

**Universidade de São Paulo
Instituto de Matemática e Estatística**

Centro de Estatística Aplicada

Relatório de Análise Estatística

RAE-CEA–11P25

RELATÓRIO DE ANÁLISE ESTATÍSTICA SOBRE O PROJETO:

“Estudo da incidência de casos de dengue no município de Campo Grande no período de 2017 a 2024 com a aplicação de diferentes métodos de análise”

Estevão Kusaba Rodrigues

Melissa Tiemi Tanaka

Chang Chiann

Mateus Gonzalez de Freitas Pinto

São Paulo, junho de 2025

CENTRO DE ESTATÍSTICA APLICADA - CEA – USP

TÍTULO: Relatório de Análise Estatística sobre o Projeto: “Estudo da incidência de casos de dengue no município de Campo Grande no período de 2017 a 2024 com a aplicação de diferentes métodos de análise”.

PESQUISADORA: Vanessa Coelho de Aquino Benjaino Ferraz

COLABORADORES: Margarete Knoch, Victor Vohryzek Ferezin

INSTITUIÇÃO: Secretaria Municipal de Saúde de Campo Grande - MS

FINALIDADE DO PROJETO: Pesquisa em serviço

RESPONSÁVEIS PELA ANÁLISE: Estevão Kusaba Rodrigues
Melissa Tiemi Tanaka
Chang Chiann
Mateus Gonzalez de Freitas Pinto

REFERÊNCIA DESTE TRABALHO: RODRIGUES, E.K.; TANAKA, M.T.; CHIANN, C.; PINTO, M.G.F. **Relatório de análise estatística sobre o projeto: “Estudo da incidência de casos de dengue no município de Campo Grande no período de 2017 a 2024 com a aplicação de diferentes métodos de análise”.** São Paulo, IME-USP, 2025. (RAE–CEA-25P11)

FICHA TÉCNICA

REFERÊNCIAS BIBLIOGRÁFICAS:

AGRESTI, A. (2013) ***Categorical data analysis***. 3. ed. Hoboken: John Wiley & Sons.

ALMEIDA, W.M.M.; FERNANDES R.O.; GUARALDO E. (2023). ***Acesso às Áreas Verdes Urbanas e Equidade Verde: um estudo em Campo Grande, MS***. Interações (Campo Grande), 24(1), 281–297.

BUSSAB, W.O.; MORETTIN, P.A. (2024). ***Estatística Básica***. São Paulo: Ed. Saraiva, 10^a ed.

CHOLLET, F. (2021) ***Deep Learning with Python***. 2. ed. New York: Manning Publications.

BRASIL (2023). Ministério da Saúde. Secretaria de Vigilância em Saúde. Departamento de Imunização e Doenças Transmissíveis. Coordenação-Geral de Vigilância de Arboviroses. ***Nota Técnica nº 1/2023-CGAR/DEIDT/SVS/MS: Orientações para o monitoramento de casos de arboviroses urbanas com o uso de diagrama de controle utilizando a medida de tendência central mediana***. Disponível em: <https://www.gov.br/saude/pt-br/centrais-de-conteudo/publicacoes/notas-tecnicas/2023/nota-tecnica-no-01-2023-cgarb-deidt-svs-ms/> Acesso em: 11 de maio de 2025.

BRASIL (2024). Ministério da Saúde. Secretaria de Vigilância em Saúde. Departamento de Imunização e Doenças Transmissíveis. Coordenação-Geral de Vigilância de Arboviroses. ***Boletim Epidemiológico: Situação da dengue no Brasil***. Disponível em: <https://www.gov.br/saude/pt-br/centrais-de-conteudo/publicacoes/boletins/epidemiologicos/edicoes/2024/boletim-epidemiologico-volume-55-no-11.pdf>. Acesso em: 11 de junho de 2025.

IBGE (2022). ***Censo Demográfico (2022) - Tabelas - Agregados por Setores Censitários preliminares: População e Domicílios***. Disponível em: <https://www.ibge.gov.br/estatisticas/sociais/populacao/22827-censo-demografico-2022.html?edicao=39499&t=resultados> Acesso em: 03 de maio de 2025

PROGRAMAS COMPUTACIONAIS UTILIZADOS:

Microsoft Word Online (Microsoft 365)

R for Windows 4.3.3 (2025-02-28 ucrt)

RStudio for Windows 2024.04.2+764

Python for Windows 3.13.3

TÉCNICAS ESTATÍSTICAS UTILIZADAS

Análise de Dados Categorizados (06:030)

Séries Temporais (11:010)

Análise Descritiva Multidimensional (03:020)

ÁREA DE APLICAÇÃO

Medicina - Epidemiologia (14:040)

Resumo

Neste trabalho, foi realizada uma análise estatística abrangente sobre a incidência de casos de dengue no município de Campo Grande, utilizando o banco de dados de notificações individuais de 2008 a 2024. O estudo teve como objetivos caracterizar o padrão epidemiológico da doença, avaliar métodos de detecção de epidemias por meio do canal endêmico e aplicar técnicas modernas de modelagem preditiva com redes neurais Long Short-Term Memory (LSTM).

A análise descritiva sugere importantes padrões espaciais e temporais, destacando distritos com maior incidência absoluta e relativa, além de evidenciar a sazonalidade anual e a diferença entre anos epidêmicos e não epidêmicos. Também foram observadas variações no perfil demográfico dos casos e na distribuição dos registros ao longo das semanas epidemiológicas, com maior concentração de notificações no primeiro semestre dos anos epidêmicos.

Na etapa de vigilância, foram construídos canais endêmicos pelos métodos de média móvel e de mediana com intervalos interquartis, conforme recomendações do Ministério da Saúde. A comparação entre os métodos mostrou que o canal de medianas apresenta maior sensibilidade para detecção precoce de epidemias, enquanto o canal de médias é mais específico, com ambos apresentando desempenho global semelhante na discriminação entre semanas epidêmicas e não epidêmicas, segundo as curvas ROC (Receiver Operating Characteristic) os valores de AUC (Area Under the ROC Curve).

Por fim, a modelagem das séries temporais com redes LSTM demonstrou ser eficaz para captar tanto a sazonalidade quanto a variabilidade dos casos, permitindo previsões robustas para o município e seus distritos. Os modelos ajustados apresentaram bom desempenho na reconstituição da dinâmica da série e indicaram a possibilidade de novos picos epidêmicos para o ano de 2025, reforçando o potencial das LSTM's como ferramenta de apoio à vigilância epidemiológica e ao planejamento em saúde pública.

Sumário

1. Introdução.....	8
2. Objetivos	8
3. Descrição do estudo	9
4. Descrição das variáveis	10
5. Análise descritiva	11
6. Canal endêmico	13
7. Séries temporais	15
8. Conclusões.....	17
APÊNDICE A	20
APÊNDICE B	23
ANEXO.....	41

1. Introdução

A dengue representa um desafio recorrente para a saúde pública em Campo Grande, impulsionada por fatores como adensamento populacional, condições ambientais e a presença do vetor *Aedes aegypti*. O monitoramento e a resposta a esse agravo são coordenados por órgãos como o Centro de Informações Estratégicas de Vigilância em Saúde de Campo Grande (CIEVS-CG), cuja atuação visa garantir a detecção precoce e a rápida disseminação de informações para subsidiar decisões em emergências. Nesse contexto, torna-se fundamental identificar de forma ágil e precisa a ocorrência de epidemias, permitindo que as intervenções de controle e prevenção sejam implementadas de maneira oportuna.

A análise sistemática dos casos de dengue no município é viabilizada pelo Sistema de Informação de Agravos de Notificação (SINAN), que reúne dados essenciais para o acompanhamento da dinâmica da doença. Considerando as particularidades dos diferentes distritos sanitários de Campo Grande, este estudo busca contribuir para o aprimoramento da vigilância epidemiológica local, proporcionando subsídios para estratégias mais eficazes de enfrentamento da dengue e para a redução do impacto das epidemias sobre a população.

2. Objetivos

O objetivo principal deste estudo é analisar a incidência de dengue no município de Campo Grande, com ênfase na identificação e previsão de períodos epidêmicos, a partir de dados registrados entre 2008 e 2024 pelo Sistema de Informação de Agravos de Notificação (SINAN).

De forma mais detalhada, o projeto busca:

- Apresentar o canal endêmico da dengue com base nos cálculos da mediana e da distribuição em quartis;
- Comparar diferentes métodos de cálculo do canal endêmico;

- Analisar a distribuição dos casos de dengue por distrito sanitário do município de Campo Grande;
- Construir séries históricas de incidência da dengue considerando o crescimento populacional;
- Buscar e avaliar modelos preditivos para a incidência de dengue;
- Comparar a incidência e sua influência nos modelos preditivos em anos epidêmicos e não epidêmicos.

3. Descrição do Estudo

O presente estudo foi realizado a partir de um banco de dados composto por notificações de casos de dengue registrados no município de Campo Grande (MS). Os dados utilizados foram obtidos junto ao Centro de Informações Estratégicas de Vigilância em Saúde de Campo Grande (CIEVS-CG), órgão responsável pelo monitoramento epidemiológico local e vinculado à Secretaria Municipal de Saúde. O CIEVS-CG integra a Rede Nacional de Monitoramento e Respostas às Emergências em Saúde Pública e atua na detecção, análise e rápida disseminação de informações sobre agravos de importância para a saúde pública.

A coleta dos dados ocorre de forma padronizada nas unidades de saúde do município. Sempre que um paciente é atendido com suspeita de dengue em uma unidade de pronto atendimento, os profissionais de saúde são orientados a preencher uma ficha de notificação específica. Esta ficha contém informações como data da notificação, classificação final do caso, evolução clínica, bairro e distrito de residência do paciente, entre outros dados relevantes. As notificações são então inseridas no Sistema de Informação de Agravos de Notificação (SINAN), plataforma nacional responsável por centralizar e armazenar os dados de doenças de notificação compulsória em todo o Brasil. Consolidam-se essas informações no SINAN, disponibilizando-as para análise pelos órgãos de vigilância epidemiológica em diferentes esferas administrativas.

O banco de dados fornecido é composto por todas as notificações de dengue referentes ao município de Campo Grande no período de 2008 até 2024. Apesar de o título

do projeto mencionar o período de 2017 a 2024, serão utilizados os dados desde 2008, totalizando aproximadamente 290 mil registros, em que cada registro corresponde a uma notificação individual. Essa decisão foi tomada para garantir uma maior quantidade de dados durante o treinamento dos modelos preditivos e, assim, obter resultados mais robustos. Para garantir a qualidade e a consistência dos dados analisados, foram incluídas apenas notificações que atendiam aos critérios de caso suspeito ou confirmado de dengue.

A variável que será utilizada para medir a evolução da incidência da doença é o número de notificações de dengue por semana. Para isso, os dados foram organizados e agrupados por semanas epidemiológicas, isto é, a unidade padrão de tempo utilizada internacionalmente em vigilância em saúde para dados de doenças e eventos epidemiológicos. Cada semana epidemiológica começa no domingo e termina no sábado, e cada ano possui 52 ou 53 semanas epidemiológicas, dependendo do calendário do ano.

4. Descrição das variáveis

Variáveis que compõem o banco de dados:

- Número da notificação: código de identificação da notificação
- Data da notificação: data na qual o atendimento de dengue foi notificado
- Ano: ano da notificação de dengue
- Código do distrito sanitário: código do distrito provável de infecção do caso
 - 264: Segredo
 - 265: Anhanduizinho
 - 266: Bandeira
 - 267: Imbirussu
 - 268: Centro
 - 269: Prosa
 - 270: Lagoa
 - Outros valores: fora de Campo Grande
- Gênero:
 - M: Masculino

- F: Feminino

5. Análise descritiva

Nesta seção, apresentamos a análise descritiva dos dados, que nos fornece uma visão inicial das características do conjunto de dados e orienta a interpretação dos resultados do estudo (Bussab e Morettin, 2024).

A distribuição espacial dos casos foi analisada considerando os diferentes distritos sanitários do município, o que permitiu uma avaliação mais detalhada da incidência da doença em distintas regiões urbanas e rurais de Campo Grande. A Tabela A.1 e a Figura B.1 mostram que o distrito sanitário de Anhanduizinho possui o maior volume de casos entre todos os distritos com 25,14% dos casos, e que há uma distribuição homogênea entre os outros distritos, entre 14,11% e 16,00%, com exceção de Prosa e Centro com 8,81% e 4,92%, respectivamente. No entanto na Figura B.2, que mostra a quantidade de casos a cada mil habitantes de cada distrito, nota-se que há uma concentração maior de casos em Segredo e Imbirussu, e que os distritos Bandeira, Lagoa e Anhanduizinho possuem uma quantidade similar de casos no total, com apenas Prosa e Centro possuindo um valor bem menor de casos a cada 1000 habitantes.

A Figura B.3 apresenta a evolução anual da distribuição do gênero das notificações de dengue. Percebe-se uma proporção estável ao longo dos anos, com uma leve prevalência do sexo feminino. Existe a hipótese de esta prevalência estar relacionada com uma possível tendência da população do sexo masculino em ser mais negligente com a própria saúde em comparação à população do sexo feminino, fazendo com que procurem atendimento médico com menos frequência do que as mulheres.

A distribuição dos casos de dengue entre os distritos por ano se mantém estável ao longo dos anos com exceção de 2022, onde em Prosa houve mais casos em relação aos outros distritos, e em Anhanduizinho houve menos, conforme a Figura B.4.

