

EDUARDO JUN SHINOHARA

Modelagem do comportamento espaço-temporal de veículo rastreado

SÃO PAULO
2013

EDUARDO JUN SHINOHARA

Modelagem do comportamento espaço-temporal de veículo rastreado

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

Área de concentração:
Engenharia de Transportes
Subárea: Informações Espaciais

Orientador: Professor Livre Docente
José Alberto Quintanilha

SÃO PAULO
2013

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

São Paulo, de novembro de 2013.

Assinatura do autor _____

Assinatura do orientador _____

FICHA CATALOGRÁFICA

Shinohara, Eduardo Jun

**Modelagem do comportamento espaço-temporal de veículo rastreado / E.J. Shinohara. – versão corr. -- São Paulo, 2013.
72 p.**

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

1. Sistemas inteligentes de transportes 2. Veículos (Monitoramento) 3. Sistemas de informações geográficas I. Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia de Transportes II. t.

AGRADECIMENTOS

Ao Prof. José Alberto Quintanilha pela orientação, amizade e pela confiança, à Professora Linda Lee e ao Prof. Marcos Rodrigues, que sem o preciso apoio esta dissertação não seria concretizada.

Aos colegas e amigos do Departamento de Engenharia de Transportes e do Laboratório de Geoprocessamento pelo apoio em especial a Claudia, á Mariana, á Alessandra, ao Manfre, à Arlete, à Janaina, à Martinha, à Eliane, à Patrícia e a Rosangela.

Aos amigos em especial ao Jorge Lopes, pelo auxílio com os dados, pelas risadas e boas conversas.

RESUMO

No Brasil existe a perspectiva de crescimento expressivo do volume de dados a ser processado pelas prestadoras de serviços de rastreamento em decorrência do aumento natural do uso de sistemas de rastreamento e também para atender a Resolução 330 de 2009 e Deliberação 135 de 30/01/2013 do Conselho Nacional de Trânsito (CONTRAN). Este crescimento gera a necessidade da incorporação de ferramentas analíticas nos sistemas de gerenciamento do rastreamento e monitoramento de veículos e na gestão de risco, para aumentar a sua eficiência e atender o crescimento do mercado.

O objetivo desta dissertação é de propor uma metodologia que permita caracterizar o comportamento de movimentação de um veículo, com a finalidade de auxiliar o processo de tomada de decisão no gerenciamento e monitoramento de veículos.

A caracterização do comportamento de movimentação do veículo foi feita pela geração de um modelo analítico do comportamento de movimentação, coletando os dados pretéritos da posição espacial e temporal. Este modelo baseia-se na movimentação e considera os aspectos comportamentais espaciais e temporais de forma independente. A caracterização do comportamento gera informações para identificar o comportamento espacial e temporal do veículo monitorado para um determinado nível de confiabilidade.

Palavras Chave: Sistema de Informações Geográficas, Análise Espacial, Sistemas Inteligentes de Transportes, Rastreamento de Veículos

ABSTRACT

In Brazil there is the prospect of growth in the volume of data to be processed by the tracking service providers due to the natural increase of the use of tracking systems and also to meet the Resolution 330 of 2009 and Resolution 135 of 01.30.2013 of the National Traffic Council (CONTRAN). due to this growth the need of incorporation of analytical tools in systems management tracking and monitoring of vehicles and risk management are created, to increase their efficiency and meet market growth.

This study aims is to propose a methodology to characterize the moving vehicle behavior, in order to assist the process of decision making in management and vehicle tagging.

The vehicle handling behavior will be characterized by generating an analytical model of the vehicle movement, collecting bygone data of spatial position and time. This model will consist of a motion model taking into account that the spatial and temporal aspects of behavior are taken independently. The behavior characterization generates reports able to identify the spatial and temporal behavior of the monitored vehicle for a given level of reliability.

Keyword: Geographic Information System (GIS), Spatial Analysis, Intelligent Transportation System (ITS) , Automatic Vehicle Location (AVL), Vehicle Monitoring

Sumário

1.	Introdução	1
2.	Objetivo	2
3.	Revisão bibliográfica	2
3.1.	Mercado das empresas de Rastreamento de veículos	2
3.2.	Rastreamento de veículos	3
3.2.1.	Posicionamento.....	4
3.2.2.	Comunicação	7
3.2.3.	Gestão da informação	8
3.3.	Estudo do comportamento espaço-temporal	9
3.3.1.	Comportamento espacial - conceitos.....	9
3.3.2.	Objetos móveis	12
3.3.3.	Padrões de trajetória	13
3.3.4.	Análise das trajetórias.....	16
3.3.5.	Análise das atividades diárias.....	17
3.3.6.	Análise de tempo de viagem.....	18
3.3.7.	Ferramentas de análise.....	20
3.3.8.	Padrões de distribuição espacial – transporte publico	21
3.3.9.	Equipamentos para coleta de dados.....	23

3.3.10. Técnicas e métodos selecionados	25
4. Metodologia.....	31
4.1. Obtenção dos dados de rastreamento dos dados	31
4.2. Ferramentas utilizadas para a análise.....	31
4.3. Análise dos dados de movimentação de um veículo.	32
4.4. Pré-processamento dos dados de rastreamento.....	34
4.5. Comportamento Temporal	36
4.5.1. Estimação dos coeficientes das preditoras do modelo.....	37
4.6. Comportamento Espacial	40
4.6.1. Análise das distribuições espaciais.....	41
5. Experimento	46
5.1. Caracterização dos dados coletados.....	46
5.2. Modelagem temporal	47
5.2.1. Avaliação da significância das preditoras	48
5.2.2. Avaliação para agregação das preditoras.....	51
5.2.3. Constatações da modelagem temporal	54
5.3. Modelagem espacial.	58
6. Conclusão	60
7. Referencias Bibliografias	61

Lista de Figuras

Figura 1 - Componentes do sistema de AVL	4
Figura 2 Esquema da Constelação de GPS	5
Figura 3- Esquema de Posicionamento GPS	5
Figura 4 - Sistemas híbridos – A-GPS	7
Figura 5 - Posições espaciais de um veículo.....	9
Figura 6 - Representação em prisma (temporal e espacial)	10
Figura 7 - Prisma Espaço Tempo fonte: Yu(2007).....	10
Figura 8 - Probabilidade de ocorrência do evento	26
Figura 9 - Estimador de Kernel.....	26
Figura 10 - Função densidade de probabilidade, para valores de a 1 (\equiv); 0,7($--$) e 0,5 (\cdots) (adaptado de	28
Figura 11 - Função distribuição de probabilidade – distribuição Weibull para valores de α e γ	29
Figura 12 - Função distribuição de probabilidade Log normal	30
Figura 13 - Esquema para análise da movimentação do veículo	33
Figura 14 - Exemplo do método gráfico – gerado no software minitab	37
Figura 15 - Representação dos pontos amostrados	41
Figura 16 - Exemplos de análise de envoltórias	42
Figura 17 - Mapa de isoprobabilidade de ocorrência do evento.....	42
Figura 18 - Padrões de distribuição espacial da movimentação de veículos	43

Figura 19 - