

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ

BRENDON LEONAM PASQUIM

**QUANDO CHEGA? ANÁLISE DE PREVISIBILIDADE DE TEMPOS DE
ESPERA EM TRANSPORTE PÚBLICO URBANO UTILIZANDO
MODELOS MISTOS**

CURITIBA

2025

BRENDON LEONAM PASQUIM

**QUANDO CHEGA? ANÁLISE DE PREVISIBILIDADE DE TEMPOS DE
ESPERA EM TRANSPORTE PÚBLICO URBANO UTILIZANDO
MODELOS MISTOS**

**When does it arrive? An Analysis of Passenger Waiting Time Predictability
in Urban Public Transport Using Mixed Models**

Dissertação apresentada como requisito para obtenção do título de Mestre em Ciências do Programa de Pós-Graduação em Engenharia Elétrica e Informática Industrial da Universidade Tecnológica Federal do Paraná.

Orientador(a): Prof^ª. Dr^ª. Keiko Veronica Ono Fonseca

Coorientador(a): Prof. Dr. Luiz Ledo Mota Melo Junior

CURITIBA

2025



[4.0 Internacional](https://creativecommons.org/licenses/by-nc-sa/4.0/)

Esta licença permite remixe, adaptação e criação a partir do trabalho, para fins não comerciais, desde que sejam atribuídos créditos ao(s) autor(es) e que licenciem as novas criações sob termos idênticos. Conteúdos elaborados por terceiros, citados e referenciados nesta obra não são cobertos pela licença.



BRENDON LEONAM PASQUIM

QUANDO CHEGA? ANÁLISE DE PREVISIBILIDADE DE TEMPOS DE ESPERA EM TRANSPORTE PÚBLICO URBANO UTILIZANDO MODELOS MISTOS

Trabalho de pesquisa de mestrado apresentado como requisito para obtenção do título de Mestre Em Ciências da Universidade Tecnológica Federal do Paraná (UTFPR). Área de concentração: Telecomunicações E Redes.

Data de aprovação: 04 de Agosto de 2025

Dra. Keiko Veronica Ono Fonseca, Doutorado - Universidade Tecnológica Federal do Paraná

Dra. Angelica Maria Tortola Ribeiro, Doutorado - Universidade Tecnológica Federal do Paraná

Dr. Ernani Francisco Choma, Doutorado - Harvard University

Dr. Ricardo Luders, Doutorado - Universidade Tecnológica Federal do Paraná

Documento gerado pelo Sistema Acadêmico da UTFPR a partir dos dados da Ata de Defesa em 04/08/2025.

AGRADECIMENTOS

O quanto se sacrifica por um sonho? Quem se sacrifica por um sonho? O sucesso acadêmico, o diploma na parede, um abraço e um sorriso das pessoas queridas conseguem apagar muitas das mágoas e tristezas passadas, deixando para trás apenas a serenidade e a resiliência do caminho difícil.

Dedico essa conquista a minha família, em especial minha mãe, Maria Alice, que sempre me fez acreditar que era capaz; aos meus amigos, que, a cada piada e risada, me davam forças para não desistir; aos meus colegas de pesquisa e de laboratório, que me faziam acreditar que havia um fim naquela estrada em que eu estava. E a meus orientadores, Prof^a. Dr. Keiko V. O. Fonseca e Prof. Dr. Luiz Ledo M. Melo, baluartes da paciência, que me ajudaram a transformar o caos de minhas ideias em ciência, em avanço, em benefício a nossa sociedade.

Por fim, agradeço à educação pública, gratuita e de qualidade, que transformou um garoto em técnico, um jovem em engenheiro e um homem em mestre. Se ela fez isso por mim, o que mais ela pode fazer pelo Brasil?

"As cidades têm a capacidade de proporcionar algo a todos, apenas porque, e somente quando, são criadas por todos."(JACOBS, 1961).

RESUMO

Este trabalho apresenta uma análise exploratória de dados abertos de transporte público urbano e propõe modelos para a previsão do tempo de espera de passageiros em pontos de parada de ônibus. Para isso, foram utilizadas abordagens baseadas em modelos mistos e intervalos de confiança, com o objetivo de identificar padrões operacionais do sistema de transporte em um estudo de caso na cidade de Curitiba. Os resultados demonstram que é possível desenvolver modelos preditivos de forma incremental, considerando as particularidades das linhas analisadas e adaptando-os às necessidades e à realidade dos dados disponíveis. As contribuições do estudo incluem: (1) a aplicação de modelos lineares mistos e seus potenciais benefícios para usuários e operadores; (2) a integração de métodos e ferramentas adequados à análise de dados de transporte público; (3) uma discussão sobre o uso dos modelos para avaliação da qualidade do serviço; e (4) recomendações sobre variáveis relevantes a serem incorporadas aos modelos e seu impacto nos tempos de espera.

Palavras-chave: previsão de tempo de espera; modelos mistos; transporte público urbano; dados abertos.

ABSTRACT

This study presents an exploratory analysis of open urban public transportation data and proposes models for predicting passenger waiting times at bus stops. To achieve this, mixed models and confidence intervals were employed to identify operational patterns in a case study focused on the city of Curitiba. The results show that it is possible to develop predictive models incrementally, taking into account the specific characteristics of the analyzed bus lines and adapting the models to the needs and realities of the available data. The contributions of this study include: (1) the implementation of linear mixed models and their potential benefits for both users and system operators; (2) the integration of methods and tools specifically designed to handle public transportation data; (3) a discussion on how model outputs can be used to assess service quality; and (4) recommendations on relevant variables to be included in the models and their impact on waiting times.

Keywords: wait time forecast; mixed models; public urban transport; open data.

LISTA DE FIGURAS

Figura 1	– Etapas da metodologia proposta	25
Figura 2	– Conceitos de Itinerário S e Tempo de Espera de Itinerário T	26
Figura 3	– Dados extraídos do repositório do C3SL	26
Figura 4	– Parâmetros de entrada e saída para a aplicação “urbs-data-processing” desenvolvida em (PEIXOTO <i>et al.</i>, 2020)	27
Figura 5	– Parâmetros de entrada e saída para a aplicação “busanalysis” desenvolvida em (BORGES <i>et al.</i>, 2023)	28
Figura 6	– Parâmetros de entrada e saída para o script R	29
Figura 7	– Dados e ferramentas utilizadas	29
Figura 8	– Mapa da linha 863 em Curitiba, itinerário 12494	30
Figura 9	– Dados observados das 20 execuções do itinerário 12494 pelo ônibus HC306	33
Figura 10	– Comparação de previsões dos modelos M1 e M2 considerando todas as 20 execuções do itinerário 12494 de um dia de operação do veículo HA310	36
Figura 11	– 5 maiores e 5 menores tempos de espera em pontos de parada. Efeito Fixo é plotado em preto	37
Figura 12	– Gráfico de Autocorrelação do modelo M2	38
Figura 13	– Gráfico de Autocorrelação do modelo M3	38
Figura 14	– Gráfico de Autocorrelação do modelo M3 Sazonal	39
Figura 15	– Gráfico de Previsão do modelo M2 com Intervalo de Confiança considerando as 20 execuções do itinerário 12494	40
Figura 16	– Gráfico de Previsão do modelo M3 com Intervalo de Confiança considerando as 20 execuções do itinerário 12494	41
Figura 17	– Gráfico de Previsão do modelo M3 Sazonal com Intervalo de Confiança considerando as 20 execuções do itinerário 12494	41
Figura 18	– Dados observados para o dia 11/09/2024 e intervalo de confiança para o dia 11/09/2024	43
Figura 19	– Dados observados para o dia 18/09/2024 e intervalo de confiança para o dia 11/09/2024	44

LISTA DE TABELAS

Tabela 1 – Comparação entre horários da tabela horária de referência e da execução do Itinerário.....	34
Tabela 2 – EQM Modelos	42

LISTA DE QUADROS

Quadro 1 – Definição dos termos do modelo misto no contexto de transporte público	16
Quadro 2 – Visão geral da literatura.....	23
Quadro 3 – Outras datas analisadas	30
Quadro 4 – Linha analisada (Referência).....	31
Quadro 5 – Modelos de previsão desenvolvidos	35
Quadro 6 – Destaques dos modelos M1 e M2	36

LISTA DE ABREVIATURAS E SIGLAS

UTFPR	Universidade Tecnológica Federal do Paraná
STPU	Sistema de Transporte Público Urbano
URBS	Urbanização de Curitiba
IPPUC	Instituto de Pesquisa e Planejamento Urbano de Curitiba
LLM	Linear Mixed Models
AR	Autoregressivo
EQM	Erro Quadrático Médio
REQM	Raiz quadrada do Erro Quadrático Médio
ARIMA	Autoregressive Integrated Moving Average
KNN	k-Nearest Neighbours
KF	Kalman Filtering
KKM	Markov chain
LR	Linear Regression
DTR	Decision Tree Regressor
ESDR	Exponential Smoothing Deep Residual
MAPE	Mean absolute percentage error
t-Full	Student's t-Full
LSTM	Long Short-Term Memory Recurrent
IWOA	Improved Whale Optimization Algorithm
GA-BP	Genetic Algorithm - Backpropagation
C3SL	Centro de Computação Científica e Software Livre
JSON	JavaScript Object Notation
CSV	Comma-separated values
HTML	Hypertext Markup Language
ADF	Augmented Dickey–Fuller test

SUMÁRIO

1	INTRODUÇÃO	12
1.1	Objetivos.....	12
2	REFERENCIAL TEÓRICO.....	14
2.1	Dados Públicos Abertos.....	14
2.2	Sistemas de transporte público	14
2.3	Modelos Mistos	15
2.3.1	Aplicação de Modelos Mistos ao Transporte Público	16
2.3.2	Modelagem do Erro	17
2.4	Modelos Mistos no R	17
2.4.1	Forma final do Modelo por Grupo.....	18
2.4.2	Modelo combinado para Todos os Grupos	18
2.4.3	Exemplo: Matriz R Bloco-Diagonal	19
2.5	Correlograma e seleção de modelos	20
2.5.1	Métricas de Qualidade	20
3	TRABALHOS RELACIONADOS.....	22
3.1	Trabalhos Publicados.....	23
4	METODOLOGIA.....	24
4.1	Formalização.....	24
4.2	Dados do sistema de Transporte Público	25
4.3	Identificação de itinerários	27
4.4	Predição de Tempo de Espera com Modelos Mistos.....	29
4.5	Estudo de Caso	30
5	DISCUSSÃO E RESULTADOS	32
5.1	Análise de aderência da tabela horária aos dados	32
5.1.1	Metodologia	32
5.1.2	Análise dos Dados.....	32
5.1.3	Resultados	33
5.2	Modelos de previsão	35
5.3	Análise de correlação dos resíduos.....	37
5.4	Análise de propriedades dos resíduos	39
5.4.1	Teste de Resíduos	39
5.4.2	Análise Gráfica	40
5.4.3	Erro Quadrático Médio	42
5.5	Análise de aderência do modelo aos dados	42
5.5.1	Metodologia	42
5.5.2	Resultados	42
6	CONCLUSÃO	45
7	TRABALHOS FUTUROS.....	46
	REFERÊNCIAS.....	47

1 INTRODUÇÃO

Os efeitos das mudanças climáticas têm demandado estudos para o transporte público de massa (WIMBADI; DJALANTE; MORI, 2021) (GALLIVAN *et al.*, 2011). A adoção do transporte público pelo usuário depende não só da sua condição financeira, mas também da sua percepção sobre a relação custo-benefício. Se apresentar custos e benefícios melhores ou equivalentes comparada com o uso do veículo individual aumenta-se a chance do usuário decidir pelo uso do sistema de transporte público urbano (STPU) (HABIB; KATTAN; ISLAM, 2011).

Em termos de benefícios, em geral se considera a qualidade da prestação do serviço, por exemplo, medida pelo tempo de deslocamento do veículo ou tempo de espera do passageiro, acessibilidade, comodidade, conforto e usabilidade, pontualidade (aqui entendida como aderência à tabela horária fornecida previamente), previsibilidade (aqui entendida como o estudo da redução da variabilidade prevista com a aproximação temporal de um evento previsto) (BüCHEL; CORMAN, 2022), entre outros (HABIB; KATTAN; ISLAM, 2011).

A partir de um estudo exploratório, busca-se contribuir com (1) um melhor entendimento sobre a operação do STPU a partir de padrões nos dados abertos disponíveis de uma cidade, (2) o uso de modelos mistos para a análise do tempo de espera e previsibilidade da prestação de serviços do STPU (3) e sugestões e novas hipóteses a serem investigadas em estudos confirmatórios.

Um estudo de caso com dados abertos de Curitiba, em especial, de ônibus do seu STPU é apresentado e os resultados discutidos à luz da qualidade de dados, modelos e estudos anteriores. Em específico, considera-se a tabela horária de referência do STPU a partir de informações geotemporais públicas e avalia-se a aderência dessa tabela a partir de dados históricos. Investiga-se ainda outros possíveis dados públicos que possam ser relacionados ao sistema de transporte público e sua operação, possíveis fatores de impacto na aderência à tabela horária, bem como estratégias de melhor previsão da chegada de ônibus em seus pontos de parada.

Este documento está assim organizado: a partir desta introdução, a próxima seção apresenta os conceitos chave de STPU, de modelos mistos e de ferramentas de manipulação de dados aplicados neste trabalho. A Seção 3 apresenta uma revisão da literatura, discutindo trabalhos relacionados. A Seção 4 descreve o método utilizado, detalhando o estudo de caso e a Seção 5 traz os resultados obtidos, uma abordagem implementada para sua validação e uma discussão sobre os mesmos. A Seção 6 apresenta as considerações finais e alguns estudos futuros.

1.1 Objetivos

Dois são os objetivos gerais deste trabalho: (1) Analisar a aderência da tabela horária de referência aos dados observados do transporte público e, (2) desenvolver um modelo misto capaz de prever tempos de espera de passageiros em pontos de ônibus de um STPU. Estes objetivos são subdivididos em quatro objetivos específicos:

1. Coletar e preparar os dados públicos do STPU de Curitiba;
2. Identificar itinerários de ônibus e compará-los a tabela horária de referência do STPU de Curitiba;
3. Implementar um modelo misto extensível para calcular o tempo de espera geral e ponderado em pontos de ônibus;
4. Avaliar a aderência do modelo misto proposto utilizando intervalos de confiança e análise de resíduos.