Analisando a evolução anual da distribuição de casos a cada 1000 habitantes por distrito, na Figura B.5, observa-se que os distritos Segredo, Imbirussu e Bandeira tendem a ter mais casos por habitante. Observa-se também que Anhanduizinho, Prosa e Lagoa não aparentam seguir um padrão visual e que Centro sempre possui menos casos a cada

mil habitantes do distrito. Adicionalmente, pode-se perceber pela figura os anos com mais incidência de dengue como 2010, 2013, 2015, 2016, 2019, 2020, 2022, 2023 e 2024, marcando prováveis anos epidêmicos, e anos com baixa incidência de casos como 2008, 2009, 2011, 2012, 2014, 2017, 2018 e 2021. Vale ressaltar que, devido à pandemia de COVID-19, pode ter havido subnotificações no ano de 2021.

Com o objetivo de separar a parte estocástica da parte de tendência linear e da parte sazonal dos dados e, assim, facilitar a modelagem, foram feitas as decomposições STL (Seasonal-Trend decomposition using LOESS) multiplicativas para as séries de notificações de cada distrito. Como as séries estão organizadas em semanas epidemiológicas, foi utilizado como período sazonal o período de 52 semanas. Assim, é possível avaliar os padrões de longo prazo (tendência), os ciclos regulares (sazonalidade) e as flutuações aleatórias que compõem cada série, permitindo análises mais precisas mesmo em séries mais complexas. As séries dos distritos com suas respectivas decomposições constam nas Figuras B.6 a B.13, sendo a Figura B.13 a série do município com todos os distritos agregados.

Ao avaliar os gráficos das séries de notificações dos distritos e do total do município, nota-se que tanto as curvas originais quanto as decomposições apresentam padrões muito semelhantes entre si. Todos os distritos exibem picos sincronizados, tendências parecidas e a sazonalidade anual bem definida, com pouca variação estrutural de um distrito para outro. Isso sugere que os processos que ditam a dinâmica da doença são compartilhados entre os distritos devido a fatores comuns.

A Figura B.14 apresenta a quantidade de casos ao longo das semanas epidemiológicas que compõem um ano, ao longo dos diferentes anos disponíveis na base de dados. O gráfico da Figura B.14 destaca que, em anos epidêmicos como 2010, 2013, 2015, 2016, 2019, 2020, 2022, 2023 e 2024, há picos acentuados de notificações entre as semanas 2 e 25, enquanto nos anos não epidêmicos as curvas permanecem mais baixas e estáveis durante todo o ano. Essa diferença sugere que o comportamento da incidência de dengue ao longo das semanas epidemiológicas depende fortemente do contexto epidêmico de cada ano.

O *box plot* da Figura B.15 complementa essa análise ao mostrar a distribuição das notificações semanais ao longo dos anos. Nota-se que a variabilidade e os valores mais altos se concentram entre as semanas 2 e 25, período em que aumentos expressivos de casos podem ocorrer devido à sazonalidade da doença. Após a semana 25, mesmo nos anos epidêmicos, o número de casos cai drasticamente e permanece baixo durante o inverno, com pouca variabilidade, voltando a apresentar maior variação nas semanas finais do ano, no início do verão.

6. Canal endêmico

O canal endêmico ou diagrama de controle é um gráfico que representa o padrão regular da frequência dos casos de uma doença em um determinado local e período, composto por três linhas principais: o limite superior (limiar epidêmico), que indica a frequência máxima esperada de casos; a medida de tendência central, que representa a frequência esperada; e o limite inferior, que indica a frequência mínima esperada. A construção do canal endêmico é fundamental para a vigilância epidemiológica, pois permite monitorar a incidência da doença e identificar rapidamente a ocorrência de epidemias, definidas como o excesso de casos em relação ao esperado. Isso possibilita a tomada de decisões oportunas para o controle e prevenção da doença, além de ajudar a estabelecer o comportamento ou tendência da doença ao longo do tempo.

Na etapa de análise, para a construção dos canais endêmicos, foram utilizados os dados dos sete anos anteriores ao ano de interesse, excluindo-se os anos classificados como epidêmicos pelo Ministério da Saúde, sendo estes: 2010, 2013, 2015, 2016, 2019, 2020, 2022 e 2023 (Brasil, 2024). Dessa maneira obtemos um canal representativo de situações de não epidemia. Essa abordagem está alinhada com as Diretrizes Nacionais para a Prevenção e Controle de Epidemias de Dengue, que recomendam o uso da média móvel como medida de tendência central e um período de 10 anos para a construção do canal endêmico. Contudo, na Nota Técnica N° 1/2023-CGAR/DEIDT/SVS/MS do Ministério da Saúde (Brasil, 2023), essa recomendação foi atualizada, orientando o uso do diagrama de controle baseado na mediana semanal e nos intervalos interquartis (Q1 e Q3) para o monitoramento de arboviroses urbanas, como dengue, Chikungunya e Zika.

O método alternativo proposto pelo Ministério da Saúde, cuja avaliação da eficácia é de interesse dos pesquisadores, consiste na construção do canal endêmico utilizando a mediana e os intervalos interquartis calculados a partir de um período histórico de cinco a sete anos, excluindo anos epidêmicos. Essa metodologia é considerada mais sensível para a detecção precoce de epidemias, pois a mediana não é influenciada por valores extremos, ao contrário da média móvel tradicional.

Os canais endêmicos baseados nos últimos 7 anos, excluindo-se anos epidêmicos, construídos com os métodos de média e mediana constam nas Figuras B.16 e B.17, respectivamente. Analisando os canais, vemos que o canal de médias gera, em geral, intervalos com limites maiores do que o canal de medianas, já que a média costuma ser mais afetada por valores extremos do que a mediana. Isso sugere que o canal de médias é mais tolerante a pequenas variações na quantidade de casos, enquanto o canal de medianas seria mais sensível a variações pontuais.

As Tabelas A.2 e A.3 mostram as matrizes de confusão para as classificações realizadas pelo canal endêmico baseado em médias e pelo canal endêmico baseado em medianas, respectivamente. É possível avaliar o desempenho de cada abordagem pela quantidade de vezes que cada canal classificou corretamente ou incorretamente as semanas como epidêmicas ou não epidêmicas. A partir das tabelas, podemos calcular a sensibilidade e a especificidade de cada canal; tais medidas se encontram na Tabela A.4. Considerando a sensibilidade, o canal de medianas (0,801) se sobressai sobre o canal de médias (0,657), isto é, o canal de medianas classifica corretamente 80,1% das semanas epidêmicas reais contra 65,7% do canal de médias. Em relação à especificidade, o canal de medianas (0,548) é superado pelo canal de médias (0,731), isto é, o canal de medianas classifica corretamente 54,8% das semanas não epidêmicas reais contra 73,1% do canal de médias.

Tratando cada canal construído como um classificador binário (epidemia/não epidemia), podemos construir a curva ROC para comparar o desempenho de cada método. Na Figura B.18, podemos ver as curvas ROC de cada canal e suas respectivas AUC's. Em geral, as curvas possuem comportamento parecido; no caso analisado, a AUC do

canal de médias foi de 0,758 e a do canal de medianas foi de 0,775, valores muito próximos. Isso sugere que os canais apresentam desempenho semelhante na discriminação de períodos epidêmicos e não epidêmicos. É importante pontuar que, atualmente, o canal de medianas é o método utilizado oficialmente para classificar e detectar epidemias (Nota Técnica nº 1/2023-CGARB/DEIDT/SVS/MS), o que pode estar relacionado à leve diferença nas AUC's a favor do canal de medianas. Neste sentido, cabe posterior análise da utilização de ambos os métodos para detecção de epidemias, já que não foi possível concluir com certeza a superioridade de qualquer um dos métodos.

7. Séries temporais

A modelagem da série de incidência foi feita considerando como unidade amostral a quantidade de casos por semana epidemiológica. Como pode ser visto na Figura B.13, a série apresenta indícios de sazonalidade, pois os grandes picos acontecem sempre nas primeiras e últimas semanas do ano, possivelmente por condições climáticas e/ou regionais. No entanto, esse comportamento da série não se repete em todos os anos; em anos como 2008, 2011, 2014, 2017 e 2018, a quantidade de casos se manteve baixa e estável durante todas as semanas do ano. Isso sugere que a série apresenta dois comportamentos diferentes: um para anos epidêmicos, com maior variabilidade nas épocas mais quentes do ano, e outro para anos não epidêmicos, no qual a quantidade de casos se mantém estável durante todo o ano.

Para captar o comportamento não-linear da série, foram utilizados modelos de redes neurais Long Short-Term Memory (LSTM), um tipo de rede neural recorrente (RNN) desenvolvida para lidar com dados sequenciais, como séries temporais. A principal diferença delas em relação às RNN's tradicionais se dá na capacidade de aprender padrões e dependência de longo prazo e utilizar esse contexto para melhorar as previsões. Isso é possível por conta da sua arquitetura interna composta por células de memória e mecanismos comumente chamados de “portões”, que regulam o fluxo de informações e decidem o que deve ser lembrado, atualizado ou descartado ao longo do tempo. Essas características as tornam eficazes na tarefa de modelar séries temporais complexas,

como as que apresentam sazonalidade e eventos inesperados, pois conseguem aprender tanto padrões de curto quanto de longo prazo nos dados. Dessa forma as LSTM's se mostraram uma escolha adequada para a modelagem de casos de dengue, onde o comportamento da série pode variar significativamente dependendo se o ano é epidêmico ou não. Para mais detalhes sobre o funcionamento das LSTM's, consultar Chollet (2021).

Dessa maneira, foram treinados modelos com horizonte de previsão de 12 e 26 (3 e 6 meses, aproximadamente) passos para a série de incidência geral de Campo Grande, e com horizonte de previsão de 26 passos para cada série de incidência dos distritos. Os hiperparâmetros utilizados foram escolhidos por meio da biblioteca Optuna do Python, uma ferramenta de otimização que utiliza métodos numéricos para buscar os melhores parâmetros que minimizam o erro do modelo. Neste caso, o processo buscou reduzir tanto o erro no conjunto de validação quanto a diferença entre os erros de validação e teste. Para essa escolha dos hiperparâmetros, o conjunto de dados foi dividido em 80% para treino e 20% para teste. Os hiperparâmetros de cada modelo se encontram na Tabela A.5.

Com os hiperparâmetros escolhidos para cada modelo, o conjunto de dados foi dividido em 64% para treino, 16% para validação e 20% para teste (Figura B.19), e os modelos foram treinados por meio da biblioteca PyTorch. As Figuras B.20 a B.28 mostram as previsões do primeiro ponto de cada janela do conjunto de teste de cada modelo com seus respectivos RMSE's (*Root Mean Squared Error* - Raíz do Erro Quadrático Médio), assim como o gráfico do RMSE ao longo das janelas do conjunto de teste. Nessas figuras, é possível ver como os modelos preveem no curto prazo, isto é, a acurácia do modelo ao prever exatamente uma semana à frente, e como a variabilidade das previsões se comporta ao longo das janelas. Em geral, para este fim, o modelo obteve desempenho satisfatório, captando a variabilidade da série tanto nos picos quanto nos trechos estáveis. Nota-se também que o RMSE dos modelos é baixo na média, mas tende a ser maior quando acontecem picos na série; esse comportamento pode ser visto na Figura B.29, que mostra todas as janelas de previsão no conjunto de teste para o modelo de horizonte 26 da série geral de Campo Grande, apresentando maior variabilidade nas épocas em que há maior incidência da doença.

Com indícios de que os modelos estão bem ajustados, podemos ajustar os modelos com todos os dados e fazer previsões para o ano de 2025. As Figuras B.30 a B.38 mostram as previsões para cada distrito e horizontes de previsão ajustados. Nota-se que as previsões refletem o comportamento observado nos últimos anos de 2022 até 2024, sugerindo mais um pico para o início do ano de 2025.

Destaca-se que o modelo utilizado possui implementação computacional simples e pode ser facilmente incorporado a uma rotina diária de monitoramento. Com atualizações periódicas dos dados, é possível criar um fluxo automatizado para execução regular do modelo, permitindo não apenas a geração contínua de previsões, mas também o acompanhamento sistemático de sua performance ao longo do tempo.

8. Conclusões

A análise descritiva revelou importantes padrões espaciais, demográficos e temporais nas notificações de dengue em Campo Grande. Observou-se que o distrito de Anhanduizinho concentra o maior volume absoluto de casos, enquanto Segredo e Imbirussu se destacam em incidência relativa por habitante. A distribuição proporcional dos casos entre os distritos é relativamente homogênea, com exceção do Centro, que apresenta valores consistentemente mais baixos. No perfil demográfico, a série histórica mostra uma leve predominância do sexo feminino entre os casos notificados, possivelmente relacionada a diferenças no comportamento de busca por atendimento de saúde. Em relação à evolução temporal, os distritos exibem padrões muito semelhantes, com picos de casos sincronizados e sazonalidade anual bem definida, sugerindo influência de fatores comuns em toda a cidade. Os anos de maior incidência (2010, 2013, 2015, 2016, 2019, 2020, 2022, 2023 e 2024) coincidem com picos agudos de notificações entre as semanas 2 e 25, enquanto anos não epidêmicos apresentam curvas mais estáveis e baixas. A decomposição das séries temporais confirmou a presença de tendência, sazonalidade e flutuações aleatórias semelhantes entre os distritos, reforçando a ideia de processos epidemiológicos compartilhados. Por fim, a análise das notificações semanais evidenciou que a maior variabilidade e os maiores valores concentram-se no primeiro

semestre, especialmente durante os períodos epidêmicos, com redução significativa dos casos no inverno.

A análise dos canais endêmicos evidenciou algumas diferenças entre os métodos baseados em médias e em medianas, tanto na construção dos limites quanto no desempenho como classificadores de epidemias. Observou-se que o canal de médias, por ser mais influenciado por valores extremos, gera intervalos superiores mais amplos, tornando-se mais conservador e menos sensível a oscilações pontuais no número de casos. Em contrapartida, o canal de medianas, ao utilizar a mediana e os quartis, apresenta limites mais estreitos e maior sensibilidade, facilitando a detecção precoce de aumentos atípicos nas notificações. As análises das matrizes de confusão e das métricas de desempenho confirmaram esse comportamento: o canal de medianas apresentou sensibilidade superior, enquanto o canal de médias destacou-se em especificidade. Apesar dessas diferenças, as curvas ROC e os valores de AUC mostraram que ambos os métodos oferecem desempenho global semelhante na discriminação entre semanas epidêmicas e não epidêmicas.

É importante destacar que a escolha do método para construção do canal endêmico deve ser orientada pelo objetivo do monitoramento: se a prioridade é identificar rapidamente possíveis epidemias e minimizar o risco de não detectar surtos reais, recomenda-se priorizar a sensibilidade, favorecendo o uso do canal de medianas. Por outro lado, se o objetivo é evitar classificações equivocadas de epidemia, pode ser mais adequado priorizar a especificidade, optando pelo canal de médias. Assim, a decisão entre sensibilidade e especificidade envolve necessariamente um equilíbrio entre o risco de não detectar epidemias e o risco de respostas desnecessárias, devendo ser feita de acordo com o contexto epidemiológico e as necessidades dos pesquisadores. Destaca-se, ainda, que o canal de medianas é atualmente o método oficial recomendado pelo Ministério da Saúde, o que reforça a relevância de sua adoção e pode justificar a leve diferença a seu favor observada nas AUC's.

A modelagem das séries temporais com redes LSTM demonstrou capacidade de capturar os padrões complexos da incidência de dengue em Campo Grande, incluindo sua sazonalidade característica e a distinção entre anos epidêmicos e não epidêmicos.

Os modelos, configurados com horizontes de 12 e 26 semanas, reproduziram adequadamente a variabilidade da série, identificando tanto os picos abruptos quanto os períodos de estabilidade, com erros (RMSE) consistentes ao longo das janelas de teste. As previsões para 2025 sugerem a persistência de ciclos epidêmicos similares ao comportamento recente (2022-2024), com projeções indicando novo pico de casos, alinhado aos padrões históricos. A coerência entre as previsões distritais e a série agregada do município reforça a robustez da abordagem, evidenciando que as LSTM's são ferramentas promissoras para a previsão de dengue em contextos com sazonalidade marcada e heterogeneidade espacial.

APÊNDICE A

Tabelas

Tabela A.1 Distribuição de frequências da quantidade de casos de dengue por distrito sanitário

Distrito	Quantidade de casos	Percentual (%)	População (2022)	Casos a cada 1000 habitantes
Anhanduizinho	67.999	25,14	218.585	311,087
Lagoa	43.250	16,00	134.964	320,456
Segredo	42.268	15,62	128.312	329,416
Bandeira	41.669	15,40	136.691	304,841
Imbirussu	38.178	14,11	108.031	353,399
Prosa	23.829	8,81	95.589	249,286
Centro	13.312	4,92	61.653	215,918

Tabela A.2 Matriz de confusão do canal endêmico baseado em médias

Real / Previsto	Epidemia	Não Epidemia
Epidemia	205	107
Não Epidemia	56	152

Tabela A.3 Matriz de confusão do canal endêmico baseado em medianas

Real / Previsto	Epidemia	Não Epidemia
Epidemia	250	62
Não Epidemia	94	114

Tabela A.4 Sensibilidade e especificidade dos canais endêmicos

Canal endêmico	Sensibilidade	Especificidade
Médias	0,657	0,731
Medianas	0,801	0,548

Tabela A.5 Hiperparâmetros dos modelos de redes LSTM

Série	Horizonte de previsão	Janela de entrada	Nº de neurônios por camada	Nº de camadas	Dropout	Taxa de aprendizado	Batch size
Campo Grande	12	29	136	4	0,108	0,00046	8
Campo Grande	26	51	104	1	0,133	0,00561	8
Anhanduizinho	26	52	40	3	0,260	0,00330	32
Bandeira	26	49	46	3	0,143	0,00174	16
Centro	26	50	69	3	0,163	0,00193	32
Imbirussu	26	52	130	1	0,478	0,00110	16
Lagoa	26	52	20	2	0,208	0,00460	8
Prosa	26	52	138	2	0,073	0,00140	32
Segredo	26	48	86	3	0,222	0,00044	16

APÊNDICE B

Figuras

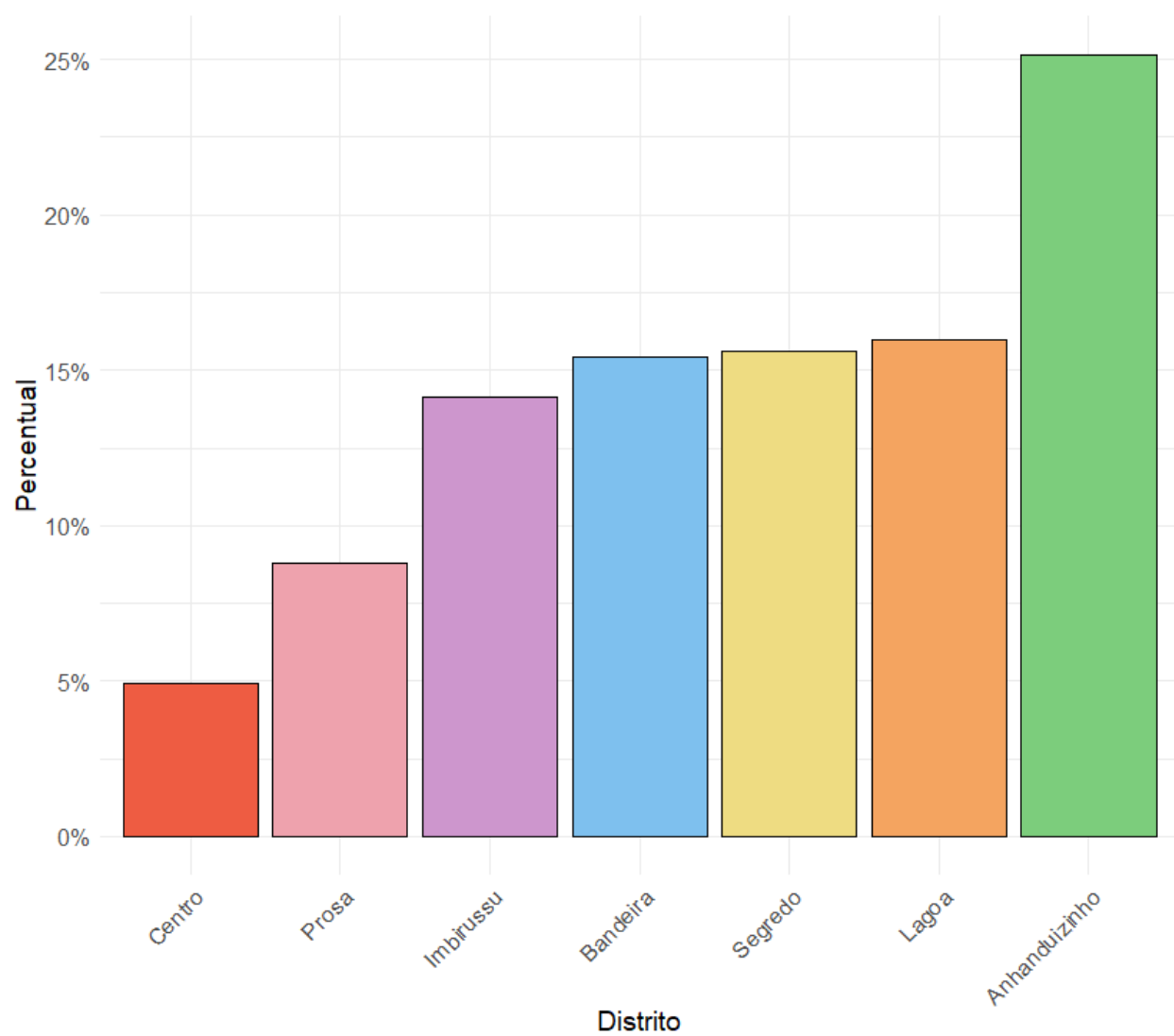


Figura B.1 Quantidade de casos de dengue por distrito sanitário

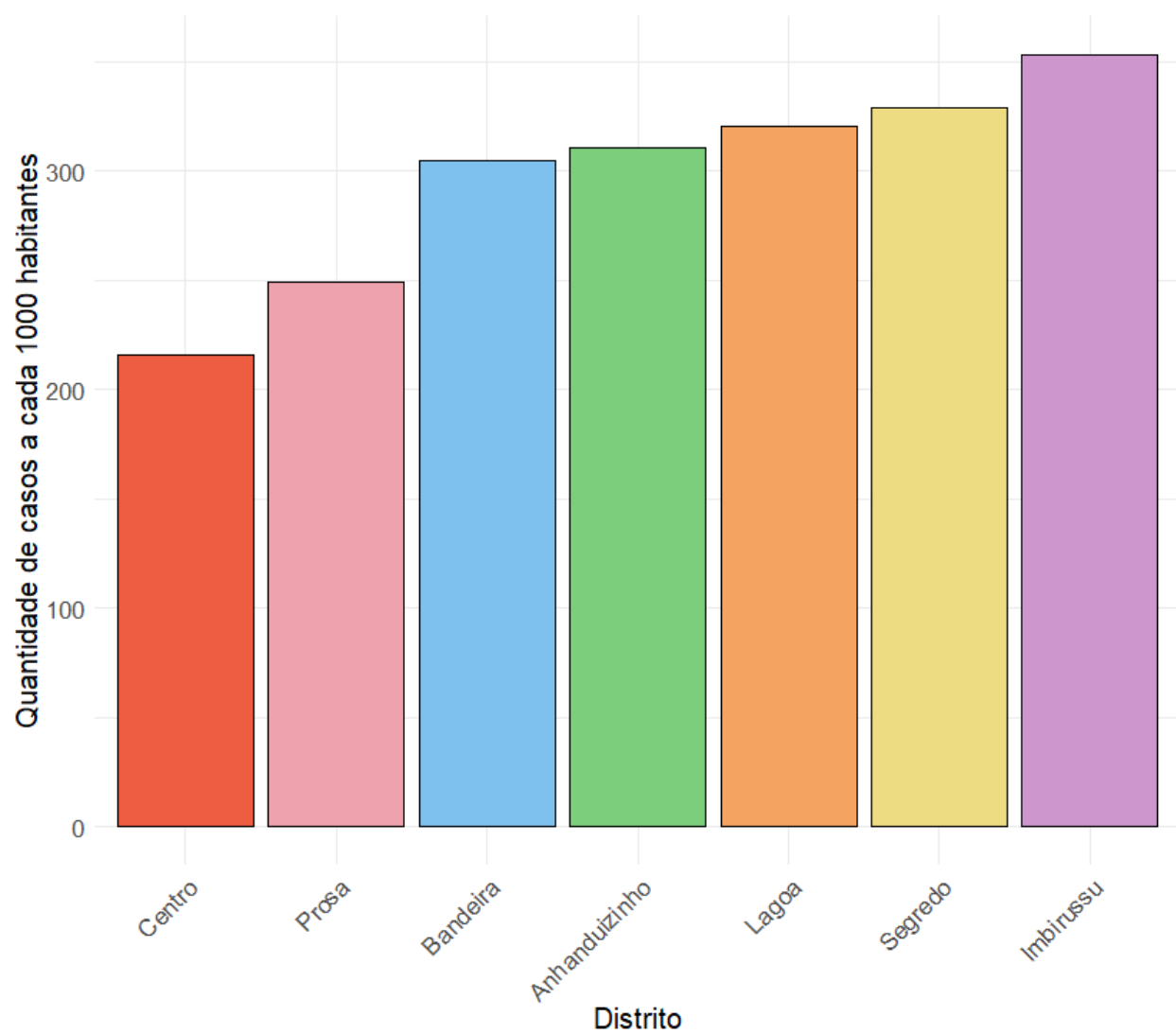


Figura B.2 Quantidade de casos de dengue a cada 1000 habitantes por distrito sanitário

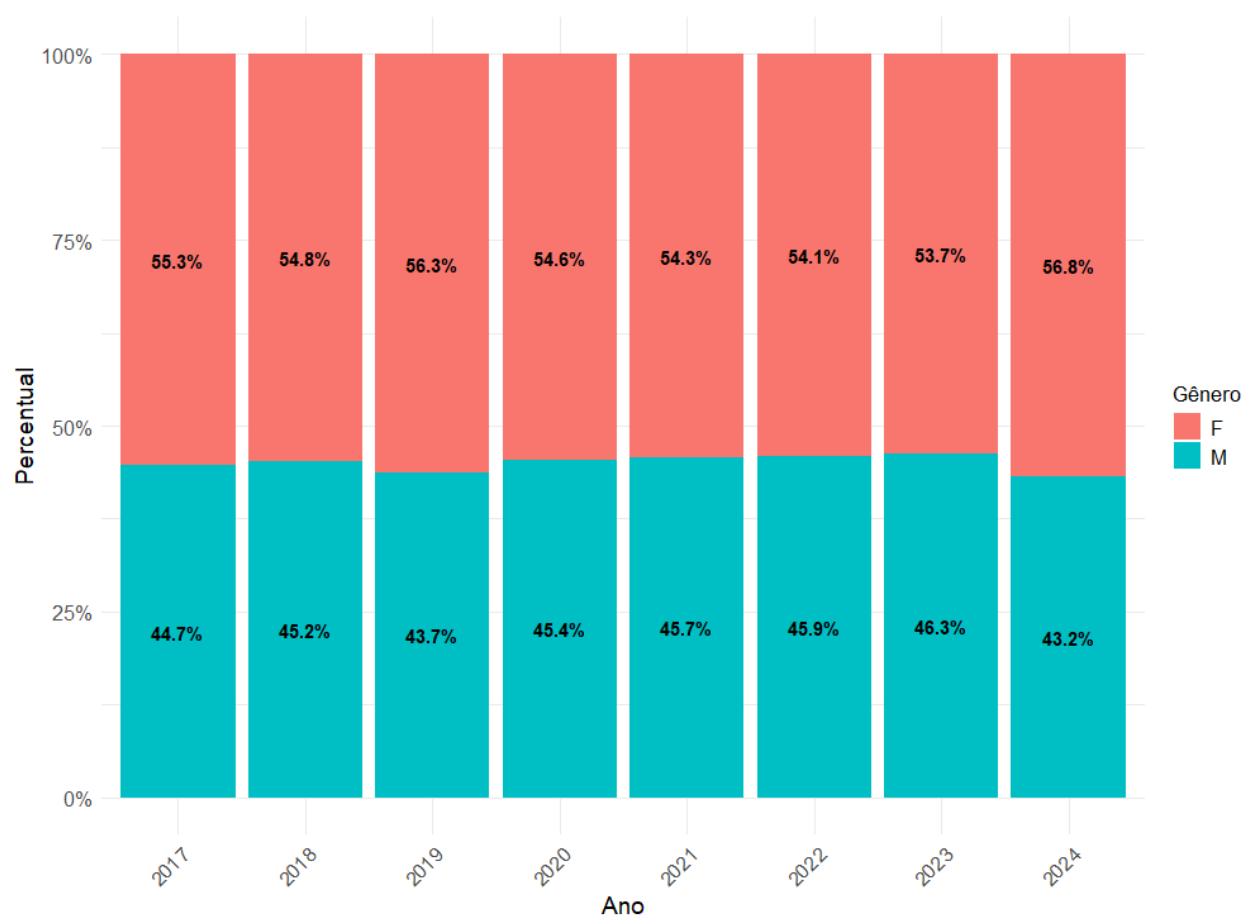


Figura B.3 Distribuição de frequências de gênero dos pacientes por ano

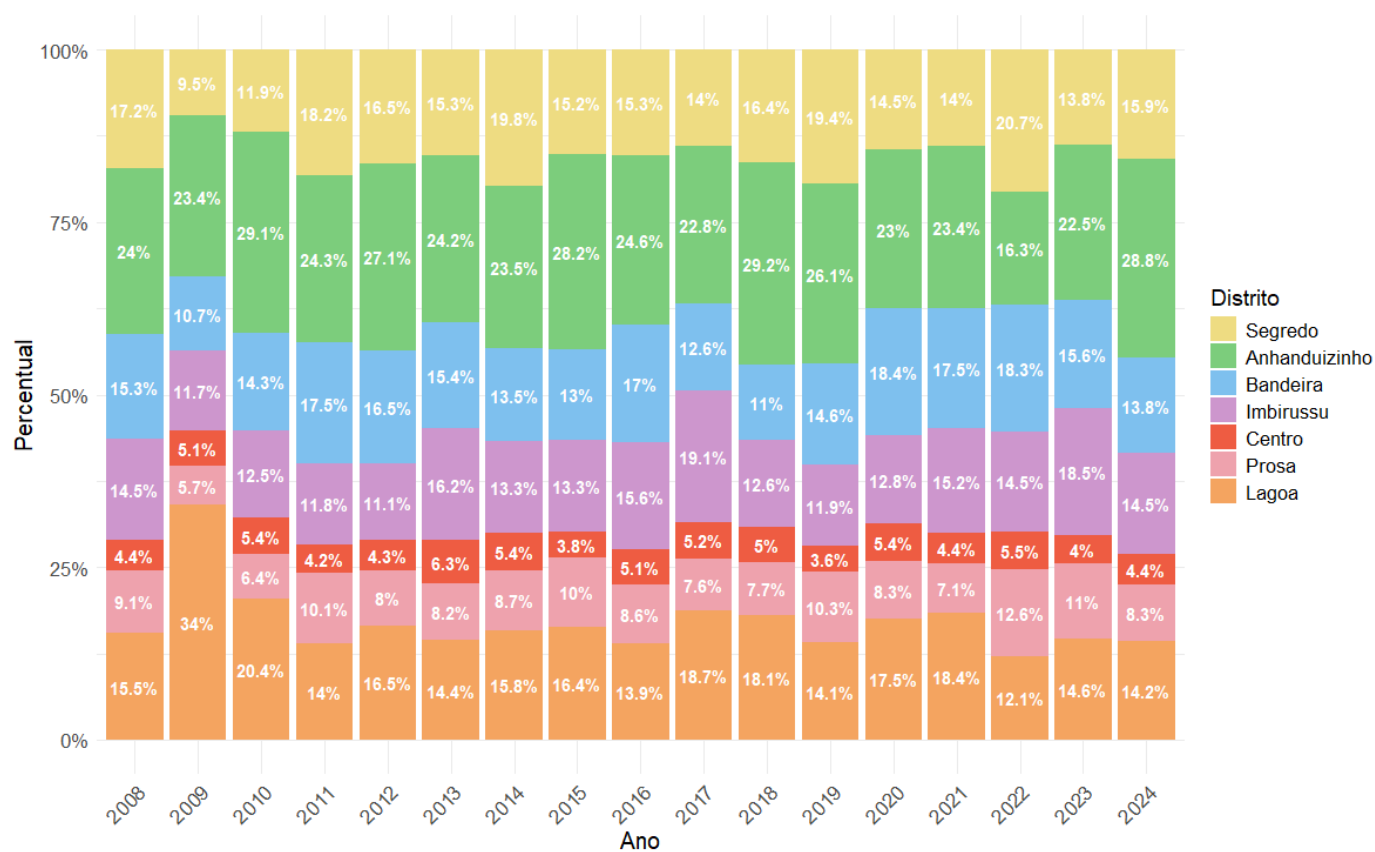


Figura B.4 Distribuição da quantidade de casos de dengue por distrito sanitário e ano

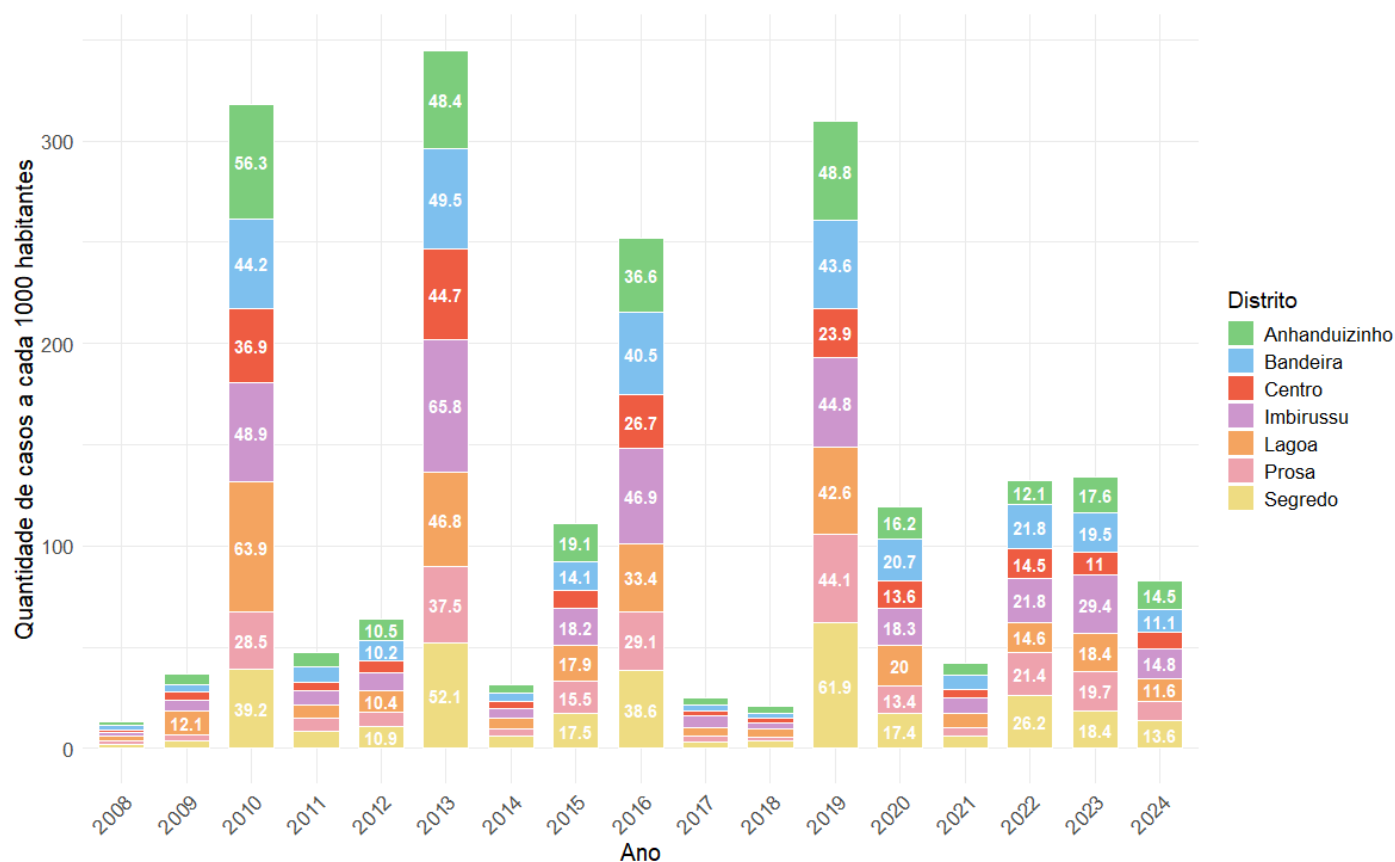


Figura B.5 Quantidade de casos de dengue a cada 1000 habitantes por distrito sanitário e ano

