

MARIANA MARTINS DE BRITO SOUSA

**Previsão espaço-temporal de entregas urbanas na etapa de *last-mile*  
utilizando o modelo STARMA**

Versão Corrigida

Dissertação apresentada à Escola Politécnica  
da Universidade de São Paulo para obtenção  
do título de Mestre em Ciências.

Área de Concentração: Engenharia de  
Sistema Logístico

Orientador: Prof. Dr. Daniel de Oliveira Mota

São Paulo

2021



MARIANA MARTINS DE BRITO SOUSA

**Previsão espaço-temporal de entregas urbanas na etapa de *last-mile*  
utilizando o modelo STARMA**

São Paulo

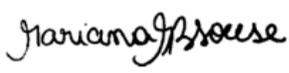
2021

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Este exemplar foi revisado e corrigido em relação à versão original, sob responsabilidade única do autor e com a anuência de seu orientador.

São Paulo, 05 de Outubro de 2021

Assinatura do autor: 

Assinatura do orientador: 

#### Catálogo-na-publicação

Sousa, Mariana Martins de Brito

Previsão espaço-temporal de entregas urbanas na etapa de last-mile utilizando o modelo STARMA / M. M. B. Sousa -- versão corr. -- São Paulo, 2021.

82 p.

Dissertação (Mestrado) - Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Produção.

1.Veículos autônomos 2.Last-mile 3.Previsão de demanda 4.Modelo STARMA 5.Séries temporais I.Universidade de São Paulo. Escola Politécnica.

## **Agradecimentos**

Agradeço imensamente ao meu orientador, Prof. Dr. Daniel de Oliveira Mota, pelos ensinamentos, orientações e dedicação em todo o período do meu mestrado.

À Prof. Dra. Linda Lee HO, por me auxiliar no desenvolvimento do trabalho.

À Prof. Dra. Denise Aparecida Botter e ao Prof. Dr. Lino Guimarães Marujo, pela disponibilidade de participar e me avaliar na banca examinadora.

Aos Professores Doutores, Cláudio Barbieri da Cunha, Hugo Tsugunobu Yoshida Yoshizaki, Rui Carlos Botter, José Geraldo Vidal, Marco Antonio Brinati, e a Professora Doutora Débora Pretti Ronconi que estiveram presente durante meu mestrado me ensinando e apoiando durante suas aulas.

Ao Prof. Dr. Cauê Sauter Guazzelli por me incentivar a iniciar o mestrado e me auxiliar durante a graduação.

E aos meus familiares e amigos pelo apoio ao longo de toda nossa jornada.



## Resumo

O nível de urbanização do mundo está aumentando nos últimos anos e conseqüentemente cresce o número de movimentos de carga e pessoas nos centros urbanos, desafiando ainda mais a infraestrutura de mobilidade das cidades, durante a etapa de *last-mile*. Uma alternativa importante em relação ao transporte convencional é a utilização de uma frota de veículos autônomos, que interfere positivamente na sustentabilidade, diminuindo o consumo de combustíveis e as emissões de carbono. A proposta desse trabalho é criar um modelo com a função de realizar a previsão da demanda espaço-temporal das viagens de *Yellow Taxi*, na cidade de Nova Iorque, com o intuito de reduzir a quantidade de taxis vazios nas ruas, economizando energia e diminuindo o congestionamento de veículos nos grandes centros urbanos. Para modelar o problema foi utilizado o modelo STARMA, de séries temporais, considerando a correlação espaço-temporal dos dados. Os resultados indicaram um erro percentual absoluto médio de aproximadamente 45% para as previsões, demonstrando que a correlação espacial exerce papel importante nos dados.

Palavras-chaves: Veículos autônomos, *Last-mile*, Previsão, Séries Temporais, modelo STARMA

## Abstract

The level of urbanization in the world has been increasing in recent years and consequently the number of cargo movements and people in urban centers has grown, further challenging the mobility infrastructure of cities during the last-mile stage. An important alternative in relation to conventional transport is the use of a fleet of autonomous vehicles, which positively interferes with sustainability, reducing fuel consumption and carbon emissions. The purpose of this work is to create a model to forecast the space-time demand for Yellow Taxi trips in New York City, in order to reduce the amount of empty taxis on the streets, saving energy and decreasing the vehicle congestion in large urban centers. To model the problem, it will be used a STARMA model, of time series, considering the spatio-temporal correlation of the data. The results indicated a mean absolute percentage error of approximately 45% for the predictions, demonstrating that the spatial correlation plays an important role in the data.

Keywords: Autonomous vehicles, Last-mile, Forecast, Time Series, STARMA model

## SUMÁRIO

<b>1. INTRODUÇÃO</b> .....	<b>1</b>
1.1. INFORMAÇÕES ADICIONAIS .....	3
<b>2. REVISÃO BIBLIOGRÁFICA</b> .....	<b>4</b>
2.1. TRANSPORTE E LOGÍSTICA NAS CIDADES .....	4
2.2. VEÍCULOS AUTÔNOMOS .....	6
<b>2.2.1. Mobilidade Compartilhada</b> .....	<b>9</b>
<b>2.2.2. Entregas urbanas na etapa de <i>last-mile</i></b> .....	<b>10</b>
<b>2.2.3. Desafios no transporte de pessoas e cargas</b> .....	<b>12</b>
2.3. PREVISÃO DE DEMANDA .....	14
<b>2.3.1. Aplicação de previsão de demanda na tecnologia dos veículos autônomos</b> .....	<b>15</b>
2.4. SÉRIES TEMPORAIS .....	17
2.5. MODELOS PARA SÉRIES TEMPORAIS .....	20
<b>2.5.1. Modelo ARIMA e suas variações</b> .....	<b>21</b>
2.6. MODELOS PARA DADOS COM CORRELAÇÃO ESPAÇO-TEMPORAL .....	22
<b>2.6.1. Matriz de vizinhança e variograma</b> .....	<b>23</b>
<b>2.6.2. Modelo STARMA e STARIMA</b> .....	<b>25</b>
<b>2.6.3. Função espaço-temporal de autocorrelação e autocorrelação parcial</b> .....	<b>26</b>
<b>3. METODOLOGIA</b> .....	<b>29</b>
3.1. MODELO STARMA .....	36
<b>3.1.1. Identificação</b> .....	<b>37</b>
<b>3.1.2. Estimação e Verificação</b> .....	<b>45</b>
3.2. MODELO ARIMA .....	45
<b>3.2.1. Identificação</b> .....	<b>46</b>

<b>3.2.2. Estimação e verificação .....</b>	<b>49</b>
<b>4. RESULTADOS.....</b>	<b>51</b>
4.1. Modelo STARMA.....	51
4.2. Modelo SARIMA.....	55
4.3. Comparação STARMA x SARIMA .....	57
<b>5. CONCLUSÃO.....</b>	<b>60</b>
<b>REFERÊNCIAS.....</b>	<b>62</b>
<b>ANEXO A.....</b>	<b>66</b>

## LISTA DE FIGURAS

Figura 1 - Processo de previsão .....	15
Figura 2 - Tipos de séries temporais.....	18
Figura 3 - Ordem espacial em sistemas de uma e duas dimensões.....	23
Figura 4 - Mapa da cidade de Nova Iorque segmentado pelas zonas de táxi e boroughs.....	30
Figura 5 - Gráfico da quantidade de pick-up ocorrido por mês em 2019 .....	30
Figura 6 - Gráfico da quantidade de pick-up ocorrido por dia no mês de maio/2019.....	31
Figura 7 - Gráfico da quantidade de pick-ups(a) ocorridos no mês de maio por zona .....	31
Figura 8 - Zonas de táxi em Manhattan.....	32
Figura 9 - Variograma .....	34
Figura 10 - Plotagem dos dados .....	38
Figura 11 - Função de autocorrelação espaço temporal (para os lags 0 e 1).....	39
Figura 12 - Histograma dos dados indicando as transformações .....	40
Figura 13 - Plotagem dos Dados com Transformação Deviance Residual .....	41
Figura 14 - Função de autocorrelação espaço temporal dos Dados com Transformação Deviance Residual .....	42
Figura 15 - STACF dos dados com transformação e uma diferenciação .....	43
Figura 16 – Função de autocorrelação parcial espaço temporal com os dados transformados e com diferenciação .....	44
Figura 117 - Gráfico da quantidade de pick-ups(a) e drop-off(b) ocorridos no mês de maio por zona .....	46
Figura 18 - Plotagem dos dados .....	46
Figura 19 - Função de autocorrelação.....	47
Figura 20 - Histograma dos dados indicando as transformações .....	47
Figura 21 - ACF dos dados com transformação e uma diferenciação .....	48
Figura 22 - Função de autocorrelação parcial com os dados transformados e com diferenciação .....	49

Figura 23 - Diagnóstico gráfico do Modelo .....	53
Figura 24 - Comparação dos dados observados pelos dados ajustados pelo modelo .....	54
Figura 25 - Diagnóstico gráfico do Modelo SARIMA .....	57

Tabela 1 - Definição da SAE dos níveis de automação dos veículos .....	8
Tabela 2 - Identificação do Modelo STARMA e suas variações.....	28
Tabela 3 - Matriz de Distâncias.....	32
Tabela 4 - Banco de dados.....	33
Tabela 5 - Matriz de Ponderação da Vizinhança W1.....	34
Tabela 6 - Matriz de ordem dos parâmetros para construção do modelo STARMA .....	44
Tabela 7 - Valores estimados dos parâmetros do modelo STARMA (15,16,15,16) .....	51
Tabela 8 - Parâmetros do modelo SARIMA .....	55



## 1. INTRODUÇÃO

O nível de urbanização do mundo está aumentando significativamente ao longo dos anos. Projeta-se a inclusão de aproximadamente 2,5 bilhões de pessoas aos centros urbanos até 2050 (ONU, 2014). Esse crescimento gera consequências que impactam na vida de boa parte da população como, por exemplo, o aumento da demanda por serviços e por infraestrutura das cidades. Esse aumento implica na necessidade de um número de veículos cada vez maior circulando nas vias urbanas resultando em congestionamentos. Segundo Beirigo et al. (2018) espera-se um cenário desastroso caso nenhuma ação seja tomada no sentido de racionalizar o uso das vias.

Na tentativa de reduzir o impacto do aumento do número de veículos nas cidades, o termo *Smart City*, tem se tornado cada vez mais relevante quando se busca atingir tal racionalização. Segundo XIAN (2019) o objetivo das *Smart Cities* é promover o desenvolvimento sustentável e a alta qualidade de vida por meio do gerenciamento inteligente de recursos.

Apoiado no contexto de *Smart City*, alguns autores defendem que o transporte de pessoas e cargas pode se beneficiar bastante dos veículos autônomos. De acordo com CHONG (2011) se trata de uma importante alternativa em relação ao transporte convencional e tem potencial para mudar a maneira como as pessoas se movimentam pelas cidades. Principalmente quando se fala de *last-mile*, já que essa etapa está intrinsecamente relacionada ao ambiente urbano, além de ser considerada a menos eficiente e mais poluente de toda a cadeia de suprimentos (GEVAERS, et al., 2014). LITMAN (2017) diz que os veículos autônomos podem tornar o tráfego mais eficiente, reduzindo congestionamentos, consumo de energia e emissão de poluentes. Mas, ele ressalta que esses impactos dependerão de como os veículos irão interagir com outras tendências, como: (1) a mudança de veículos particulares para veículos compartilhados; e (2) o uso inteligente desses veículos.

Com a chegada dos veículos autônomos é importante analisar os impactos que serão causados na quantidade total de viagens realizadas nos centros urbanos. De acordo com Miller e Kang (2019), alguns fatores poderão influenciar no aumento das viagens como: (1) deslocamentos com

os veículos vazios para realizar o *pick-up* de um passageiro, *drop-off* de bens de consumos e viagens para as estações de manutenção, (2) maior comodidade e (3) passageiros não habilitados que poderão utilizar os veículos como meio de transporte. Porém, com o auxílio de estudos prévios, espera-se que não haja aumento na quantidade de viagens e sim uma redução. Entre esses estudos, destaca-se a previsão de demanda, estimando onde e quando serão as próximas viagens, a fim de, otimizar a maneira com que os veículos irão circular. De acordo com Moreira et al. (2013) existem aspectos que dificultam a modelagem e previsão da demanda espacial. Por isso, a maioria dos métodos de previsões já existentes utilizam somente dados de fluxos de veículos em locais específicos desconsiderando características importantes da demanda como os dados de contagem que se correlacionam no domínio espacial.

Sendo assim, o objetivo desse trabalho foi incorporar as particularidades dos veículos autônomos para viabilizar o uso desta tecnologia nos centros urbanos como solução de mobilidade, tendo em vista, contribuir de forma racional para a qualidade de vida nas cidades. Essa racionalização foi feita por meio da proposta de um sistema de transporte público urbano que visa o dimensionamento adequado da frota de veículos durante a etapa de *last-mile*, otimizando a quantidade de veículos necessários para atender a população e a alocação dos mesmos a fim de minimizar as distâncias percorridas.

Com o intuito de atingir o objetivo foi indicado um modelo de previsão de demanda, que estimou as solicitações por transporte público individual no ambiente urbano utilizando séries temporais e levando em consideração aspectos espaço-temporal. Para o desenvolvimento desse trabalho foram utilizados dados de transporte público individual (táxi), que representam uma parcela importante no deslocamento das pessoas entre os transportes públicos e privados sendo que uma quantidade significativa da população viaja de táxi diariamente em todo o mundo (YUAN et al., 2011). Com o intuito de realizar uma previsão precisa, foi utilizado um modelo que correlaciona aspectos temporais e espaciais.

O trabalho foi dividido em capítulos. No Capítulo 2, foram apresentados os conceitos introdutórios para construção do conhecimento e melhor compreensão do trabalho e seu objetivo. Exibindo, inicialmente, assuntos sobre veículos autônomos, mobilidade e entregas

urbanas na etapa de *last-mile*. Depois foram abordados, os conceitos a respeito da previsão de demanda e sua importância. E por fim, o conceito e as propriedades das séries temporais assim como, os modelos que podem ser utilizados para previsão. No Capítulo 3, foi apresentada a metodologia utilizada no trabalho, com a aplicação dos dados das viagens de táxi da cidade de Nova Iorque, e em seguida foram indicados os resultados do modelo gerado. Já o Capítulo 4 foi responsável por concluir o trabalho e apresentar os próximos passos que foram desenvolvidos.

### 1.1. INFORMAÇÕES ADICIONAIS

Inicialmente o trabalho foi desenvolvido com o intuito de criar um modelo de previsão de demanda que pudesse ser implementado em uma simulação de veículos autônomos (VA) focada em entregas urbanas, visando aumentar o nível de serviço e reduzir o tráfego nos grandes centros urbanos, desenvolvida por mim em parceria com o *Massachusetts Institute of Technology* (MIT). O laboratório Media Lab do MIT criou um veículo autônomo próprio, conhecido como *Persuasive Electric Vehicle* (PEV) e por isso houve um grande interesse em desenvolver uma simulação que unisse os benefícios de um VA com a logística *last-mile*, através do estudo desses temas percebeu-se a importância de incluir a previsão de demanda do transporte de pessoas e cargas.

Porém ao longo do desenvolvimento do trabalho, nos deparamos com uma pandemia e infelizmente o Media Lab, alterou o foco de seus projetos para o Corona Virus, sendo assim, houve uma pausa nos estudos relacionados à VA. Mas como meu trabalho estava parcialmente desenvolvido resolvi seguir com meu tema, porém redirecionado para um novo objetivo central.

Sendo assim, a proposta desse trabalho foi criar um modelo de previsão de demanda que considere aspectos espaço-temporais, cujo principal objetivo fosse auxiliar na tomada de decisão em um sistema de transporte de *last-mile* operando em grandes centros urbanos. Comparando dois modelos de previsão afim de comentar sobre as principais diferenças e vantagens de cada um deles.

Para isso, utilizou-se dados de taxi da cidade de Nova Iorque para compor o banco de dados do trabalho, já que foram os únicos dados confiáveis e em grande quantidade que se pode encontrar online, focando então apenas no transporte de pessoas.

## 2. REVISÃO BIBLIOGRÁFICA

Neste capítulo foram apresentados conceitos importantes para a compreensão do trabalho: (1) transporte e logística nas cidades, abrangendo a história e os desafios encontrados nesse assunto; (2) veículos autônomos, inserido no contexto da mobilidade urbana e entregas realizadas na etapa de *last-mile*; (3) previsão de demanda, incluindo sua definição e com foco na previsão de viagens considerando origem e destino; e finalmente (4) séries temporais e seus modelos.

Na modelagem de séries temporais foram apresentados e abordados os modelos que apresentam correlação espacial nos dados, STARMA (*Spatio-Temporal Autoregressive Moving Average*) e STARIMA (*Spatio-Temporal Autoregressive Integrated Moving Average*).

### 2.1. TRANSPORTE E LOGÍSTICA NAS CIDADES

Problemas em logística e no transporte de pessoas e mercadorias são enfrentados desde antes dos computadores serem inventados e da Pesquisa Operacional (PO) se tornar uma disciplina que visa desenvolver métodos e técnicas para tomadas de decisão.

De acordo com Speranza (2018) após o desenvolvimento dos primeiros modelos de otimização, a Pesquisa Operacional tem contribuído substancialmente para tornar os sistemas de transporte mais eficientes. Capturando a complexidade dos problemas e as interações entre as partes de um sistema para melhorar a qualidade da tomada de decisão. Os métodos de PO dependem da disponibilidade de dados e de alta capacidade computacional.

Sendo assim, através de avanços tecnológicos, da internet, do uso individual de dispositivos de informação e comunicação, e da ampla disponibilidade de grandes quantidades de dados surgiram novos desafios e oportunidades para sistemas de transporte e logística.

De acordo com Waller (2013) os avanços tecnológicos mais recentes estão relacionados à explosão dos dados digitais, os chamados "*big data*", e à expansão do conceito de internet para a chamada Internet das Coisas (Internet of Things - IoT). O número de pesquisas no Google por "*big data*" excedeu em 2013 o número de pesquisas por "supply chain management". Isso não significa que os dados sejam mais importantes do que o gerenciamento da cadeia de suprimentos,

mas certamente é um sinal da crescente percepção de que a disponibilidade de grandes quantidades de dados é relevante para empresas e serviços, para os setores privado e público.

IoT permite que os objetos sejam detectados e controlados remotamente em uma infraestrutura de rede existente, criando oportunidades para uma integração mais direta entre o mundo físico e os sistemas baseados em computador. A IoT abrange redes inteligentes, casas inteligentes, transporte inteligente e cidades inteligentes. Cada coisa é identificada por sua tecnologia incorporada e é capaz de interoperar com a infraestrutura de Internet existente. Espera-se que a IoT gere no futuro maiores quantidades de dados do que os disponíveis atualmente. *Big data* e IoT estão abrindo enormes oportunidades para um grande número de novas aplicações e projetos de pesquisa (Speranza, 2018).

O desenvolvimento dessas novas tecnologias alterou o ambiente das atividades de supply chain e transportes e diversas estratégias estão sendo aplicadas para resolver os problemas de pesquisa operacional.

Embora os carros particulares continuem sendo o meio de transporte dominante para a grande maioria das pessoas, o conjunto de opções de mobilidade está crescendo, startups deste setor se estabeleceram em pouco tempo. Uber, BlaBlaCar, Zipcar são apenas alguns dos nomes correspondentes a empresas que oferecem um meio de transporte alternativo às pessoas, umas para curtas, outras para longas distâncias. Os jovens tendem a usar essas novas opções e a atrasar a compra de um carro e a obtenção da carteira de habilitação.

Porter (2015), cita as seis principais tendências para o transporte de pessoas e mercadorias que irão mudar a maneira que ocorre a logística last-mile nas grandes cidades:

- (I) Veículos autônomos: veículos que podem se movimentar sem o auxílio do ser humano
- (II) Veículos elétricos: principalmente os ônibus e os veículos de curto alcance; os veículos elétricos estão se tornando mais econômicos e podem viajar mais tempo sem serem abastecidos.
- (III) Veículos conectados: os dados de tráfego estão se tornando disponíveis nos veículos, os veículos estão equipados com conectividade com a Internet.

- (IV) Mobilidade compartilhada: as opções de mobilidade sob demanda estão crescendo, as opções colaborativas permitem a mobilidade sem a utilização de veículos individuais.
- (V) Redes multimodais eficientes: o crowdsourcing de dados de trânsito adaptará os horários às necessidades dos viajantes, e várias opções de viagem serão oferecidas aos viajantes.
- (VI) Novos materiais: Serão projetados veículos mais leves, também para aumentar a distância percorrida pelos veículos elétricos.

Uma frota de veículos autônomos compartilhados, conectados à infraestrutura rodoviária, à internet e à uma rede mais ampla de opções de transporte público, já está sendo discutida e estudada por pesquisadores.

Os incentivos ao uso de veículos elétricos estão crescendo nos últimos anos, devido aos diversos impactos positivos em relação ao meio ambiente. Enquanto o impacto global de uma substituição massiva de veículos tradicionais por elétricos ainda está sendo viabilizada, especialmente em termos de produção de eletricidade, a tendência para o uso de veículos elétricos parece ser irreversível, acarretando inúmeros benefícios ambientais, especialmente em áreas densamente povoadas. Vários artigos já abordaram problemas de otimização que surgiram especificamente para veículos elétricos (ver, por exemplo, Adler & Mirchandani, 2014, Schneider, Stenger, & Goeke, 2014, Pelletier, Jabali, & Laporte, 2014, Yang & Sun, 2015). Tais problemas podem dizer respeito à localização de estações de recarga, ao roteamento de veículos limitado pela autonomia, à reserva de uma bateria e à gestão de frota.

## 2.2. VEÍCULOS AUTÔNOMOS

De acordo com Stocker e Shaheen (2017) veículos autônomos (*AVs – Autonomous Vehicles*) são aqueles utilizados para transportar passageiros ou cargas com algum nível de automação, cujo objetivo é substituir o controle humano. Ou seja, são veículos preparados para realizar tarefas de controle e direção, tomando decisões, inspiradas em padrões humanos, por meio de plataformas e sensores eletrônicos.

Há alguns anos os AVs já estão funcionando, principalmente para o uso em sistemas controlados ou com trilhos fixos, como no caso de trens ou transportadores de pessoas em aeroportos. Porém, atualmente esses veículos estão sendo desenvolvidos para uso em vias públicas por diversos fabricantes de automóveis e empresas de tecnologia (LI et al., 2018).

O termo “automatização de veículos” já está sendo discutido e estudado desde o uso em massa do automóvel no início do século XX. A primeira tentativa de um veículo automatizado foi em 1925, quando uma empresa de equipamentos de rádio chamada *Houdina Radio Control* conduziu um veículo pela cidade de Nova York com um carro atrás controlando seu movimento através de antenas transmissoras. Uma das primeiras vezes em que a ideia de AVs ganhou ampla exposição ao público foi durante a exibição do Futurama da General Motors na Feira Mundial de Nova York de 1939. A empresa imaginou um futuro em que os carros navegariam em rodovias usando o "controle automático de rádio" para manter distâncias seguras em altas velocidades (STOCKER & SHAHEEN, 2017). Nos anos seguintes, tecnologias mais avançadas começaram a ser desenvolvidas em todo o mundo.

A Sociedade de Engenheiros Automotivos (SAE – *Society of Automotive Engineers*) definiu os diferentes níveis de funcionalidade autônoma em veículos variando do nível 0 (sem autonomia) até o nível 5 (completamente autônomo) (SAE International, 2020). A Tabela 3 indica os diferentes níveis de autonomia e suas definições. A maior distinção está entre os níveis 0-2 e 3-5, baseados em definir se o humano ou o sistema automatizado será o principal responsável pelo monitoramento do ambiente de direção. Além disso, nos quatro primeiros níveis é obrigatório a presença de um motorista dentro do veículo, para que durante algum incidente o motorista assumo o controle, já no quinto nível o motorista não tem nenhum controle sobre o veículo.

Tabela 1 - Definição da SAE dos níveis de automação dos veículos

Nível de Automação	Descrição
Nível 0	Somente o motorista: Sem automação
Nível 1	Assistido: Automação de uma função de controle principal, ex: estacionamento automático ou frenagem automática
Nível 2	Automação Parcial: Automação de duas ou mais funções de controle principal, projetado para aliviar o motorista no controle dessas funções
Nível 3	Condução Autônoma Limitada: o motorista pode ceder o controle de todas as funções críticas de segurança sob certas condições de tráfego ou ambientais, mas deve estar disponível para controle caso haja necessidade
Nível 4	Alta Condução Autônoma: sem a necessidade de controle humano, com capacidade operacional mesmo se o motorista não responder apropriadamente a um pedido de intervenção
Nível 5	Condução Autônoma Total: sem controle humano, em qualquer ambiente

Fonte: Adaptado de STOCKER & SHAHEEN (2017)

De acordo com estatísticas (CARSTEN, 2015) cerca de 10 a 30% dos acidentes em rodovias acontecem devido à desatenção ou distração dos motoristas. Devido à automação, acidentes causados por ineficiência humana como inexperiência, fadiga, influência de drogas, problemas de saúde, entre outros são completamente excluídos. Também é esperado que haja redução nas colisões de alta velocidade devido a um elevado tempo de reação, além de reduzir erros visuais e acidentes com veículos únicos. Neste contexto, acredita-se que os AVs são mais seguros que veículos manuais, uma vez que não estão sujeitos a distrações ou fadigas. Porém, o nível de segurança esperado para esse tipo de veículo é muito maior e com mais restrições sendo que deverão ser feitos ajustes na legislação de trânsito para tratar desses veículos, definindo responsabilidades legais sobre os eventos provocados por AVs.

Atualmente a adoção de qualquer nova tecnologia está associada com seus aspectos de sustentabilidade. Em particular, os veículos autônomos podem ser julgados em termos de 3 áreas: (1) o impacto social, esse tipo de carro irá permitir o deslocamento de todos os grupos de pessoas, incluindo, idosos e pessoas com deficiência; (2) o impacto ambiental, os carros autônomos que estão sendo desenvolvidos em sua maioria são elétricos, portanto, reduzindo a emissão dos gases poluentes, além disso, os veículos estão programados de modo que sua condução seja menos agressiva, diminuindo o índice de frenagens e acelerações desnecessárias, sendo assim reduzindo

a queima de combustíveis; por fim, (3) o impacto econômico, haverá uma diminuição significativa nos custos generalizados, como em fretes de cargas (FAHEEM,2017).

### 2.2.1. Mobilidade Compartilhada

Entende-se por mobilidade compartilhada o uso comum de veículos, como: bicicleta, patinete ou outro modo de transporte, que permite ao usuário ter acesso em um período curto de tempo, aos meios de transporte, de acordo com as suas necessidades (SHAHEEN et al., 2015).

De acordo com Shaheen, Martin e Cohen (2013), o sistema de mobilidade compartilhada opera em três diferentes tipos de *business models*: (1) *Business-to-Consumer (B2C)*, em que as empresas geralmente possuem uma frota de veículos e permitem que os usuários o utilizem em troca do pagamento de taxas de serviço e uso, alguns exemplos desse tipo de compartilhamento são, os carros compartilhados (*carsharing*) com benefícios de um carro particular, porém sem custos de compra e manutenção, e as bicicletas compartilhadas (*bikesharing*) permitindo que as pessoas possam ter acesso à bicicletas de acordo com suas necessidades em diversos pontos distribuídos pelas cidades. (2) *Peer-to-peer (P2P)*, se trata do serviço em que as empresas supervisionam as transações entre proprietários e locatários dos veículos, fornecendo a plataforma e os recursos necessários. O *P2P carsharing*, por exemplo, realiza o compartilhamento de veículos particulares para os usuários da companhia. Outro exemplo, desse tipo de serviço são as caronas compartilhadas (*ridesharing*), facilitando a comunicação de motoristas e passageiros com origens e destinos similares. E por fim, (3) *For-Hire Service*, que se refere à contratação de um motorista de acordo com as necessidades do cliente, os mais conhecidos são os serviços de *ridesourcing*, como Uber, e os táxis.

Sabe-se que a mobilidade compartilhada está crescendo rapidamente ao redor do mundo. Sendo assim, o avanço da tecnologia de veículos autônomos e o crescimento da oferta de serviços compartilhados, pode gerar alternativas importantes e inovadoras de maneira a substituir o transporte convencional, e ter o potencial de alterar a forma como as pessoas se movimentam nas cidades. A convergência dessas duas inovações está sendo desenvolvida, através de experimentos com os veículos autônomos compartilhados (*SAV – Shared Autonomous Vehicles*) emergindo pelo mundo. Utilizar veículos autônomos como parte de uma frota pode trazer

benefícios relevantes com alta eficiência, baixo nível de emissão de gases poluentes e baixo custo de operação (GREENBLATT & SHAHEEN, 2015).

Existem alguns desenvolvimentos recentes em mobilidade compartilhada que incluem sistemas autônomos parciais ou condicionais (SAE Nível 2 e 3, respectivamente), que fornecem funções automatizadas em um ambiente operacional controlado. Em várias regiões, o serviço de transporte automatizado, como o uso de ônibus autônomos, está em fase de teste. Por estar em pequena escala os benefícios são de difícil aferição uma vez que os impactos dos resultados destas experiências piloto ainda não foram documentados (STOCKER & SHAHEEN, 2017).

Além do uso de AVs parciais ou condicionais, houve uma explosão de interesse no uso de frotas totalmente automatizadas (SAE Níveis 4 e 5) nos últimos anos. Esse interesse, deve-se, principalmente, à grande divulgação do desenvolvimento de veículos autônomos, assim como à popularidade dos serviços de transporte e à percepção de que o custo operacional por quilômetro de serviço pode diminuir consideravelmente em comparação aos preços atuais. Sendo assim, trata-se de uma alternativa viável para a etapa de *first/last-mile*, que atualmente apresenta escala limitada, já que a única solução viável é o transporte convencional.

A inclusão de opções adicionais de alta qualidade que resolvem o problema de acessibilidade para uma variedade maior de usuários pode permitir que mais pessoas aproveitem os serviços de transporte em massa. Sendo assim, os veículos autônomos compartilhados contribuirão para o lançamento de soluções inovadoras para a mobilidade nos centros urbanos (STOCKER & SHAHEEN, 2017).

### 2.2.2. Entregas urbanas na etapa de *last-mile*

Segundo Gevaers, et al. (2014) logística a *last-mile* é considerada o último elo da logística urbana, e está relacionada à otimização dos sistemas de transporte urbano de mercadorias. O mesmo autor afirma que atualmente a logística no *last-mile*, é uma das etapas mais caras (correspondem à 28% dos custos de *delivery*), menos eficiente e mais poluente de toda a cadeia de suprimentos. O aumento da urbanização e a conscientização dos impactos do transporte de mercadorias enfatizam a importância da logística urbana como uma abordagem que visa diminuir os efeitos

negativos das atividades de distribuição sem penalizar questões sociais, culturais e econômicas (DE MARCO et al., 2014).

De acordo com Guerrero e Díaz-Ramírez (2017) o termo, *last-mile*, é comumente utilizado quando lida-se com a distribuição ou processos de *delivery* dentro de uma cidade, ou seja, se refere a última etapa do *supply chain*, e é uma etapa importante por envolver a transação final entre a empresa e o cliente. Portanto, a logística *last-mile* é o último trecho de uma carga B2C (*business-to-customer*). *City Logistics* é descrito por Crainic et al (2009) como o processo de otimização total das atividades logísticas e de transporte realizado por empresas privadas em áreas urbanas. Este conceito leva em consideração a rede de tráfego, os congestionamentos e o consumo de energia no âmbito de uma economia com mercado livre. Além disso, o transporte de pessoas dentro da cidade também exerce um fator importante quando o assunto é a logística das cidades. Baseada em uma abordagem que envolve grande quantidade de processos técnicos incluindo modelagem, avaliação e aplicação de informações tecnológicas. Existem três aspectos característicos que condicionam a distribuição urbana: infraestrutura, estratégia de distribuição e os veículos (SAENZ, 2016). Nesta Seção foi abordada como a inclusão de veículos autônomos na etapa de *last-mile* pode diminuir os custos e a poluição, além do aumento da eficiência.

Veículos autônomos podem ser a solução para os problemas de *last-mile* por facilitar o acesso do transporte de cargas e pessoas, principalmente o uso de veículos autônomos compartilhados, melhorando o fluxo do tráfego, diminuindo o consumo de combustíveis e as emissões de carbono, reduzindo o número de acidentes e aumentando a produtividade (MANKYA et al, 2013).

A demanda do transporte de passageiros varia bastante ao longo do dia, de modo que qualquer frota, sendo ela compartilhada ou não, fique, inevitavelmente, ociosa fora do horário de pico. Uma alternativa para aproveitar todo o potencial de um veículo de transporte urbano seria utilizá-lo também para o transporte de cargas durante os períodos de ociosidade, atendendo as demandas de frete sempre que necessário (BEIRIGO,2018).

### 2.2.3. Desafios no transporte de pessoas e cargas

De acordo com Rosana, M. (2020) o transporte urbano refere-se à mobilidade de pessoas e mercadorias conectando pontos de origem e destino dentro das áreas urbanas, incluindo, transporte público e privado, pedestres, modos de transporte não motorizados (por exemplo, bicicletas) e distribuição de carga.

Transporte urbano de carga, (GRUCHMANN, T, 2019), definido como todo movimento de mercadorias para dentro, fora, através ou dentro da área urbana, feito por veículos leves ou pesados, incluindo: (i) Entrega de encomendas (comercial e residencial); (ii) Transporte de serviço e tráfego de demolição; (iii) Viagens de compras feitas por famílias; (iv) Logística reversa para retirada de resíduos e gerenciamento de devoluções; (v) Carrinhas de serviço para manutenção, abastecimento e retirada de peças.

No contexto da gestão de cargas urbanas, o objetivo é alcançar a integração máxima da movimentação de cargas nas operações e atividades urbanas que permitam às pessoas obter os bens de que precisam, tendo em vista o desenvolvimento sustentável.

O desenvolvimento sustentável é mais comumente definido como "atender às necessidades do presente sem comprometer a capacidade das gerações futuras de atender às suas próprias necessidades" (Ito, P. 2019). Com isso, está implícita a harmonização conjunta de três questões: crescimento econômico, igualdade social e proteção dos recursos ambientais (Mak, H., 2018). Assim são várias as áreas em que as pesquisas têm sido realizadas, visando o consumo eficiente de energia em todo o sistema logístico urbano, melhorar o meio ambiente da cidade alcançando melhor qualidade do ar e reduzindo o ruído e aumentar a satisfação do cliente por meio da entrega eficiente no last-mile.

O transporte urbano pode ficar congestionado facilmente devido ao tráfego pesado, o frete urbano é responsável por cerca de 25% das emissões de CO<sub>2</sub> e 30 a 50% de outros poluentes relacionados ao transporte (material particulado, óxido de nitrogênio). Além disso, o tráfego urbano é responsável por uma parte significativa do ruído ambiente (Rosana, M., 2020). Por isso nos últimos anos as entregas B2C são de 70 a 75% em ambiente urbano " (Ito, P. 2019), sendo

assim, o objetivo das pesquisas atuais deve ser aumentar a eficiência dos sistemas de logística e entrega dentro das áreas urbanas.

De acordo com Habibi, M.R., 2017 o transporte inteligente de pacotes em uma cidade é fundamental para a implementação eficiente de Smart Cities. A logística da cidade deve considerar o processo de otimização logística e as atividades de last-mile em uma área urbana levando em consideração três pilares: econômico, social e ambiental. Conseqüentemente, o planejamento de rotas é extremamente confuso, mentalmente desgastante e dependente de uma série de fatores. Eficiência de combustível, congestionamento de tráfego e regulamentações ambientais são considerações básicas que a empresa deve considerar com antecedência.

Com 86% dos clientes prontos para pagar por entregas mais rápidas (GRUCHMANN, T, 2019), ainda é uma tarefa desafiadora para as empresas executar as entregas de last-mile de maneira eficiente, considerando o aumento da demanda e as penalidades por atrasos nas entregas. Além disso a volatilidade da demanda torna a entrega last-mile extremamente complexa. Sendo assim, se torna importante conhecer a demanda futura do uso dos veículos para transporte de cargas e pessoas.

Para entender o comportamento do transporte de pessoas na etapa de last-mile nas grandes cidades, é possível estudar os padrões de demanda de viagens de táxi. Os táxis na cidade de Nova York realizam 172 milhões de viagens anualmente (11% de todas as viagens), o que torna os táxis um importante meio de transporte na cidade (Schaller, 2013)

Todos os táxis da cidade de Nova York são regulamentados pela *Taxi and Limousine Commission*, definindo regras de tarifas, embora os motoristas de táxi escolham onde circular para pegar passageiros. Para um planejamento e gerenciamento eficazes da frota de táxis, é necessário entender quais fatores impulsionam a demanda por táxis, como o uso de táxis está relacionado à disponibilidade de transporte público e como esses padrões variam no espaço e no tempo. Um modelo de geração de viagem que relaciona a demanda de táxi às características observáveis de um bairro (por exemplo, demografia, emprego e acessibilidade de trânsito) é útil para planejadores e formuladores de políticas para gerenciar serviços de táxi de forma eficaz, aumentando assim o nível de serviço e reduzindo custos para os taxistas.

### 2.3. PREVISÃO DE DEMANDA

O termo, previsão se refere ao ato de prever os eventos futuros. Atualmente, as previsões estão se tornando um desafio importante em vários campos, incluindo negócios, indústria, governo, economia, ciências ambientais, medicina, ciências sociais, política e finanças. Existem três classes de previsão, definidas em curto, médio e longo prazo. Os problemas de curto prazo envolvem previsão de eventos que estejam a apenas alguns períodos do futuro (dias, semanas ou meses). Já as previsões à médio prazo se estende de 1 a 2 anos, e as previsões de longo prazo, podem durar muitos anos (MONTGOMERY et al., 2015).

Se por um lado, as previsões de curto e médio prazo são necessárias em atividades produtivas, podendo ocorrer no gerenciamento das operações, as previsões de longo prazo normalmente impactam em questões de planejamento estratégico encaminhando decisões de orçamento e seleção de novos projetos e pesquisas. As previsões são baseadas na identificação, modelagem e extrapolação dos padrões comportamentais encontrados nos dados históricos, sendo comumente empregados métodos estatísticos (MONTGOMERY et al., 2015).

Dentre os tipos de previsão, tem-se a previsão de demanda, definida pelo processo de antever o nível futuro de procura por um determinado produto ou serviço. No contexto urbano de transporte individual, o entendimento dos padrões de demanda é essencial para compreender a movimentação dos potenciais usuários de transportes inteligentes em *Smart Cities*. Um modelo de previsão preciso (termo usado para avaliar a dispersão de resultados entre ensaios independentes repetidos de uma mesma amostra ou padrões em condições definidas) pode ajudar na alocação de recursos para identificar novas viagens e reduzir a frota de taxis circulando nas ruas, economizando energia e diminuindo o congestionamento de veículos nos grandes centros urbanos, responsável por afetar a vida cotidiana de milhões de pessoas (XIAN et al., 2019).

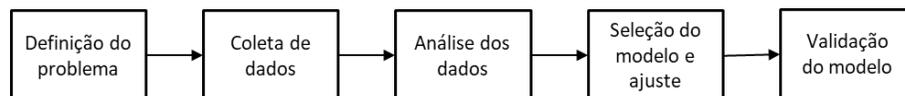
Existem diversos aspectos que dificultam a modelagem e previsão da demanda de uma matriz origem-destino (OD). Segundo Caceres et al. (2008), esse modelo de matriz é utilizado para quantificar a mobilidade associada a pessoas e mercadorias, indicando o número de viagens ocorridas entre cada zona de origem e destino em um determinado período de tempo, representando a demanda por transporte em uma região.

Por razão da dificuldade na modelagem a maioria dos métodos de previsões já existentes utilizam somente dados de fluxos de veículos em locais específicos desconsiderando características importantes da demanda como os dados de contagem que se correlacionam nos domínios espacial (ASHOK & BEN-AKIVA, 2002 e MOREIRA-MATIAS et al., 2013).

Além disso, modelar as correlações espaciais na rede de tráfego envolve a estimativa de um grande número de parâmetros correlacionados, podendo causar problemas computacionais e afetando a precisão da previsão. Portanto, para fornecer estimativas precisas e evitar o excesso de ajustes, é crucial incorporar o conhecimento físico da rede de tráfego e considerar a escassez da estrutura de correlação (XIAN et al., 2019).

A tradicional abordagem para prever a demanda da rede de veículos, especificamente o fluxo de táxis em centros urbanos, se refere aos métodos de previsão a partir de séries temporais. Entre eles os mais frequentes são o ARIMA, para os modelos puramente temporais e STARIMA, para modelos espaço-temporais. Em ambos os casos a análise de previsão de demanda é a mesma, e ocorre de acordo com o processo indicado na Figura 1:

Figura 1 - Processo de previsão



Fonte: Adaptado de MONTGOMERY et al., 2015

### 2.3.1. Aplicação de previsão de demanda na tecnologia dos veículos autônomos

O avanço inevitável da tecnologia dos veículos autônomos irá afetar as cidades e diversos aspectos da vida de todos os cidadãos, alterando de maneira significativa a infraestrutura de toda a cidade. Porém ainda serão enfrentadas inúmeras barreiras para a adoção plena dessa tecnologia. De acordo com, Bezai (2021) essas barreiras podem ser categorizadas em: (1) Perspectiva do usuário e do governo, considerando aceitação e comportamento do usuário, segurança e legislação; e (2) Tecnologias de informação e comunicação, que incluem, software e hardware, sistemas de comunicação, posicionamento e mapeamento precisos.

Visando compreender melhor o posicionamento e mapeamento dos AVs no last-mile, pensou-se na utilização de estudos que possam integrar o conhecimento já existente de padrões de movimentação, através da análise de dados, e ferramentas que prevejam padrões futuros para serem inseridos na tecnologia dos veículos autônomos, garantindo o conhecimento prévio do roteiro desses veículos, auxiliando em diversos fatores de tráfego. Dentre elas o estudo da previsão de demanda por viagens considerando uma matriz origem-destino.

Tendo em vista esse objetivo e a importância desse estudo, realizou-se um levantamento bibliográfico, para compreender melhor o andamento desses estudos pelos diversos pesquisadores da área. Utilizando o “Google Acadêmico” foram escolhidos alguns conjuntos de palavras chaves para a pesquisa. Inicialmente foi utilizado “*autonomous vehicles*” + “*urban logistics*”, percebendo que dentre os 201 artigos encontrados, 84 deles se referiam à transporte e entregas last-mile, e dentre esses artigos, somente 22 descreviam algoritmos com a função de otimizar as rotas dos veículos e nenhum deles propunha um modelo de previsão de demanda a fim de entender os padrões de movimentação dos veículos atuais para serem empregados nos veículos autônomos.

Após o resultado da primeira pesquisa, buscou-se por palavras chaves mais precisas e relacionadas ao tema em questão “*forecast*” + “*taxi trips*” + “*spatial-temporal*”, com o intuito de verificar quantos artigos seriam encontrados relacionando os 3 temas, ou seja, artigos que dissertassem sobre previsão de demanda espaço-temporal para viagens de táxi, e foram encontrados 142 artigos. Foi feito um levantamento para verificar quantos desses artigos utilizam o modelo STARMA para realizar suas previsões e de 94 artigos que realizam a construção de um modelo espaço temporal 12 são referentes ao modelo citado.

Depois foi analisado a quantidade de resultados obtidos quando unidas as palavras: “*forecast*” + “*taxi trips*” + “*autonomous vehicles*”, para verificar quantos dos artigos anteriores tinham relação com veículos autônomos. E foram obtidos 102 resultados, concluindo que 15 artigos tratavam realmente sobre demanda, sendo apenas 3 deles referente à previsão de demanda. Ou seja, quando os assuntos são tratados separadamente existem centenas de artigos que dissertam

sobre os dois temas (previsão de demanda e veículos autônomos), porém quando os assuntos são mesclados, ainda não existem muitas pesquisas sendo realizadas.

#### 2.4. SÉRIES TEMPORAIS

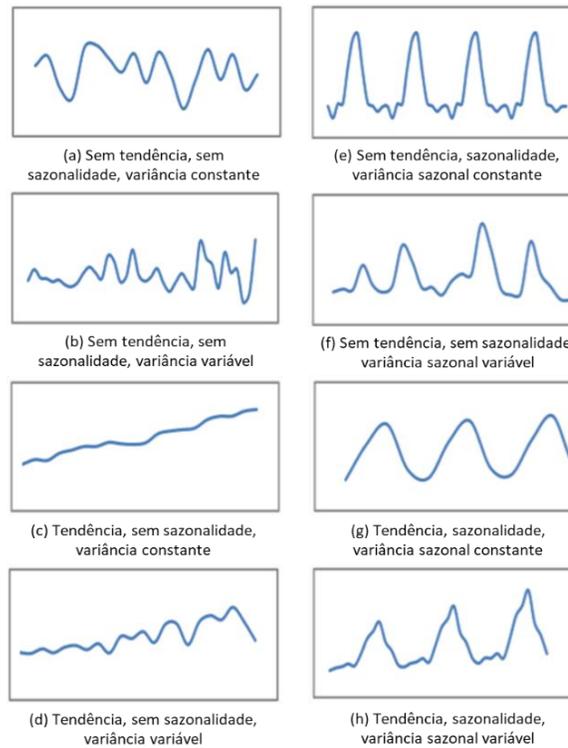
Série temporal é uma sequência de observações cronológicas ou temporalmente orientadas, de uma determinada variável de interesse, ou seja, se refere a qualquer conjunto de dados aleatórios organizados de acordo com a ordem em que são obtidas no tempo.

O objetivo principal da análise de séries temporais é o desenvolvimento de modelos matemáticos que forneçam descrições plausíveis dos dados amostrais. Os modelos utilizados para descrever séries temporais são considerados processos estocásticos, isto é, processos controlados por leis probabilísticas. Por tanto, um processo estocástico é uma família de variáveis aleatórias, formada por uma série temporal discreta,  $\{Z_t\} = Z_t, Z_{t-1}, Z_{t-2}, \dots$ , para cada  $t \in T$  períodos de tempo (SHUMWAY e STOFFER, 2000).

Uma das suposições mais frequentes que se faz a respeito a uma série temporal é de ser estacionária, quando as propriedades dos dados não são afetados pela mudança do tempo de origem, isto é, se o conjunto de distribuição probabilística das observações  $Z_t, Z_{t+1}, \dots, Z_{t+n}$  é exatamente igual ao conjunto  $Z_{t+k}, Z_{t+k+1}, \dots, Z_{t+k+n}$ . Estacionariedade implica em um tipo de equilíbrio estatístico, portanto a série temporal deve conter média e variância constante.

Porém a maior parte das séries encontradas na prática apresentam alguma forma de não estacionariedade, contendo tendência, quando a série flutua ao redor de uma reta com inclinação positiva ou negativa, e sazonalidade, quando os fenômenos apresentam um padrão cíclico (por exemplo, ano a ano ou semana a semana). Na Figura 2 são apresentados alguns modelos de séries temporais (MORETTIN, TOLOI 2004).

Figura 2 - Tipos de séries temporais



Fonte: Adaptado de DONKOR et al. (2014)

Portanto, quando os dados não forem estacionários, é necessário transformá-los. A transformação mais comum se refere a realizar diferenciações sucessivas da série original, até se obter uma série temporal estacionária, utilizando o operador de diferenciação,  $\nabla$ :

$$\nabla z_t = z_t - z_{t-1} = z_t^* \quad (2.1)$$

Outra maneira de escrever a operação de diferenciação será em termos do operador de translação ao passado (*backshift operator*),  $B$ .

$$\nabla z_t = (1 - B)z_t = z_t - z_{t-1} = z_t^* \quad (2.2)$$

Com  $\nabla = (1 - B)$  e  $B$  definido por  $Bz_t = z_{t-1}$  e estendido para potencializações,  $B^2 z_t = B(Bz_t) = Bz_{t-1} = z_{t-2}$ , sendo assim,  $B^d z_t = z_{t-d}$ .

Por tanto, em geral a diferenciação de ordem  $d$  não-sazonal para eliminar a tendência dos dados, pode ser definida como:

$$\nabla^d = (1 - B)^d \quad (2.3)$$

A diferenciação apresenta algumas vantagens em relação a outras maneiras de ajustes do modelo. Primeiro, por não requerer a estimação de nenhum parâmetro, se torna uma abordagem mais simples, e segundo, o ajuste do modelo pressupõe que a tendência seja fixa ao longo do histórico das séries temporais e permanecerá assim no futuro, portanto, o componente de tendência uma vez estimado é assumido como determinístico. Normalmente são requeridas uma ou duas diferenciações nos dados, a fim de remover a tendência.

Além disso, a diferenciação também é utilizada para eliminar a sazonalidade dos dados, definido para um operador sazonal de diferenciação  $d$  como

$$\nabla_d z_t = (1 - B^d) z_t = z_t - z_{t-d} \quad (2.4)$$

Por exemplo, se os dados forem mensais com sazonalidade anual, é comumente utilizado  $d = 12$ , portanto os dados com diferenciação sazonal serão

$$\nabla_{12} z_t = (1 - B^{12}) z_t = z_t - z_{t-12} \quad (2.5)$$

Quando tendência e sazonalidade estão presentes, pode-se diferenciar os dados sequencialmente para eliminar os dois efeitos, ou seja, a primeira diferença sazonal para remover o componente sazonal e depois utilizando o operador de diferenciação regular uma ou mais vezes para remover a tendência.

Informações importantes a respeito da estacionariedade dos dados podem ser obtidas ao plotar o diagrama de dispersão de todos os pares de dados  $z_t, z_{t+k}$ , separados pelo mesmo intervalo  $k$ , usualmente chamado de *lag*. Se não houver uma estrutura, que se assemelhe a uma reta, ou seja, se os dados estiverem totalmente dispersos, pode-se verificar que não há correlação entre eles. Mas caso haja, então pode-se concluir que existe correlação.

A covariância entre  $z_t$  e o seu valor em outro período,  $z_{t+k}$ , é chamado de autocovariância no *lag*  $k$ , definido por

$$\gamma_k = Cov(z_t, z_{t+k}) = E[(z_t - \mu)(z_{t+k} - \mu)] \quad (2.6)$$

Sendo  $\mu$  a média dos dados  $\{z_t\}$ , o conjunto de valores de  $\gamma_k$ , para  $k = 0, 1, 2, \dots$  é chamado de função de autocovariância. A autocovariância no *lag*  $k=0$  é a variância da própria série temporal, isto é,  $\gamma_k = \sigma_z^2$ , que é constante para séries temporais estacionárias. Já o coeficiente de autocorrelação no *lag*  $k$  é:

$$\rho_k = \frac{E[(z_t - \mu)(z_{t+k} - \mu)]}{\sqrt{E[(z_t - \mu)^2]E[(z_{t+k} - \mu)^2]}} = \frac{Cov(z_t, z_{t+k})}{Var(z_t)} = \frac{\gamma_k}{\gamma_0} \quad (2.7)$$

O conjunto de valores de  $\rho_k$ , para  $k = 0, 1, 2, \dots$  é chamado de função de autocorrelação (ACF). E por definição o valor de  $\rho_0 = 1$  e  $\rho_k = \rho_{-k}$ . Além disso, o ACF é adimensional.

É necessário estimar os valores de autocovariância e ACF para uma série finita. A estimação usual da função de autocovariância é expressa por

$$c_k = \hat{\gamma}_k = \frac{1}{T} \sum_{t=1}^{T-k} (z_t - \bar{z})(z_{t+k} - \bar{z}), \quad k = 0, 1, 2, \dots, K \quad (2.8)$$

onde  $\bar{z}$  se refere a média dos dados amostrais, e a ACF é estimada pela função de autocorrelação amostral

$$r_k = \hat{\rho}_k = \frac{c_k}{c_0}, \quad k = 0, 1, 2, \dots, K \quad (2.9)$$

## 2.5. MODELOS PARA SÉRIES TEMPORAIS

Modelo se refere a uma representação ou interpretação da realidade. Os modelos utilizados para descrever séries temporais são processos estocásticos, esses modelos podem ser classificados em duas classes, segundo o número de parâmetros envolvidos. O método paramétrico, para uma quantidade finita de parâmetros e o método não-paramétrico para infinitos parâmetros.

Na classe de modelos paramétricos, a análise é feita no domínio do tempo. Dentre eles os mais frequentes são os modelos de regressão, os autorregressivos integrado de média móvel (ARIMA) e modelos espaciais autorregressivos integrados de média móvel (STARIMA), os estruturais e os

não-lineares. Já a classe de modelos não-paramétricos os mais utilizados são a função de auto covariância e sua transformada de Fourier (MORETTIN, 2004).

### 2.5.1. Modelo ARIMA e suas variações

A Autorregressão Integrada e de Média Móvel (ARIMA – *Autorregressive Integreted Moving Average*) é uma metodologia bastante conhecida (BOX e JENKINS, 1967), cujo objetivo é modelar e prever dados de séries temporais univariadas, como dados de fluxo de tráfego e outros problemas de previsão de curto prazo. A vantagem do ARIMA em relação a outros algoritmos é a sua versatilidade para representar diferentes tipos de séries temporais: autorregressivas (AR), média móvel (MA), a combinação das duas (ARMA), e até mesmo a modelagem de séries temporais não-estacionárias.

#### 2.5.1.1. Modelo Autorregressivo de Média Móvel

Esse modelo representa a união dos modelos autorregressivos e de média móvel e é utilizado para séries temporais estacionárias. Abreviado por ARMA (p, q) e representador por

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \dots + \phi_p z_{t-p} + a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \dots + \theta_q a_{t-q} \quad (2.10)$$

Onde,  $z_t$  é estacionário,  $a_t$  é uma sequência de valores aleatórios que seguem uma distribuição normal, com média zero e variância comum  $\sigma_a^2$ , conhecido como *white noise* ou inovações. Os parâmetros p e q são chamados de ordem autorregressiva e de média móvel, respectivamente, em que  $\phi_1, \phi_2, \dots, \phi_p$  e  $\theta_1, \theta_2, \dots, \theta_q$  são parâmetros constantes e diferentes de zero. (MONTGOMERY, 2015).

Normalmente o modelo ARMA (p, q) é representado utilizando o operador autorregressivo ( $\phi(B) = 1 - \phi_1 B^1 - \dots - \phi_p B^p$ ) e de média móvel ( $\theta(B) = 1 - \theta_1 B^1 - \dots - \theta_q B^q$ ), portanto, pode ser escrito como:

$$\phi(B)z_t = \theta(B)a_t \quad (2.11)$$

#### 2.5.1.2. Modelo Autorregressivo Integrado de Média Móvel (ARIMA)

Quando a série temporal não for estacionária será necessário utilizar a modelagem ARIMA que contém o termo de diferenciação, discutido na Seção 3.2.

De acordo com Box & Pierce (1970) o modelo autorregressivo integrado de média móvel, ou modelo de Box-Jenkins, com ordem  $p$  (autorregressiva),  $d$  (diferenciação) e  $q$  (média móvel), é abreviado como ARIMA ( $p, d, q$ ) e expresso por

$$\phi(B)\nabla^d z_t = \theta(B)a_t \quad (2.12)$$

## 2.6. MODELOS PARA DADOS COM CORRELAÇÃO ESPAÇO-TEMPORAL

A família de modelos ARIMA, apresentados na Seção 2.4. faz parte de uma classe de modelos empíricos e flexíveis que seguem o procedimento conhecido como Box-Jenkins, esses modelos são muito úteis para o espectro das análises estatísticas (PFEIFER & DEUTSCH, 1980). Se tratam de modelos univariados, aplicáveis somente a dados que apresentem uma série temporal única. Por mais que estatísticas frequentemente estão disponíveis para regiões no espaço, modelos univariados podem lidar somente com o histórico de uma região em particular ou com dados de diversas regiões agregados formando uma única região.

Uma alternativa para modelagem de séries temporais univariadas, são os modelos multivariados (GRANGER & NEWBOLD, 1977). Esses modelos descrevem e preveem simultaneamente um conjunto de  $N$  observações. Quando essas  $N$  séries representam padrões das  $N$  regiões, as interrelações entre as diferentes regiões devem ser consideradas, e assim, resultar em uma descrição melhor do sistema.

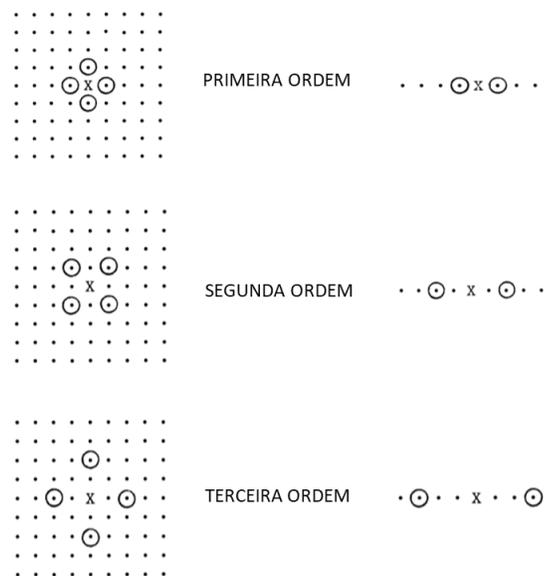
Além disso, os modelos de séries temporais multivariadas podem ser mais precisos se o sistema for modelado exibindo dependência sistemática entre a observação de cada região e as observações vizinhas a esta, ou seja, se a presença de algum parâmetro em um determinado local torna mais provável sua presença em locais vizinhos. Esse fenômeno é conhecido por autocorrelação espacial (CLIFF E ORD 1975).

Portanto, a extensão dos modelos de séries temporais univariadas ARIMA no domínio espacial resulta em um modelo de classe geral conhecido como STARIMA (*Space-time Autorregressive Integrated Moving Average*) (CLIFF & ORD, 1975), que foi abordado na Seção 2.5.2.

### 2.6.1. Matriz de vizinhança e variograma

A matriz,  $W_l$  é uma matriz  $N \times N$ , cuja função é incorporar um termo que descreva a relação de vizinhança na modelagem espacial, se trata de uma matriz padronizada, ou seja, as linhas devem estar padronizadas para que a soma seja igual a 1, interpretado como média da vizinhança. Quando  $l=0$  a matriz de vizinhança torna-se neutra,  $W_0 = I$ , por isso os termos  $\phi_{k0}$  e  $\theta_{k0}$  serão puramente temporais. Ela apresenta elementos não-nulos somente nos pares de locais que foram definidos como vizinhos de  $l$ -ésima ordem. Os vizinhos de primeira ordem devem estar mais próximos que os de segunda ordem, que por sua vez devem estar mais próximos que aqueles de terceira ordem e assim sucessivamente (Figura 3). Os termos da diagonal são iguais a zero, para não existir uma relação de vizinhança com o próprio elemento em análise. A tarefa de especificar a qual ordem cada vizinho pertence, é do construtor do modelo, a fim de capturar da melhor maneira as propriedades físicas do sistema espacial de interesse (PFEIFER & BODILY, 1990).

Figura 3 - Ordem espacial em sistemas de uma e duas dimensões



Fonte: Adaptado de PFEIFER & DEUTSCH (1980a)

Os elementos  $w_{ij}$  indicam a dependência na vizinhança entre o  $i$ -ésimo e o  $j$ -ésimo local:

$$\sum_{j=1}^N w_{ij} = 1, \quad i = 1, \dots, N$$

Para a construção da matriz de vizinhança pode-se utilizar alguns tipos de critérios, (I) a formação matriz de vizinhança através da adjacência dos pontos. (II) Outra maneira comumente utilizada é pela relação de distância entre os pontos. Nesse caso a matriz de ordem 1 terá valores iguais a 1 para regiões que estão até uma determinada distância e zero para aqueles que ultrapassarem, a matriz de ordem 2 parte dos elementos cuja distância é superior à de ordem anterior até o limite de interesse. (III) Uma outra proposta é considerar o inverso da distância entre os elementos, se estiverem próximos a relação será perto de 1, mas conforme a distância aumenta a relação entre o par de elementos tende a zero. Desse método surgem outros como o inverso da distância ao quadrado ou até mesmo a adoção de outras potências. A finalidade é que as regiões mais próximas tenham uma influência maior que aquelas consideradas distantes.

Nos estudos com análise espacial utiliza-se frequentemente a análise do variograma, responsável por indicar se há relação espacial entre os dados. Para sua construção, admite-se que a variância depende da distância entre os pontos. Sendo uma posição no espaço e  $s+h$  uma nova posição cuja distância a partir de  $s$  resulta em  $h$ . Cressie (1993) indica duas condições de estacionariedade a partir da primeira diferença de  $Z_{s+h} - Z_s$ , com  $E(Z_s) = \mu$  para qualquer posição espacial  $s$ :

$$E(Z_{s+h} - Z_s) = 0 \text{ e } Var(Z_{s+h} - Z_s) = 2\gamma(h)$$

O valor de  $2\gamma(h)$  é chamado variograma e  $\gamma(h)$  é o semivariograma. Na prática consiste em calcular a matriz de distância dos pontos, em seguida são separadas as faixas de interesse do estudo, e por fim, é obtida a correlação entre os pontos de cada faixa estabelecida previamente. No total existem  $N$  localidades, e, portanto,  $N_h$  localidades quando são agrupadas em faixas de distâncias até  $h$ . O estimador mais usual da variância é:

$$2\gamma(h) = \frac{1}{N_h} \sum_{s=1}^{N_h} (Z_{s+h} - Z_s)^2$$

Gráficos com comportamento crescente indicam a correlação espacial dos dados, porém quando o gráfico é formado por uma linha paralela ao eixo das distâncias não há dependência.

### 2.6.2. Modelo STARMA e STARIMA

O modelo STARMA (*Spatial-time Autorregressive Moving Average*) é caracterizado pela dependência linear no *lag* espacial e temporal, utilizado para séries temporais estacionárias, foi estudado por Cliff e Ord (1975) e Pfeifer e Deutsch (1980). Assumindo que  $z_t$  representa o vetor de observações  $N \times 1$ , nas  $N$  localizações fixas no espaço ao longo de  $T$  períodos de tempo, o modelo STARMA pode ser expresso por

$$z_t = \sum_{k=1}^p \sum_{l=0}^{\lambda_k} \phi_{kl} W_l z_{t-k} - \sum_{k=1}^q \sum_{l=0}^{m_k} \theta_{kl} W_l a_{t-k} + a_t \quad (2.13)$$

onde  $p$  é a ordem autorregressiva,  $q$  a ordem da média móvel,  $\lambda_k$  representa a ordem espacial do  $k$ -ésimo termo autorregressivo,  $m_k$  a ordem espacial do  $k$ -ésimo termo de média móvel,  $\phi_{kl}$  e  $\theta_{kl}$  são os parâmetros dos termos autorregressivos e de média móvel, respectivamente,  $W_l$  é a matriz de vizinhança, e  $a_t$  representa o erro (as inovações) que seguem uma distribuição normal multivariada com média zero e matriz de variância-covariância  $G = \sigma^2 I_N$ , onde  $I_N$  é a matriz identidade com dimensão  $N \times N$ .

O modelo espaço-temporal autorregressivo de média móvel, é comumente representado por STARMA ( $p_{\lambda_1, \lambda_2, \dots, \lambda_p}, q_{m_1, m_2, \dots, m_q}$ ). Duas subclasses do modelo STARMA podem ser notadas, quando  $q = 0$ , haverá somente o termo autorregressivo, se tornando um modelo espaço-temporal autorregressivo (STAR). Já os modelos que não contém o termo autorregressivo ( $p=0$ ) são referidos como STMA, modelo espaço-temporal de média móvel (PFEIFER & BODILY, 1990).

Uma forma de visualizar o modelo STARMA é a partir da extensão do modelo univariado ARMA em dimensões espaciais. Para cada uma das  $N$  séries temporais, o modelo STARMA representa a observação atual como uma combinação linear das observações passadas e das inovações (equivalente ao ARMA) assim como combinações ponderadas das observações e inovações dos locais vizinhos. São esses termos espaciais que unem as  $N$  séries e diferenciam o modelo STARMA de uma sequência de  $N$  modelos separados univariados ARMA ((PFEIFER & DEUTSCH, 1980b).

Quando os dados a serem analisados se tratarem de uma série temporal não estacionária, será necessário acrescentar o termo de diferenciação no modelo (assim como no modelo ARIMA, visto

na Seção 3.3.1.) Sendo assim, necessário o uso do modelo espaço temporal autorregressivo integrado de média móvel (STARIMA), definido por

$$\nabla^d z(t) = \sum_{k=1}^p \sum_{l=0}^{\lambda_k} \phi_{kl} W_l \nabla^d z_{t-k} - \sum_{k=1}^q \sum_{l=0}^{m_k} \theta_{kl} W_l a_{t-k} + a_t \quad (2.14)$$

onde  $\nabla$  se refere a matriz do operador de diferenciação, sendo  $\nabla z_t = z_t - z_{t-1}$  e  $\nabla^2 z_t = z_t - 2z_{t-1} + z_{t-2}$ . E  $d$  é o número de diferenciações. Nesse caso, quando  $d=0$ , o modelo STARIMA, volta a ser uma modelo STARMA (PFEIFER & DEUTSCH, 1980a).

### 2.6.3. Função espaço-temporal de autocorrelação e autocorrelação parcial

Nos modelos de séries temporais são utilizadas as funções de autocorreção ACF e autocorrelação parcial PACF. Para os modelos espaço-temporais existem as funções de autocorrelação espaço-temporal (STACF - Spatio-time Autocorrelation Function) e de autocorrelação espaço-temporal parcial (STPACF - Spatio-time Partial Autocorrelation Function). De acordo com Morettin e Tolo (2006), autocorrelação é a correlação entre os valores de uma série temporal em diferentes momentos no tempo. E a autocorrelação parcial de uma série de tempo, indicada por  $\phi_{k,k}$ , é definida como a correlação restante entre  $z_t$  e  $z_{t+k}$  após a retirada da influência das variáveis intermediárias.

Segundo Pfeifer e Deutsch (1980a) a covariância entre os *lags* espaciais  $r$  e  $s$  no *lag* temporal  $k$  é expresso por

$$\gamma_{r,s}^k = E \{ [W_r Z_t]' [W_s Z_{t+k}] \} \quad (2.15)$$

O cálculo ainda pode ser expresso de forma simplificada, seja  $\Gamma(k) = E [Z_t Z_{t+k}']$ , então a covariância assume

$$\gamma_{r,s}^k = tr \{ W_r' W_s \Gamma(k) \} \quad (2.16)$$

A estimativa amostral para  $\hat{\gamma}_{r,s}^k$  é obtida trocando-se  $\Gamma(k)$  por  $\hat{\Gamma}(k)$

$$\hat{\Gamma}(k) = \sum_{t=1}^{T-k} \frac{(Z_t - \bar{Z})(Z_{t+k} - \bar{Z})'}{T - k} \quad (2.17)$$

Sendo  $\bar{Z}$  um vetor de tamanho  $N \times 1$  e cada elemento igual à média sobre todas as localidades e tempo. Com isso, autocorrelação fica expressa como

$$\rho_{r,s}^k = \frac{\gamma_{r,s}^k}{\sqrt{\gamma_{r,r}^0 \gamma_{s,s}^0}} \quad (2.18)$$

E a variância da autocorrelação é

$$\hat{V}ar(\rho_{r,0}^k) \approx \frac{1}{N(T-k)} \quad (2.19)$$

Já a autocorrelação parcial é obtida do modelo STAR

$$z_t = \sum_{k=1}^p \sum_{l=0}^{\lambda_k} \phi_{kl} W_l z_{t-k} + a_t \quad (2.20)$$

Multiplicando a Equação 2.40 por  $[W_h Z_{t-u}]'$ , com  $u$  sendo um *lag* temporal e  $h$  uma ordem espacial

$$Z'_{t-u} W'_h z_t = \sum_{k=1}^p \sum_{l=0}^{\lambda_k} \phi_{kl} Z'_{t-u} W'_h W_l z_{t-k} + Z'_{t-u} W'_h a_t \quad (2.21)$$

As autocovariâncias são derivadas da Expressão 3.23 para o cálculo da expectativa  $E$

$$E(Z'_{t-u} W'_h z_t) = \sum_{k=1}^p \sum_{l=0}^{\lambda_k} \phi_{kl} E(Z'_{t-u} W'_h W_l z_{t-k}) + E(Z'_{t-u} W'_h a_t) \quad (2.22)$$

Da Equação 2.36 é possível obter

$$E(Z'_{t-u} W'_h z_t) = \gamma_{h,0}^u \quad (2.23)$$

$$E(Z'_{t-u} W'_h W_l z_{t-k}) = \gamma_{h,l}^{u-k} \quad (2.24)$$

$$E(Z'_{t-u} W'_h a_t) = 0 \quad (2.25)$$

Reescrevendo a Equação 2.43

$$\gamma_{h,0}^u = \sum_{k=1}^p \sum_{l=0}^{\lambda_k} \phi_{kl} \gamma_{h,l}^{u-k} \quad (2.26)$$

A equação acima compõe o sistema de Yule-Walker, utilizado para estimar os parâmetros  $\phi_{kl}$  que formam a autocorrelação parcial.

As estruturas STPACF e STACF foram utilizadas repetidas vezes no processo de modelagem até que não exista mais correlação residual nas componentes de espaço e tempo. Pfeifer e Deutsch (1980) apresentam a Tabela 2 para identificação dos modelos de acordo com o comportamento das correlações

Tabela 2 - Identificação do Modelo STARMA e suas variações

	STAR(p)	STMA(q)	STARMA(p,q)
STACF	Decai gradualmente	Zero após o do lag q	Decai gradualmente
STPACF	Zero após o lag p	Decai gradualmente	Decai gradualmente

Fonte: Adaptado de PFEIFER E DEUTSCH (1980)

### 3. METODOLOGIA

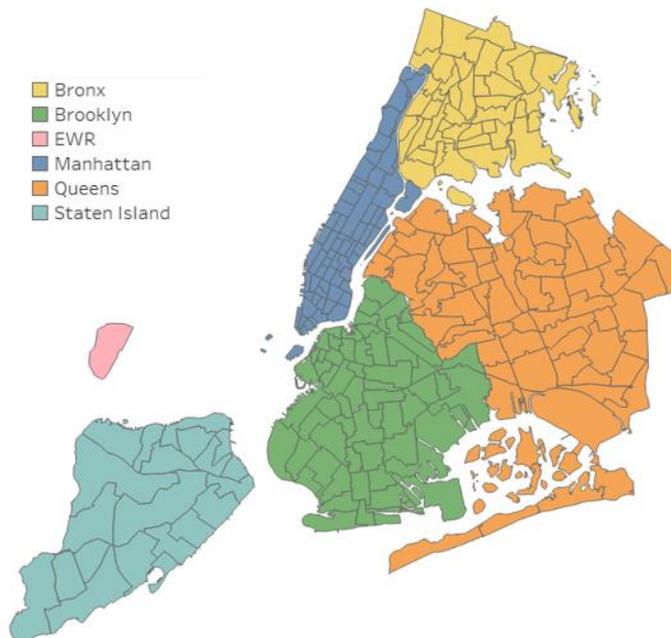
A proposta deste trabalho foi prever a demanda espaço-temporal do número de viagens origem-destino (O-D) na etapa de *last-mile*, utilizando séries temporais. A previsão de demanda de viagens de taxi O-D são muito importantes para ajudar na alocação dinâmica de recursos, reduzindo táxis vazios nas ruas que gastam energia e contribuem para piorar o congestionamento nas grandes cidades.

A primeira etapa do trabalho foi reunir dados confiáveis que representassem o problema em questão. Tendo como ponto de partida esses dados, o intuito foi ser capaz de criar um modelo que o melhor represente, podendo prever a demanda futura de ocorrências de forma precisa.

Os dados utilizados se referem aos dados de táxi da cidade de Nova Iorque, a cidade mais populosa dos Estados Unidos, contendo aproximadamente 8,5 milhões de pessoas, com densidade populacional de cerca de 27 mil pessoas por milha quadrada (*NYC Open Data, 2020*). A cidade de Nova Iorque é dividida em cinco *boroughs*, equivalente às grandes regiões da cidade. Além disso, os *boroughs* são divididos em 263 zonas de táxi diferentes, que se referem a *conjuntos* de endereços definidos pela TLC (*Taxi and Limousine Commission*) para demarcar as zonas de *pick-up* e *drop-off* das corridas de táxi, baseado nos bairros da cidade.

Na Figura 4 pode-se observar a divisão das zonas de táxi, agrupadas por *borough* em Nova Iorque, sendo eles Bronx, Brooklyn, Manhattan, Queens e Staten Island. O Aeroporto Internacional de Newark Liberty (EWR) apresenta uma denominação específica, pois não está contido em nenhum *borough*.

Figura 4 - Mapa da cidade de Nova Iorque segmentado pelas zonas de táxi e *boroughs*



Fonte: Elaborado pelo autor

Os dados, dos chamados *yellow taxis*, foram coletados e disponibilizados pela TLC através de tecnologias fornecidas pela *Taxicab & Livery Passenger Enhancement Programs (TPEP/LPEP)*. Eles incluem data/horário e localização do *pick-up* e *drop-off*, distância das viagens, tipos de pagamento, quantidade de passageiros, tarifas discriminadas e tipos de taxas. Os dados estão disponibilizados mês a mês, desde o ano de 2009. Em 2019 houve aproximadamente 84,4 milhões de viagens de táxi na cidade de Nova Iorque, a Figura 5 mostra um gráfico com a quantidade de viagens OD realizadas por mês.

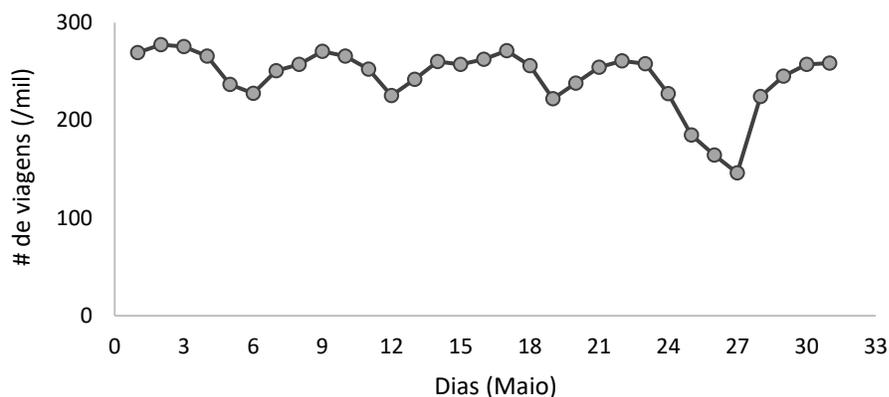
Figura 5 - Gráfico da quantidade de pick-up ocorrido por mês em 2019



Fonte: Elaborado pelo autor

O método proposto foi aplicado aos dados de maio de 2019 (com aproximadamente 7,5 milhões de viagens OD), na Figura 6 é indicado a quantidade de viagens realizadas por dia.

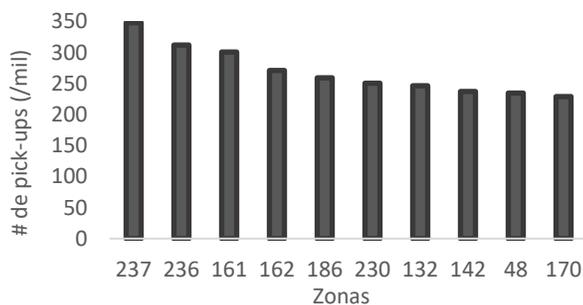
Figura 6 - Gráfico da quantidade de pick-up ocorrido por dia no mês de maio/2019



Fonte: Elaborado pelo autor

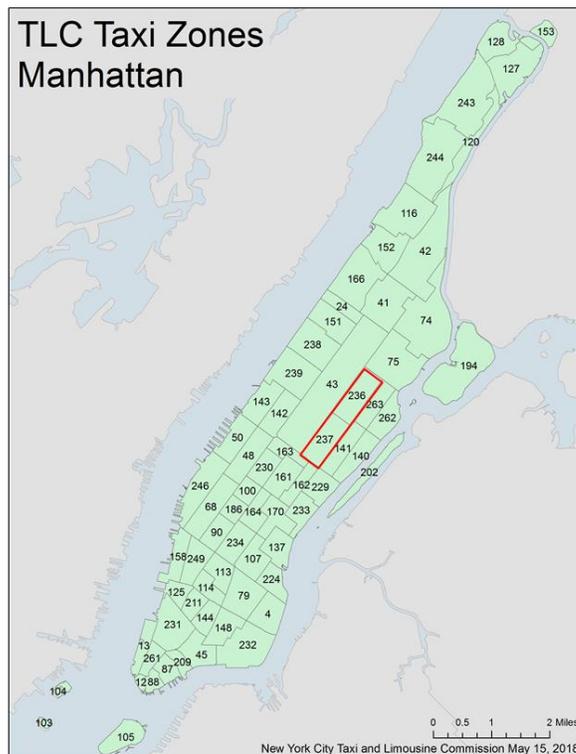
Foi selecionado o par origem-destino que continha a maior quantidade de viagens realizadas, para que posteriormente fosse possível admitir a vizinhança dessas zonas. A Figura 7 mostra o gráfico com as 10 zonas que apresentam maior quantidade de *pick-ups*. Sendo assim, foi escolhido as duas zonas com maior quantidade de viagens, a zona 237 e 236 com 347.987 e 312.153 *pick-ups*, respectivamente. Além disso, na Figura 8 é possível verificar que essas zonas estão localizadas em Manhattan, ao lado do *Central Park*.

Figura 7 - Gráfico da quantidade de *pick-ups(a)* ocorridos no mês de maio por zona



Fonte: Elaborado pelo autor

Figura 8 - Zonas de táxi em Manhattan



Fonte: TLC, 2018

Após verificar as zonas com maior quantidade de viagens, foi estipulado quais zonas seriam consideradas vizinhas à 237 e 236. Para esse estudo, adotou-se como vizinhança as zonas: 43, 75, 141, 162, 163 e 263. O método de escolha foi a utilização de zonas que tangenciassem as zonas 237 e 236. Depois de escolher às zonas foi verificado a distância euclidiana entre cada par de zonas. Encontrando assim, uma matriz de distâncias (Tabela 3).

Tabela 3 - Matriz de Distâncias

	<b>43</b>	<b>75</b>	<b>141</b>	<b>162</b>	<b>163</b>	<b>236</b>	<b>237</b>	<b>263</b>
<b>43</b>	0,00	1,88	1,88	2,95	1,07	0,80	1,61	1,34
<b>75</b>	1,88	0,00	1,34	4,29	3,76	1,34	2,95	1,34
<b>141</b>	1,88	1,34	0,00	1,61	1,61	1,61	0,54	1,61
<b>162</b>	2,95	4,29	1,61	0,00	1,07	2,95	1,34	2,95
<b>163</b>	1,07	3,76	1,61	1,07	0,00	2,41	1,07	2,68
<b>236</b>	0,80	1,34	1,61	2,95	2,41	0,00	1,61	0,54
<b>237</b>	1,61	2,95	0,54	1,34	1,07	1,61	0,00	1,61
<b>263</b>	1,34	1,34	1,61	2,95	2,68	0,54	1,61	0,00

Fonte: Elaborado pelo autor

Após selecionar as zonas de *pick-up* e *drop-off*, o banco de dados foi organizado, como indicado na Tabela 4, as colunas referentes ao *pickup\_datetime*, *dropoff\_datetime*, *PULocationID* e *DOLocationID*, fazem parte do banco de dados da *TLC*, e representam, respectivamente, o dia e horário que ocorreu o *pick-up*, o dia e horário de *drop-off*, a zona em que ocorreu o *pick-up* e a zona de *drop-off*. Já as colunas de *ID*, *DIA* e *HORA*, foram criadas pelo autor, com a intenção de numerar as viagens (*ID*) e ter acesso a quantidade de viagens que ocorreram por dia e por hora, respectivamente.

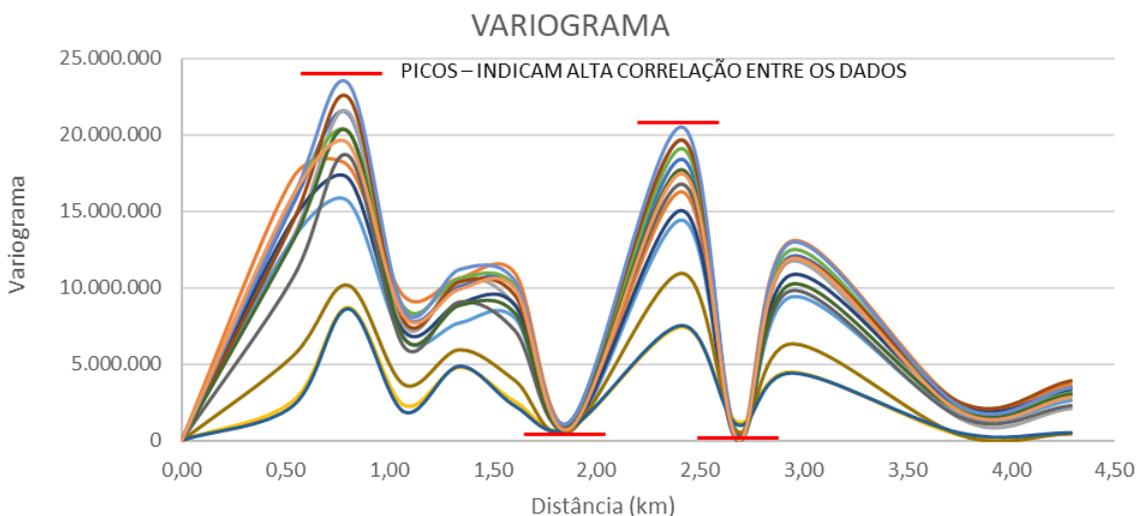
Tabela 4 - Banco de dados

ID	pickup_datetime	dropoff_datetime	PULocationID	DOLocationID	DIA	HORA
1	01/05/2019 00:06	01/05/2019 00:12	237	236	1	0
2	01/05/2019 00:04	01/05/2019 00:08	43	163	1	0
3	01/05/2019 00:28	01/05/2019 00:30	162	75	1	0
4	01/05/2019 00:42	01/05/2019 00:43	141	236	1	0
5	01/05/2019 00:15	01/05/2019 00:18	75	237	1	0

Fonte: Elaborado pelo autor

Foi realizado um estudo do variograma para verificar e analisar a presença de correlação entre os dados e quais os vizinhos que exerceriam maior influência entre si afim de ponderar a matriz de vizinhança, a análise foi feita, assim como mencionado na Seção 2.6.1. O gráfico da Figura 9 mostra o variograma dos dados entre os dias 01 e 15 de maio de 2019 (cada dia do mês está indicado com um cor diferente), em faixas de 0,5km de distância. É possível notar que algumas curvas apresentam valores menores, representando uma quantidade menor de viagem aos finais de semana, porém o comportamento é o mesmo. Além disso, é possível notar que nas faixas de distância de 1,5 a 2,0 e 2,5 a 3,0 km a correlação pode ser desconsiderada, já que os valores são baixos. E a partir de 4,5km a tendência é de que também não haja correlação entre os dados. Foi ocultado no gráfico os dados entre os dias 16 e 31 de maio, pois apresentam o mesmo padrão da primeira quinzena do mês.

Figura 9 - Variograma



Fonte: Elaborado pelo autor

A partir da análise do variograma foi possível criar a matriz de vizinhança. Como critério, foi adotado uma ponderação em relação a distância entre os pares de vizinhos e a correlação entre eles. Ou seja, verificou-se qual a distância entre o par de zonas e relacionou-se com o valor do variograma. Sendo assim, vizinhos que estavam a distâncias de 0 à 1,5 e 2,0 à 2,5 obtiveram pesos maiores na matriz de ponderação da vizinhança. Foi utilizado uma única matriz de ponderação de vizinhança,  $W_1$  (Tabela 5), relacionando todos os vizinhos.

Tabela 5 - Matriz de Ponderação da Vizinhança  $W_1$ 

	<b>43</b>	<b>75</b>	<b>141</b>	<b>162</b>	<b>163</b>	<b>236</b>	<b>237</b>	<b>263</b>
<b>43</b>	0,000	0,019	0,019	0,191	0,125	0,330	0,146	0,169
<b>75</b>	0,024	0,000	0,213	0,058	0,039	0,213	0,240	0,213
<b>141</b>	0,020	0,170	0,000	0,147	0,147	0,147	0,221	0,147
<b>162</b>	0,180	0,044	0,138	0,000	0,118	0,180	0,159	0,180
<b>163</b>	0,145	0,035	0,169	0,145	0,000	0,351	0,145	0,009
<b>236</b>	0,219	0,112	0,097	0,127	0,202	0,000	0,097	0,146
<b>237</b>	0,128	0,167	0,192	0,148	0,110	0,128	0,000	0,128
<b>263</b>	0,161	0,161	0,139	0,182	0,008	0,209	0,139	0,000

Fonte: Elaborado pelo autor

É importante lembrar também que não se considera relação de vizinhança de uma zona com ela própria, sendo assim a diagonal principal da matriz W, é igual a zero.

Depois de realizar esse estudo, foi possível criar uma matriz que indicasse a quantidade de viagens ocorridas nos dias do mês de maio a cada hora (Figura 9). Onde é possível perceber a forte sazonalidade ao longo dos dias e ao longo da semana, os dias úteis apresentam maior quantidade de viagens quando comparado aos finais de semana, além disso as viagens se concentram no período das 7 às 19hs todos os dias.

Figura 10 - Matriz da quantidade de viagens

HORA	DIA																														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
0	10	11	10	21	24	3	13	9	11	13	19	22	6	6	11	11	15	18	20	8	6	14	8	11	7	3	1	2	5	4	13
1	4	2	4	10	11	1	8	0	1	4	6	15	0	3	2	3	10	2	15	1	2	3	5	2	4	4	0	2	2	2	4
2	2	1	3	6	4	0	4	1	1	2	2	1	0	1	1	2	1	7	2	0	2	1	1	1	2	0	2	1	1	1	3
3	0	0	1	1	3	0	1	1	0	1	2	1	0	1	1	1	1	0	1	1	2	1	3	3	0	3	0	0	0	3	0
4	2	0	1	0	3	2	1	2	1	0	4	0	0	0	0	1	0	2	0	1	0	4	1	0	0	1	0	0	1	0	0
5	5	7	11	3	0	2	4	6	7	2	2	2	4	6	8	9	7	1	0	5	8	4	3	0	0	0	0	2	6	8	3
6	14	22	19	9	2	12	17	22	18	12	8	1	17	15	13	18	21	7	5	15	20	16	26	11	2	2	7	16	13	15	16
7	132	129	113	30	18	136	121	137	117	107	19	14	114	112	117	115	97	30	7	92	109	115	101	78	10	8	6	107	107	118	106
8	127	114	138	32	26	101	121	112	129	132	41	28	116	129	108	153	117	45	32	130	138	121	145	106	17	8	17	126	139	119	122
9	102	120	97	70	65	92	108	109	103	113	56	48	104	103	112	111	117	74	61	98	105	112	113	73	30	22	25	110	97	98	87
10	90	107	116	130	87	106	108	107	120	111	94	95	101	109	115	110	105	75	64	89	108	105	106	84	33	28	24	114	91	105	90
11	115	108	141	118	88	90	86	116	124	112	77	72	111	93	121	126	113	76	60	93	115	130	116	106	27	25	29	121	102	114	118
12	106	115	136	107	90	123	119	128	125	120	84	85	118	129	141	134	142	85	75	121	114	127	117	102	49	39	39	123	119	119	104
13	121	103	137	93	111	111	119	114	109	125	79	106	121	123	128	126	119	91	84	111	110	136	141	100	35	41	28	114	131	154	116
14	147	201	166	103	92	160	187	153	158	147	131	107	169	197	186	156	128	104	74	160	151	164	143	113	48	45	49	155	163	148	168
15	138	171	145	117	82	146	161	137	144	128	126	90	117	160	178	151	128	107	95	179	173	177	148	103	64	44	51	156	170	176	141
16	138	155	128	122	89	135	153	147	160	122	132	82	133	153	158	199	161	107	87	160	156	155	132	85	43	34	43	137	153	131	119
17	171	188	157	128	92	150	163	135	154	132	88	75	118	155	155	153	119	85	70	127	134	164	144	73	45	44	59	133	167	162	133
18	165	130	125	69	87	130	151	158	164	139	91	65	125	149	174	138	137	85	94	167	188	173	131	63	31	33	52	132	141	126	127
19	142	156	114	75	62	131	135	140	125	87	63	63	92	147	137	121	84	94	65	110	132	168	87	54	33	27	42	115	129	107	88
20	94	126	83	72	40	94	121	84	119	82	71	55	82	102	105	112	74	60	39	78	117	80	65	59	33	19	29	94	83	102	79
21	105	106	88	53	32	85	112	106	83	107	61	26	89	113	92	119	103	66	34	102	100	113	72	50	28	12	26	104	107	74	79
22	65	68	70	64	15	41	73	73	77	50	57	20	42	48	61	82	68	62	23	48	69	65	42	29	19	13	7	35	50	64	48
23	30	43	34	44	8	13	33	37	23	39	38	8	20	22	29	43	30	37	11	15	13	25	31	17	13	7	8	20	20	39	35

Fonte: Elaborado pelo autor

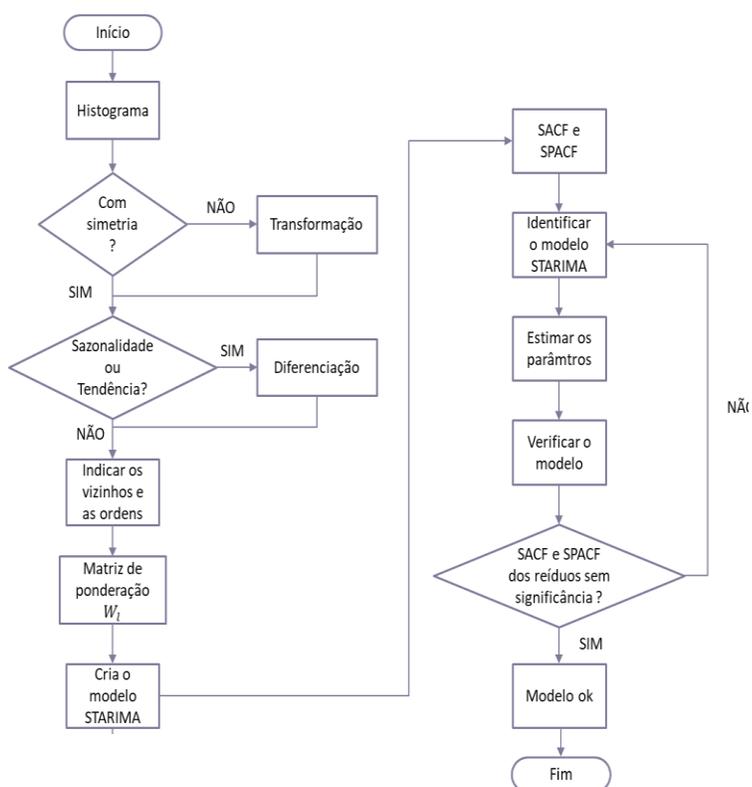
Depois de gerar os dados, eles foram organizados em forma matricial H x N, em que H representa a quantidade de horas dentro do mês, sendo  $H = 24 \times 31 = 744$  horas, portanto, o dia 1 foi indicado pela hora 0 até a hora 23, o dia 2 estará entre 24 e 48 horas e assim por diante e N a quantidade de zonas que foram analisadas. Com os dados já separados é possível iniciar a modelagem do problema, que foi estruturada no próximo tópico.