2 REFERENCIAL TEÓRICO

2.1 Dados Públicos Abertos

Adota-se aqui o conceito de dados abertos como aqueles disponíveis de forma livre, podendo ser utilizados ou republicados sem restrições de direitos autorais ou patentes (BRAUNSCHEWIG *et al.*, 2012).

No Brasil, desde 2011 está em vigor a Lei de Acesso à Informação Pública (Lei 12.527/2011). Através desta, o acesso a dados públicos é regulamentado e divulgado independentemente de solicitações de terceiros (BRASIL, 2011).

Algumas cidades oferecem dados abertos do Transporte Público tais como Linhas e Pontos de ônibus, itinerários e a geolocalização dos ônibus para consulta e consumo. Em Curitiba, o “Portal de Dados Abertos” disponibiliza dados e informações de domínio público para livre utilização pela sociedade¹. Além do objetivo de atender a legislação vigente sobre transparência da gestão pública, os dados abertos podem ser uma fonte para a geração de evidências científicas a serem usadas no estabelecimento, acompanhamento e avaliação de políticas públicas.

Neste trabalho utiliza-se dados abertos fornecidos pela URBS (Urbanização de Curitiba) e IPPUC (Instituto de Pesquisa e Planejamento Urbano de Curitiba)² descritos na seção 4.2.

2.2 Sistemas de transporte público

Transporte público pode ser definido como qualquer forma de transporte de passageiros ou cargas disponível para locação e remuneração. Geralmente se refere ao transporte terrestre de passageiros através de ônibus, trens e suas variantes (PRESTON, 2009). Em geral, um sistema de transporte público tem sua qualidade associada aos serviços prestados, ao número de usuários atendidos, seu custo, além de seus impactos sociais e ambientais.

O presente estudo limita-se ao transporte público de passageiros em ônibus urbanos, sem perda de generalidade. Define-se aqui itinerário ou percurso de um veículo associado a uma rota e horários pré-determinados como uma sequência ordenada de pontos que o veículo deve obedecer para embarque e desembarque de passageiros. Passageiros são aqueles que utilizam os veículos de transporte público para se deslocar. Pontos de Ônibus ou Pontos de Parada de Ônibus são locais dedicados para o embarque e desembarque de passageiros. Por fim, um ônibus é um veículo automotor de grande porte dedicado ao transporte em massa de passageiros.

A avaliação de desempenho do serviço de transporte público, em geral, considera o comportamento temporal da oferta do serviço ao usuário e de dados abertos (MOUWEN, 2015).

Embora o desempenho de STPU possa ser associado a vários parâmetros, entre eles, ocupação do ônibus, aderência à tabela horária, conforto interno, frequência do serviço, etc; nosso estudo foca na avaliação da aderência à tabela horária fornecida pelo provedor do serviço

¹ <https://www.curitiba.pr.gov.br/dadosabertos/> (acesso em 01/04/2025)

² <https://dadosabertos.c3sl.ufpr.br/curitiba/TransporteColetivo> (acesso em 01/04/2025)

de transporte público, bem como fornecer previsões de tempos de espera para todos os pontos de parada do itinerário a partir dos dados abertos disponíveis.

2.3 Modelos Mistos

Em muitas áreas a investigação científica busca estabelecer um modelo geral para uma população, onde dentre as várias características em comum dessa população, seleciona-se uma variável de interesse Y e um vetor de variáveis explicativas independentes $X = [X_1, \dots, X_n]$ com objetivo de prever e explicar Y .

Um possível candidato de modelo geral é o da regressão múltipla, definida por:

$$Y = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p + \epsilon$$

onde $\epsilon \sim N(0, \sigma^2)$ e N uma distribuição normal de média zero e variância σ^2 . β s são os coeficientes angulares em relação à variável explicativa independente e o ϵ os fatores residuais (valores não previstos pelo modelo), mais os possíveis erros de medição.

Note que se faz uma suposição de homocedasticidade, ou seja, assume-se que em $Y|x \sim N(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p, \sigma^2)$, a variância é constante para qualquer valor x e foi fixada para toda população. Esta é uma suposição assumida ao longo da população descartando a possibilidade da formação de grupos na população que podem ter diferença na variabilidade (MATLOFF, 2017).

Uma forma de levar em conta a diferença entre grupos na população é com o uso de modelos mistos (mais especificamente, modelos lineares mistos ou *Linear Mixed Models*, LLM), uma metodologia que incorpora a regressão múltipla como efeito fixo da população e também permite que os coeficientes e interceptos β_0 variem em função dos grupos (como efeitos aleatórios) (FARAWAY, 2006).

Em aplicações onde ao longo do tempo ocorrem medidas repetidas de indivíduos ou grupos na população sob análise, é possível estabelecer uma análise de dados longitudinais (FITZMAURICE; LAIRD; WARE, 2012). Um caso particular de modelos mistos é definido na Equação 1:

$$Y_{it} = \beta_0 + b_{0i} + \beta_1 t_{1it} + b_{1i} t_{1it} + \dots + \beta_p t_{pit} + b_{pi} t_{pit} + \epsilon_{it} \quad (1)$$

Os termos b_{0i} e b_{pi} representam efeitos aleatórios específicos do grupo i . O termo b_{0i} afeta o intercepto β_0 , ou seja, permite que cada grupo tenha um valor médio diferente. Já os termos b_{pi} afetam a inclinação de cada variável explicativa t_{pit} , permitindo que os coeficientes variem entre os grupos.

2.3.1 Aplicação de Modelos Mistos ao Transporte Público

O Quadro 1 define os termos utilizados na Equação 1, indicando seu significado no contexto da previsão de tempos de espera em transporte público urbano.

Quadro 1 – Definição dos termos do modelo misto no contexto de transporte público

Símbolo / Termo	Definição estatística	Significado no contexto de transporte público
Y_{it}	Variável resposta para o grupo i no tempo t	Tempo de espera previsto (em segundos) no ponto de ônibus i no instante t
β_0	Intercepto fixo da população	Média geral de tempo de espera para todos os pontos de ônibus de um mesmo itinerário
b_{0i}	Efeito aleatório do intercepto para o grupo i	Diferença específica do ponto i em relação à média geral (positiva se o ponto tende a ter espera maior, negativa se tende a ter espera menor)
β_p	Coefficiente fixo da variável explicativa p	Impacto médio (na população) da variável explicativa p no tempo de espera (ex.: chuva, horário de pico, dia da semana)
t_{pit}	Valor da variável explicativa p para o grupo i no tempo t	Medida da condição associada ao ponto i no momento t (ex.: intensidade da chuva, volume de tráfego, número de passageiros embarcando)
b_{pi}	Efeito aleatório do coeficiente para a variável p no grupo i	Variação no efeito da variável p específica do ponto i (ex.: um ponto próximo a um cruzamento pode ser mais sensível ao trânsito que outros)
ϵ_{it}	Erro aleatório, com distribuição $N(0, \sigma^2)$	Variações imprevisíveis no tempo de espera devido a fatores não medidos ou aleatórios (ex.: falha mecânica, evento inesperado no trajeto)
$i = 1, \dots, G$	Índice do grupo	Identifica cada ponto de ônibus
$t = 1, \dots, T$	Índice de tempo	Instante ou período de observação
$N = G \times T$	Total de observações	Número total de registros de tempo de espera coletados

Fonte: Autoria Própria (2025).

Modelos Mistos são utilizados neste trabalho para modelar o comportamento do STPU (que pode ser imprevisível), no qual a composição do efeito fixo aliado ao efeito aleatório modela as diversas variáveis que afetam o tempo de espera em pontos de parada de ônibus (por exemplo, chuva, congestionamento, embarque e desembarque de passageiros em pontos de parada de grande fluxo, acidentes de trânsito, etc).

Neste contexto, β_0 representa a média geral de tempo de espera (em segundos), isto é, de um mesmo ônibus para todos os pontos de ônibus em um dado itinerário. Ou seja, é um valor constante dentro do itinerário.

Já b_{0i} representa o fator aleatório específico do ponto de ônibus, sendo responsável por ponderar o tempo de espera de cada ponto de ônibus em relação a média (sendo um número positivo ou negativo). Por isso, é um valor variável por ponto de ônibus do itinerário. Um valor positivo significa que o tempo de espera no ponto de parada é superior a média geral de espera do itinerário. Da mesma forma, um valor negativo representa um tempo de espera inferior a média geral do itinerário.

É importante notar que os dados disponíveis (vide a Seção 2.1) fornecem o tempo total decorrido para um ônibus transitar de um ponto de parada a outro, mas não detalha quais variáveis afetaram esse tempo. Portanto, os termos β_p , b_{pi} e t_{pit} são desconsiderados na construção do modelo de previsão (para detalhes sobre variáveis, vide Seção 7).

2.3.2 Modelagem do Erro

Duas propostas de modelagem do erro ϵ_{it} do grupo i no tempo t são apresentadas, baseado-se na existência de correlação temporal entre os dados.

$$\text{Corr}(\epsilon_{it}, \epsilon_{ik}) = 0, \forall t \neq k \quad (2)$$

A Equação 2 modela o erro quando não existe correlação temporal entre os dados que são avaliados através dos resíduos. Caso contrário, isso é, se houver uma estrutura temporal dos resíduos, então:

$$\text{Corr}(\epsilon_{it}, \epsilon_{ik}) = \rho, \forall t \neq k \quad (3)$$

Identificando-se a estrutura de correlação, como na Equação 3, pode-se ajustar modelos AR ou Box e Jenkins (BOX *et al.*, 2015).

2.4 Modelos Mistos no R

No software R, utilizou-se a função `lme`, disponível no pacote `nlme`, que permite ajustar modelos lineares mistos assumindo uma estrutura de erros com distribuição normal multivariada (PINHEIRO *et al.*, 2018).

Para cada grupo i , o “Modelo Hierárquico por Grupo” pode ser escrito como:

$$\mathbf{y}_i = \mathbf{X}_i \beta + \mathbf{Z}_i b_i + \epsilon_i$$

onde:

- $\mathbf{y}_i \in \mathbb{R}^T$: vetor de respostas para o grupo i ,
- $\mathbf{X}_i = \mathbf{1}_{T \times 1}$: matriz de covariáveis (somente o intercepto fixo),
- $\beta \in \mathbb{R}$: intercepto fixo comum a todos os grupos,

- $\mathbf{Z}_i = \mathbf{1}_{T \times 1}$: matriz de efeitos aleatórios (também apenas intercepto),
- $b_i \sim \mathcal{N}(0, \sigma_b^2)$: efeito aleatório de intercepto do grupo i ,
- $\boldsymbol{\epsilon}_i \sim \mathcal{N}(0, \mathbf{R}_i)$: vetor de erros autocorrelacionados com estrutura $AR(1)$.

e, além disso, os erros são independentes entre grupos, ou seja:

$$\text{Cov}(\boldsymbol{\epsilon}_i, \boldsymbol{\epsilon}_j) = \mathbf{0}, \quad \text{para } i \neq j$$

A estrutura $AR(1)$ dos erros define a matriz de covariância como:

$$\text{Cov}(\boldsymbol{\epsilon}_i) = \mathbf{R}_i = \sigma^2 \cdot \boldsymbol{\Lambda}_i$$

onde $\boldsymbol{\Lambda}_i$ é uma matriz $T \times T$ com elementos:

$$(\boldsymbol{\Lambda}_i)_{jk} = \rho^{|j-k|}, \quad \text{com } |\rho| < 1$$

2.4.1 Forma final do Modelo por Grupo

Pela não inclusão de variáveis explicativas (como distância, trânsito, etc), a forma final do modelo por grupo i programado no R é:

$$\mathbf{y}_i = \mathbf{1}_T \beta + \mathbf{1}_T b_i + \boldsymbol{\epsilon}_i$$

com:

$$b_i \sim \mathcal{N}(0, \sigma_b^2), \quad \boldsymbol{\epsilon}_i \sim \mathcal{N}(0, \sigma^2 \boldsymbol{\Lambda}_i)$$

2.4.2 Modelo combinado para Todos os Grupos

Empilhando todos os \mathbf{y}_i em um vetor $\mathbf{y} \in \mathbb{R}^N$, o modelo completo fica:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{b} + \boldsymbol{\epsilon}$$

onde:

$$\mathbf{b} = (b_1, \dots, b_G)^T \sim \mathcal{N}(0, \sigma_b^2 \cdot \mathbf{I}_G)$$

$$\boldsymbol{\epsilon} \sim \mathcal{N}(0, \mathbf{R})$$

com \mathbf{R} sendo uma matriz bloco-diagonal:

$$\mathbf{R} = \text{blockdiag}(\sigma^2 \mathbf{\Lambda}_1, \dots, \sigma^2 \mathbf{\Lambda}_G)$$

Cada bloco $\sigma^2 \mathbf{\Lambda}_i$ representa a covariância temporal dos erros dentro do grupo i com estrutura $AR(1)$.

2.4.3 Exemplo: Matriz \mathbf{R} Bloco-Diagonal

A matriz total de covariância dos erros, \mathbf{R} , pode ser escrita como uma matriz bloco-diagonal, onde cada bloco corresponde à matriz de covariância dos resíduos de um grupo:

$$\mathbf{R} = \begin{bmatrix} \sigma^2 \mathbf{\Lambda}_1 & 0 & \cdots & 0 \\ 0 & \sigma^2 \mathbf{\Lambda}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \mathbf{\Lambda}_G \end{bmatrix} = \text{blockdiag}(\sigma^2 \mathbf{\Lambda}_1, \dots, \sigma^2 \mathbf{\Lambda}_G)$$

Se cada grupo tem $T = 3$ observações e $\rho = 0.5$, então cada bloco é:

$$\mathbf{\Lambda}_i = \begin{bmatrix} 1 & 0.5 & 0.25 \\ 0.5 & 1 & 0.5 \\ 0.25 & 0.5 & 1 \end{bmatrix}$$

e a matriz \mathbf{R} para $G = 2$ grupos ficaria:

$$\mathbf{R} = \sigma^2 \cdot \begin{bmatrix} 1 & 0.5 & 0.25 & 0 & 0 & 0 \\ 0.5 & 1 & 0.5 & 0 & 0 & 0 \\ 0.25 & 0.5 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0.5 & 0.25 \\ 0 & 0 & 0 & 0.5 & 1 & 0.5 \\ 0 & 0 & 0 & 0.25 & 0.5 & 1 \end{bmatrix}$$

Aqui considerou-se que atrasos em pontos de ônibus se acumulam ao longo do percurso, configurando uma estrutura autorregressiva nos resíduos do modelo. Essa dependência temporal reflete que o atraso em um ponto afeta os atrasos subsequentes, caracterizando um processo do tipo $AR(P)$.

Embora o software R ofereça suporte nativo para estruturas de erros $AR(1)$ em modelos mistos (por exemplo, através da função `corAR1` no pacote `nlme`), o problema em discussão neste trabalho requer a modelagem com estruturas autorregressivas de ordem superior, $AR(P)$ (considerando P o número de pontos de parada no itinerário, ou seja, sempre superior a um).

Para contornar essa limitação, adotou-se uma abordagem baseada na estimação por Mínimos Quadrados Estendidos (*Extended Least Squares*), conforme descrito em (AGUIRRE, 2007). Nesta metodologia, inicialmente estimou-se os efeitos fixos do modelo, obtendo os resíduos correspondentes, que são então modelados separadamente pela estrutura autorregressiva desejada.

No repositório Git do autor³, apresenta-se uma prova de conceito utilizando um exemplo com resíduos $AR(1)$, demonstrando que a abordagem por Mínimos Quadrados Estendidos apresenta resultados muito próximos aos obtidos com o modelo $AR(1)$ nativamente implementado no R.

2.5 Correlograma e seleção de modelos

Correlograma é a representação gráfica da correlação entre múltiplas variáveis (correlação cruzada), ou até mesmo, da mesma variável ao longo do tempo (autocorrelação). No contexto da análise de autocorrelação o correlograma é também conhecido como “gráfico de correlação serial”.

Correlogramas possuem diversas aplicações, dentre elas, aprimoramento de modelos e resíduos, análise de dependência em séries temporais e escolha de modelos de previsão. No presente trabalho, correlogramas foram utilizados para definir parâmetros autorregressivos para resíduos de modelos de predição. Para isso, calculou-se a correlação entre observações defasadas por j períodos de tempo para se obter os coeficientes de autocorrelação (Equação 4):

$$r_j = \frac{\sum_{k=1}^{n-j} (t_{i'k} - \bar{t}_{i'k})(t_{i'k+j} - \bar{t}_{i'k+j})}{\sum_{k=1}^n (t_{i'k} - \bar{t}_{i'k})^2} \quad (4)$$

Os coeficientes de autocorrelação são dados por $r_j = \frac{c_j}{c_0}$, onde c_j é a covariância entre observações defasadas em j períodos de tempo. $\bar{t}_{i'k}$ representa o valor previsto e $t_{i'k}$ o valor observado.

Um gráfico com os j primeiros coeficientes de autocorrelação como função de j é chamado de correlograma. Modelos autorregressivos tem correlogramas que caem exponencialmente (BOX *et al.*, 2015).

2.5.1 Métricas de Qualidade

A seleção do modelo mais adequado, ou seja, aquele com melhor ajuste aos dados reais, se fez através do “Erro Quadrático Médio” (EQM). EQM é uma métrica usada para avaliar a qualidade de previsões em modelos estatísticos, medindo a diferença média quadrática entre os valores previstos $\bar{t}_{i'k}$ e observados $t_{i'k}$. EQM é definido na Equação 5:

³ <https://github.com/Brendonpasquim/mixed-model-bus-wait-time-forecast>

$$EQM = \frac{1}{n} \sum_{k=1}^n (t_{i'k} - t_{i'k}^-)^2 \quad (5)$$

A raiz quadrada do EQM, REQM, é dada na Equação 6. Essa métrica é especialmente útil quando se deseja analisar o erro do modelo na mesma unidade de medida da variável sob análise.

$$REQM = \sqrt{EQM} \quad (6)$$

3 TRABALHOS RELACIONADOS

Um indicador associado à qualidade do serviço de transporte público é a aderência à tabela horária (caracterizada por instantes de tempo de chegada e partida dos veículos do transporte público em um ponto específico de seu itinerário), oferecendo ao usuário previsibilidade para embarcar, desembarcar, fazer conexões e assim alcançar seu destino em um intervalo de tempo limitado (BüCHEL; CORMAN, 2022).

A previsibilidade advinda da predição do tempo de chegada é afetada pela qualidade e abrangência dos dados reais utilizados. Erros, incertezas e dados repetidos no conjunto de dados afetam os resultados das previsões, exigindo a aplicação proativa de estratégias de tratamento, como em (MARTINS *et al.*, 2022).

Modelos para predição de tempos de chegada já foram propostos anteriormente. (SUWARDI; NAPIAH; KAMARUDDIN, 2010) investiga o uso de modelos ARIMA para conjuntos de dados em que poucos parâmetros estão disponíveis, aplicando séries temporais e dados históricos, mas utiliza um conjunto de dados pequeno.

(DONG; ZOU; ZHANG, 2013) propõe o uso de modelos de predição distintos para curtas e longas distância, utilizando dados em tempo-real do tráfego para curta distância e dados históricos e *K-Nearest Neighbors* para longa distância. Entretanto, é fortemente dependente de dados em tempo real.

(LI *et al.*, 2017) busca lidar com a variação dos atrasos propondo um modelo em 3 estágios, sendo o primeiro o treinamento de padrões com dados históricos utilizando *K-Nearest Neighbor* e *K-means*, o segundo um filtro Kalman para predição e terceiro, uma cadeia de Markov para refinar o resultado.

(SERIN; ALISAN; ERTURKLER, 2022) também propõe um arquitetura em três estágios, em que a predição do tempo de viagem (primeiro estágio) é calculado separadamente da predição do resíduo (segundo estágio), sendo então combinados (terceiro estágio) para produzir o resultado final. Dos algoritmos testados, *Linear regression* para o primeiro, *Decision tree regression* para o segundo e *Exponential smoothing deep residuals* para o terceiro estágio apresentaram o melhor MAPE.

Apesar dos resultados, ambas as arquiteturas em três camadas propostas são custosas em termos de recursos computacionais.

Considerando a cidade de Curitiba, (HASHIGUCHI *et al.*, 2020) investiga a cidade especificamente, contribuindo ao indicar diferenças no padrão do horário de pico, como considerado pelo operador do STPU, mas não oferece informações de tempo de espera em paradas. (CURZEL *et al.*, 2019) propõe uma metodologia baseada em *link streams* para modelar e analisar temporalmente a operação de STPU com dados reais de Curitiba. A análise considera um terminal de ônibus e suas conexões entre linhas mas não as possíveis causas de acúmulo de atrasos na operação das linhas.

O Quadro 2 apresenta um resumo não exaustivo dos trabalhos revisados. (RODRIGUEZ-DENIZ; VILLANI, 2022) indica o código-fonte como público, entretanto, em

Quadro 2 – Visão geral da literatura

Técnica usada	Código Público	Usa dados em Tempo-real	Tipo de Via	Autor
ARIMA	N	N	Rodovia	(SUWARDO <i>et al.</i> , 2010)
KNN	N	Y	Urbana	(DONG <i>et al.</i> , 2013)
KNN/KF/KKM	N	Y	Urbana	(LI <i>et al.</i> , 2017)
LR/DTR/ESDR	N	Y	Rodovia	(SERIN <i>et al.</i> , 2022)
t-Full	N*	Y	Urbana	(RODRIGUEZ-DENIZ <i>et al.</i> , 2022)
KF/LSTM	N	N	Urbana	(LIU <i>et al.</i> , 2023)
IWOA-LSTM	N	N	Urbana	(ZHANG <i>et al.</i> , 2024)
GA-BP	N	N	Urbana	(SUN <i>et al.</i> , 2025)
Modelos Mistos	Y	N	Urbana	Este trabalho

Fonte: Autoria Própria (2025).

Junho de 2025 o mesmo não está mais disponível. Nota-se que o trabalho aqui proposto é o único com código-fonte público.

Portanto, o presente trabalho busca superar limitações dos artigos estudados ao tratar de forma estruturada e extensível os dados do STPU através de (PEIXOTO *et al.*, 2020), identificar itinerários com (BORGES *et al.*, 2023) e implementar um modelo misto baseado em dados offline históricos, aplicando-o a um estudo de caso na cidade de Curitiba.

3.1 Trabalhos Publicados

O presente trabalho resultou na publicação do artigo (PASQUIM; FONSECA; JR., 2025):

- PASQUIM, B.; FONSECA, K.; JR., L. M. Quando chega? Análise de previsibilidade de tempos de espera em transporte público urbano utilizando modelos mistos. In: Anais do IX Workshop de Computação Urbana. Porto Alegre, RS, Brasil: SBC, 2025. p. 15–28.

4 METODOLOGIA

A análise do comportamento temporal de ônibus de transporte público utilizando dados abertos e modelos mistos se baseou em avaliar a aderência dos modelos propostos aos dados observados de operação de um STPU. Pela disponibilidade de dados e ferramentas de processamento específicas para o conjunto de dados existente, o sistema integrado de transporte público de Curitiba foi escolhido.

O método consistiu em:

1. Estudar um conjunto limitado de dados de transporte público para um período de tempo específico e linha de ônibus arbitrária;
2. Identificar itinerários utilizando o algoritmo proposto em (BORGES *et al.*, 2023);
3. Analisar e relacionar os dados observados de itinerário com as tabelas horárias de referência.
4. Calcular, através de modelos mistos, a média geral de tempo de espera para todos os pontos de ônibus e ponderar o tempo de espera de cada ponto em relação à média para uma linha e itinerário de ônibus arbitrário;
5. Validar os resultados do modelo de previsão em relação aos dados observados, em uma data arbitrária, avaliando sua aderência.

O item 1 objetivou coletar, limpar e normalizar um conjunto reduzido de dados de transporte público de Curitiba, permitindo assim analisar sua estrutura. O item 2 expandiu o item 1, resolvendo problemas relacionados a lacunas e inconsistências nos dados através de uma estratégia robusta para a identificação de itinerários.

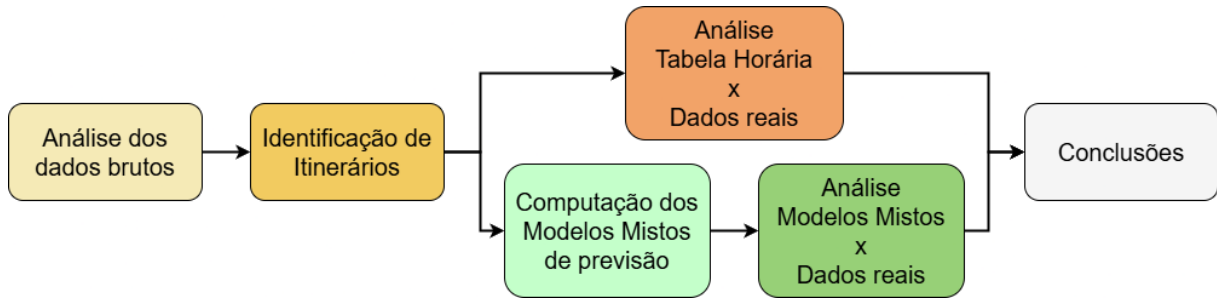
O item 3 analisou os dados obtidos em 2, comparando-os aos dados de referência fornecidos pelo operador do STPU. O item 4 buscou analisar em detalhes o comportamento do itinerário através de modelos mistos identificando quais pontos de ônibus possuem tempos de espera reduzidos ou elevados e o quanto estes se distanciam da média geral de espera.

Por fim, o item 5 objetivou aplicar o modelo de previsão na mesma linha modelada, mas em uma data arbitrária distinta, avaliando se os resultados produzidos para os tempos de espera em pontos de parada de ônibus refletem o comportamento da linha de ônibus. A Figura 1 resume o método proposto.

4.1 Formalização

Entende-se aqui por Itinerário um conjunto S de termos ordenados s_i onde cada s refere-se a geolocalização de um ponto de ônibus (localização pré-especificada para embarque e desembarque de passageiros no ônibus) e i o número de sequência do ponto no percurso, formando assim uma sequência ordenada de pontos de ônibus (definida na Equação 7):

Figura 1 – Etapas da metodologia proposta



Fonte: Autoria Própria (2025).

$$S = (s_1, s_2, s_3, \dots, s_n) \quad (7)$$

Na Equação 7 $i, n \in \mathbb{N}_{>0}$ sendo que n representa o último ponto de ônibus do percurso. A localização geográfica dos pontos de ônibus, bem como a ordem de parada em cada ponto na rota definida, são determinados pelo operador do STPU.

Ainda, dado o Itinerário S e os pontos de ônibus s_i e s_j , se define T como o Tempo de Espera do Itinerário, expresso na Equação 8, como uma sequência de tempos de espera $t_{i \rightarrow j}$ entre cada par de pontos de ônibus, onde $t \in \mathbb{R}_{>0}$ e $i, j \in \mathbb{N}^*$, sendo $i < j, j \leq n$.

$$T = (t_{1 \rightarrow 2}, t_{2 \rightarrow 3}, t_{3 \rightarrow 4}, \dots, t_{(n-1) \rightarrow n}) \quad (8)$$

Interpreta-se esse tempo como aquele que o usuário do sistema precisa esperar em um ponto de parada em relação ao ponto anterior do itinerário. Por exemplo, estando o usuário no ponto de sequência 2, $t_{1 \rightarrow 2}$ representa o tempo para o ônibus trafegar de 1 para 2. Nota-se que o número de intervalos de tempo de espera é igual ao o número de pontos de parada do itinerário menos um.

A partir da Equação 8 calcula-se *Tempo*, o tempo total necessário para um ônibus completar um Itinerário. Esse valor é expresso na Equação 9:

$$Tempo = \sum_{\substack{i=1 \\ j=i+1}}^{n-1} t_{i \rightarrow j} \quad (9)$$

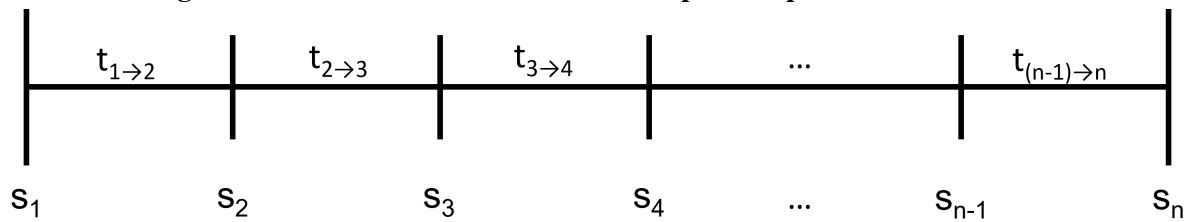
A Figura 2 ilustra os conceitos de *Itinerário* e *Tempo de Espera do Itinerário*:

4.2 Dados do sistema de Transporte Público

Como exposto na Seção 2.1, a cidade de Curitiba publica dados brutos de transporte público como dados abertos. Tais dados são reunidos e armazenados pelo C3SL¹. Os dados disponibilizados são:

¹ <https://www.c3sl.ufpr.br/> (acesso em 22/02/2025)

Figura 2 – Conceitos de Itinerário S e Tempo de Espera de Itinerário T



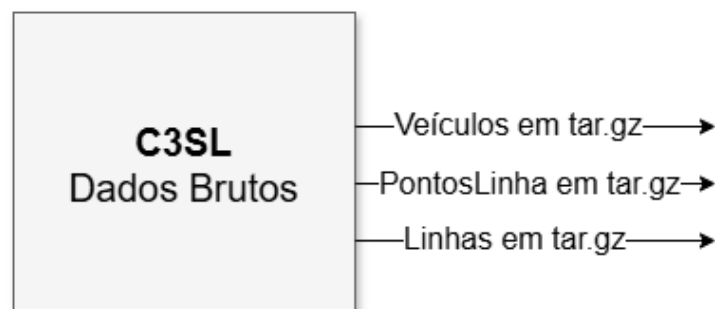
Fonte: Autoria Própria (2025).

- **Veículos** - As coordenadas geográficas dos veículos nas linhas de transporte;
- TrechosItinerarios - Os trechos dos itinerários das linhas de transporte;
- TabelaVeiculo - O número da tabela horária que o veículo estava executando;
- TabelaLinha - A tabela horária das linhas de transporte;
- ShapeLinha - As coordenadas por onde as linhas de transporte passam;
- **PontosLinha** - Os pontos de ônibus que pertencem as linhas de transporte;
- Pois - Pontos de referência da Cidade de Curitiba;
- **Linhas** - As linhas da Rede Integrada do Transporte Coletivo de Curitiba.

Os itens em negrito são aqueles necessários para se identificar itinerários de ônibus. “Veículos” contém dados relacionados a posição geográfica dos ônibus no itinerário, “PontosLinha” a posição geográfica e endereço dos pontos de parada e “Linhas” o nome e identificador numérico ID da linha. Estes dados, quando ainda não processados, são aqui denominados “dados brutos”.

Os dados brutos são disponibilizados em D+1, ou seja, um dia após serem coletados pelo STPU. Estes dados estão empacotados e comprimidos em formato *tar.gz*. Após descomprimidos, os dados ficam disponíveis em formato *json*. A Figura 3 mostra os dados brutos extraídos do repositório do C3SL.

Figura 3 – Dados extraídos do repositório do C3SL



Fonte: Autoria Própria (2025).

Nota-se que os dados de geolocalização dos ônibus também são disponibilizados em tempo-real, a cada 2 minutos, através de um *webservice*². Entretanto, se optou por utilizar os dados em D+1 pela sua granularidade, visto que estão mapeados a cada 20 segundos, em média.

O processo de baixar e preparar os dados brutos em um formato adequado para análise foi efetuado em diversos trabalhos anteriores, como em (HASHIGUCHI *et al.*, 2020) e (MANIKA, 2022), que utilizam o formato JSON. Já o formato parquet³, que dispõe os dados em colunas e é otimizado para compressão e recuperação, foi empregado em (BORGES *et al.*, 2023) e (PEIXOTO *et al.*, 2020).

Este trabalho utilizou o formato parquet para o processamento dos dados públicos de Curitiba, obtendo-os através da aplicação “*urbs-data-processing*”⁴, do mesmo autor de (PEIXOTO *et al.*, 2020).

A Figura 4 exemplifica o processo de conversão dos dados brutos para o formato parquet. Nota-se a necessidade de se fornecer uma data para baixar e converter os dados (seta cinza “Data”).

Figura 4 – Parâmetros de entrada e saída para a aplicação “*urbs-data-processing*” desenvolvida em (PEIXOTO *et al.*, 2020)



Fonte: Autoria Própria (2025).

Com os dados de transporte em formato parquet, se iniciou o processo de identificação de itinerários, descrito na Seção 4.3.

4.3 Identificação de itinerários

Identificar itinerários a partir dos dados brutos não é trivial: não há uma relação clara entre a posição geográfica (geolocalização) de um ônibus e do ponto de parada pelo qual ele vai passar, em um dado momento. Além disso, os dados podem apresentar lacunas, não havendo informações disponíveis em certo intervalo de tempo.

Esse trabalho utilizou a aplicação “*busanalysis*” desenvolvida em (BORGES *et al.*, 2023) para identificar itinerários. A aplicação executa três passos: (1) Associa a posição geográfica de pontos de parada com ônibus, identificando o horário (*Map Matching*); (2) Ordena

² https://dadosabertos.c3sl.ufpr.br/curitiba/TransporteColetivo/Documentaç~ao_WEB-SERVICE_\T1\textendash_TRANSPORTE_COLETIVO_DE_CURITIBA.pdf (acesso em 10/06/2024)

³ <https://parquet.apache.org/> (acessado em 05/06/2025)

⁴ <https://github.com/altierispeixoto/urbs-data-processing.git> (acesso em 20/04/2025)

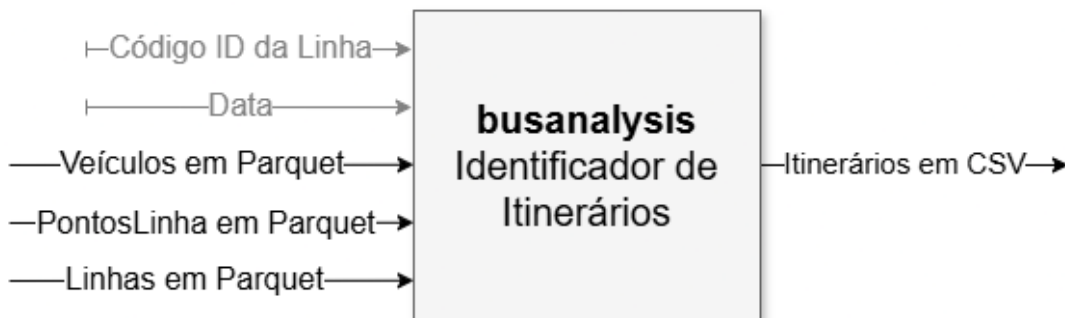
os registros a partir do horário (Sequenciamento Temporal); (3) Associa a sequência temporal com um itinerário. Como entrada, são necessários:

- Data a ser processada;
- Código ID da Linha de ônibus;
- Veículos;
- PontosLinha;
- Linhas.

Como saída, a aplicação produz um arquivo CSV (Figura 5) contendo:

- Código ID da Linha;
- Código ID do Veículo;
- Código ID do Ponto de Parada;
- Código ID do Itinerário ativo;
- Horário de passagem do ônibus pelo ponto;
- Número de sequência do ponto dentro do itinerário;

Figura 5 – Parâmetros de entrada e saída para a aplicação “busanalysis” desenvolvida em (BORGES *et al.*, 2023)



Fonte: Autoria Própria (2025).

Lacunas nos dados brutos podem ocorrer devido a problemas de comunicação dos dados de geolocalização. Por isso, durante o passo 3 a aplicação implementa interpolações e ajustes de temporização quando necessário, garantindo uma sequência de horários sem falhas. O código fonte pode ser consultado no repositório Git “busanalysis”⁵.

Por fim, o arquivo CSV de saída é utilizado como entrada para o script R que implementa o Modelo Misto proposto. Este modelo é detalhado na Seção 4.4.

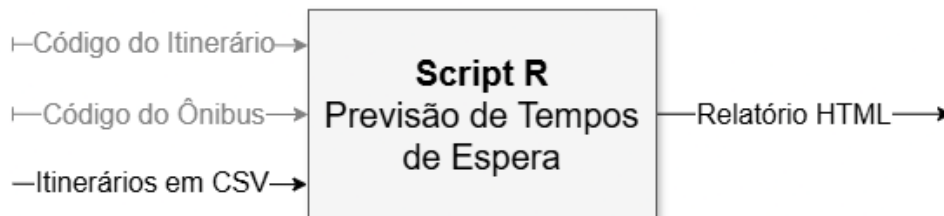
⁵ <https://github.com/jcnborges/busanalysis.git> (acesso em 20/04/2025)

4.4 Predição de Tempo de Espera com Modelos Mistos

O modelo misto proposto tem como objetivo gerar previsões para o tempo de espera de usuários do transporte público em pontos de parada. O Tempo de Espera é definido na Equação 8. Este modelo é implementado através de um script R⁶.

O script R implementado⁷ recebe como entrada o Código do Itinerário, o Código do Ônibus e o arquivo CSV gerado durante o processo descrito na Seção 4.3. Como saída, um relatório em formato HTML é gerado, contendo valores de tempo de espera médio e ponderado por ponto, além de representações gráficas dos tempos de espera total por ponto e previsões considerando intervalos de confiança. A Figura 6 exemplifica essa relação.

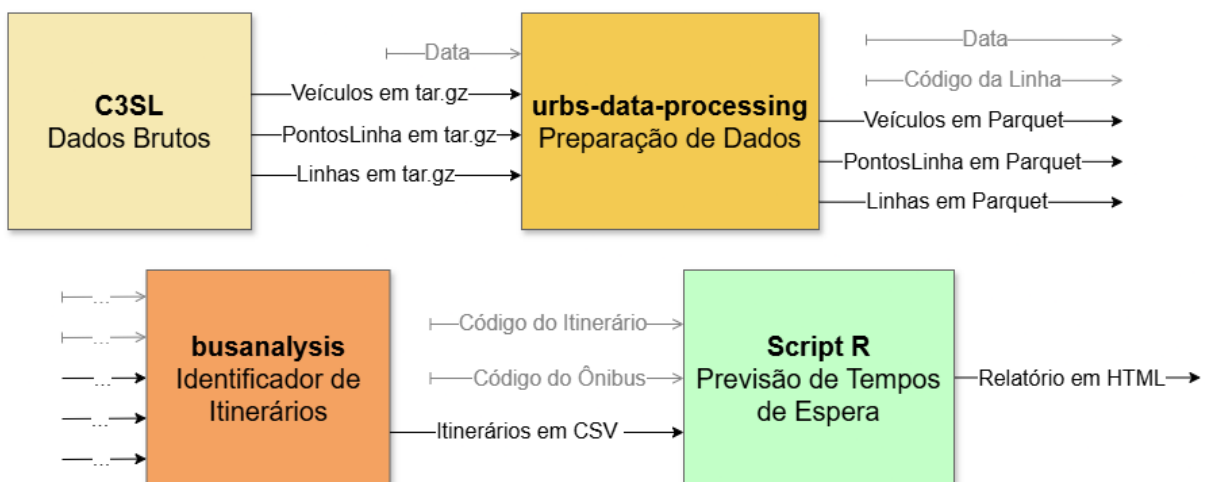
Figura 6 – Parâmetros de entrada e saída para o script R



Fonte: Autoria Própria (2025).

A Figura 7 resume o método usado desde a coleta dos dados brutos até o cálculo das previsões de tempo de espera. As setas cinzas iniciadas com uma linha vertical representam parâmetros arbitrários.

Figura 7 – Dados e ferramentas utilizadas



Fonte: Autoria Própria (2025).

⁶ <https://www.r-project.org/about.html>

⁷ <https://github.com/Brendonpasquim/mixed-model-bus-wait-time-forecast> (acesso em 20/04/2025)

4.5 Estudo de Caso

A linha Convencional 863-Água Verde foi escolhida para avaliação. Esta trafega entre os bairros Centro e Água Verde da cidade de Curitiba e foi escolhida por cruzar a região central (Figura 8).

Figura 8 – Mapa da linha 863 em Curitiba, itinerário 12494



Fonte: Cittamobi (2025).

Outras linhas (como 323-V. Autódromo) e datas (vide Quadro 3) foram testadas, sem mudança significativa nos resultados.

Quadro 3 – Outras datas analisadas

Data	Dia da Semana	Características
10/05/2023	Quarta-Feira	Referência; Sem chuva
16/06/2024	Domingo	Partida de Futebol, Arena da Baixada
08/07/2024	Segunda-Feira	Chuva, 33mm
15/08/2024	Quinta-Feira	Partida de Futebol, Arena da Baixada
04/09/2024	Quarta-Feira	Sem chuva
11/09/2024	Quarta-Feira	Sem chuva
18/09/2024	Quarta-Feira	Sem chuva
22/01/2025	Quarta-Feira	Pouca chuva (> 6mm)

Fonte: Autoria Própria (2025).

A linha 863 tem dois itinerários para dias úteis, 12494 (Sentido Praça Tiradentes/Centro) e 12477 (Sentido Água Verde). Na data de 10/05/2023 (Quarta-Feira) dois ônibus estavam em operação, HA310 e HC303. O veículo HA310 e itinerário 12494 foram escolhidos para análise.

O Quadro 4 apresenta as características da linha e itinerário utilizados no decorrer deste trabalho (com exceção a Seção 5.1, em que uma data diferente é analisada).

Quadro 4 – Linha analisada (Referência)

Métrica	Valor
Linha	863 - ÁGUA VERDE
Categoria	Convencional
Pontos de Parada	18
Intervalos de espera	17
Data de Análise	10/05/2023
Amostras (Filtrado)	343
Identificador do Itinerário	12494
Execuções do Itinerário	20
Início da Operação	5h20min
Término da Operação	23h30min

Fonte: Autoria Própria (2025).

A categoria “Convencional” significa que os passageiros fazem o embarque e desembarque a nível de pavimento, sem a utilização de estações-tubo ou plataformas elevadas em terminais de integração. O itinerário 12494 contém 18 pontos de parada, sendo o primeiro no bairro Água Verde e o último no bairro Centro, o que resulta em 17 intervalos de tempo em que o ônibus trafega de um ponto ao outro.

“Execuções do Itinerário” representa o número de vezes que o veículo completou o itinerário de 18 pontos de parada. Para o dia 10/05/2023 o itinerário foi completado 20 vezes. Destaca-se que o itinerário 12494 não é circular, ou seja, não inicia e termina no mesmo ponto de parada. Para uma análise completa do comportamento de ida e volta do ônibus durante o dia de operação é necessário analisar o itinerário 12477, que não faz parte do presente estudo de caso.