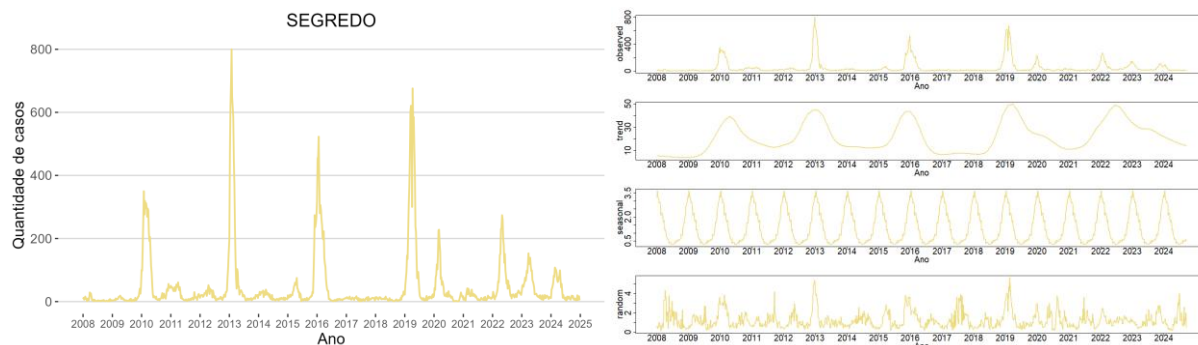


Figura B.6 Quantidade de casos de dengue e decomposição STL – Segredo

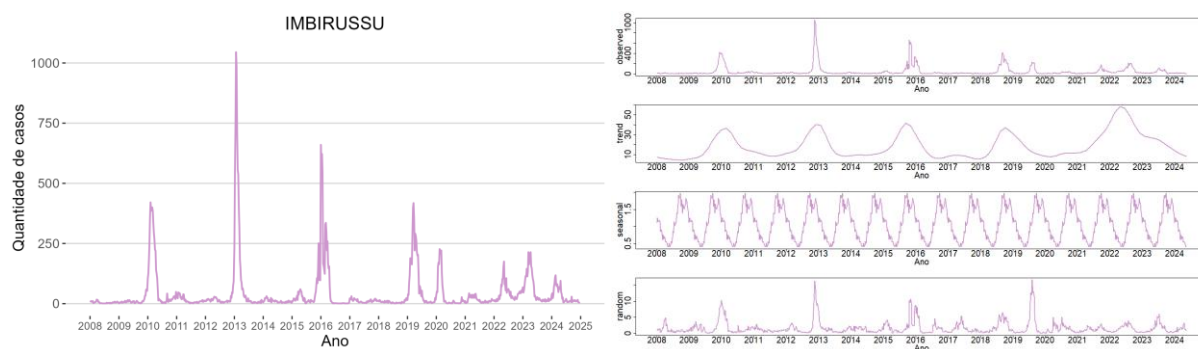


Figura B.7 Quantidade de casos de dengue e decomposição STL – Imbirussu

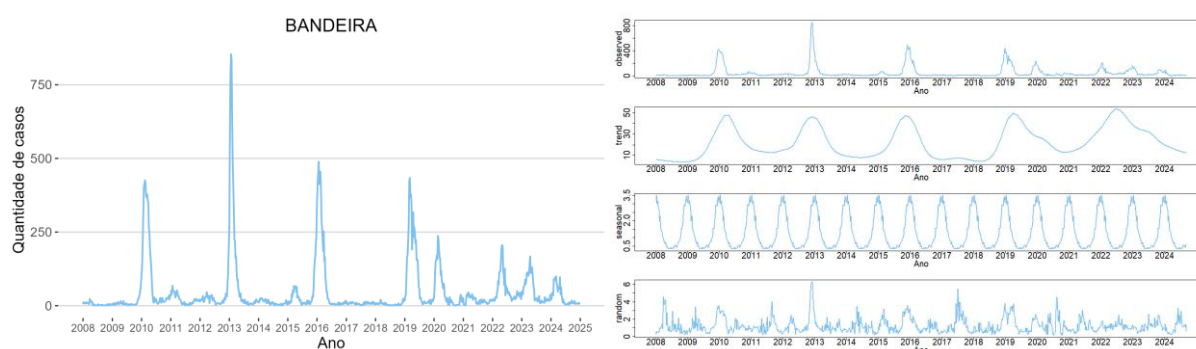


Figura B.8 Quantidade de casos de dengue e decomposição STL – Bandeira

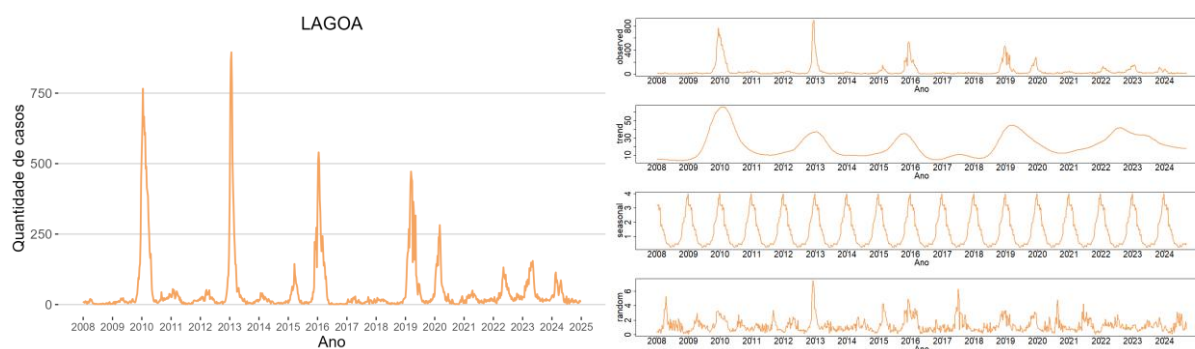


Figura B.9 Quantidade de casos de dengue e decomposição STL – Lagoa

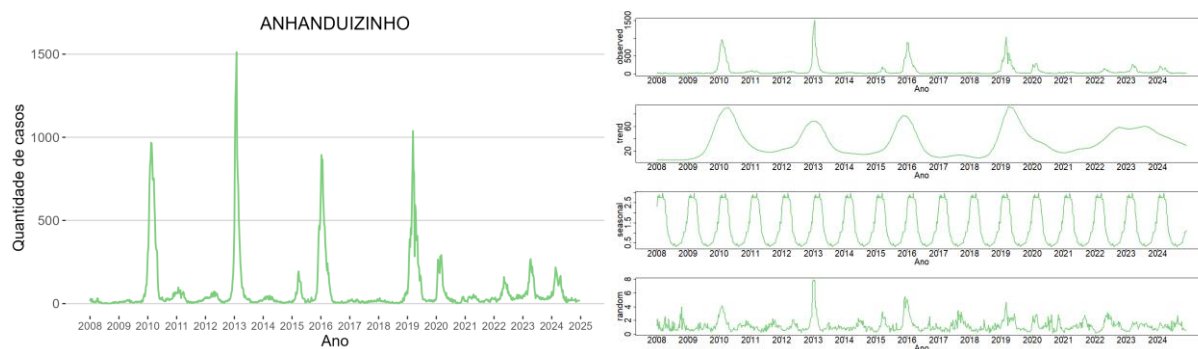


Figura B.10 Quantidade de casos de dengue e decomposição STL – Anhanduizinho

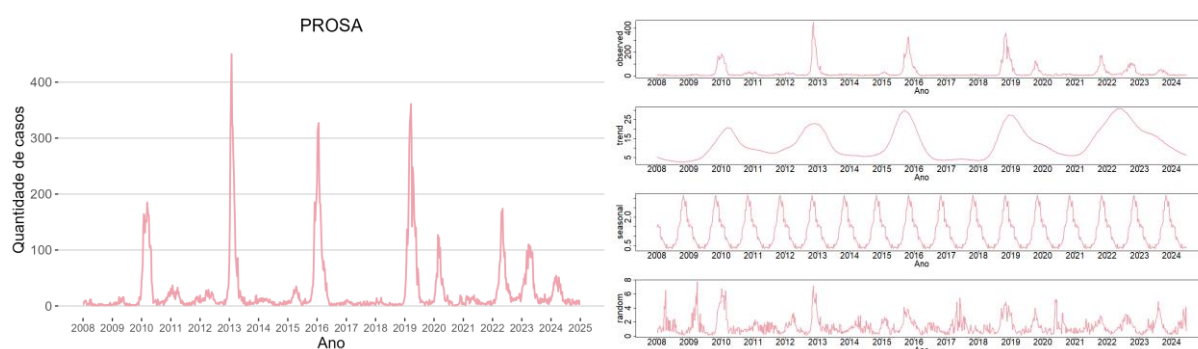


Figura B.11 Quantidade de casos de dengue e decomposição STL – Prosa

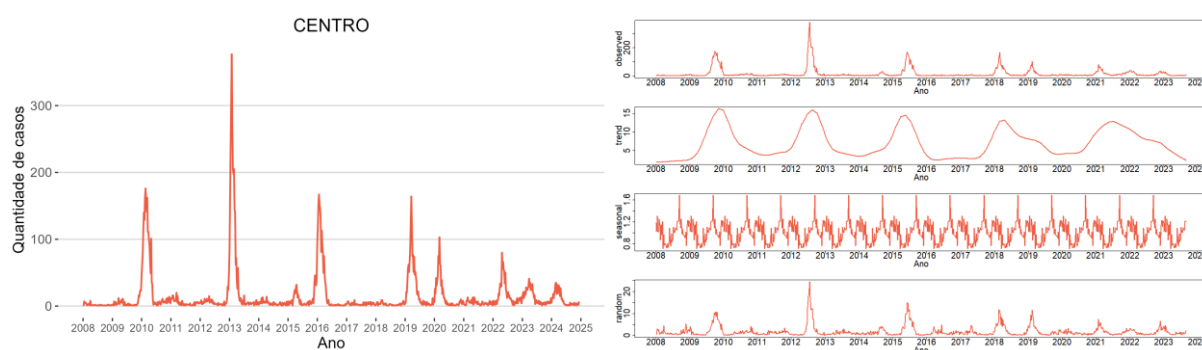


Figura B.12 Quantidade de casos de dengue e decomposição STL – Centro

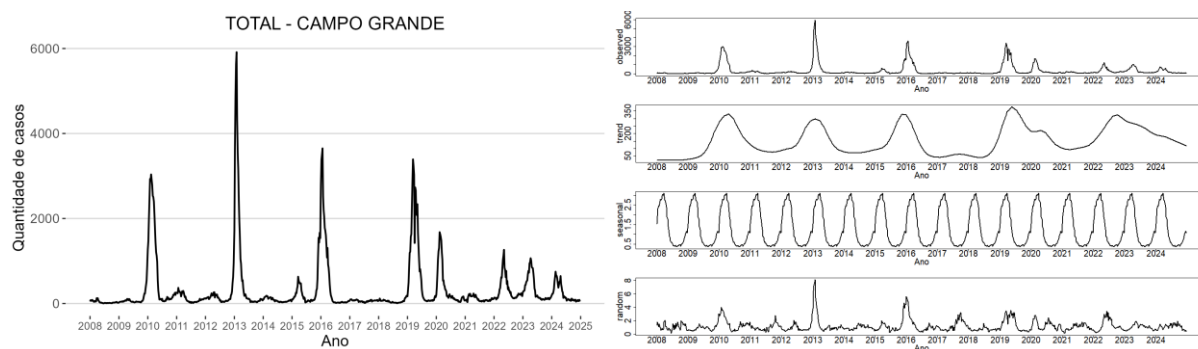


Figura B.13 Quantidade de casos de dengue e decomposição STL - Campo Grande

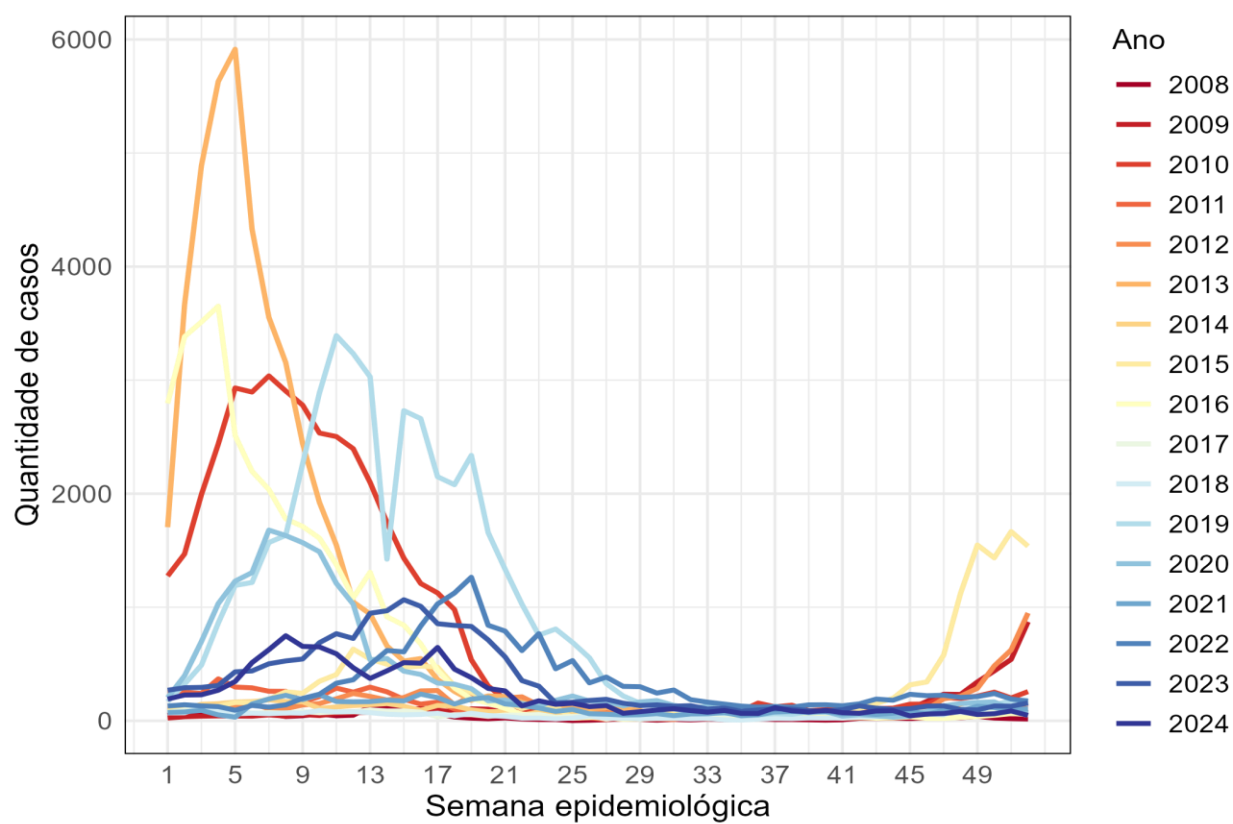


Figura B.14 Quantidade de casos de dengue por semana epidemiológica e ano