### 3.1. MODELO STARMA

A estratégia para a construção do modelo foi baseada nos conceitos apresentados por CLIFF e ORD (1975) e PFEIFER e DEUTSCH (1980), indicados na Seção 2.6.2. A escolha da estrutura do modelo é baseada nos próprios dados. Para a modelagem, são considerados três estágios, são eles: identificação de um modelo, com base na análise de autocorrelações e autocorrelações parciais espaço-temporais; estimação, na qual os parâmetros do modelo identificado são estimados; e finalmente a verificação ou diagnóstico do modelo ajustado, através da análise dos resíduos e do erro gerado pela previsão.

Para melhor compreensão da metodologia que foi seguida para a construção do modelo STARMA foi desenvolvido o Fluxograma 1, indicando desde a análise inicial dos dados para escolha do modelo até as etapas de modelagem do modelo.

Fluxograma 1 – Modelo Conceitual STARMA



### 3.1.1. Identificação

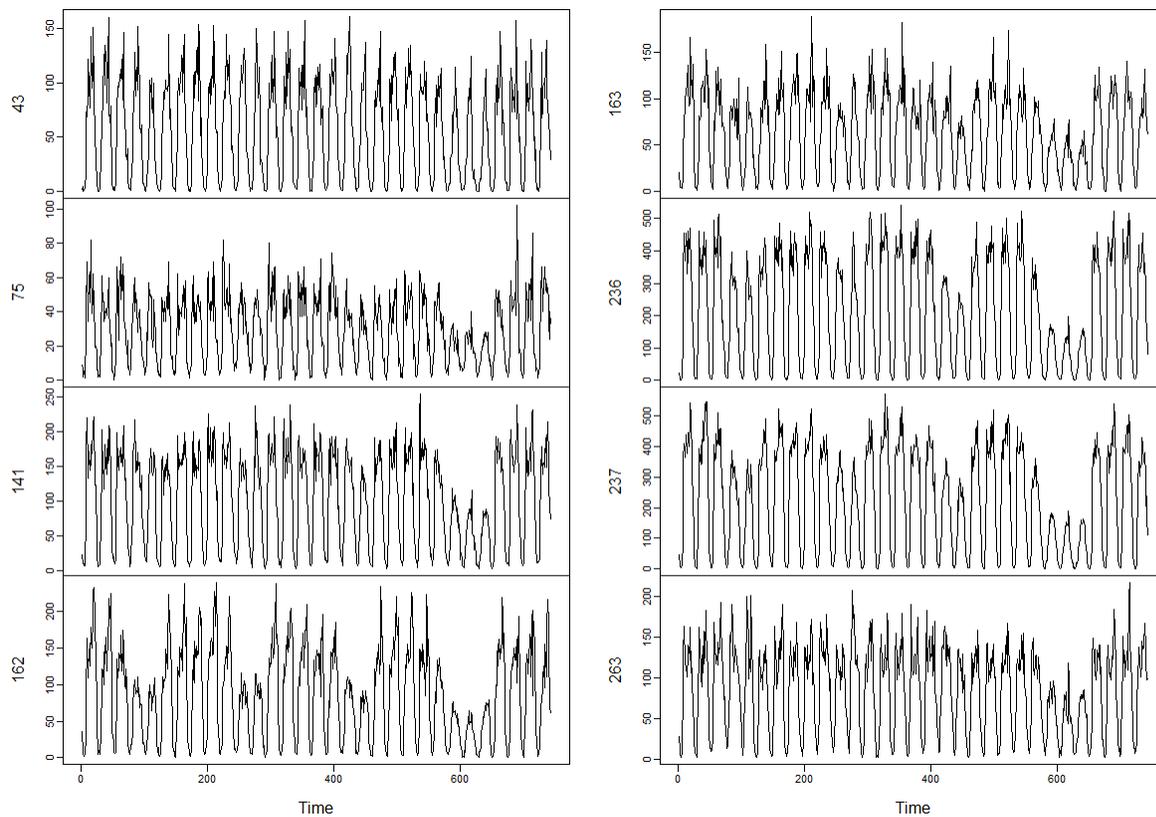
O procedimento de identificação consiste em três etapas:

- a) Verificar se existe uma necessidade de transformação na série original, essa identificação pode ser realizada utilizando o auxílio de gráficos

Inicialmente foi feita a plotagem dos dados, para compreender o comportamento dos chamados de táxi ao longo do mês. A Figura 10 indica a quantidade de viagens que ocorreram em cada uma das zonas ao longo do tempo, sendo possível notar sazonalidade diária, ou seja, a cada 24 horas os efeitos se repetem. Além disso, verificou-se a ausência de tendência, indicando que ao longo do mês não houve aumento na quantidade média de viagens ocorridas por dia.

Todas as zonas apresentam comportamento similar, sendo, picos entre às 7hs – 10hs e 15hs e 18hs. As zonas 162, 236 e 237 tem uma queda na quantidade de viagens maior aos fins de semana, quando comparadas as outras zonas. E todas elas, exceto a zona 43, apresentam uma redução expressiva no último final de semana (por volta da hora 600), pois se trata de um feriado ocorrido em Nova Iorque.

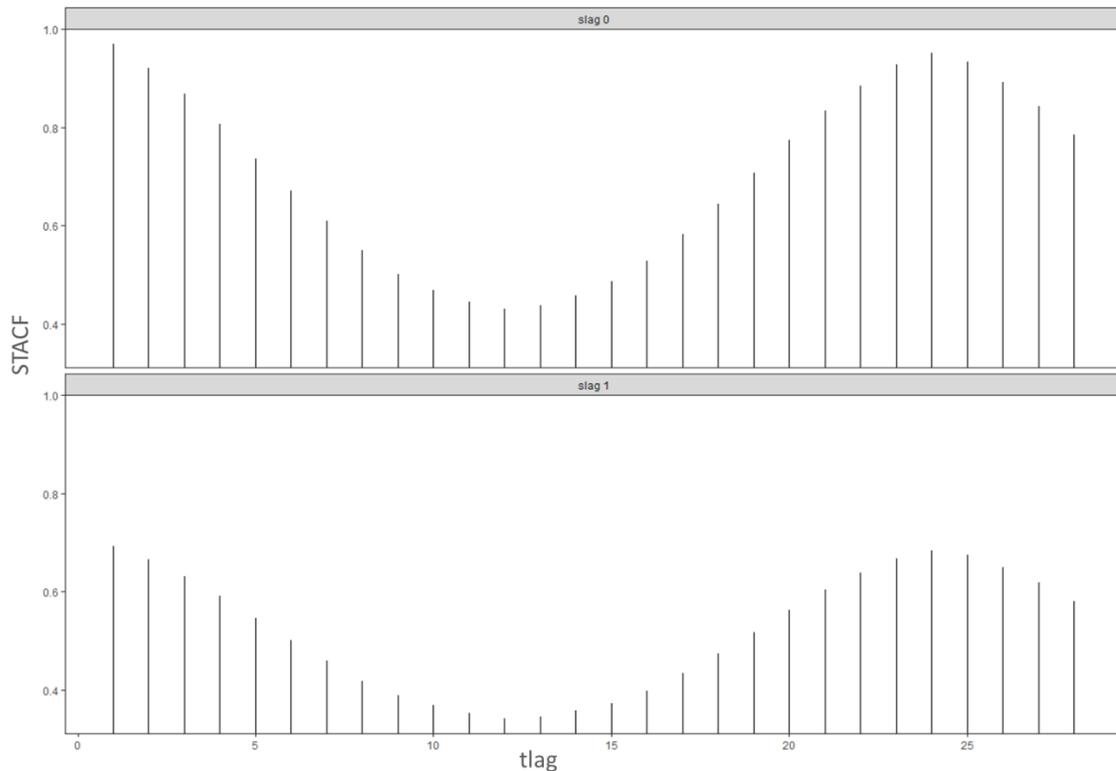
Figura 11 - Plotagem dos dados



Fonte: Elaborado pelo autor

Em seguida, com o auxílio do software R, foi gerado o gráfico da função de autocorrelação espaço-temporal (STACF), demonstrado na Figura 11, para a série temporal formada pelos dados indicados anteriormente. A ciclicidade no STACF, indica e confirma a forte sazonalidade dos dados (conforme já havia sido discutido anteriormente).

Figura 12 - Função de autocorrelação espaço temporal (para os lags 0 e 1)



Fonte: Elaborado pelo autor

No modelo STARMA (e em suas variações) adota-se como premissa que a normalidade nos resíduos e isso é alcançado de forma mais simples se os próprios dados tiverem uma estrutura simétrica, no entanto o item (a) da Figura 12 mostra a assimetria da variável “número de viagens”. A primeira tentativa de modelagem indicou que essa assimetria não poderia ser ignorada e por isso foram propostas algumas transformações, dentre elas a raiz quadrada, raiz cúbica, logaritmo, box-cox e deviance residual.

A transformação de *Deviance Residual* (ALENCAR et al., 2015) indicada pela Equação 3.1, sendo  $Z_t$  a quantidade de viagens (dados originais) e  $Z_t^D$  a variável transformada, tem-se

$$Z_t^D = \text{sign}(Z_t - \mu) \sqrt{d_t^2} \quad (3.1)$$

Onde  $d_t^2$  é calculado por

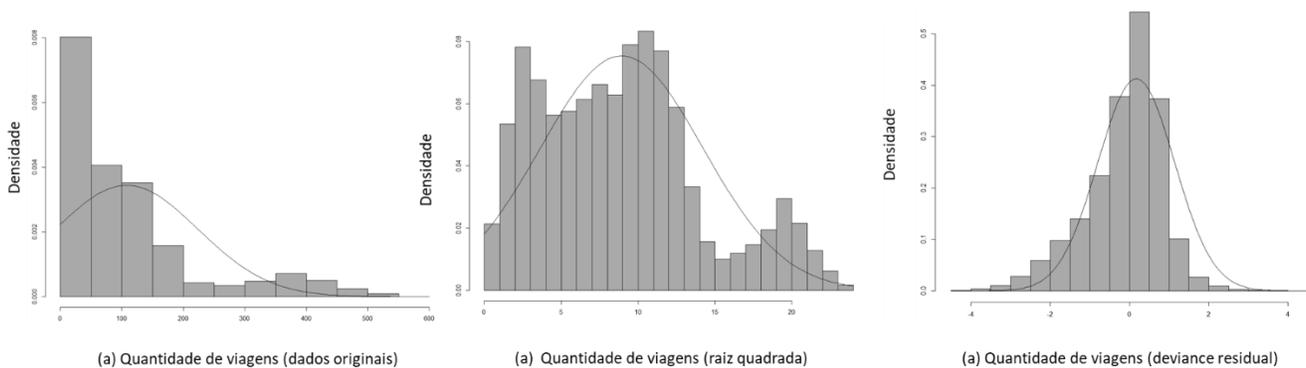
$$d_t^2 = \begin{cases} 2 \tau \ln(1 + \mu/\tau) & \text{se } Z_{*t} = 0 \\ 2 Z_t \ln(Z_{*t}/\mu) - 2 \tau (1 + Z_t/\tau) \ln\left(\frac{1 + Z_{*t}/\tau}{1 + \mu/\tau}\right) & \text{se } Z_{*t} > 0 \end{cases} \quad (3.2)$$

O valor de  $\mu$  representa a média da quantidade de viagens realizadas por hora nos 31 dias do mês de maio e o parâmetro  $\tau$  é estimado previamente através de  $Var(Z_t) = \frac{\tau \pi_z}{(1 - \pi_z)^2}$ , com  $\pi_z = \frac{\mu}{\mu + \tau}$ .

O item (a) da Figura 12 mostra o histograma dos dados originais, o item (b) apresenta a transformação de raiz quadrada e o item (c) a transformação de *Deviance Residual*, todos eles contêm curvas que indicam uma distribuição normal com média e desvio padrão dos dados. O modelo STARMA foi ajustado na série original e em suas transformadas, para a raiz quadrada, raiz cúbica e box-cox os resíduos ainda indicaram muitos pontos fora da reta de confiança do gráfico QQ-Plot. Já os dados com *Deviance* resultaram em um melhor ajuste segundo a aderência dos resíduos.

Dessa maneira, optou-se por utilizar os dados com transformação *Deviance Residual* para as próximas etapas da modelagem.

Figura 13 - Histograma dos dados indicando as transformações

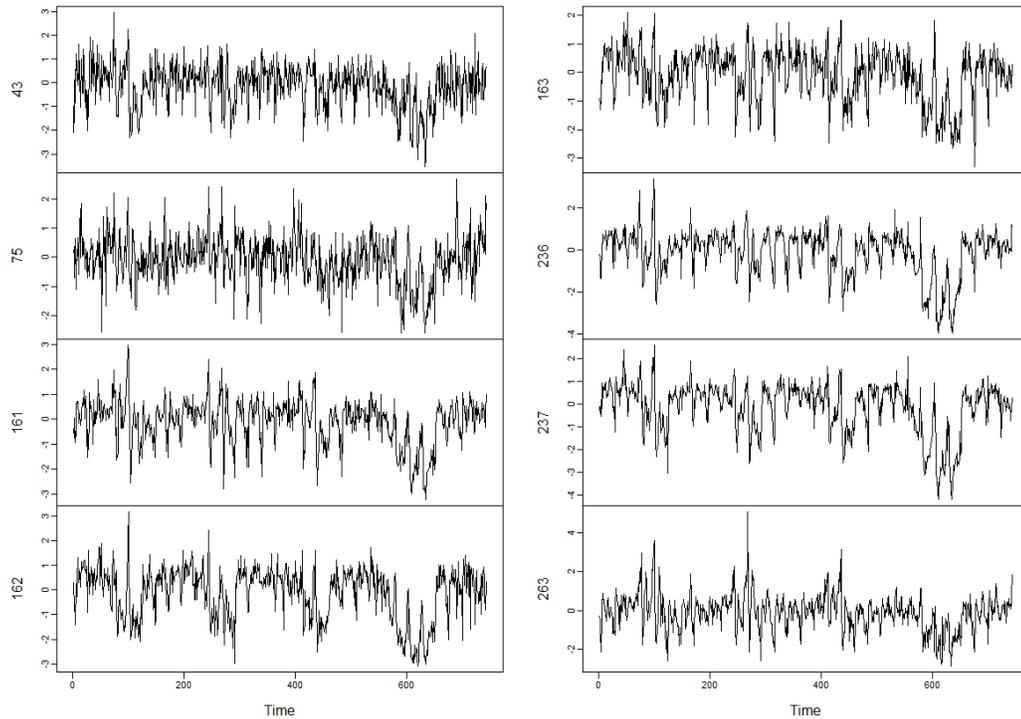


Fonte: Elaborado pelo autor

- b) Realizar diferenciações na série obtida no item (a), tantas vezes quantas necessárias para se obter uma série estacionária. O número de diferenças necessárias, é alcançado quando a função de autocorrelação espaço temporal amostral decresce rapidamente para zero.

Após a transformação de *Deviance Residual*, os dados foram novamente plotados. Como pode-se observar na Figura 13.

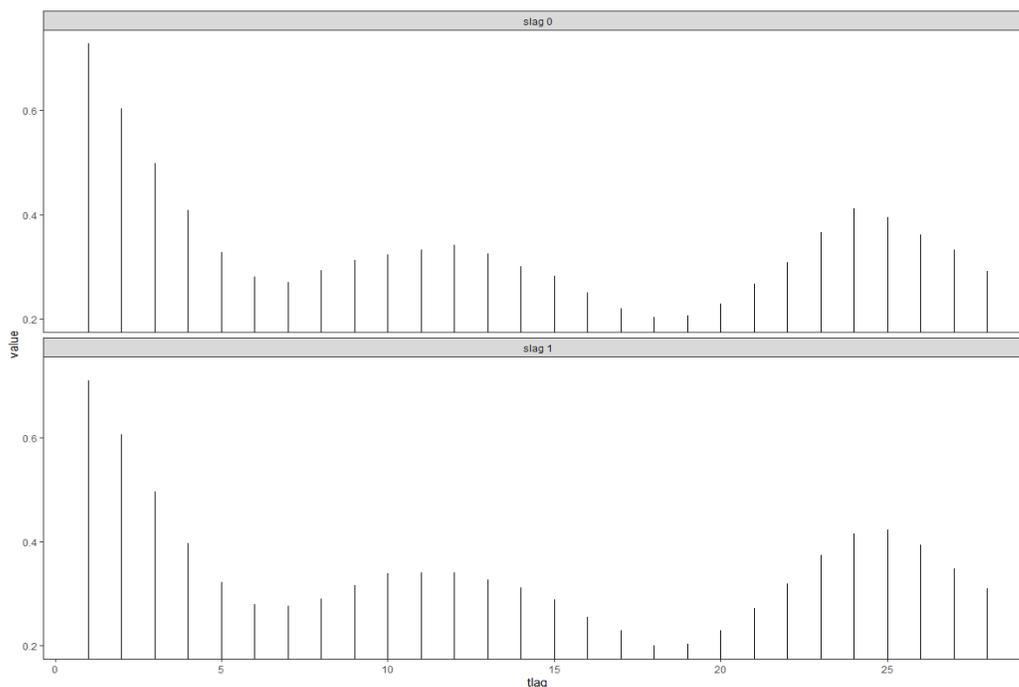
Figura 14 - Plotagem dos Dados com Transformação Deviance Residual



Fonte: Elaborado pelo Autor

Na Figura 13, foi notado a presença de sazonalidade (também verificada no STACF, da Figura 14) e ausência de tendência. Sendo necessário a realização da diferenciação dos dados transformados, como indicado na Seção 2.6, até que a série se torne estacionária. Após a primeira diferenciação, foi analisado o STACF (Figura 15), para garantir que os dados não continham sazonalidade. Além disso é possível utilizar um comando em R (*“isSeasonal”*) que retorna *“True”* caso os dados sejam sazonais e *“False”* se não forem, o que ocorreu no caso dos dados após a diferenciação.

Figura 15 - Função de autocorrelação espaço temporal dos Dados com Transformação Deviance Residual

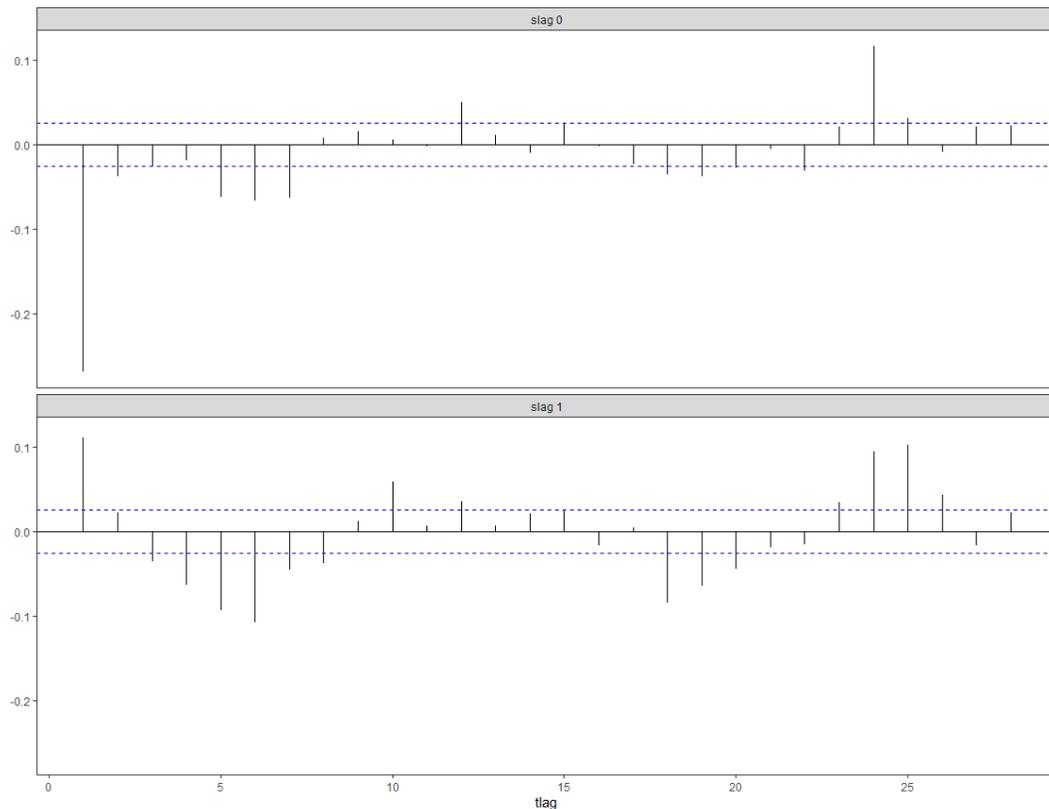


Fonte:

Elaborado pelo autor

Outro ponto importante que se pode notar no STACF (Figura 15) é que em modelos espaço temporais o primeiro gráfico indicado como “*slag 0*” irá representar somente a correlação temporal, pois quando a matriz de vizinhança é indicada por  $W_0$  ela representa a matriz identidade. E depois todos os outros *lags* espaciais que aparecerem representarão correlação espaço-temporal em cada uma das ordens. No caso do modelo desenvolvido para esse trabalho tem-se apenas uma ordem de vizinhança, portanto, “*slag 1*”.