5 DISCUSSÃO E RESULTADOS

Neste capítulo são apresentados os resultados obtidos. A Seção 5.1 apresenta a análise da aderência da tabela horária aos dados observados, indicando o quanto são compatíveis. A Seção 5.2 discute os modelos de previsão propostos e seu desempenho. A Seção 5.3 avalia a estrutura de correlação dos resíduos dos modelos e seu ajuste aos dados. A Seção 5.4 verifica as propriedades dos resíduos, indicando se as hipóteses estruturais definidas são respeitadas. Finalmente, a Seção 5.5 avalia o desempenho do melhor modelo proposto em relação aos dados observados.

5.1 Análise de aderência da tabela horária aos dados

5.1.1 Metodologia

Para avaliar a aderência da tabela horária aos dados observados se aplicou o seguinte método:

1. Tomou-se uma data arbitrária qualquer;
2. Identificou-se o itinerário para a linha 863 nesta data, utilizando a aplicação “busanalysis”;
3. Identificou-se a tabela horária utilizada pelo ônibus na mesma data arbitrária (“tabelaLinha” e “tabelaVeiculo”, discutidos na Seção 4.2);
4. Comparou-se os horários da tabela horária com os horários obtidos no item 1, avaliando se estes estão dentro do intervalo de tolerância definido.

Para esta análise se escolheu o dia 20/11/2024 e os ônibus *HC302* e *HC306*. Salienta-se que esta data é diferente da data de referência utilizada em outras seções, bem como que a tabela horária de referência para a linha 863 apresenta horários para apenas dois pontos de parada do itinerário (o primeiro e o último) e portanto, a análise se limita a estes apenas.

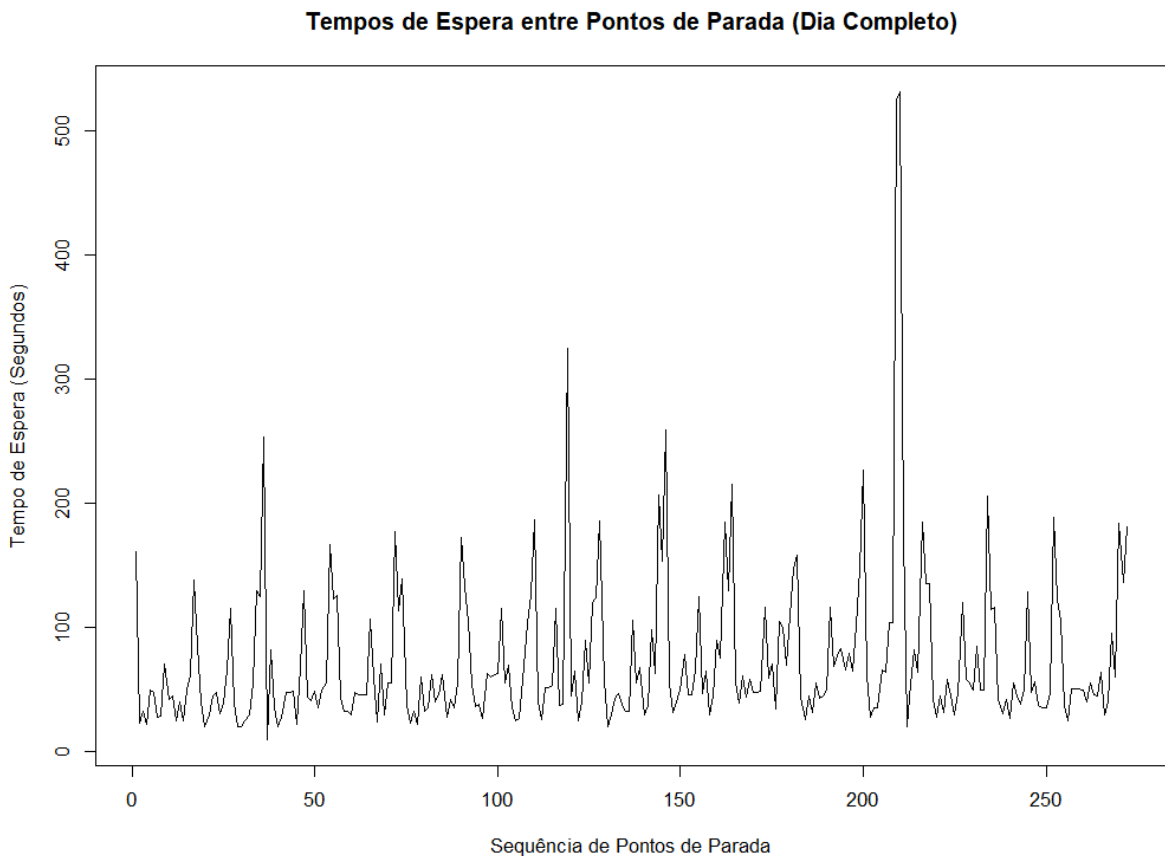
5.1.2 Análise dos Dados

Logo de início, notam-se problemas ao se correlacionar os dados de “tabelaVeiculo” e da execução do itinerário. Os códigos de ônibus indicados na “tabelaVeiculo” (*HI050* e *HC303*) não são os mesmos indicados na execução do itinerário (*HC302* e *HC306*). Além disso, há divergência em relação ao ponto inicial do itinerário, sendo que, “tabelaLinha” indica o ponto 150938 (R. Eng. Niepce da Silva, 488) como ponto inicial, enquanto que a execução do itinerário indica 150888 (R. Amazonas, 1293). Estes pontos estão distantes 200m entre si.

Devido a incerteza causada pela falta de correlação entre os dados observados e de referência, se decidiu descartar a análise para ao ponto inicial, mantendo apenas a análise para o ponto final (ID 101017, Praça Tiradentes).

A Figura 9 apresenta um exemplo de gráfico de tempos de espera para os dados observados da data sob análise (20/11/2024) para o veículo *HC306*. Este gráfico mostra um dia completo de operação, ou seja, todas as N execuções do itinerário 12494 pelo veículo *HC306*.

Figura 9 – Dados observados das 20 execuções do itinerário 12494 pelo ônibus *HC306*



Fonte: Autoria Própria (2025).

5.1.3 Resultados

Ao se associar os dados observados de tempo de espera dos ônibus *HC302* e *HC306* com a tabela horária de referência se obtêm a Tabela 1. Nota-se a existência de lacunas na coluna “Horário Itinerário”, demonstrando que não há dados observados indicando que um ônibus passou pelo ponto de parada no horário indicado pela tabela horária.

Além disso, percebe-se que há variação para mais e para menos no tempo, indicando que o ônibus está adiantado ou atrasado em relação a tabela horária de referência. Tal variação pode ser desprezível, como na linha 34 da tabela, ou superior a 15 minutos, como na linha 16.

Considerando a discussão ocorrida na Seção 4.2, se definiu dois minutos como a margem de tolerância da diferença entre o tempo observado e o tempo da tabela horária de referência. A escolha se deu porque esse é o tempo mínimo o qual um usuário de transporte público precisa esperar antes de requisitar uma nova posição GPS do ônibus ao *webservice* do STPU.

Fixando-se uma tolerância de dois minutos para mais ou para menos para classificar um ônibus como "No Horário", se obtém 2 registros no horário, 9 adiantados, 10 atrasados e 15 lacunas. É seguro afirmar, para o dia analisado, que existe uma tendência na qual não se atende a tabela de referência, entretanto, não existe um tendência indicando que os ônibus se atrasem ou adiantem com mais frequência.

Tabela 1 – Comparação entre horários da tabela horária de referência e da execução do Itinerário.

#	Ônibus	Horário Tabela	Horário Itinerário	Diferença Absoluta	Status
1	HC306	05:45	05:49:22	00:04:22	Atrasado
2	HC306	06:13	-	-	-
3	HC306	06:41	06:36:40	00:04:20	Adiantado
4	HC306	07:10	07:24:58	00:14:58	Atrasado
5	HC306	07:43	-	-	-
6	HC306	08:16	08:11:46	00:04:14	Adiantado
7	HC306	08:49	08:57:52	00:08:52	Atrasado
8	HC306	09:22	-	-	-
9	HC306	09:54	09:48:44	00:05:16	Adiantado
10	HC306	10:27	10:41:12	00:14:12	Atrasado
11	HC306	11:00	-	-	-
12	HC306	11:33	11:35:29	00:02:29	Atrasado
13	HC306	12:06	-	-	-
14	HC306	12:39	12:32:42	00:06:18	Adiantado
15	HC306	13:12	-	-	-
16	HC306	13:45	13:29:13	00:15:47	Adiantado
17	HC306	14:18	14:24:13	00:06:13	Atrasado
18	HC306	14:51	-	-	-
19	HC306	15:24	15:37:27	00:13:27	Atrasado
20	HC306	15:57	-	-	-
21	HC306	16:30	16:31:03	00:01:03	No Horário
22	HC306	17:05	17:17:23	00:12:23	Atrasado
23	HC306	17:40	-	-	-
24	HC306	18:15	18:06:43	00:08:17	Adiantado
25	HC302	18:50	-	-	-
26	HC302	19:25	19:31:03	00:06:03	Atrasado
27	HC302	19:53	-	-	-
28	HC302	20:21	-	-	-
29	HC302	20:49	20:43:04	00:05:56	Adiantado
30	HC302	21:17	-	-	-
31	HC302	21:45	21:34:32	00:10:28	Adiantado
32	HC302	22:12	22:22:54	00:10:54	Atrasado
33	HC302	22:39	-	-	-
34	HC302	23:06	23:05:43	00:00:17	No Horário
35	HC302	23:33	-	-	-
36	HC302	00:00	23:54:34	00:05:26	Adiantado

Fonte: Autoria Própria (2025).

5.2 Modelos de previsão

Três modelos de previsão foram desenvolvidos de forma incremental, no qual novos componentes foram adicionados para melhor lidar com os dados observados. Os dados observados do dia 10/05/2023 foram utilizados no treinamento dos modelos (vide a Seção 4.5). O Quadro 5 apresenta um resumo.

Quadro 5 – Modelos de previsão desenvolvidos

Nome	Descrição
M1	Efeito Fixo
M2	Efeito Fixo + Aleatório
M3	Efeito Fixo + Aleatório + Autocorrelação nos Resíduos
M3 Sazonal	M3 + Componente Sazonal no Resíduo (Lag 16)

Fonte: Autoria Própria (2025).

Ao se comparar os três modelos para estimar o tempo de espera do segmento do Itinerário ($t_{i \rightarrow j}$), denomina-se $t_{ij} = t_{i'}$, sendo i' a representação de um efeito aleatório, ou seja, representa o deslocamento de um ponto i para um ponto j no itinerário.

- Modelo 1 (M1): $t_{i'k} = \beta_0 + \epsilon_{t_{i'k}}$, onde $\epsilon_{t_{i'k}} \sim N(0, \sigma^2)$, $i' = 1, \dots, n$, $k = 1, \dots, K_{i'}$, sendo n o número de deslocamentos e $K_{i'}$ número de replicações do deslocamento i' . Assume-se um mesmo efeito fixo (média geral) para todos os tempos de espera.
- Modelo 2 (M2): $t_{i'k} = \beta_0 + b_{0i'} + \epsilon_{t_{i'k}}$, $\epsilon_{t_{i'k}} \sim N(0, \sigma^2)$, $i' = 1, \dots, n$, $k = 1, \dots, K_{i'}$, sendo n o número de deslocamentos e $K_{i'}$ o número de replicações do deslocamento i' . Assumimos um mesmo efeito fixo (média geral) e adicionamos mais um efeito aleatório representado por um intercepto por cada tempo de espera relacionado com o respectivo deslocamento.
- Modelo 3 (M3): $t_{i'k} = \beta_0 + b_{0i'} + \epsilon_{t_{i'k}}$, $\epsilon_{t_{i'k}} = \theta_1 \epsilon_{t_{i'(k-1)}} + \dots + \theta_r \epsilon_{t_{i'(k-r)}} + \Theta_1 \epsilon_{t_{i'(k-s)}} + \dots + \Theta_p \epsilon_{t_{i'(k-ps)}} + \nu_{t_{i'k}}, \nu_{t_{i'k}} \sim N(0, \sigma_e^2)$, $i' = 1, \dots, n$, $k = 1, \dots, K_{i'}$, sendo n o número de deslocamentos e $K_{i'}$ o número de replicações do deslocamento i' . Assumimos um mesmo efeito fixo (média geral), adicionamos um efeito aleatório representado por um intercepto para cada tempo de espera relacionado com o respectivo deslocamento e incorporamos uma modelo autorregressivo nos resíduos (estimativa do erro).

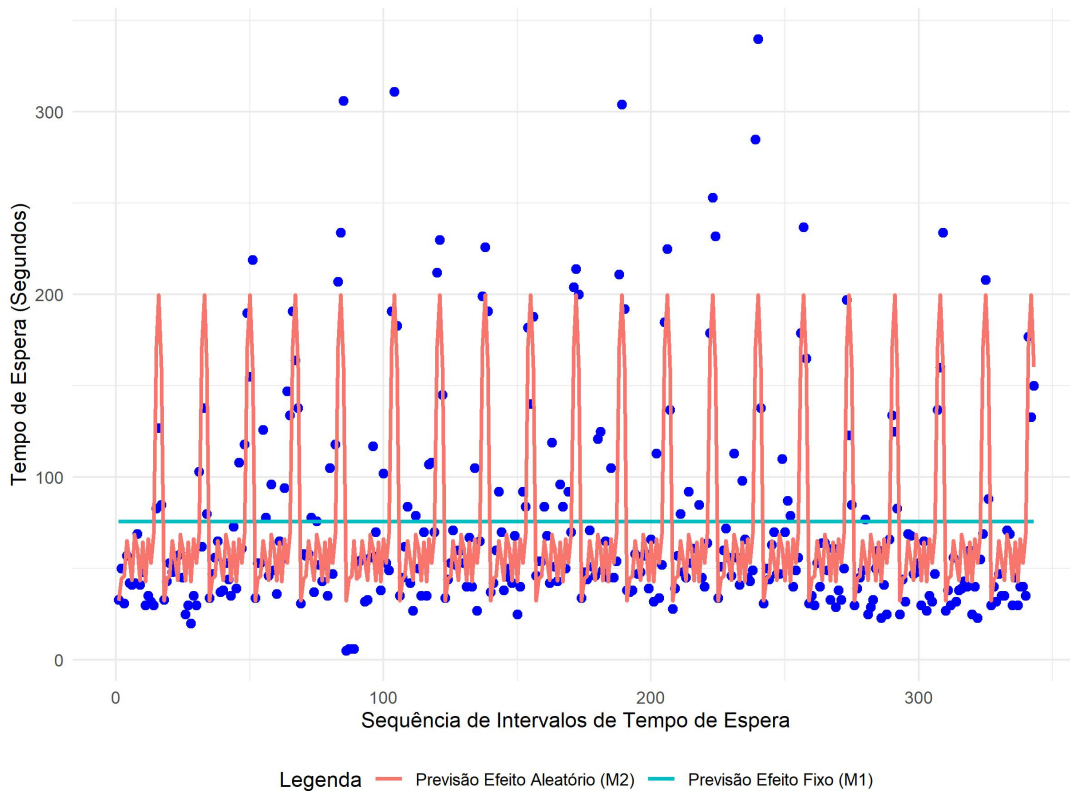
No componente autorregressivo temos o vetor de parâmetros $(\theta_1, \dots, \theta_r)$ contendo valores entre 0 e 1, que controlam a dependência do resíduo estimado no tempo t em relação aos resíduos estimados no tempo $t - 1, \dots, t - r$.

