Diamante do Norte, Floraí, Guaíra, Itambé, Jandaia do Sul, Juraanda, Mariluz, Paranacity, Paraíso do Norte, Querência do Norte, Santa Amélia, Santa Isabel do Ivaí, Tapejara.

Tipo	Cluster	Cidades
Epidemia	1	Abatiá, Alto Paraná, Alto Piquiri, Altônia, Antonina, Atalaia, Bandeirantes, Barbosa Ferraz, Cruzeiro do Oeste, Cruzeiro do Sul, Douradina, Doutor Camargo, Floresta, Guairaçá, Guaíra, Itaguajé, Ivatuba, Jandaia do Sul, Juranda, Kaloré, Lindoeste, Lupionópolis, Maria Helena, Marilena, Nossa Senhora das Graças, Paçandu, Paranaguá, Paranapoema, Paraíso do Norte, Planaltina do Paraná, Porecatu, Presidente Castelo Branco, Quatro Pontes, Quedas do Iguaçu, Querência do Norte, Santa Fé, Santa Inês, Santa Mônica, Santo Antônio do Caiuá, Santo Inácio, Sarandi, Serranópolis do Iguaçu, São João do Ivaí, Terra Rica, Terra Roxa, Tuneiras do Oeste.
	2	Alvorada do Sul, Anahy, Astorga, Barra do Jacaré, Braganey, Cafelândia, Cambará, Campina da Lagoa, Capanema, Carlópolis, Cornélio Procópio, Céu Azul, Diamante D'Oeste, Francisco Alves, Iguaçu, Iguatu, Iporã, Itaipu-lândia, Ivaté, Jaguapitã, Jardim Olinda, Loanda, Lobato, Mandaguacu, Marechal Cândido Rondon, Marmeleiro, Matinhos, Missal, Munhoz de Mello, Nova Fátima, Pato Branco, Pontal do Paraná, Porto Rico, Primeiro de Maio, Rancho Alegre, Sabáudia, Santa Cecília do Pavão, Santa Helena, Santa Mariana, São Jorge do Patrocínio, São João, São Manoel do Paraná, São Miguel do Iguaçu, São Pedro do Ivaí, São Sebastião da Amoreira, São Tomé, Tapira, Telêmaco Borba, Uraí, Ângulo.
	3	Amaporã, Ampére, Araruna, Assaí, Assis Chateaubriand, Boa Vista da Aparecida, Cafeara, Campo Mourão, Cidade Gaúcha, Corumbataí do Sul, Dois Vizinhos, Faxinal, Foz do Iguaçu, Guaraci, Itambaracá, Janiópolis, Jardim Alegre, Jussara, Mandaguari, Maripá, Nova América da Colina, Nova Esperança, Nova Santa Rosa, Ribeirão do Pinhal, Rolândia, Roncador, Salto do Lontra, Santa Cruz de Monte Castelo, Santa Lúcia, Santa Tereza do Oeste, Tapejara, Três Barras do Paraná, Tupãssi.
	4	Alto Paraíso, Andirá, Arapongas, Boa Esperança, Bom Sucesso, Brasilândia do Sul, Cafezal do Sul, Cambé, Cascavel, Catanduvas, Centenário do Sul, Cianorte, Corbélia, Entre Rios do Oeste, Formosa do Oeste, Francisco Beltrão, Fênix, Goioerê, Guaratuba, Ibiporã, Icaraíma, Indianópolis, Japurá, Jataizinho, Leópolis, Londrina, Mamborê, Marialva, Marilândia do Sul, Maringá, Marumbi, Matelândia, Medianeira, Mercedes, Mirador, Moreira Sales, Nova Aurora, Nova Londrina, Nova Olímpia, Ourizona, Ouro Verde do Oeste, Palotina, Pato Bragado, Perobal, Planalto, Pérola, Pérola D'Oeste, Ramilândia, Rancho Alegre D'Oeste, Realeza, Rondon, Santa Terezinha de Itaipu, Santo Antônio da Platina, Santo Antônio do Sudoeste, Sertaneja, Sertanópolis, São Carlos do Ivaí, São Jorge D'Oeste, São Jorge do Ivaí, Tamarana, Terra Boa, Toledo, Ubatã, Umuarama, Vera Cruz do Oeste, Verê, Xambrê.

Tipo	Cluster	Cidades
	5	Apucarana, Capitão Leônidas Marques, Colorado, Diamante do Norte, Engenheiro Beltrão, Floraí, Florestópolis, Inajá, Itambé, Ivaiporã, Jacarezinho, Mariluz, Paranacity, Paranaíba, Peabiru, Quinta do Sol, Santa Amélia, Santa Isabel do Ivaí, Santa Izabel do Oeste, São João do Caiuá, São Pedro do Paraná, Tamboara.

A.3 Distribuição espacial nos cenários de endemia e surto/epidemia para o algoritmo PAM

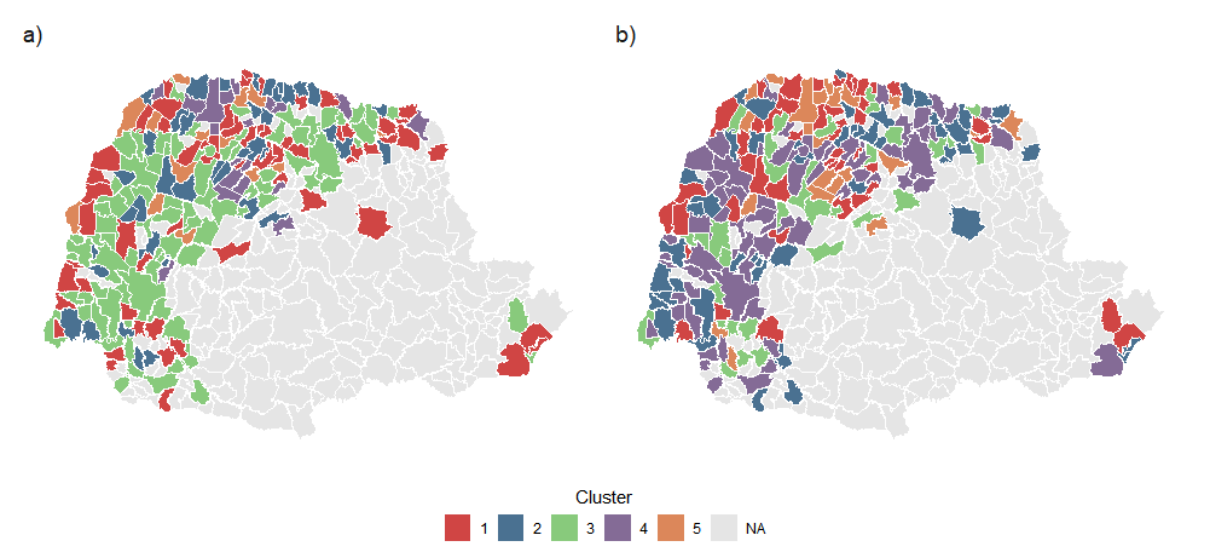


Figura 45 – Distribuição espacial dos clusters por classificação epidemiológica: a) endemia e b) surto/epidemia. Observação: as cores atribuídas são independentes e aleatória entre os mapas.