Figura 16 - STACF dos dados com transformação e uma diferenciação

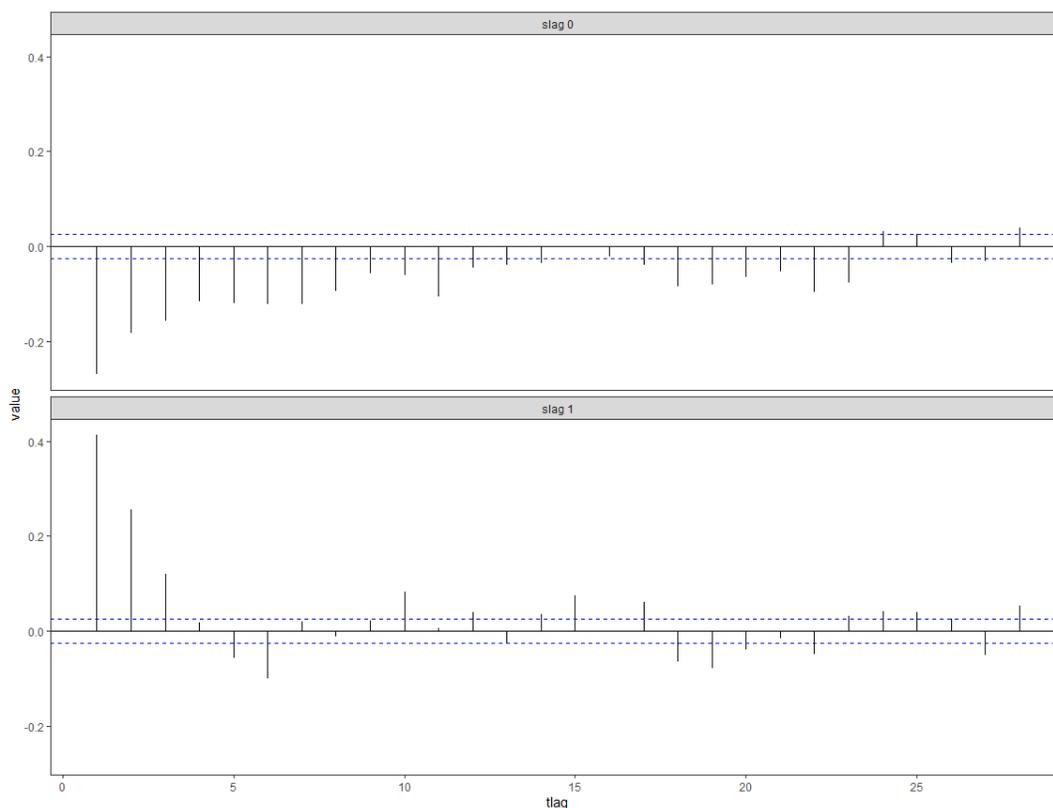


Fonte: Elaborado pelo autor

- c) A terceira etapa se trata de identificar a ordem dos termos do modelo STARMA  $(p_{\lambda_1, \lambda_2, \dots, \lambda_p}, q_{m_1, m_2, \dots, m_q})$ . Através da análise das autocorrelações e autocorrelações parciais estimadas espaço temporais.

É possível que mesmo após eliminar a componente sazonal determinística, ainda reste autocorrelação significativa em *lags* de baixa ordem, indicando que os resíduos são correlacionados podendo ajustá-los através de um modelo STARMA. Nos modelos de correlação espaço temporal, viu-se na Tabela 3 da Seção 2.6.3. que através da análise do STACF (Figura 15) e STPACF (Figura 16) é possível estimar o valor dos parâmetros e analisar os seus resíduos até que se obtenha um resultado que minimize os erros de previsão.

Figura 17 – Função de autocorrelação parcial espaço temporal com os dados transformados e com diferenciação



Fonte: Elaborado pelo autor

No caso, dos modelos espaço temporais os parâmetros de autorregressão e média móvel podem ser indicados em forma de matriz, indicando em uma dimensão o *lag* temporal que deve ser considerado relevante e na outra dimensão o *lag* espacial em que isso deve ocorrer. Após análise das funções de autocorrelação e autocorrelação parcial, chegou-se na seguinte matriz de ordem dos parâmetros, indicada na Tabela 6:

Tabela 6 - Matriz de ordem dos parâmetros para construção do modelo STARMA

		t-1	t-2	t-3	t-4	t-5	t-6	t-7	t-8	t-9	t-10	t-11	t-12	t-13	t-14	t-15	t-16
AR	Slag 0	1	1	0	0	0	1	1	0	0	0	0	1	0	0	1	0
	Slag 1	0	1	0	1	1	1	1	0	1	1	0	1	0	0	1	1
MA	Slag 0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
	Slag 1	1	1	1	0	1	1	1	1	1	1	1	0	1	1	1	1

Fonte: Elaborado pelo autor

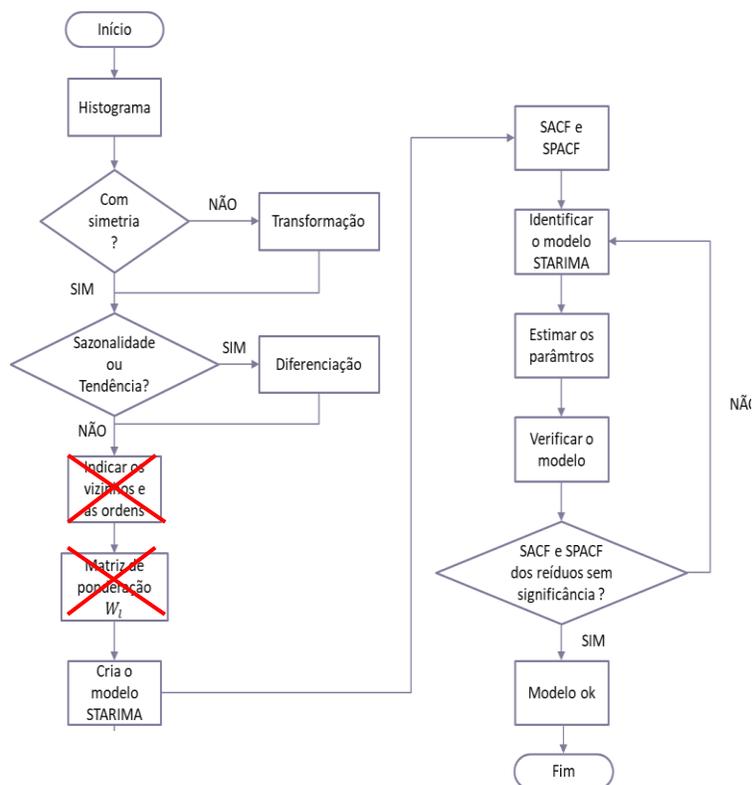
### 3.1.2. Estimação e Verificação

Após encontrar o modelo o valor dos parâmetros deve ser estimado. Nesse estágio, deve-se analisar a STACF e STPACF dos resíduos. Como mencionado na Seção 2.6.2. o modelo estará bem ajustado somente quando o *lag* 0 do STACF (sempre igual a 1) e nenhum *lag* do STPACF tiver valor maior que o intervalo de significância. Os resíduos do modelo encontrado foram analisados, tanto STACF e STPACF quanto sua aproximação de uma distribuição normal através do QQ-plot.

### 3.2. MODELO ARIMA

Além do modelo STARMA, foi desenvolvido um modelo puramente temporal, para verificar a relevância de se considerar a parte espacial no modelo de previsão de viagens contendo origem-destino. Da mesma maneira como o modelo STARMA criou-se o Fluxograma 2 para demonstrar as principais diferenças na criação de cada um dos modelos.

Fluxograma 2 – Modelo Conceitual SARIMA

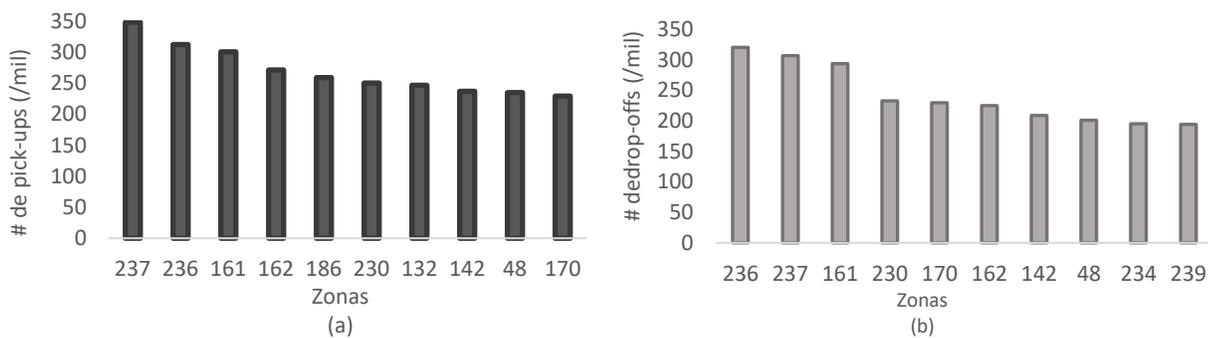


Fonte: Elaborado pelo autor

### 3.2.1. Identificação

Foi utilizada a mesma estratégia do modelo STARMA, porém como não poderia ser utilizado a correlação espacial, optou-se por escolher a zona de origem e a zona de destino com maior quantidade de viagens, indicadas na Figura 17 a e b, respectivamente, selecionando somente as viagens que tinham sido ocorridas entre essas zonas.

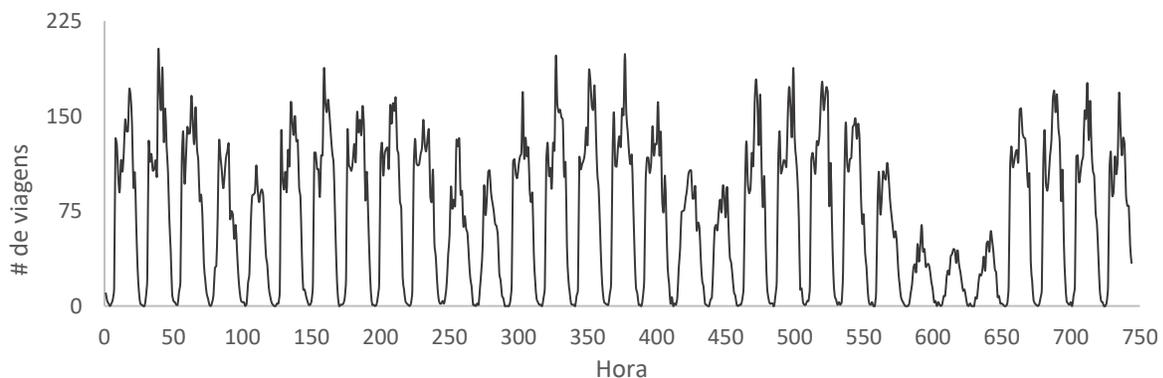
Figura 118 - Gráfico da quantidade de *pick-ups(a)* e *drop-off(b)* ocorridos no mês de maio por zona



Fonte: Elaborado pelo autor

Para a etapa de Identificação do modelo, considerada a primeira etapa de construção segundo Box-Jenkins, foi verificada a necessidade de transformação nos dados, e assim como no modelo espaço-temporal foi realizada a plotagem dos dados (Figura 18) a fim de visualizar os padrões comportamentais da série temporal, como a presença de sazonalidade ou tendência.

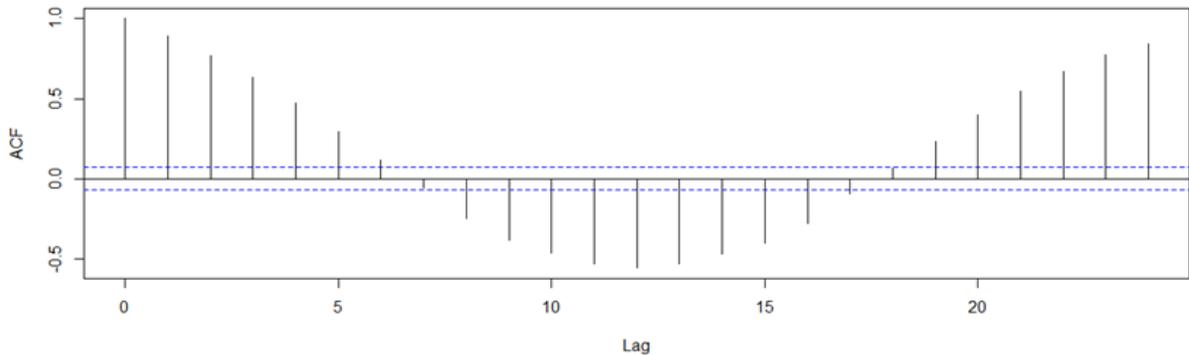
Figura 19 - Plotagem dos dados



Fonte: Elaborado pelo autor

Além disso, avaliou-se o ACF, presente na Figura 19. Como é possível notar a função de autocorrelação é cíclica, demonstrando a existência de sazonalidade.

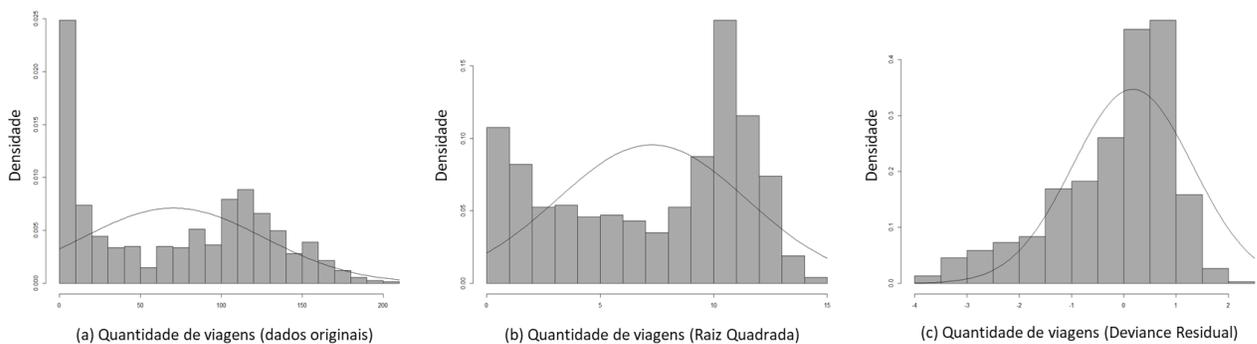
Figura 20 - Função de autocorrelação



Fonte: Elaborado pelo autor

Além disso, os dados, assim como anteriormente, seguem uma distribuição que se aproxima de uma binomial negativa, porém para o ajuste do modelo é necessário que os dados sigam uma distribuição normal, por isso, foi realizada a transformação Deviance nos dados (Figura 20). A mesma transformação realizada nos dados do modelo STARMA, para que fosse possível compatibilizar os resultados.

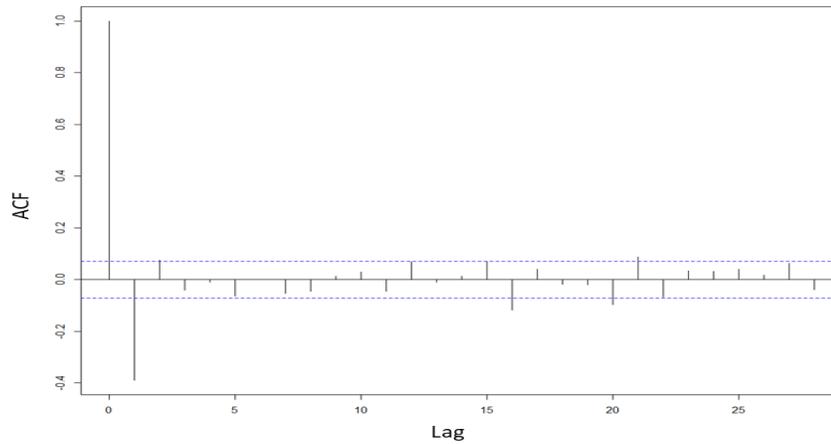
Figura 21 - Histograma dos dados indicando as transformações



Fonte: Elaborado pelo autor

Foi realizada a diferenciação nos dados e observado novamente o ACF (Figura 21), que não apresenta mais comportamento cíclico.

Figura 22 - ACF dos dados com transformação e uma diferenciação



Fonte: Elaborado pelo autor

Por fim, a última fase da primeira etapa, foi identificar os parâmetros do modelo SARIMA  $(p, 0, q, P, 1, Q)_S$ , através da análise do ACF e do PACF (Figura 21). É possível que mesmo após eliminar a componente sazonal determinística, ainda reste autocorrelação significativa em *lags* de baixa ordem, indicando que os resíduos são correlacionados podendo ajustá-los através de um modelo ARIMA; e os *lags* sazonais, isto é, múltiplos de período  $s$ . Isto significa que há necessidade de se considerar uma sazonalidade estocástica, ou seja, ajustar à série original de um modelo ARIMA sazonal (SARIMA). Que é exatamente o que pode ser percebido na Figura 21, o  $(3.3)$  *lag* 1 apresenta autocorrelação significativa e depois somente o *lag* 15, 20, 21 são significativos, indicando dessa forma a presença de sazonalidade estocástica. E dessa forma demonstrando que deve ser empregado um modelo SARIMA nos dados. Lembrando que o modelo SARIMA  $(p, d, q, P, D, Q)_S$  é expresso por

$$\phi(B)\Phi(B^S)(1 - B^S)^D(1 - B)^d z_t = \theta(B)\Theta(B^S)a_t$$

Onde

$$\phi(B) = (1 - \phi_1 B^1 - \dots - \phi_p B^p), \quad (3.4)$$

$$\theta(B) = 1 - \theta_1 B^1 - \dots - \theta_q B^q, \quad (3.5)$$

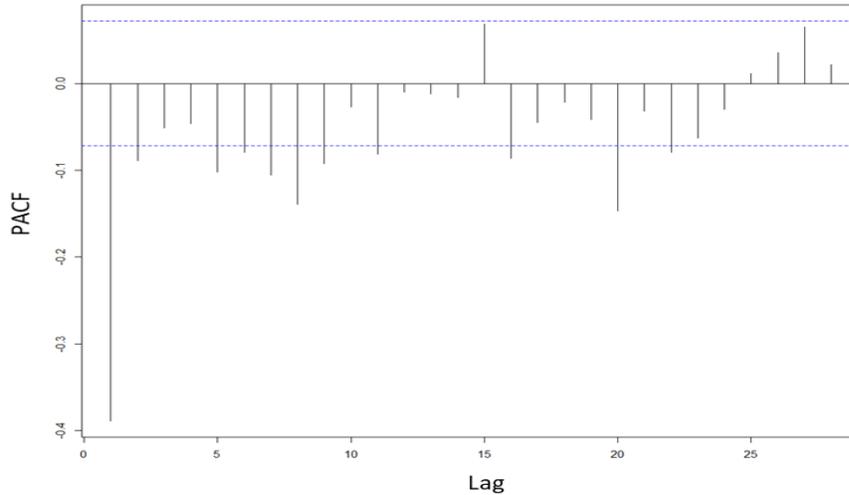
$$\Phi(B^S) = 1 - \Phi_1 B^{1*S} - \Phi_1 B^{2*S} - \dots - \Phi_P B^{P*S} e \quad (3.6)$$

$$(3.7)$$

$$\Theta(B^s) = 1 - \Theta_1 B^{1*s} - \Theta_1 B^{2*s} - \dots - \Theta_Q B^{Q*s}$$

A função de autocorrelação (Figura 16) foi responsável por auxiliar na escolha de uma ordem para o termo de média móvel sazonal e não-sazonal e a função de autocorrelação parcial (Figura 22) a ordem do termo de autorregressão sazonal e não-sazonal.

Figura 23 - Função de autocorrelação parcial com os dados transformados e com diferenciação



Fonte: Elaborado pelo autor

### 3.2.2. Estimação e verificação

Após encontrar o modelo os parâmetros devem ser estimados. Nesse estágio, deve-se analisar a ACF e PACF dos resíduos. O modelo estará bem ajustado somente quando o *lag* 0 do ACF (sempre igual a 1) e nenhum *lag* do PACF tiver valor maior que o intervalo de confiança. Além disso o critério de Informação de Akaike também foi analisado, em que Akaike (1973) sugere escolher o modelo cujo termos *p* e *q* minimizem o critério.

$$AIC(p, d, q) = N \ln \sigma_a^2 + \frac{N}{N-d} 2(p + q + 1 + \delta_{d0}) + 2 \ln 2\pi + N \quad (3.8)$$

Onde

$$\delta_{d0} = \begin{cases} 1, & \text{se } d = 0 \\ 0, & \text{se } d \neq 0 \end{cases} \quad (3.9)$$

Essas análises foram feitas a partir do fluxo de modelagem do Algoritmo 1, portanto, no software R foi realizado um loop, para encontrar todos os modelos citados e seus respectivos AICs.

Algoritmo 1: Etapas para estimação do SARIMA

---

**Algoritmo 1: Etapas para estimação do SARIMA**

---

**Início**

Calcular ACF e PACF dos dados com transformação e diferenciação

Identificação do modelo inicial

Sendo  $d=0$  e  $D=1$

**Para**  $(p, q, P, Q = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$  e  $S = \{3, 6, 12, 24\}$  **fazer**

SARIMA  $(p,d,q,P,D,Q,S)$

AIC dos resíduos

**Fim**

---

Fonte: Elaborado pelo Autor

Para os modelos que apresentaram menor AIC, seus resíduos foram analisados, tanto ACF e PACF quanto sua aproximação de uma distribuição normal através do QQ-plot.

Porém mesmo utilizando a abordagem acima, foi difícil encontrar um modelo que estivesse dentro das condições mencionadas anteriormente, por isso, decidiu-se utilizar outra abordagem, de acordo com Morettin e Toloi (2004).

Obtendo um novo modelo linear  $N_t$ , por meio de uma análise de regressão de  $z_t$  sobre  $1, \text{sen} \frac{2\pi jt}{24}$  e  $\text{cos} \frac{2\pi jt}{24}$ , para  $j = 1, \dots, 6$ . E  $t$  variando de 0 até 744 horas (24horas\*31dias)

$$N_t = \sum_{j=1}^6 \text{cos} \frac{2\pi jt}{24} + \text{sen} \frac{2\pi jt}{24} \quad (3.10)$$

Depois refazer todas as etapas mencionadas anteriormente.