O vetor de parâmetros $(\Theta_1, \dots, \Theta_p)$ avalia se ocorre uma dependência do resíduo no tempo t de maneira sazonal ou cíclica, defasados $t - s, t - ps$, unidades no tempo. Para efeito de notação, caso $(\Theta_1, \dots, \Theta_p)$ seja não significativo o modelo M3 desconsidera-se a sazonalidade, caso contrário, temos um M3 Sazonal.

A Figura 10 apresenta um exemplo de previsão para tempos de espera considerando um dia completo de operação, ou seja, as N vezes que um ônibus executou um itinerário S . Os pontos em azul representam o tempo de espera observado (vide Figura 2).

Observa-se que utilizando apenas M1 assume-se um único valor de espera para todos os pontos de parada, o que não representa a realidade dos dados. Em M2, o intercepto individualizado por cada tempo de espera adicionado ao efeito fixo se ajusta melhor aos dados.

Figura 10 – Comparação de previsões dos modelos M1 e M2 considerando todas as 20 execuções do itinerário 12494 de um dia de operação do veículo HA310



Fonte: Autoria Própria (2025).

O Quadro 6 destaca alguns números dos modelos M1 e M2. Nota-se que o maior tempo de espera ocorre entre os pontos P_{16} e P_{17} .

Quadro 6 – Destaques dos modelos M1 e M2

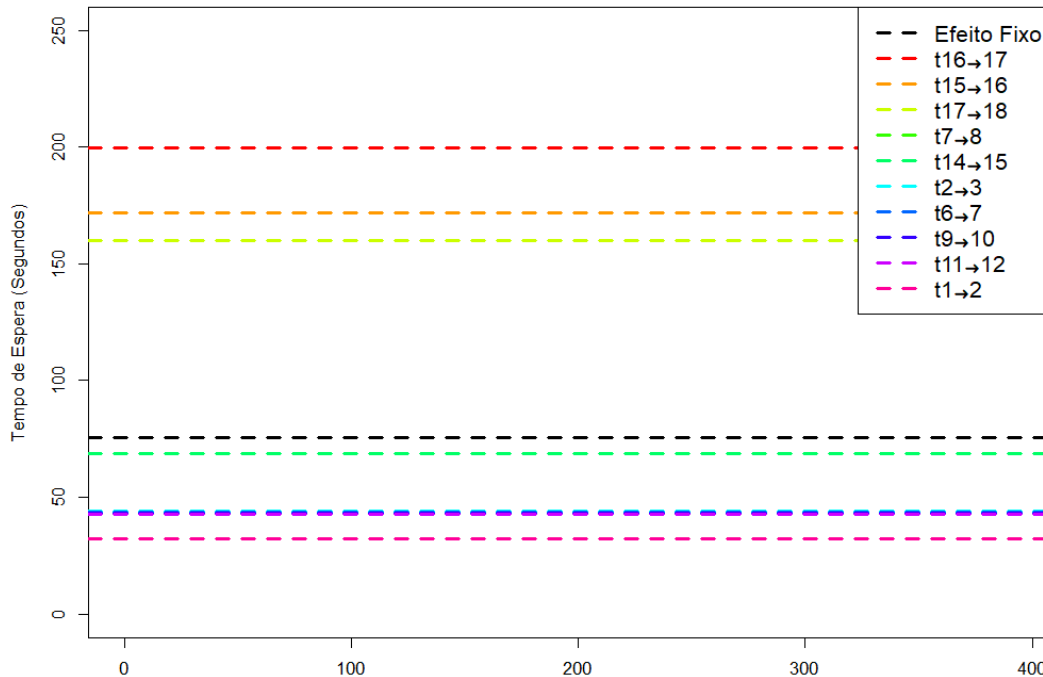
Métrica	Valor
Maior tempo de espera	200 segundos
Menor tempo de espera	32 segundos
Média tempo de espera	76 segundos
Intervalo de maior espera	Entre P_{16} e P_{17} ($t_{16 \rightarrow 17}$)
Intervalo de menor espera	Entre P_1 e P_2 ($t_{1 \rightarrow 2}$)

Fonte: Autoria Própria (2025).

Já a Figura 11 indica os cinco maiores e cinco menores tempos de espera (soma do efeito fixo e aleatório). O tempo médio (efeito fixo) é plotado em preto.

Destaca-se que os pontos de maior tempo de espera são os últimos pontos do itinerário, próximos ao centro da cidade. Os dados para as outras datas analisadas (Tabela 3) apresentaram o mesmo comportamento, indicando uma tendência.

Figura 11 – 5 maiores e 5 menores tempos de espera em pontos de parada. Efeito Fixo é plotado em preto



Fonte: Autoria Própria (2025).

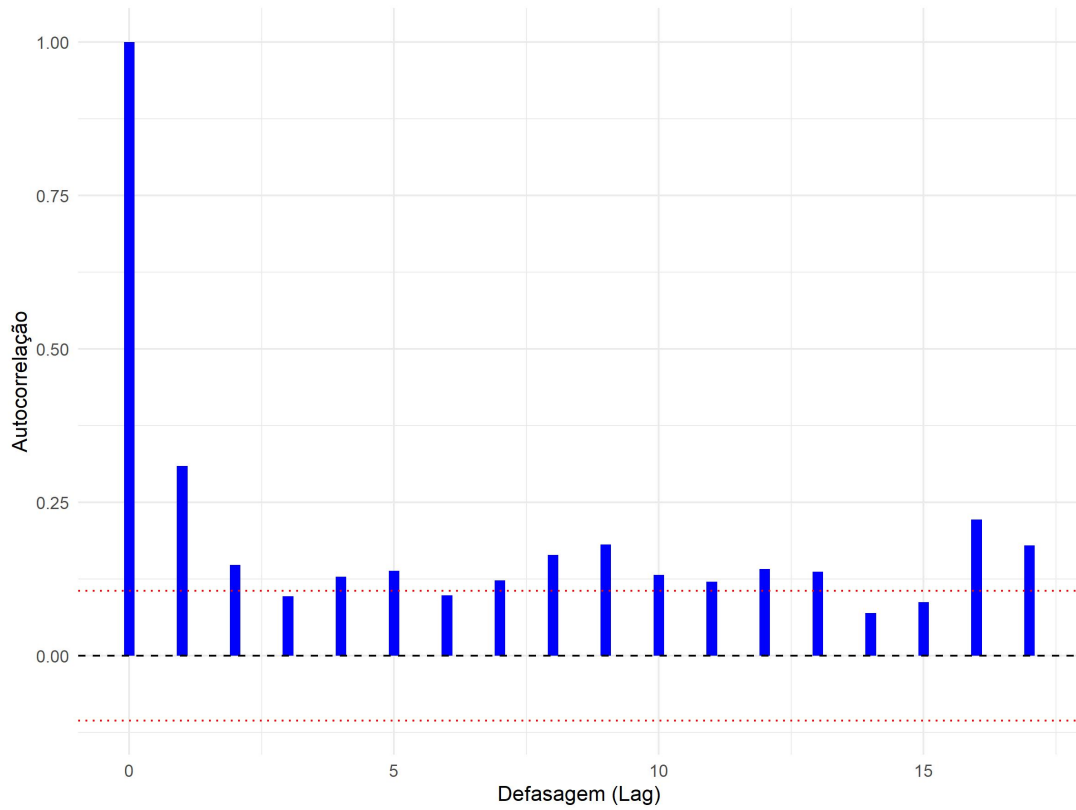
5.3 Análise de correlação dos resíduos

Após a computação dos modelos é importante a validação do ajuste através dos resíduos, verificando se os intervalos de confiança definidos são respeitados. Para tal se utilizou um correlograma nos resíduos. Caso o modelo esteja bem ajustado o correlograma não terá correlação significativa, ou seja, fora dos intervalos de confiança (correlograma de ruído branco) (FARAWAY, 2006).

Na Figura 12 aplicou-se um correlograma no modelo M2. Cada defasagem (lag) representa um intervalo de espera entre pontos de parada, em que a defasagem zero representa a autocorrelação do intervalo consigo mesmo, sendo portanto, sempre um. Se observa nos resíduos que a maior correlação está na defasagem um, indicando que os tempos de espera são influenciados pelos tempo de espera imediatamente anterior. Também é possível notar outras correlações significativas acima do intervalo de confiança.

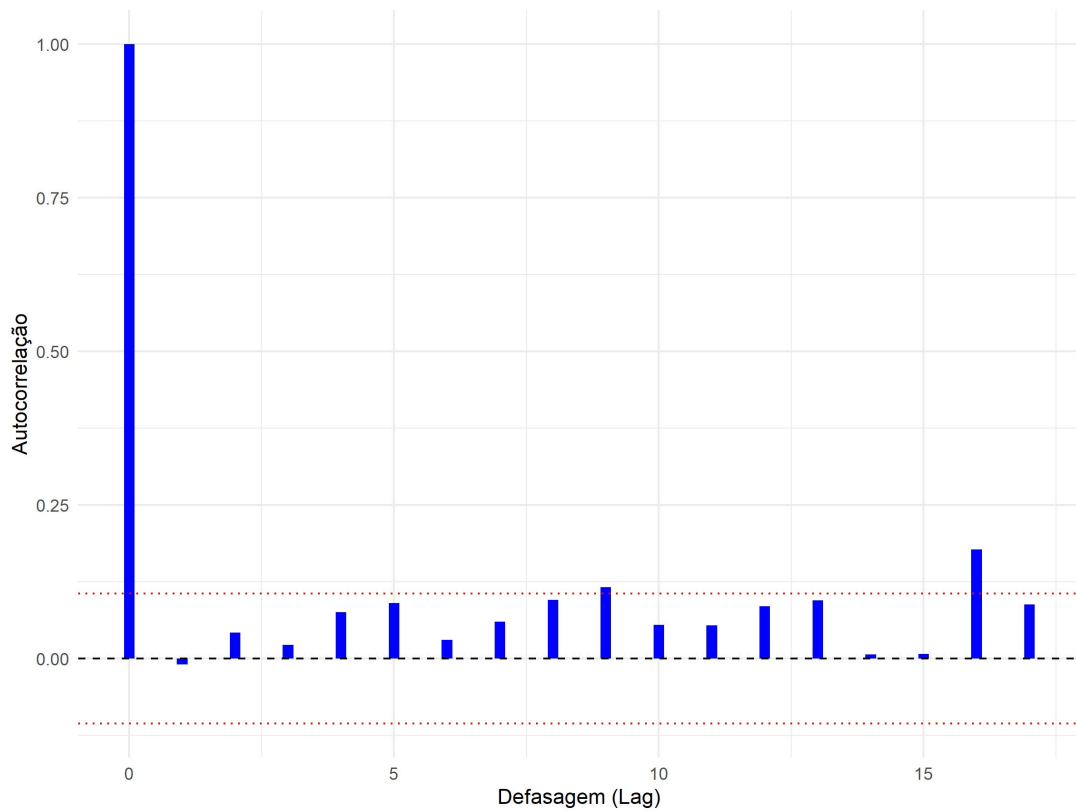
Na Figura 13 aplicou-se um correlograma no modelo M3. Se nota que boa parte dos valores de correlação se acomodaram dentro do intervalo de confiança. Entretanto, uma correlação significativa e aparentemente sazonal fora do intervalo de confiança persistiu (*Lag 16*). Para modelar essa correlação, se aplicou um modelo autorregressivo também ao componente sazonal.

Figura 12 – Gráfico de Autocorrelação do modelo M2



Fonte: Autoria Própria (2025).

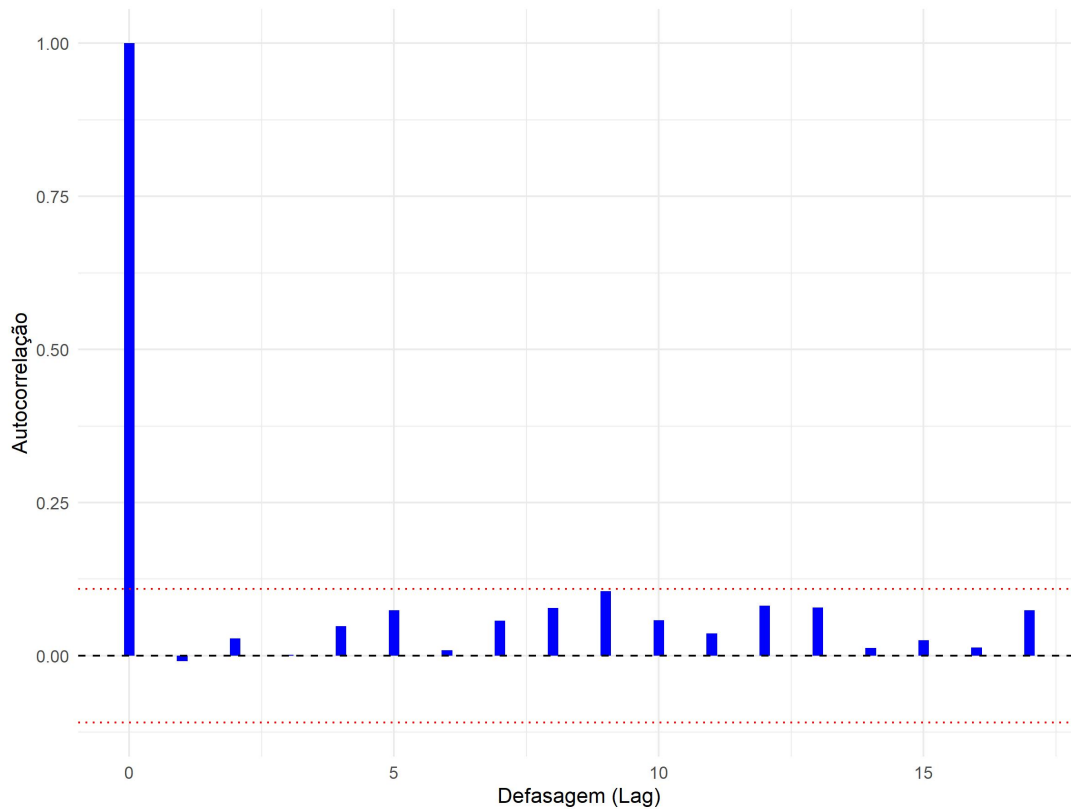
Figura 13 – Gráfico de Autocorrelação do modelo M3



Fonte: Autoria Própria (2025).