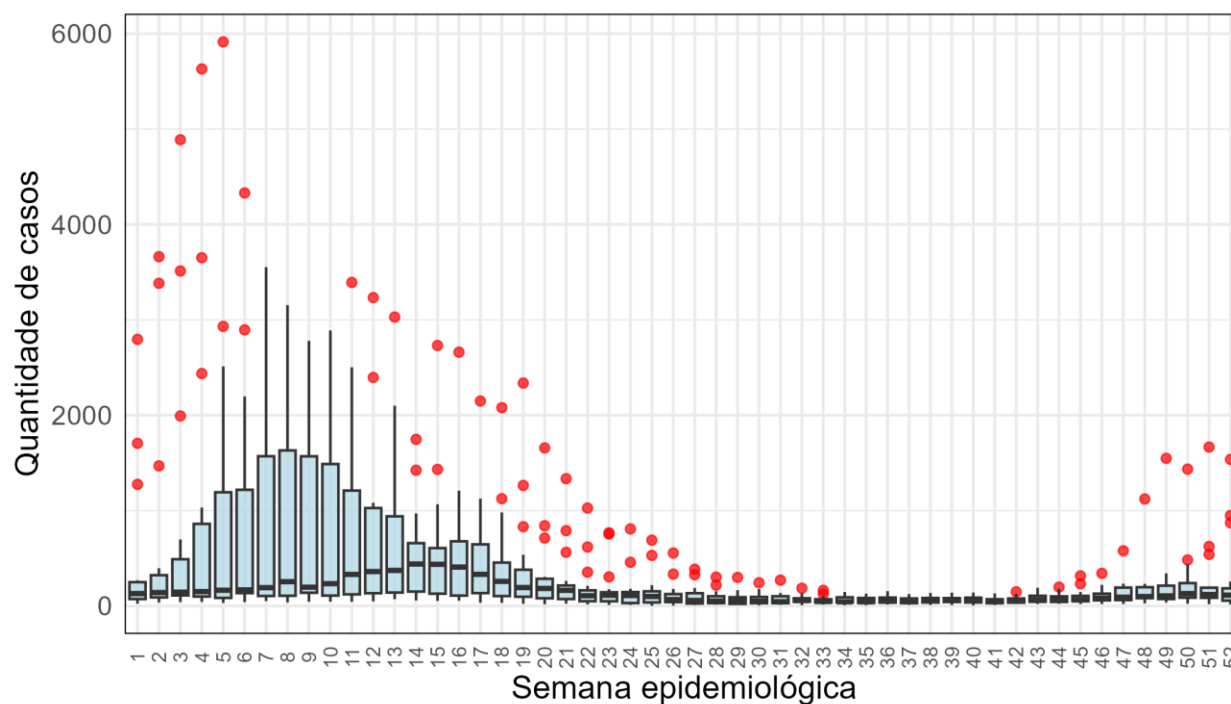


Figura B.15 Box plot da quantidade de casos de dengue por semana epidemiológica

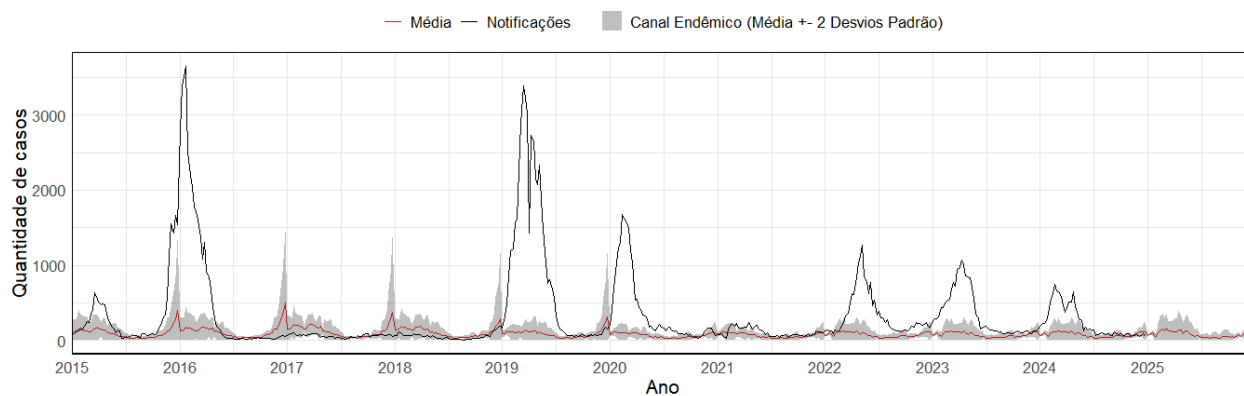


Figura B.16 Canal endêmico utilizando média e desvio padrão

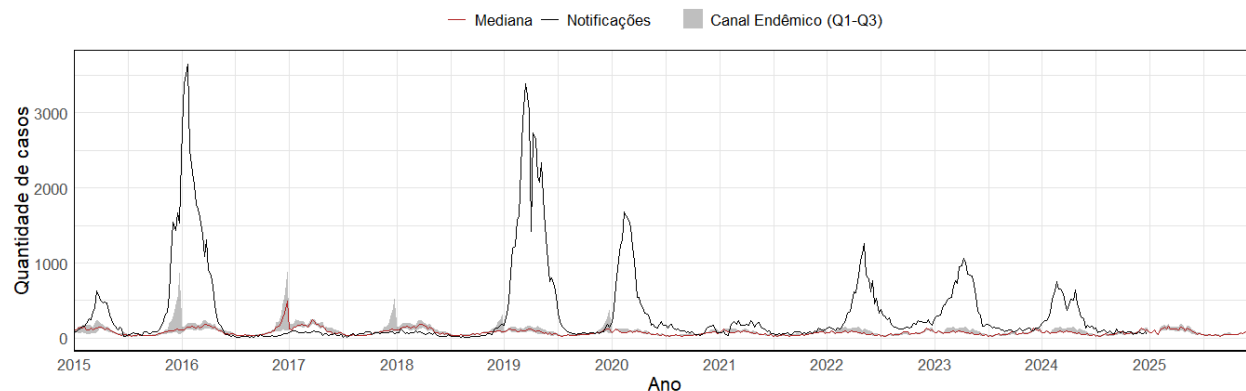


Figura B.17 Canal endêmico utilizando mediana e quartis

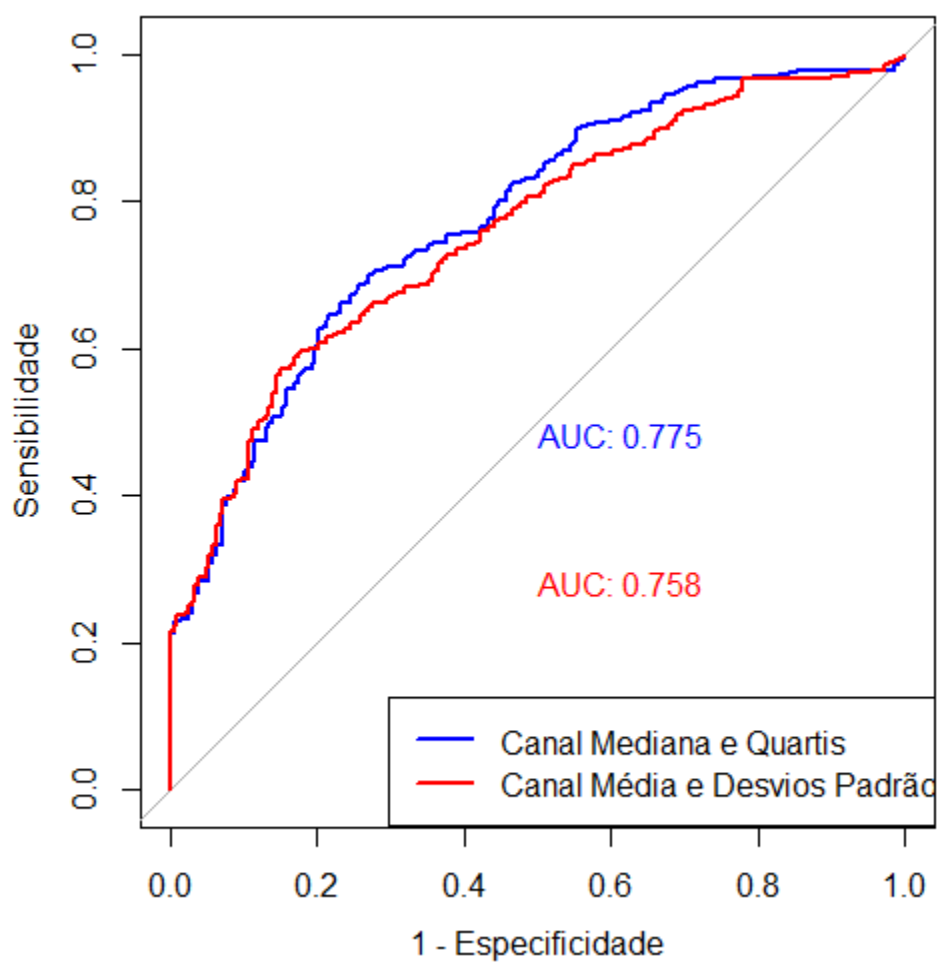


Figura B.18 Curvas ROC e AUC dos canais endêmicos

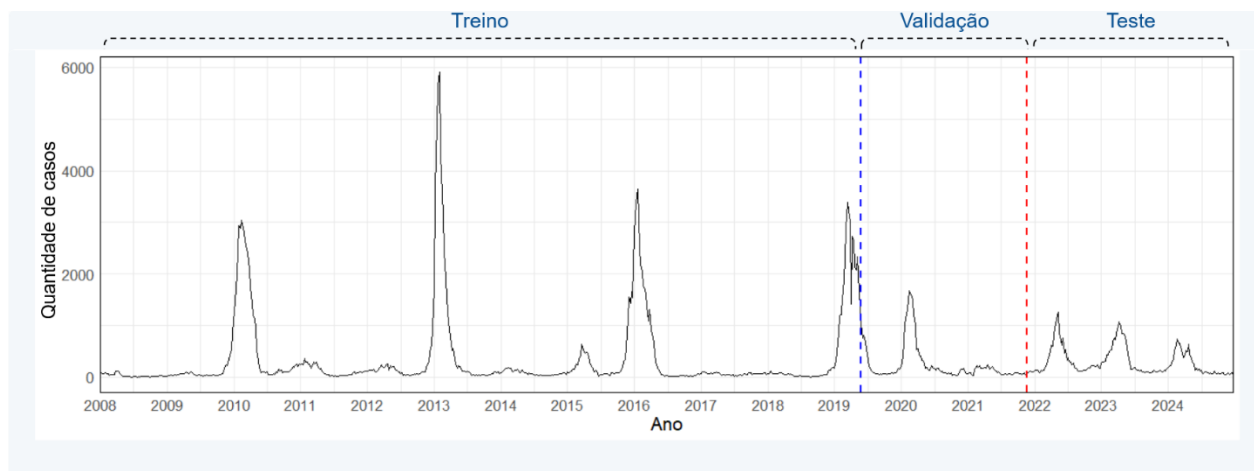


Figura B.19 Divisão do conjunto de dados em treino, validação e teste

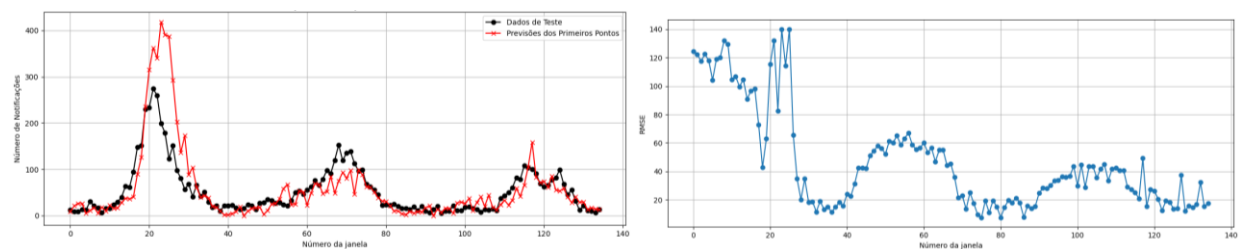


Figura B.20 Previsões do primeiro ponto das janelas de teste e RMSE das janelas com horizonte de previsão 26 – Segredo

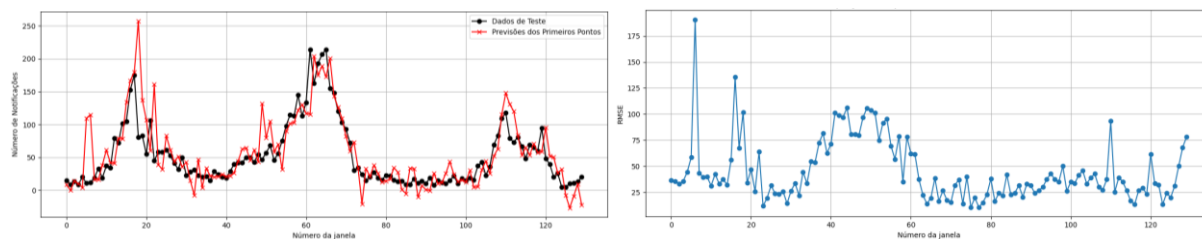


Figura B.21 Previsões do primeiro ponto das janelas de teste e RMSE das janelas com horizonte de previsão 26 – Imbirussu

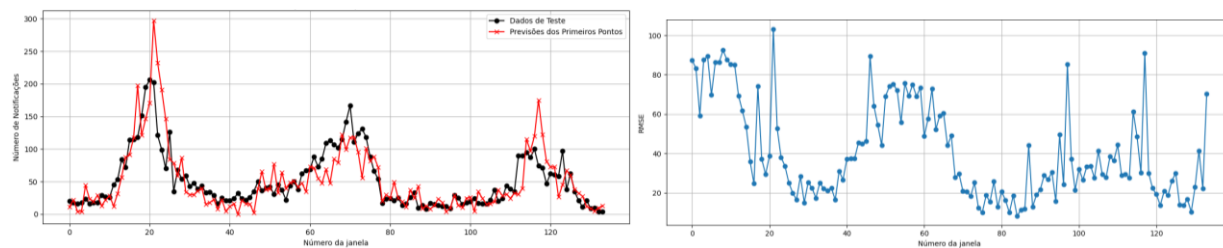


Figura B.22 Previsões do primeiro ponto das janelas de teste e RMSE das janelas com horizonte de previsão 26 – Bandeira