Após encontrar um modelo bem ajustado para os resíduos de  $N_t$ , utiliza-se o mesmo modelo para os dados que passaram pela transformação Deviance, ou algum modelo próximo a ele. A diferença desse processo, é que se parte de uma base de dados diferente inicialmente o que pode facilitar na busca por um modelo que se ajuste aos dados.

Dessa maneira foi possível encontrar um modelo que estivesse de acordo com as condições mencionadas, ou seja, cujo ACF e PACF dos resíduos não apresentasse nenhum *lag* com autocorrelação significativa, demonstrando que todos os termos importante já se encontravam no modelo.

## 4. RESULTADOS

Neste capítulo serão apresentados os resultados obtidos com cada um dos modelos apontando as principais diferenças entre eles

### 4.1. Modelo STARMA

Após os procedimentos descritos na metodologia foi possível encontrar os valores dos parâmetros do modelo STARMA ( $p_{\lambda_1, \lambda_2, \dots, \lambda_p}, q_{m_1, m_2, \dots, m_q}$ ), indicados na Tabela 7.

Tabela 7 - Valores estimados dos parâmetros do modelo STARMA (15,16,15,16)

	Estimado	Erro Padrão	P-VALUE		Estimado	Erro Padrão	P-VALUE
$\phi_{1,0}$	-0,25928	0,069917	0,0002	$\theta_{5,0}$	-0,04701	0,02072	0,0233
$\phi_{2,0}$	-0,25278	0,082377	0,0022	$\theta_{5,1}$	0,122095	0,086334	0,1574
$\phi_{2,1}$	0,209107	0,097427	0,0319	$\theta_{6,0}$	0,074327	0,085109	0,3825
$\phi_{4,1}$	-0,12063	0,033705	0,0003	$\theta_{6,1}$	-0,031	0,103551	0,7647
$\phi_{5,1}$	-0,27505	0,077525	0,0004	$\theta_{7,0}$	-0,0297	0,079325	0,7081
$\phi_{6,0}$	-0,11517	0,083574	0,1682	$\theta_{7,1}$	0,021073	0,101287	0,8352
$\phi_{6,1}$	-0,19494	0,099568	0,0503	$\theta_{8,0}$	-0,04646	0,063429	0,4639
$\phi_{7,0}$	-0,08773	0,08275	0,2891	$\theta_{8,1}$	-0,09649	0,076375	0,2065
$\phi_{7,1}$	-0,10005	0,097869	0,3067	$\theta_{9,0}$	-0,02587	0,022971	0,2601
$\phi_{9,1}$	-0,0507	0,07325	0,4889	$\theta_{9,1}$	0,03323	0,081381	0,6830
$\phi_{10,1}$	-0,15729	0,072646	0,0304	$\theta_{10,0}$	-0,01263	0,021996	0,5660
$\phi_{12,0}$	0,17024	0,05449	0,0018	$\theta_{10,1}$	0,162759	0,078542	0,0383
$\phi_{12,1}$	-0,11603	0,042367	0,0062	$\theta_{11,0}$	0,003495	0,022041	0,8740
$\phi_{15,0}$	0,056848	0,028838	0,0487	$\theta_{11,1}$	-0,12768	0,048545	0,0086
$\phi_{15,1}$	0,028066	0,05856	0,6318	$\theta_{12,0}$	-0,16079	0,055072	0,0035
$\phi_{16,1}$	0,081462	0,055539	0,1425	$\theta_{13,0}$	0,113099	0,040258	0,0050
$\theta_{1,0}$	-0,34969	0,073097	0,0000	$\theta_{13,1}$	-0,2339	0,051882	0,0000
$\theta_{1,1}$	0,546963	0,034152	0,0000	$\theta_{14,0}$	-0,00896	0,021695	0,6797
$\theta_{2,0}$	-0,00523	0,079757	0,9477	$\theta_{14,1}$	-0,03084	0,038876	0,4277
$\theta_{2,1}$	-0,06919	0,104032	0,5060	$\theta_{15,0}$	-0,05015	0,035206	0,1544
$\theta_{3,0}$	-0,22686	0,061129	0,0002	$\theta_{15,1}$	-0,01231	0,065726	0,8514
$\theta_{3,1}$	0,04789	0,064403	0,4572	$\theta_{16,1}$	-0,10763	0,065781	0,1019
$\theta_{4,0}$	-0,04973	0,023019	0,0308				

Fonte: Elaborado pelo autor

Os valores de  $\phi_{p,l}$  e  $\theta_{q,l}$  indicam os parâmetros relacionados a parte autorregressiva e de média móvel do modelo, respectivamente, sendo  $p$  e  $q$  a ordem temporal e  $l$  a ordem espacial. Como foi adotado um modelo com apenas matriz de vizinhança de ordem 1, o valor de  $l$  será 0 (indicando

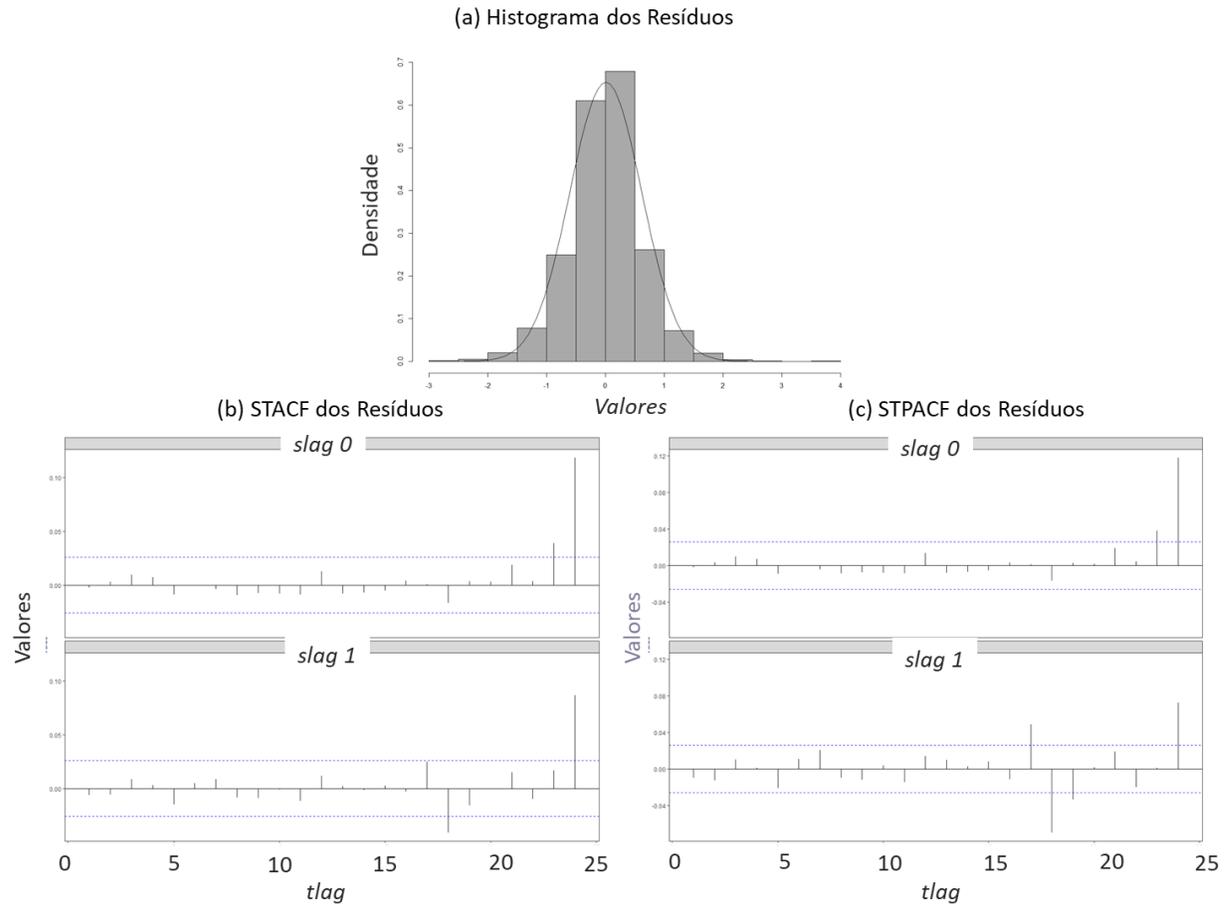
a ausência de *lag* espacial) e 1 quando corresponder a matriz  $W_1$ . Já os *lags* temporais com significância ao modelo foram indicados na Tabela 5.

Com base no modelo estimado (Equação 3.10), nota-se que o *lag* temporal puro apresenta uma correlação negativa de 25,9% do índice da hora anterior e 25,3% do índice de duas horas anteriores a prevista. Já no *lag* espacial, é possível notar que há uma correlação de 20,9% do índice de duas horas anteriores.

$$\begin{aligned}
 z_t = & -0,259Z_{t-1} - 0,253Z_{t-2} - 0,115Z_{t-6} - 0,088Z_{t-7} + 0,170Z_{t-12} \\
 & + 0,057Z_{t-15} + (0,209z_{t-2} - 0,121Z_{t-4} - 0,100Z_{t-7} - 0,051Z_{t-9} \\
 & - 0,157Z_{t-10} - 0,116Z_{t-12} + 0,028Z_{t-15} + 0,081Z_{t-16})W_1 \\
 & - (-0,350a_{t-1} - 0,005a_{t-2} - 0,227a_{t-3} - 0,050a_{t-4} - 0,047a_{t-5} \\
 & + 0,074a_{t-6} - 0,030a_{t-7} - 0,046a_{t-8} - 0,026a_{t-9} - 0,012a_{t-10} \\
 & - 0,003a_{t-11} + 0,161a_{t-12} + 0,113a_{t-13} - 0,009a_{t-14} \\
 & - 0,050a_{t-15} + (0,547a_{t-1} - 0,069a_{t-2} + 0,048a_{t-3} + 0,122a_{t-5} \\
 & - 0,031a_{t-6} + 0,021a_{t-7} - 0,096a_{t-8} + 0,033a_{t-9} + 0,163a_{t-10} \\
 & - 0,128a_{t-11} - 0,234a_{t-13} - 0,031a_{t-11} - 0,012a_{t-15} \\
 & - 0,108a_{t-16})W_1) + a_t
 \end{aligned}$$

Depois de estimar os parâmetros do modelo foi verificado sua adequação analisando seus resíduos. A Figura 17, mostra a plotagem dos resíduos, o histograma e os gráficos do STACF e STPACF. Os resíduos padronizados estimados desse modelo têm comportamento independente e identicamente distribuídos com média zero e variância constante. O histograma da Figura 23(a) indica que os resíduos se aproximam de uma distribuição normal. E o STACF e o STPACF da Figura 23(b) e (c) demonstram que ainda existe autocorrelação em alguns termos, porém, este foi o melhor encontrado.

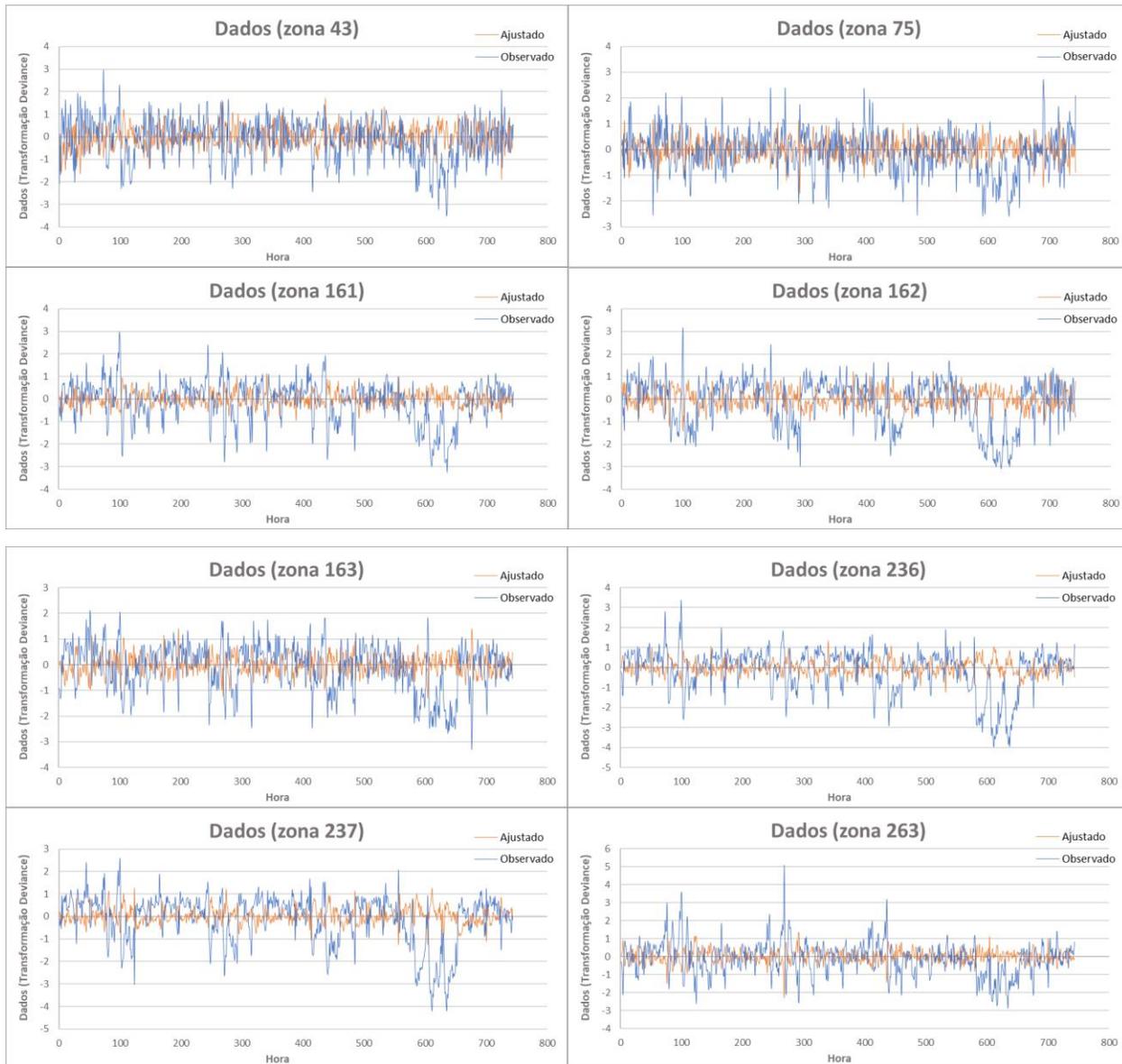
Figura 24 - Diagnóstico gráfico do Modelo



Fonte: Elaborado pelo autor

Uma vez que o modelo da série temporal foi apropriadamente ajustado, as previsões podem ser feitas, a Figura 24 mostra o número de viagens de táxi (dados com transformação *Deviance Residual*) que foram observadas no mês de maio e os valores que foram ajustados pelo modelo, através das Equações 3.10. Sendo possível prever quantas viagens irão ocorrer nas horas seguintes e onde foram essas viagens.

Figura 25 - Comparação dos dados observados pelos dados ajustados pelo modelo STARMA



Fonte: Elaborado pelo Autor

A partir dos dados ajustados foi possível calcular o erro gerado pelo modelo, para esse cálculo utilizamos o erro absoluto percentual médio, também conhecido como MAPE (do inglês, “Mean Absolute Percentage Error”), expresso por:

$$MAPE = \frac{1}{N} \sum_{i=0}^N \frac{|observado_i - previsto_i|}{observado_i} \times 100 \quad (3.11)$$

Onde, N é igual a quantidade de observações existentes. Chegando a um modelo com 46,6% de erro.

#### 4.2. Modelo SARIMA

Foi também encontrado um modelo SARIMA, puramente temporal, com sazonalidade de 12 horas. A partir da análise do ACF e do PACF (Figura 21 e 22, respectivamente) a ordem dos termos não-sazonais foram  $p = 5$  e  $q = 2$ , e dos termos sazonais foram  $P=2$  e  $Q=1$ . O valor de cada um dos parâmetros está indicado na Tabela 8, mostrando que o *lag* que apresenta maior influência sobre a previsão dos dados é o 3, com 53,9%, ou seja, a previsão apresenta grande correlação com a terceira hora que a precede (t-3).

Tabela 8 - Parâmetros do modelo SARIMA

AR não-sazonal	MA não-sazonal	AR sazonal	MA sazonal
$\phi_1 = 0,4301$	$\theta_1 = 0,9972$	$\Phi_1 = 0,1211$	$\Theta_1 = -0,9722$
$\phi_2 = -0,6581$	$\theta_2 = 0,0070$	$\Phi_2 = 0,1153$	
$\phi_3 = 0,5393$			
$\phi_4 = 0,2702$			
$\phi_5 = 0,0666$			

Fonte: Elaborado pelo autor

O modelo SARIMA  $(5,0,2)(2,1,1)_{12}$  é expresso por

$$\phi(B)\Phi(B^{12})(1 - B^{12})^1(1 - B)^0 z_t = \theta(B)\Theta(B^{12})a_t \quad (3.12)$$

$$\phi(B)\Phi(B^{12})(1 - B^{12})^1 z_t = \theta(B)\Theta(B^{12})a_t \quad (3.13)$$

$$\phi(B) = (1 - \phi_1 B^1 - \phi_2 B^2 - \phi_3 B^3 - \phi_4 B^4 - \phi_5 B^5) \quad (3.14)$$

$$\theta(B) = (1 - \theta_1 B^1 - \theta_2 B^2) \quad (3.15)$$

$$\theta(B) = 1 - \theta_1 B^1 - \theta_2 B^2 \quad (3.16)$$

$$\Phi(B^3) = 1 - \Phi_1 B^{12} - \Phi_1 B^{24} \quad (3.17)$$

$$\Theta(B^3) = 1 - \Theta_1 B^{12}$$

Substituindo os parâmetros, tem-se:

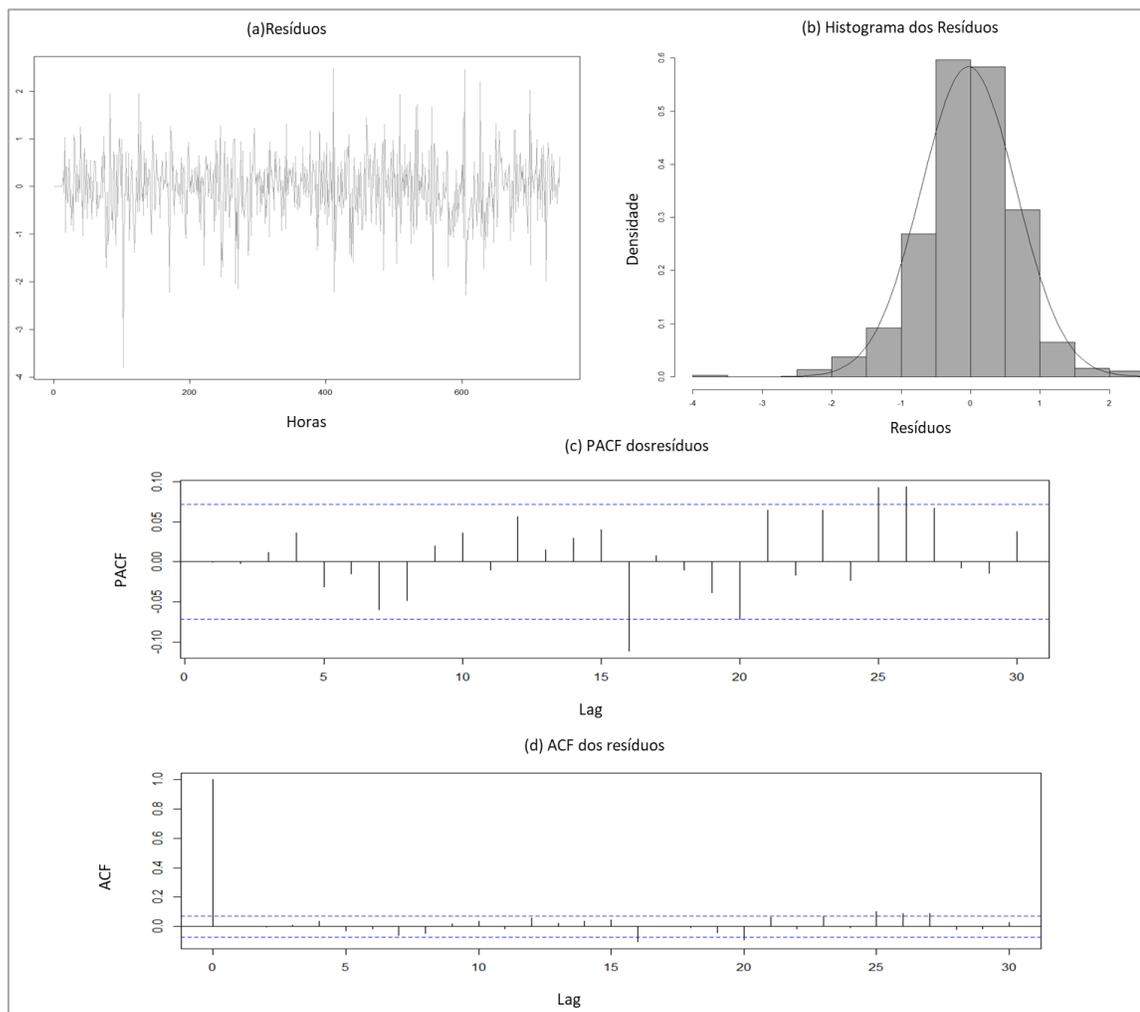
$$(1 - 0,43B^1 + 0,69B^2 - 0,54B^3 - 0,27B^4 - 0,04B^5)(1 - 0,12B^{12} - 0,11B^{24})(1 - B^{12})z_t = (1 - 0,07B^1 - 1,00B^2)(1 + 0,97B^{12})a_t \quad (3.18)$$

Lembrando que:  $B^n z_t = z_{t-n}$  e  $B^n a_t = a_{t-n}$ .

A partir da Equação 3.18 é possível perceber que o modelo irá depender das observações que ocorreram nas horas 1,2,3,4,5,12,13,14,15,16,17,24,25,26,27,28 e 29, anteriores ao atual. Além disso, também dependerá das inovações das horas 1,2,12,13 e 14, também anteriores ao atual.