A Figura 14 mostra o resíduo do modelo M3 já acrescido do componente sazonal no *Lag 16*. Nota-se que o correlograma apresentou o comportamento esperado, dentro do intervalo de confiança.

Figura 14 – Gráfico de Autocorrelação do modelo M3 Sazonal



Fonte: Autoria Própria (2025).

Esse componente sazonal no intervalo de tempo entre os pontos 15 e 16 (portando, o *Lag 16*) demonstra uma característica particular em relação aos demais pontos do itinerário 12494. Ao analisá-lo, se percebe que este ponto é uma praça no centro de Curitiba que possui um papel essencial para as conexões com outras linhas e, mesmo não possuindo um terminal de integração espacial, permite integração temporal da tarifa de transporte em determinadas linhas de ônibus.

Esse papel especial justifica o tempo de espera mais elevado, bem como sua recorrência, em comparação com os demais pontos de parada dentro do itinerário, visto que, um maior número de embarques e desembarques de passageiros é esperado.

5.4 Análise de propriedades dos resíduos

5.4.1 Teste de Resíduos

A análise da série residual foi realizada para avaliar a estacionariedade e a presença de autocorrelação significativa. Primeiramente, aplicou-se o teste de **Ljung-Box** para verificar autocorrelações significativas nos resíduos. O teste apresentou um valor de $X^2 = 11.544$, com

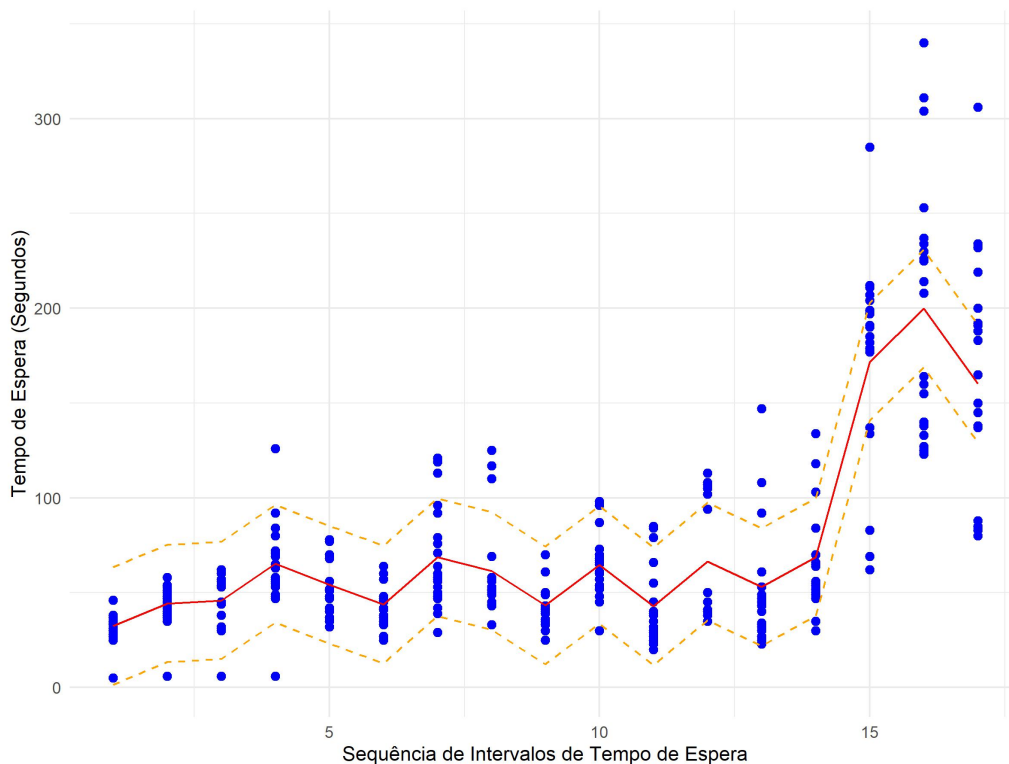
10 graus de liberdade ($df = 10$), resultando em um valor de $p = 0.3167$. Como o valor- p é maior que o nível de significância de 0,05, não há evidências suficientes para rejeitar a hipótese nula de ausência de autocorrelação significativa.

Em seguida, foi aplicado o **teste de Dickey-Fuller aumentado (ADF)** para verificar a estacionariedade dos resíduos. O teste apresentou um valor estatístico de -5.9879 , com ordem de defasagem $k = 6$, e um valor p de 0,01. Isso indica a rejeição da hipótese nula de não estacionariedade em favor da hipótese alternativa, sugerindo que a série residual é estacionária. Em resumo, os resíduos não apresentam autocorrelação significativa e são estacionários, de acordo com os testes aplicados.

5.4.2 Análise Gráfica

As Figuras 15 e 16 comparam os modelos M2 e M3, onde se plotou as 20 execuções do itinerário 12494 no mesmo eixo. O eixo das abscissas representa os intervalos de tempo de espera entre cada par de pontos de parada (vide Figura 2). Já o eixo das ordenadas representa o tempo de espera em segundos de cada intervalo.

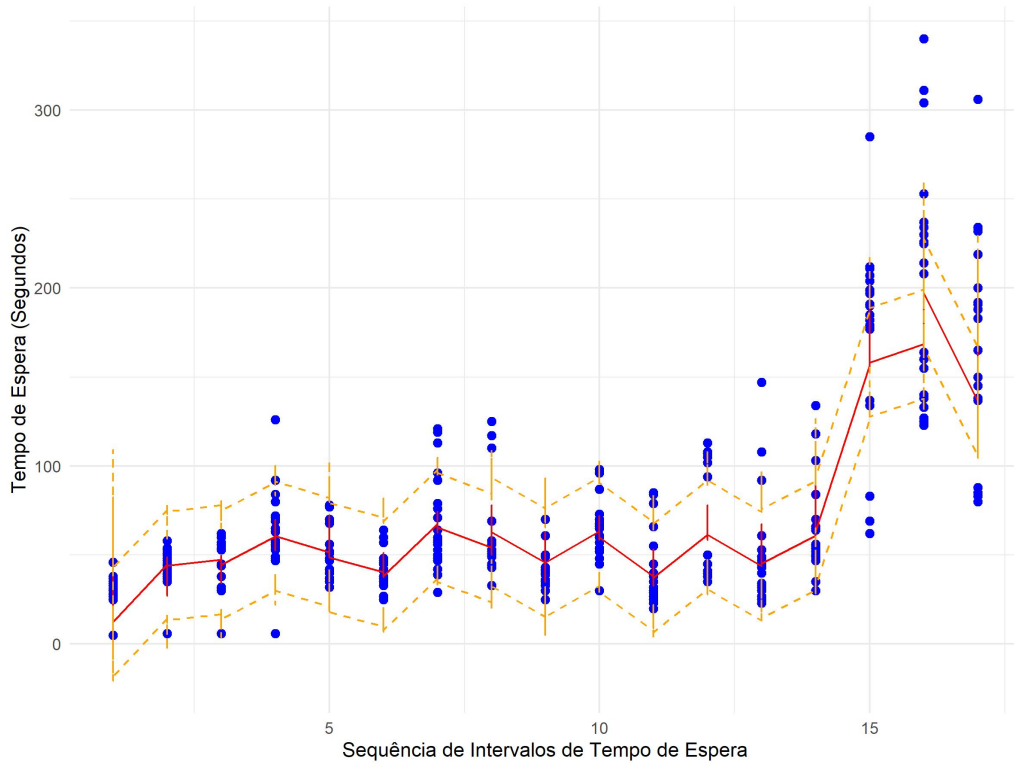
Figura 15 – Gráfico de Previsão do modelo M2 com Intervalo de Confiança considerando as 20 execuções do itinerário 12494



Fonte: Autoria Própria (2025).

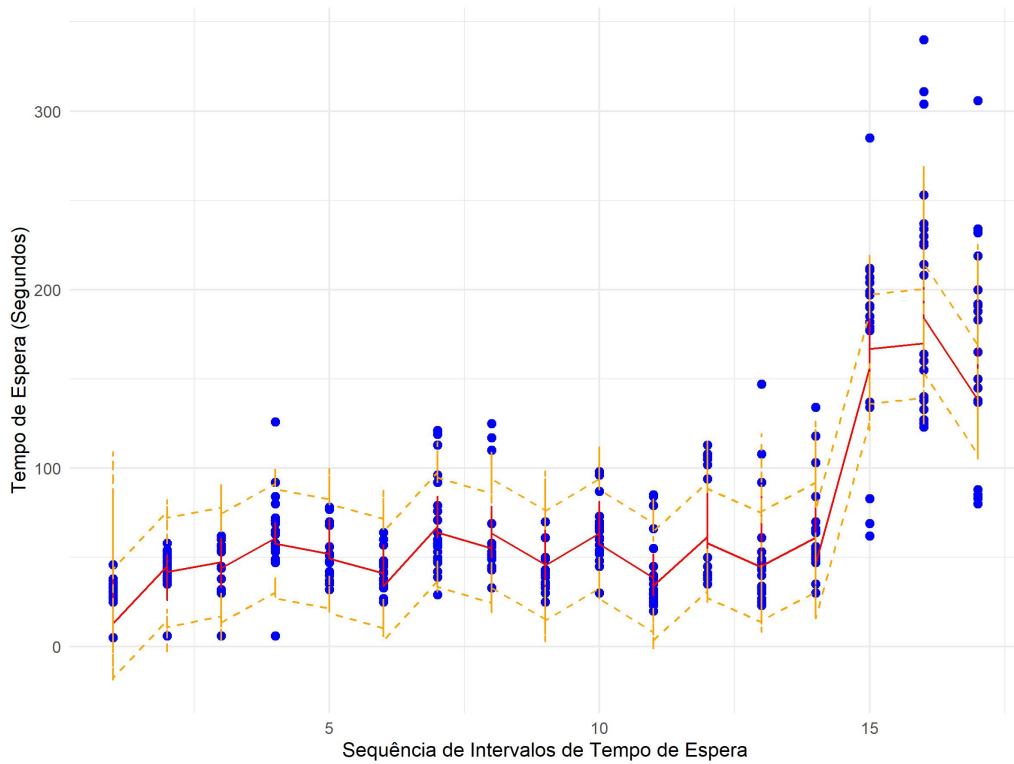
Nota-se que os dois últimos deslocamentos, $t_{16 \rightarrow 17}$ e $t_{17 \rightarrow 18}$, possuem uma maior variabilidade. O modelo M3 conseguiu ter mais valores no intervalo de confiança. O modelo M3 adicionado com a sazonalidade (Figura 17) não produziu resultados melhores que M3 pela avaliação gráfica.

Figura 16 – Gráfico de Previsão do modelo M3 com Intervalo de Confiança considerando as 20 execuções do itinerário 12494



Fonte: Autoria Própria (2025).

Figura 17 – Gráfico de Previsão do modelo M3 Sazonal com Intervalo de Confiança considerando as 20 execuções do itinerário 12494



Fonte: Autoria Própria (2025).

5.4.3 Erro Quadrático Médio

A Tabela 2 apresenta os valores da métrica EQM obtida para os três modelos implementados (inclui-se também a variante sazonal de M3). A partir dela se seleciona o modelo “M3 Sazonal” como aquele que apresentou melhor performance ($REQM = \sqrt{EQM} \approx 30$ segundos).