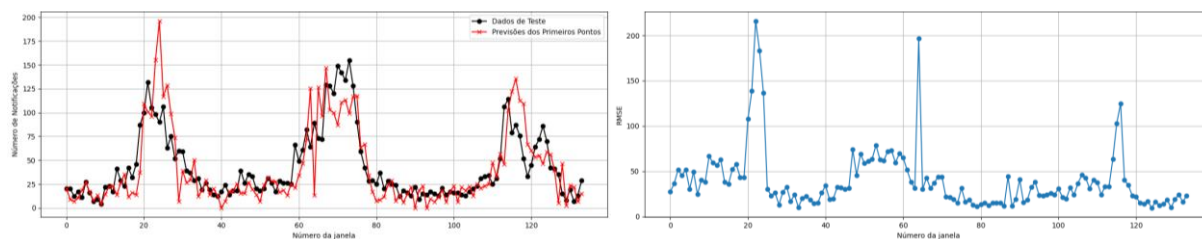


Figura B.22 Previsões do primeiro ponto das janelas de teste e RMSE das janelas com horizonte de previsão 26 – Lagoa

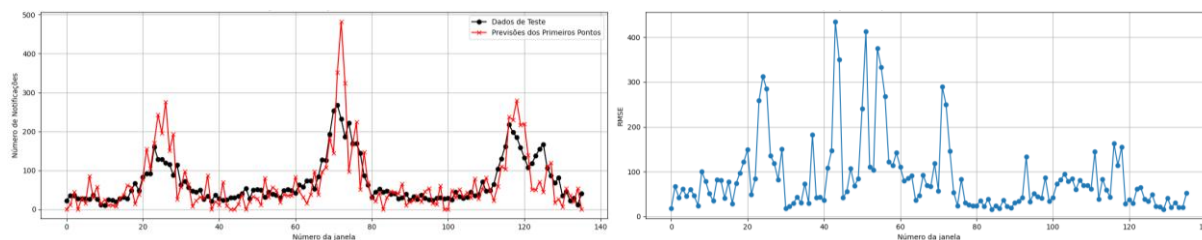


Figura B.24 Previsões do primeiro ponto das janelas de teste e RMSE das janelas com horizonte de previsão 26 – Anhanduizinho

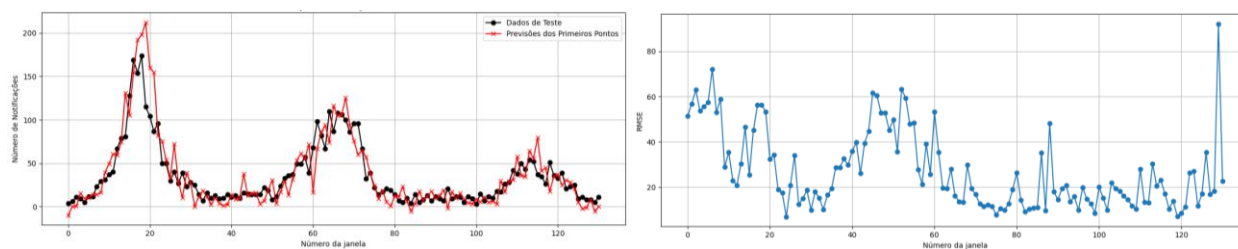


Figura B.25 Previsões do primeiro ponto das janelas de teste e RMSE das janelas com horizonte de previsão 26 – Prosa

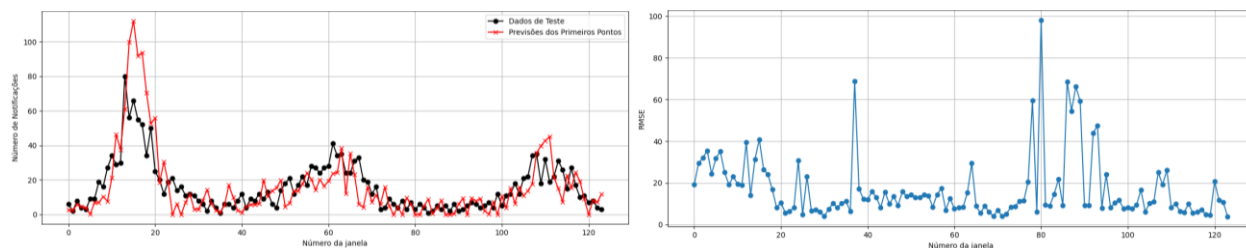


Figura B.26 Previsões do primeiro ponto das janelas de teste e RMSE das janelas com horizonte de previsão 26 – Centro

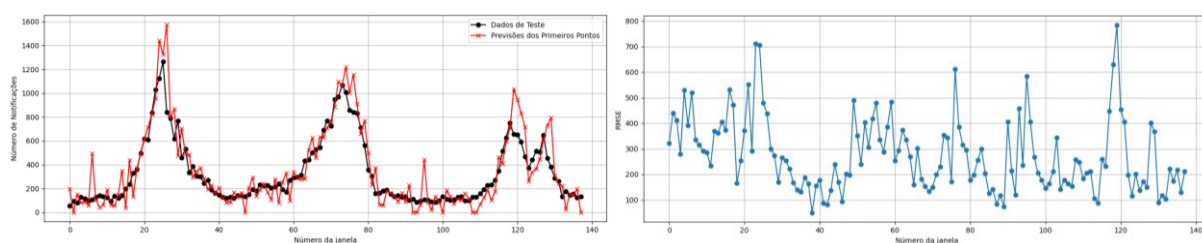


Figura B.27 Previsões do primeiro ponto das janelas de teste e RMSE das janelas com horizonte de previsão 26 – Campo Grande

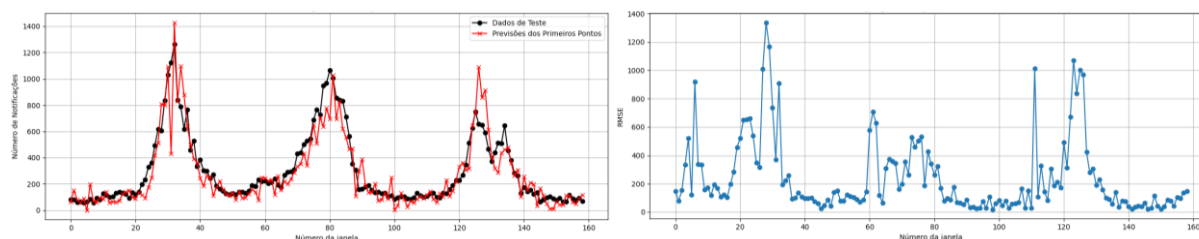


Figura B.28 Previsões do primeiro ponto das janelas de teste e RMSE das janelas com horizonte de previsão 12 – Campo Grande

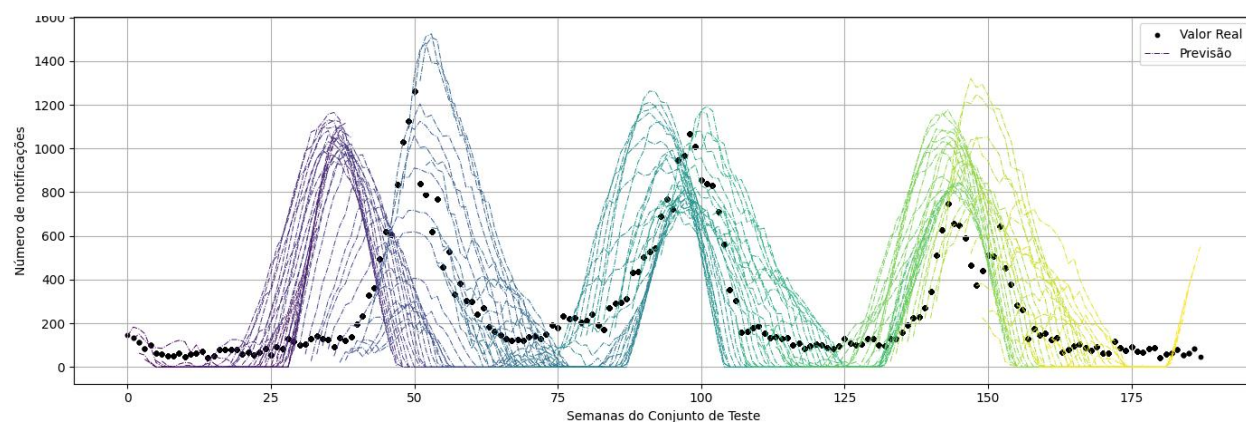
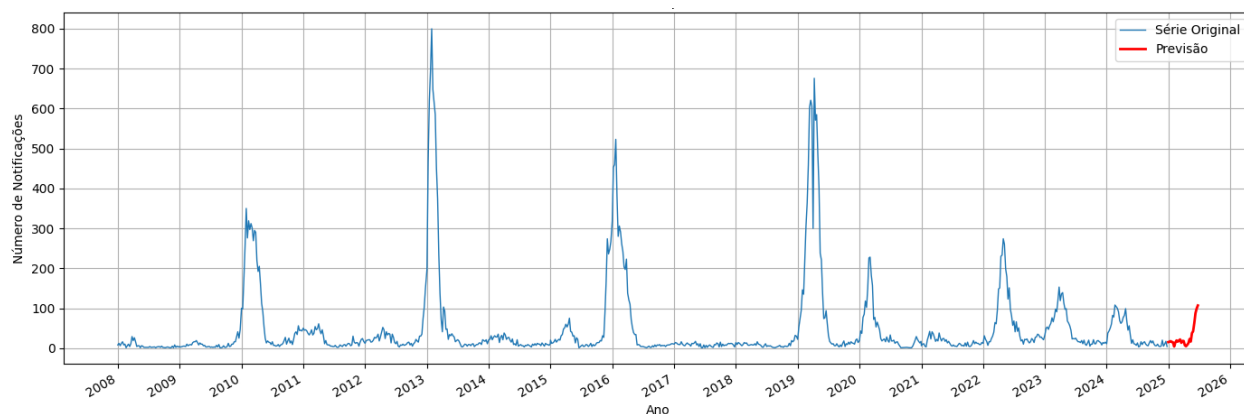
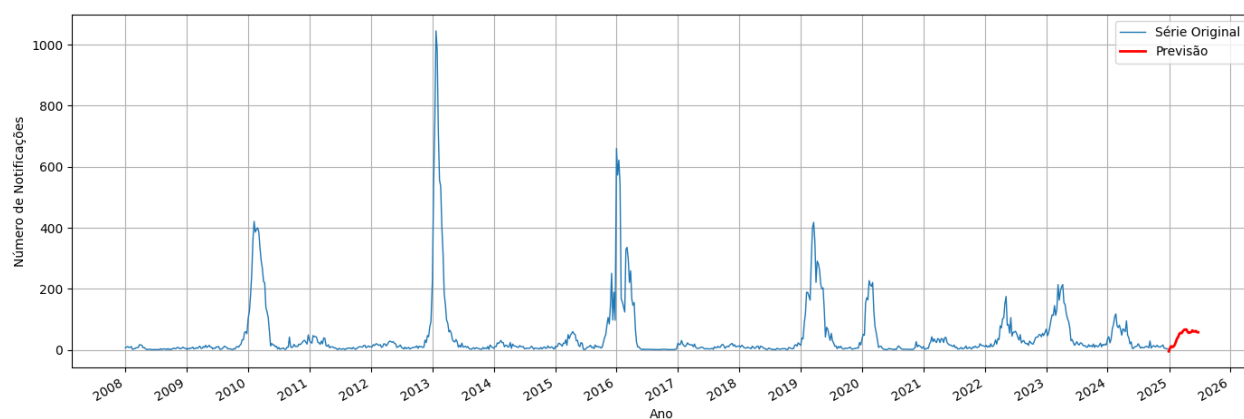
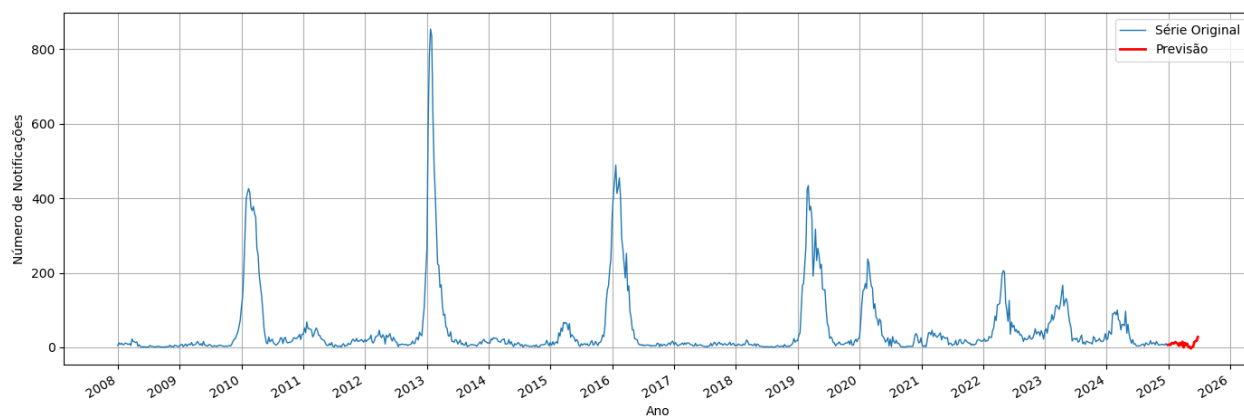