Depois de estimar os parâmetros do modelo foi verificado sua adequação analisando seus resíduos. A Figura 25, mostra a plotagem dos resíduos, o histograma e os gráficos do ACF e PACF. A Figura 25(a) sugere que os resíduos padronizados estimados desse modelo têm comportamento independente e identicamente distribuídos com média zero e variância constante. O histograma da Figura 18(b) indica que os resíduos se aproximam de uma distribuição normal. E o ACF e o PACF da Figura 18(c) e (d) demonstram que ainda existe autocorrelação em alguns termos, indicando que o modelo não é o esperado, de acordo com as condições que foram mencionadas no tópico 2.5.1. Porém, este foi o melhor encontrado dentro dessa sazonalidade. Por esse motivo, a abordagem inicial foi alterada, utilizando a regressão de  $z_t$ , descrita também no tópico 3.2.2.

Figura 26 - Diagnóstico gráfico do Modelo SARIMA



Fonte: Elaborado pelo autor

Deste modelo, também se calculou o MAPE, para que fosse feita uma comparação equivalente entre os dois modelos. Em que foi encontrado um erro de 78,3%.

#### 4.3. Comparação STARMA x SARIMA

Após a aplicação da metodologia proposta e análise dos resultados, foi possível compreender as principais diferenças entre cada um dos modelos estudados, verificando os benefícios de cada um deles (Tabela 9).

Tabela 9 - Comparação STARMA x SARIMA

	STARMA	SARIMA
<b>Modelo</b>	STARMA (16,15,16,15)	SARIMA (5,0,2)(2,1,1) <sub>12</sub>
<b>Previsão</b>	Espaço-temporal Autorregressivo de média móvel	Sazonal Autorregressivo Integrado de média móvel
<b>Matriz de ponderação</b>	✓	X
<b>Variograma</b>	✓	X
<b>Erro de previsão</b>	46,6 %	78,3%
<b>Quantidade de parâmetros</b>	42	10
<b>Complexidade</b>	✓	X
<b>Acuracidade</b>	✓	X

Como pode-se observar na Tabela 9, o modelo STARMA, apresenta uma quantidade bem maior de parâmetros quando comparado ao modelo SARIMA, indicando uma maior complexidade para encontrar os resultados de previsão. Porém outro ponto importante é que o modelo SARIMA prevê somente uma localidade em cada modelo (nesse trabalho foi possível prever somente um par origem-destino (236-237)), nesse caso, se fosse realizado um modelo para cada par de vizinhos aumentaria a quantidade de parâmetros e conseqüentemente a dificuldade de previsão.

O modelo STARMA é capaz de representar a correlação entre os vizinhos, importante para dados de origem-destino, gerando resultados com maior acuracidade, como pode-se verificar pela diferença no erro encontrado em cada um dos modelos. Modelos Sazonais Autorregressivos Integrados de média móvel, são mais utilizados para séries temporais que variam somente ao longo do tempo. Uma grande dificuldade encontrada na criação dos modelos foi a adequação dos dados, que originalmente vem de uma contagem, para um modelo com distribuição normal nos erros.

Através da revisão da literatura foi possível encontrar alguns autores que utilizaram o modelo STARMA. Safikhani, A., et al. (2018) realizou um trabalho similar a esse, onde ele previu viagens de taxi da cidade de Nova Iorque no ano de 2015, foi encontrado um erro de previsão de 27%, o

principal ponto de diferença entre o trabalho dele e este foi a quantidade de vizinhos incluídos no modelo, gerando maior confiabilidade, onde foi abrangido toda Manhattan. Além disso, ele utilizou intervalos de 15 minutos, e neste trabalho utilizou-se 1 minuto. Em sua conclusão ele discute sobre a grande dificuldade em encontrar um modelo para o banco de dados, assim como, a grande quantidade de parâmetros encontrados em seu modelo. Dentre os modelos que ele utilizou para comparar a melhor opção o modelo STARMA obteve melhor resultado.

## 5. CONCLUSÃO

A modelagem e previsão de demanda de entregas urbanas, em uma rede de tráfego, tem sido uma tarefa desafiadora na literatura, considerando a formulação de um modelo complexo exigindo grande necessidade da integração efetiva do conhecimento sobre o domínio. Nesse trabalho, foi formulado e proposto o modelo multivariado STARMA para realizar uma estimativa.

O STARMA para os dados em estudo, resultou em um modelo de previsão complexo e com grande quantidade de parâmetros, porém foi capaz de incorporar as sazonalidades diárias e horárias, pode demonstrar, também, que os dados apresentam correlação espacial em períodos de até um mês. A vantagem deste modelo é o uso das autocorrelações e autocorrelações parciais para identificar os termos mais apropriados com base nos resíduos.

Através do MAPE foi possível verificar a precisão dos dados ajustados, avaliando a magnitude do erro em relação à série histórica. Como estão sendo utilizados dados com transformação o ideal para o cálculo do erro é que a escala dos dados não influencie no resultado, o que ocorre com o erro absoluto percentual médio. O resultado do estudo e da aplicação aos dados de *yellow taxi* em Nova Iorque demonstraram que podem ser realizadas boas previsões seguindo a metodologia apresentada.

Mesmo assim, é importante ressaltar também que os dados, são heterogêneos e apresentam grande influência de componentes externos, como temperatura, feriados, acidentes de trânsito, assaltos, entre outros que corroboram para que o modelo tenha erros maiores, dificultando a previsão.

Além disso, foi possível verificar a melhora no resultado quando comparado a um modelo que considera somente a correlação temporal entre os dados, indicando a grande influência da vizinhança e a importância de ser estudado e aplicado modelos espaço-temporais quando se trata de dados com diferentes localidades, como ocorre na utilização de matrizes origem-destino.

Para os trabalhos futuros sugere-se a ampliação do banco de dados utilizado, ou seja, ampliar temporalmente os dados, para todo um ano, para que a sazonalidade do modelo possa ser analisada na dimensão mês, além das dimensões hora e dia. Além disso, incorporar uma

vizinhança maior, com diferentes ordens e incorporar ao modelo dados que indiquem as viagens que ocorrem sem passageiros, como por exemplo, as viagens canceladas e dados com transporte de carga além do transporte de pessoas. Já em relação ao modelo utilizado, pode-se realizar uma comparação com outros modelos utilizando Machine Learning, a fim de verificar o comportamento de diferentes modelos, neste mesmo banco de dados.

## REFERÊNCIAS

ASHOK, K. & BEN-AKIVA, M. E. **Estimation and prediction of time-dependent origin-destination flows with a stochastic mapping to path flows and link flows.** Transp. Sci., 2002.

BEIRIGO, B. A., SCHULTE, F., & NEGENBORN, R. R. **Integrating people and freight transportation using shared autonomous vehicles with compartments.** IFAC-PapersOnLine, 51(9), 392-397, 2018.

BOX, G. E. P. & JENKINS, G. M. **Time series analysis, forecasting and control.** 1970.

BOX, G. E. P. & PIERCE, D. A. **Distribution of residual autocorrelations in autoregressive-integrated moving average time series models.** J. Am. Stat. Assoc., 1970.

CACERES, N.; WIDEBERG, J. P.; BENITEZ, F. G. **Review of traffic data estimations extracted from cellular networks.** IET Intelligent Transport Systems, 2, 2018 p. 179-192.

CARSTEN, O. **Presentation on Automated Driving.** Australian International Driverless Vehicle Conference,

GRANGER, C. W. J.; NEWBOLD, P. **Forecasting economic time series.** Academic Press, 2014.

GRUCHMANN, T. Advanced green logistics strategies and technologies. In: **Operations, logistics and supply chain management.** Springer, Cham, 2019.

GREENBLATT, J. B. AND SHAHEEN, S. **On-Demand Mobility.** Autonomous Vehicles and Environmental Impacts, Current Sustainable/Renewable Energy Reports, Lawrence Berkeley National Laboratory: Energy Technologies Area, 2015.

GUERRERO, J. C., & DÍAZ-RAMÍREZ, J. **A review on transportation *last-mile* network design and urban freight vehicles.** In Proceedings of the 2017 International Symposium on Industrial Engineering and Operations Management Bristol (UK), 2017.

HABIBI, M.R., DAVIDSON, A., LAROCHE, M. **What managers should know about the sharing economy.** Business Horizons 60.1, 113–121, 2017.

HASEBE, K., et al. **Traffic management for *last-mile* public transportation systems using autonomous vehicles.** IEEE, 2017.

ITO P, SALVO G. **Toward an urban transport sustainability index:an European comparison.** Eur Transport Res Rev, 2019.

JAIPURIA, S; MAHAPATRA, S. S. **An improved demand forecasting method to reduce bullwhip effect in supply chains.** Expert Systems with Applications, v. 41, n. 5, p. 2395-2408, 2014.

- JINGXU, C., XUEWU, C., WEI, W. & BAOL, F. **The Demand Analysis of Bike-and-ride in Rail Transit Stations based on Revealed and Stated Preference Survey.** *Procedia - Soc. Behav. Sci.*, 2013.
- LI, Y., DÍAZ, M., MORANTES, S. & DORATI, Y. **Vehículos autónomos: Innovación en la logística urbana Autonomous vehicles: Innovation in urban logistics.** *Ric* 4, 34–39, 2018.
- LITMAN, T. **Autonomous vehicle implementation predictions.** Victoria, Canada: Victoria Transport Policy Institute, 2017.
- MAK, H. **Peer-to-peer crowdshipping as an omnichannel retail strategy.** Available at SSRN 3119687, 2018.
- MANKYA, J., et al. **Disruptive Technologies: Advances That Will Transform Life, Business and the Global Economy.** McKinsey Global Institute: New York, NY, USA, 2013.
- MARTINEZ, E. Z., da SILVA, E. A. S. & dal FABBRO, A. L. **Um modelo SARIMA para predição do número de casos de dengue em Campinas, Estado de São Paulo.** *Rev. Soc. Bras. Med. Trop.*, 2011.
- MILLER, J. S., KANG, D. **Ways to Consider Driverless Vehicles in Virginia Long Range Travel Demand Models.** Virginia Transportation Research Council, 2019.
- MOALA, A. B., HO, L. L. & QUININO, R. C. **Multivariate control charts to monitor the monthly frequency of vehicle robberies in São Paulo city.** *Spat. Stat.*, 2019.
- MONTGOMERY, D. C., JENNINGS, C. L., KULAHCI, M. **Introduction to time series analysis and forecasting.** John Wiley & Sons, 2015.
- MOORTHY, A., et al. **Shared autonomous vehicles as a sustainable solution to the last mile problem: A case study of Ann Arbor-Detroit area.** *SAE International Journal of Passenger Cars-Electronic and Electrical Systems*, 2017.
- MOREIRA-MATIAS, L., GAMA, J., FERREIRA, M. & DAMAS, L. **A predictive model for the passenger demand on a taxi network.** *IEEE Conf. Intell. Transp. Syst. Proceedings, ITSC* 1014–1019, 2012.
- MORETTIN, P. A.; TOLOI, C. M. **Análise de séries temporais.** 2004.
- OHNEMUS, M., & PERL, A. **Shared autonomous vehicles: Catalyst of new mobility for the last mile?** *Built Environment*, 42(4), 589-602, 2016.
- PFEIFER, P. E. & BODILY, S. E. **A test of space-time ARMA modelling and forecasting of hotel data.** *J. Forecast*, 1990.

PFEIFER, P. E. & DEUTSCH, S. J. **A STARIMA model-building procedure with application to description and regional forecasting.** Trans. Inst. Br. Geogr., 1980a.

PFEIFER, P. E. & DEUTSCH, S. J. **American Society for Quality A Three-Stage Iterative Procedure for Space-Time Modeling A Three-Stage Iterative.** TECHNOMETRICS, 1980b.

PORTER, B.; LINSE, M.; BARASZ, Z. **Six transportation trends that will change how we move.** Forbes, January, 2015.

QIAN, X., UKKUSURI, S. V., YANG, C., & YAN, F. **Forecasting short-term taxi demand using boosting-GCRF.** In The 6th International Workshop on Urban Computing, 2017.

ROSANO, M. **Mixing quantitative and qualitative methods for sustainable transportation in smart cities.** 2020. Tese de Doutorado. Springer Berlin Heidelberg.

SAE International 2020. < [sae.org](http://sae.org) >. Acessado em Junho, 2020.

SAFIKHANI, A., et al. **Spatio-temporal modeling of yellow taxi demands in New York City using generalized STAR models.** International Journal of Forecasting, 2018.

SAHU, P; KUMAR, R. **Demand forecasting for sales of milk product (Paneer) in Chhattisgarh.** International Journal of Inventive Engineering and Sciences, v. 1, n. 9, p. 10-13, 2013.

Schaller Consulting. **The New York City Taxicab Fact Book.** <http://www.schallerconsult.com/taxi/taxifb.pdf>. Accessed Nov. 1, 2020

SHAHEEN, S., et al. **Shared Mobility: A Sustainability & Technologies Workshop.** Definitions, Industry Developments, and Early Understanding, Transportation Sustainability Research Center, UC Berkeley, 2015.

SHAHEEN, S., MARTIN, E., and COHEN, A., **Public Bikeshaaring and Modal Shift Behavior: A Comparative Study of Early Bikeshaaring Systems in North America.** International Journal of Transportation. Vol. 1, No. 1, pp. 35-53, 2013

SHUMWAY, R. H.; STOFFER, D. S. **Time series analysis and its applications.** Springer, 2000.

SKEETE, J. P. Level 5 autonomy: **The new face of disruption in road transport.** Technol. Forecast. Soc. Change 134, 22–34, 2018.

SPERANZA, M. G. **Trends in transportation and logistics.** European Journal of Operational Research, 264(3), 2018.

STOCKER, A., & SHAHEEN, S. **Shared automated vehicles: Review of business models.** International Transport Forum Discussion Paper, 2017.

WILLIAMS, B. M. & HOEL, L. A. **Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results.** J. Transp. Eng., 2003.

WONG, S. C. & TONG, C. O. **Estimation of time-dependent origin-destination matrices for transit networks.** Transp. Res. Part B Methodol., 1998.

XIAN, X., YE, H., WANG, X. & LIU, K. **Spatiotemporal Modeling and Real-Time Prediction of Origin-Destination Traffic Demand.** Technometrics, 2019.

XU, C., JI, J. & LIU, P. **The station-free sharing bike demand forecasting with a deep learning approach and large-scale datasets.** Transp. Res. Part C Emerg. Technol., 2018.

YANG, Ci; GONZALES, Eric J. **Modeling taxi trip demand by time of day in New York City.** Transportation Research Record, 2014.

YUAN, J., ZHENG, Y., ZHANG, L., XIE, X. & SUN, G. **Where to find my next passenger?** UbiComp'11 - Proc. 2011 ACM Conf. Ubiquitous Comput., 2011.

ZHANG, S. *et al.* **Dynamic demand forecast and assignment model for bike-and-ride system.** Promet - Traffic – Traffico, 2019.

## ANEXO A

### Código utilizado para construção do modelo e posterior análise

```

1  #BIBLIOTECAS UTILIZADAS -----
2  library(starma)
3  library(spdep)
4  library(astsa)
5  library(readxl)
6  library(forecast)
7  library(Metrics)
8  library(xlsx)
9  library(qqplotr)
10 library(seastests)
11
12 #PASTA COM OS DADOS -----
13 setwd("c:/Users/maria/OneDrive - Instituto Maua de Tecnologia/Mestrado/Dissertação/TESE/Dados")
14 getwd()
15
16 #LEITURA DOS DADOS -----
17 dados_starma <- read_excel("dados.xlsx", sheet = "Dados", col_names = TRUE)
18 dados_starma <- as.matrix(dados_starma)
19 dados_starma
20
21 #GRÁFICOS (PLOTAGEM E HISTOGRAMA) -----
22 par(mfrow=c(1,1))
23 dados_starma_ts <- ts(dados_starma) #TRANSFORMAR EM SÉRIE TEMPORAL
24 plot(dados_starma_ts)
25 hist(dados_starma_ts)
26
27 #ESTADÍSTICA -----
28 xb <- mean(dados_starma) #107,83
29 sx <- sd(dados_starma) #116,03
30
31 #MATRIZ DE DISTÂNCIA -----
32 identidade <- as.matrix(read_excel("dados.xlsx", sheet = "Identidade", col_names = TRUE))
33 identidade
34 ordem1 <- as.matrix(read_excel("dados.xlsx", sheet = "Ordem1", col_names = TRUE))
35 ordem1
36 lista <- list(identidade,ordem1)
37 lista
38
39 #FUNÇÃO DE AUTOCORRELAÇÃO -----
40 par(mfrow=c(2,1))
41 stacf(dados_starma, lista) #COMPORTAMENTO CICLICO
42 stpacf(dados_starma, lista)
43

```

```

44 ## TRANSFORMAÇÃO DOS DADOS -----
45
46 # RAIZ QUADRADA
47 dados_sr <- sqrt(dados_starma)
48 hist(dados_sr)
49 qqnorm(dados_sr)
50 qqline(dados_sr)
51
52 # RAIZ CUBICA
53 dados_cr <- (dados_starma)^(1/3)
54 hist(dados_cr)
55 qqnorm(dados_cr)
56 qqline(dados_cr)
57
58 # BOX COX
59 lambda <- BoxCox.lambda(dados_starma)
60 dados_bc <- (dados_starma)^(lambda)
61 hist(dados_bc)
62 qqnorm(dados_bc)
63 qqline(dados_bc)
64
65 # LOGARITMO
66 dados_log <- log(dados_starma)
67 hist(dados_log)
68 qqnorm(dados_log)
69 qqline(dados_log)
70
71 # DEVIANCE
72 dados_deviance <- as.matrix(read_excel("dados.xlsx", sheet = "Deviance", col_names = TRUE))
73 par(mfrow=c(1,1))
74 dados_deviance_ts <- ts(dados_deviance)
75 plot(dados_deviance_ts)
76 hist(dados_deviance)
77 qqnorm(dados_deviance)
78 qqline(dados_deviance)
79
80
81 #DADOS DEVIANCE -----
82
83 hist(dados_starma)
84 hist(dados_deviance)
85 plot(dados_deviance_ts)
86
87
88 #TESTE DE NÃO CORRELAÇÃO DOS DADOS
89 stcor.test(dados_deviance_diff1, lista, fitdf=0) #OS DADOS SÃO CORRELACIONADOS ESPACIALMENTE
90
91 #FUNÇÃO DE AUTOCORRELAÇÃO DADOS DEVIANCE
92 stacf(dados_deviance, lista)
93 stpacf(dados_deviance, lista)
94
95 #DIFERENCIAÇÃO
96 dados_deviance_diff1 <- (diff(dados_deviance))
97 hist(dados_deviance_diff1)
98 qqnorm(dados_deviance_diff1)
99 qqline(dados_deviance_diff1)
100 plot(ts(dados_deviance_diff1))
101
102 #FUNÇÃO DE AUTOCORRELAÇÃO DADOS DEVIANCE DIFERENCIADOS
103 stacf(dados_deviance_diff1, lista) #NÃO É MAIS CÍCLICO
104 stpacf(dados_deviance_diff1, lista)
105
106 #HISTOGRAMA -----
107
108 par(mfrow=c(1,1))
109 hist(dados_starma_ts, col = "darkgrey", main = "Dados", breaks = 20, freq = F,
110      xlab = "Dados Iniciais Sem Transformação", ylab = "Densidade")
111 curve(dnorm(x, mean = mean(dados_starma_ts), sd = sd(dados_starma_ts)), add = T)
112
113 hist((dados_starma_ts^(0.5)), col = "darkgrey", main = "Dados", breaks = 20, freq = F,
114      xlab = "Transformação Raiz Quadrada", ylab = "Densidade")
115 curve(dnorm(x, mean = mean(dados_starma_ts^(0.5)), sd = sd(dados_starma_ts^(0.5))), add = T)
116
117
118 hist(dados_deviance_ts, col = "darkgrey", main = "Dados", freq = F, xlab = "Transformação Deviance",
119      ylab = "Densidade", nclass = 20)
120 curve(dnorm(x, mean = 0.17, sd = sd(dados_deviance_ts)), add = T)
121
122

```

```

123 #MODELO -----
124
125 #MODELO 1
126 ar <- matrix(c(1, 1, 1, 0), 2, 2)
127 ma <- matrix(c(0, 1), 1, 2)
128 model1 <- starma(dados_deviance, lista, 2, 2)
129 model1$bic
130 model1
131 summary(model1)
132 stacf(model1$residuals, lista)
133 stpacf(model1$residuals, lista)
134
135 #MODELO 2
136 ar2 <- matrix(c(2, 3, 4, 5), 2, 2)
137 ma2 <- matrix(c(0, 1), 1, 2)
138 model2 <- starma(dados_deviance_diff2, lista, ar = ar2, ma = ma2)
139 model2$bic
140 model2
141 summary(model2)
142 stacf(model2$residuals, lista)
143 stpacf(model2$residuals, lista)
144
145 #MODELO 3
146 ar3 <- matrix(c(1,1,0,0,0,1,1,0,0,0,0,1,0,0,1,0,0,1,0,1,1,1,1,0,1,1,0,1,0,0,1,1), 16, 2)
147 ma3 <- matrix(c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,0,1,1,1,1,1,1,1,1,1,0,1,1,1,1), 16, 2)
148 model3 <- starma(dados_deviance_diff1, lista, ar = ar3, ma = ma3)
149 model3$bic
150 model3
151 summary(model3)
152 #ANÁLISE DOS RESÍDUOS
153 stacf(model3$residuals, lista, tlag.max = 24)
154 stpacf(model3$residuals, lista, tlag.max = 24)
155 #QQ-PLOT
156 qqnorm(model3$residuals, main='Normal QQ plot - Model3')
157 qqline(model3$residuals)
158 hist(model3$residuals)
159
160 hist(model3$residuals, col = "darkgrey",main = "Resíduos", breaks = 20, freq = F, xlab = "Resíduos do Modelo",
161      ylab = "Densidade")
162 curve(dnorm (x, mean = mean(model3$residuals), sd = sd(model3$residuals)),add = T)
163
164
165 #RESUMO DO MODELO -----
166
167 #DADOS
168 dados_deviance_diff1
169 plot(ts(dados_deviance_diff1))
170
171 #RESÍDUOS DO MODELO
172 residual <- model3$residuals
173
174 #VALORES AJUSTADOS
175 fit_values <- dados_deviance_diff1 - residual
176 plot(ts(fit_values))
177
178 #TESTE DE NÃO CORRELAÇÃO
179 stcor.test(model3$residuals, lista, fitdf=16) #OS RESÍDUOS SÃO CORRELACIONADOS ESPACIALMENTE
180
181 #ESCREVER UM EXCEL COM OS DADOS
182 write.csv(fit_values,"fit_values.csv", row.names = TRUE)
183
184 #CÁLCULO DOS ERROS
185 mae <- mae(dados_deviance_diff1,fit_values) #0.4611
186 rmse <- rmse(dados_deviance_diff1,fit_values) #0.6103
187 mape #0,4666
188 mse(dados_deviance_diff1,fit_values) #0,3724

```