Tabela 2 – EQM Modelos

Modelo	EQM
<i>M1</i>	3419.7
<i>M2</i>	969.2
<i>M3 NãoSazonal</i>	881.3
<i>M3 Sazonal</i>	851.4

Os parâmetros para o modelo “M3 Sazonal” são expressos na Equação 10. Nela temos $\beta_0 = 75.65$ e $\epsilon_{t_{i'k}} = 0.2842\epsilon_{t_{i'k-1}} + 0.1828\epsilon_{t_{i'k-16}}$. $b_{0i'}$ é a parte do modelo que se modifica em função do ponto de parada no itinerário. Caso $t_{1 \rightarrow 2}$, temos $b_{0i'} = -43.24$ (menor tempo de espera), se $t_{16 \rightarrow 17}$, $b_{0i'} = 124.07$ (maior tempo de espera)

$$t_{i'k} = \beta_0 + b_{0i'} + \epsilon_{t_{i'k}} \quad (10)$$

5.5 Análise de aderência do modelo aos dados

5.5.1 Metodologia

Para avaliar a aderência do modelo “M3 Sazonal” aos dados observados se aplicou uma análise gráfica dos intervalos de confiança seguindo o método:

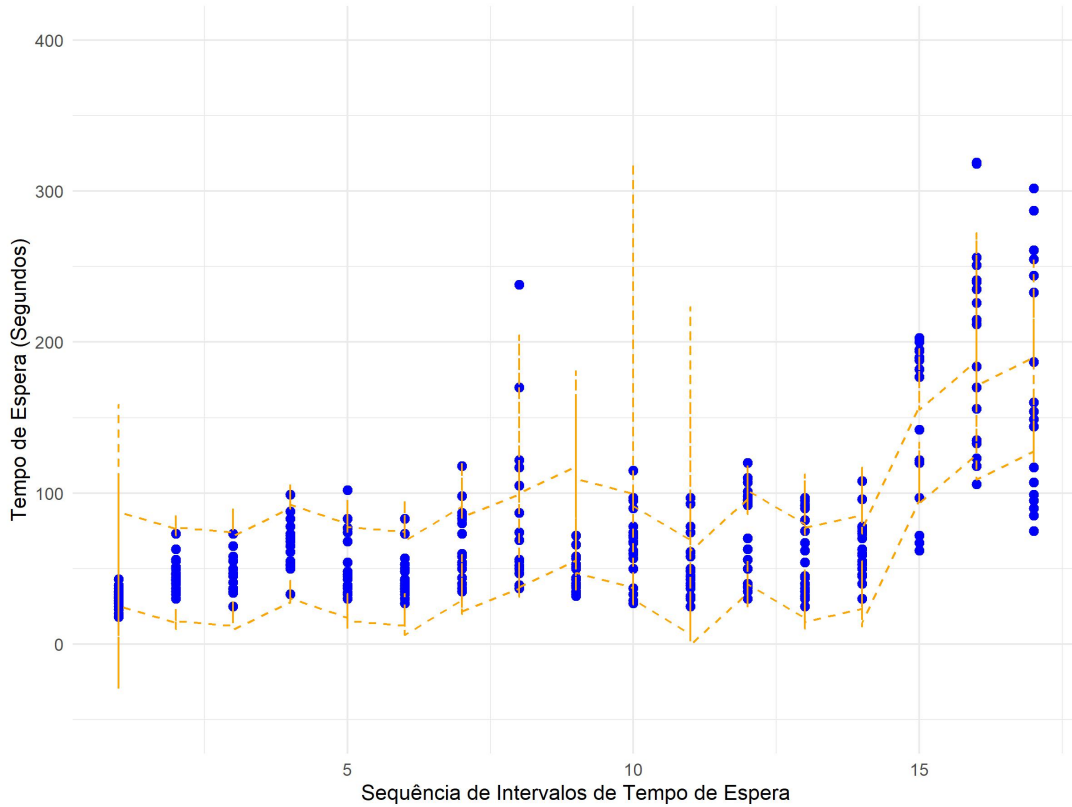
1. Treinou-se o modelo “M3 Sazonal” com os dados de uma data arbitrária. A data escolhida foi 11/09/2024 (Quarta-Feira, sem chuva);
2. Escolheu-se uma segunda data arbitrária que tivesse as mesmas características do item 1 (Quarta-Feira, sem chuva). A data escolhida foi 18/09/2024.
3. Plotou-se gráficos para ambas as datas. Através da análise gráfica, procurou-se identificar se os dados do item 2 se encontravam dentro do intervalo de confiança definido no item 1.

5.5.2 Resultados

O resultado do item 3 é exposto nas Figuras 18 e 19. Analisando-as se nota um tendência equivalente em ambos os gráficos, no qual a maioria dos pontos se encontra dentro do intervalo

de confiança. Os valores fora do intervalo de confiança são poucos em ambas as datas analisadas, não formando uma tendência por si só. Desta forma, é razoável assumi-los como valores atípicos advindos de uma flutuação pontual do comportamento do ônibus/motorista no dia em questão.

Figura 18 – Dados observados para o dia 11/09/2024 e intervalo de confiança para o dia 11/09/2024



Fonte: Autoria Própria (2025).

6 CONCLUSÃO

A partir de um estudo de caso na cidade de Curitiba, utilizando dados históricos de geolocalização de veículos de transporte coletivo percorrendo um itinerário estabelecido pela operação do sistema de transporte público, foram propostos três modelos para buscar representar o comportamento temporal. Além disso, analisou-se a aderência da tabela horária de referência com os dados reais dos itinerários calculados.

Os resultados obtidos indicam que não é possível afirmar que a tabela horária tem aderência aos dados reais na data analisada, em que o número de lacunas e adiantamentos de horários de chegada é superior a 65% das entradas de dados disponíveis. Apesar da análise considerar apenas um dia de operação, percebe-se que a metodologia proposta pode ser aplicada de maneira iterativa, generalizando os resultados. Ainda, percebe-se que a falta de informações oficiais (advindas do operador do STPU) detalhando os critérios técnicos utilizados para se construir a tabela de referência limita a análise, não sendo possível discutir as razões por trás da falta de aderência em relação aos dados reais.

Os modelos de previsão M1, M2 e M3 foram construídos de forma incremental, permitindo escolher o modelo ideal de acordo com a complexidade de implementação e requisitos de precisão de previsão. Nota-se ainda que os modelos não exigem conhecimento específico da linha sob análise, permitindo sua aplicação em qualquer linha o qual se tenha dados.

Adicionalmente, implementou-se a versão “Sazonal” para o modelo M3, em que se inclui o conhecimento específico da linha como um parâmetro adicional no modelo. Essa versão se destaca por duas razões: (1) Permite modelar comportamentos atípicos não explicados por autocorrelação e, (2) evidência como adicionar novas variáveis ao modelo, comprovando sua extensibilidade.

O modelo M3 Sazonal apresentou o melhor resultado, em termos de previsão de tempos de espera em pontos de ônibus ($REQM < 30s$), bem como na modelagem dos resíduos. Ainda, apresentou-se como uma solução robusta capaz de prever boa parte do comportamento da linha analisada, respeitando o intervalo de confiança estipulado.

Além disso, nota-se que o modelo desenvolvido tem potencial para beneficiar não apenas o usuário final do STPU, disponibilizando tempos de espera e possibilitando a previsão de chegada, mas pode ser utilizado pelo operador do STPU como uma ferramenta de controle de qualidade para identificar tendências e comportamentos atípicos, permitindo a implementação de alertas e controle proativo.

Portanto, o presente trabalho não apenas desenvolve modelos de previsão, permitindo a construção de uma tabela de referência automática baseado nos dados observados, mas habilita pesquisadores e agentes do poder público com ferramentas para explorar as nuances e fatores específicos que impactam o usuário final e as empresas operadoras do sistema.

7 TRABALHOS FUTUROS

Como trabalhos futuros e oportunidades de melhoria, se indica a possibilidade de incluir novas variáveis ao modelo, no intuito de melhor explicar as causas de atrasos. Sugere-se:

1. A adequação e uso de dados de Boletins de transporte, que informam interrupções programadas pelo operador do STPU na operação da linha do ônibus (devem ser disponibilizados em um formato legível para consumo por computadores);
2. Incluir variáveis de possível impacto na operação da linha do ônibus, como aquelas relacionadas a: (1) Pluviometria; (2) Locais de maior fluxo de passageiros; (3) Pontos de integração com outras linhas, ampliando o escopo proposto por (CURZEL *et al.*, 2019); (4) Itinerários de dias úteis, final de semana e feriados; (5) Horários de pico, cruzamentos viários e semáforos.

Adicionalmente, propõe-se expandir a análise de tempos de espera em pontos de parada, focando no ponto de vista do usuário final. Neste caso, considera-se o tempo de espera em relação a todos os ônibus executando o itinerário em dado dia, fornecendo uma previsão com intervalo de confiança. Este cenário complementa aquele discutido na seção 4.4, em que um ônibus é analisado isoladamente, privilegiando a análise do ponto de vista do operador do STPU.

Por fim, propõe-se aplicar os modelos mistos desenvolvidos na investigação do impacto específico das variáveis que afetam os tempos de espera (inferência causal) e possíveis formas de intervenção (quando aplicável) para minimizar seu impacto. Por exemplo, em dias de chuva intensa, desviar os ônibus de transporte público que passam por uma rua propensa a inundação para outra próxima, evitando que passageiros fiquem expostos a um ambiente insalubre. Ou ainda, ajustar a temporização de semáforos em cruzamentos para que ônibus tenham preferência em horários de pico.

REFERÊNCIAS

- AGUIRRE, L. A. **Introdução à Identificação de Sistemas - Técnicas Lineares e Não Lineares Aplicadas à Sistemas Reais**. Belo Horizonte, MG: Editora UFMG, 2007.
- BORGES, J. C. *et al.* Algoritmo para detecção de itinerários do transporte público usando dados de gps dos ônibus. *In: SBC. Anais do VII Workshop de Computação Urbana*. [S.l.], 2023. p. 1–14.
- BOX, G. E. *et al.* Time series analysis: forecasting and control. **John Wiley & Sons**, Wiley, 2015.
- BRASIL. Lei nº 12.527, de 18 de novembro de 2011. regula o acesso a informações previsto no inciso xxxiii do art. 5º, no inciso ii do § 3º do art. 37 e no § 2º do art. 216 da constituição federal; altera a lei nº 8.112, de 11 de dezembro de 1990; revoga a lei nº 11.111, de 5 de maio de 2005, e dispositivos da lei nº 8.159, de 8 de janeiro de 1991; e dá outras providências. **Diário Oficial da República Federativa do Brasil**, Brasília, DF, 2011. Disponível em: http://www.presidencia.gov.br/ccivil_03/_Ato2011-2014/2011/Lei/L12527.htm.
- BRAUNSCHWEIG, K. *et al.* The state of open data. **Limits of current open data platforms**, v. 1, p. 72–72, 2012.
- BüCHEL, B.; CORMAN, F. What do we know when? modeling predictability of transit operations. **IEEE Transactions on Intelligent Transportation Systems**, v. 23, n. 9, p. 15684–15695, 2022.
- CURZEL, J. L. *et al.* Temporal performance analysis of bus transportation using link streams. **Mathematical Problems in Engineering**, Wiley Online Library, v. 2019, n. 1, p. 6139379, 2019.
- DONG, J.; ZOU, L.; ZHANG, Y. Mixed model for prediction of bus arrival times. *In: 2013 IEEE Congress on Evolutionary Computation*. [S.l.: s.n.], 2013. p. 2918–2923.
- FARAWAY, J. J. Extending the linear model with r: generalized linear, mixed effects and nonparametric regression models. **Chapman and Hall/CRC**, Taylor & Francis, 2006.
- FITZMAURICE, G. M.; LAIRD, N. M.; WARE, J. H. Applied longitudinal analysis. **John Wiley & Sons**, Wiley, 2012.
- GALLIVAN, F. *et al.* Cost-effective approaches to reduce greenhouse gas emissions through public transportation in los angeles, california. **Transportation research record**, SAGE Publications Sage CA: Los Angeles, CA, v. 2217, n. 1, p. 19–29, 2011.
- HABIB, K. M. N.; KATTAN, L.; ISLAM, M. T. Model of personal attitudes towards transit service quality. **Journal of Advanced Transportation**, Wiley Online Library, v. 45, n. 4, p. 271–285, 2011.
- HASHIGUCHI, K. K. *et al.* Exploratory analysis of public transportation data of curitiba, brazil. *In: IEEE. 2020 IEEE Symposium on Computers and Communications (ISCC)*. [S.l.], 2020. p. 1–6.
- JACOBS, J. **The Death and Life of Great American Cities**. New York: Random House, 1961.
- LI, J. *et al.* Bus arrival time prediction based on mixed model. **China Communications**, v. 14, n. 5, p. 38–47, 2017.

- LIU, Y. *et al.* Understanding urban bus travel time: Statistical analysis and a deep learning prediction. **International Journal of Modern Physics B**, v. 37, n. 04, p. 2350034, 2023. Disponível em: <https://doi.org/10.1142/S0217979223500340>.
- MANIKA, E. R. **Ambiente para geração de trajetos de linhas de transporte público para análise de mobilidade e viabilidade de comunicação**. 2022. Dissertação (Mestrado) — Universidade Tecnológica Federal do Paraná, 2022.
- MARTINS, T. S. *et al.* **Map matching: uma análise de dados streaming de trajetórias de GPS no transporte público**. 2022. Dissertação (Mestrado) — Universidade Tecnológica Federal do Paraná, 2022.
- MATLOFF, N. **Statistical regression and classification: from linear models to machine learning**. [S.l.]: Chapman and Hall/CRC, 2017. 75 p.
- MOUWEN, A. Drivers of customer satisfaction with public transport services. **Transportation Research Part A: Policy and Practice**, Elsevier, v. 78, p. 1–20, 2015.
- PASQUIM, B.; FONSECA, K.; JR., L. M. Quando chega? análise de previsibilidade de tempos de espera em transporte público urbano utilizando modelos mistos. *In: Anais do IX Workshop de Computação Urbana*. Porto Alegre, RS, Brasil: SBC, 2025. p. 15–28. ISSN 2595-2706. Disponível em: <https://sol.sbc.org.br/index.php/courb/article/view/35248>.
- PEIXOTO, A. M. *et al.* Plataforma computacional para construção de um banco de dados de grafo do sistema de transporte de Curitiba. *In: SBC. Anais do IV Workshop de Computação Urbana*. [S.l.], 2020. p. 125–137.
- PINHEIRO, J. *et al.* **nlme: Linear and Nonlinear Mixed Effects Models**. 3.1-137. ed. [S.l.], 2018. R Package Version 3.1-137. Disponível em: <http://CRAN.R-project.org/package=nlme>.
- PRESTON, J. Transport, public. *In: KITCHIN, R.; THRIFT, N. (Ed.). International Encyclopedia of Human Geography*. Oxford: Elsevier, 2009. p. 452–459. ISBN 978-0-08-044910-4. Disponível em: <https://www.sciencedirect.com/science/article/pii/B9780080449104010270>.
- RODRIGUEZ-DENIZ, H.; VILLANI, M. Robust real-time delay predictions in a network of high-frequency urban buses. **IEEE Transactions on Intelligent Transportation Systems**, v. 23, n. 9, p. 16304–16317, 2022.
- SERIN, F.; ALISAN, Y.; ERTURKLER, M. Predicting bus travel time using machine learning methods with three-layer architecture. **Measurement**, v. 198, p. 111403, 2022. ISSN 0263-2241. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0263224122006364>.
- SUN, B. *et al.* Analysis and dynamic prediction of bus dwell time under rainfall conditions. **Promet-Traffic&Transportation**, Sveučilište u Zagrebu Fakultet prometnih znanosti, v. 37, n. 1, p. 105–121, 2025.
- SUWARDO, W.; NAPIAH, M.; KAMARUDDIN, I. Arima models for bus travel time prediction. **J. Inst. Eng. Malaysia**, v. 71, n. 2, p. 49–58, 2010.
- WIMBADI, R. W.; DJALANTE, R.; MORI, A. Urban experiments with public transport for low carbon mobility transitions in cities: A systematic literature review (1990–2020). **Sustainable Cities and Society**, Elsevier, v. 72, p. 103023, 2021.
- ZHANG, B. *et al.* Bus arrival time prediction based on the optimized long short-term memory neural network model with the improved whale algorithm. **Journal of Advanced Transportation**, v. 2024, n. 1, p. 6997338, 2024. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1155/2024/6997338>.