Figura B.29 Janelas completas de previsões no conjunto de teste com horizonte de

previsão 26 - Campo Grande

**Figura B.30** Quantidade de casos com previsão de 26 semanas – Segredo**Figura B.31** Quantidade de casos com previsão de 26 semanas – Imbirussu**Figura B.32** Quantidade de casos com previsão de 26 semanas – Bandeira

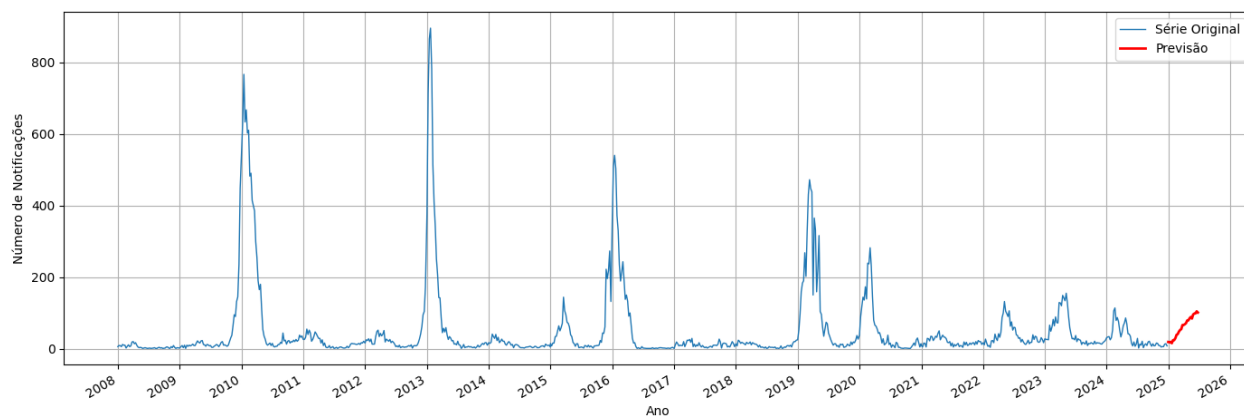


Figura B.33 Quantidade de casos com previsão de 26 semanas – Lagoa

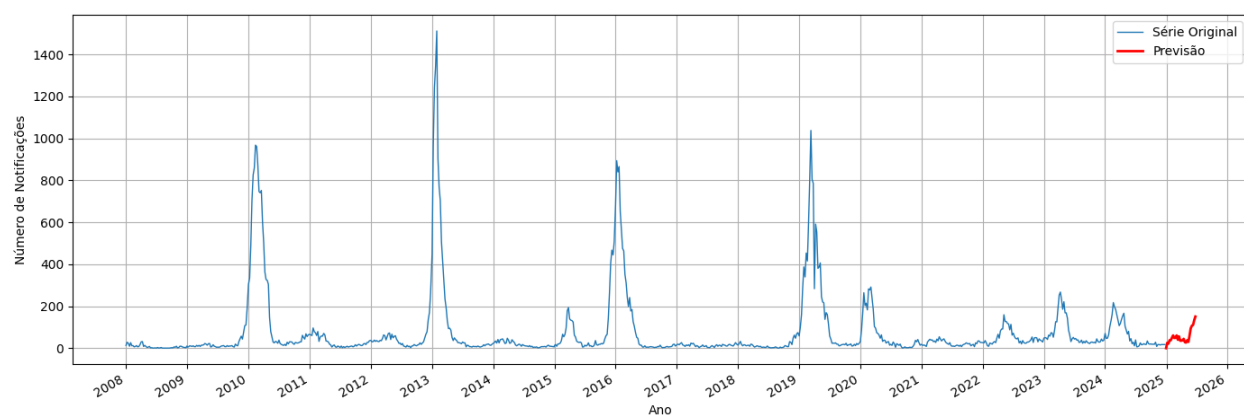


Figura B.34 Quantidade de casos com previsão de 26 semanas – Anhanduizinho

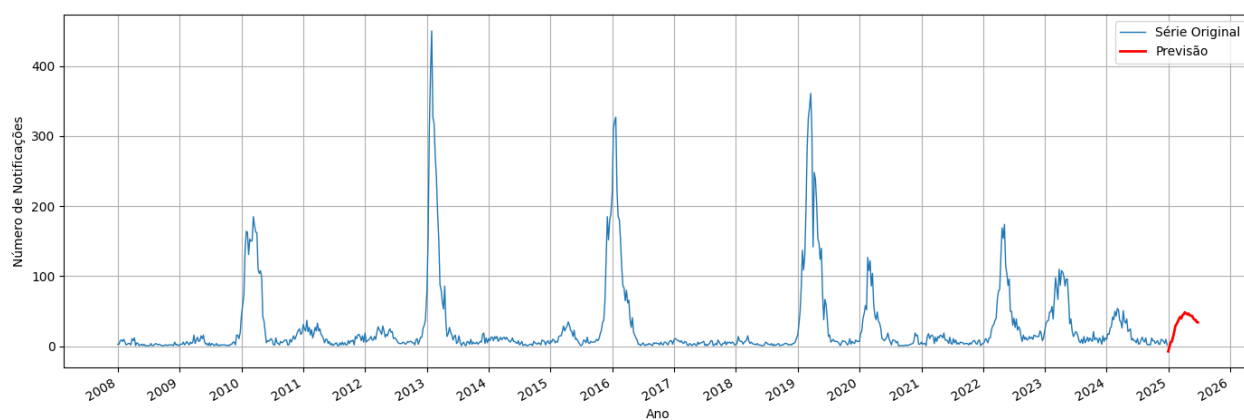


Figura B.35 Quantidade de casos com previsão de 26 semanas – Prosa

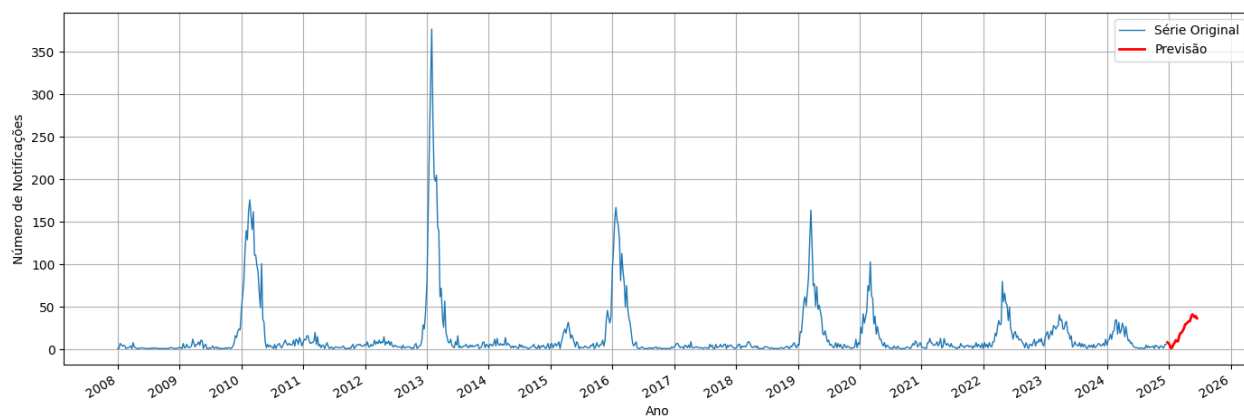


Figura B.36 Quantidade de casos com previsão de 26 semanas – Centro

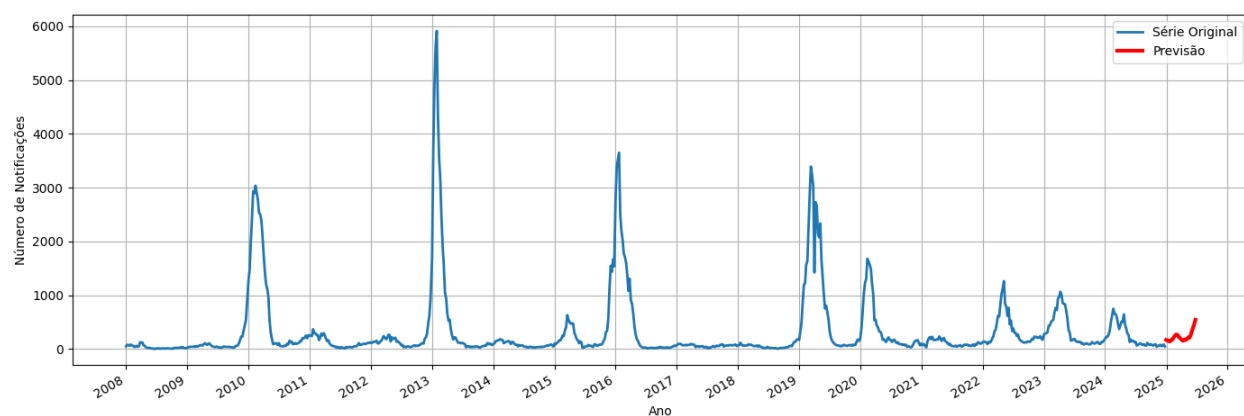


Figura B.37 Quantidade de casos com previsão de 26 semanas - Campo Grande

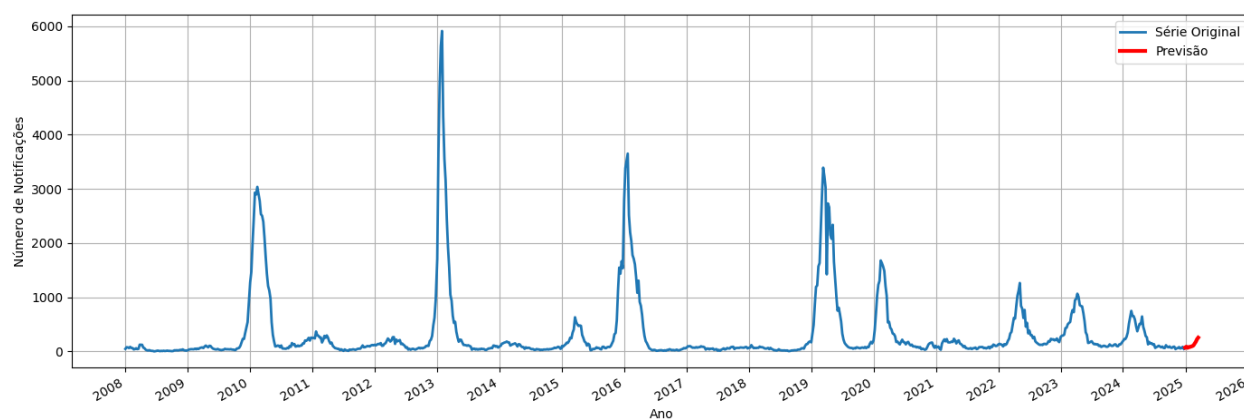
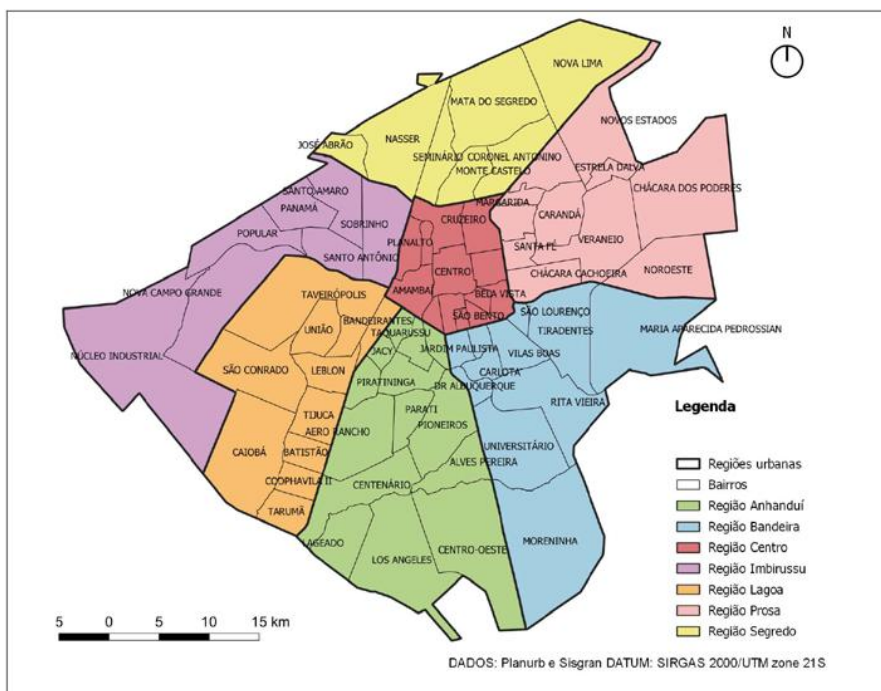


Figura B.38 Quantidade de casos com previsão de 12 semanas - Campo Grande

ANEXO



Anexo C.1 Mapa dos distritos sanitários de Campo Grande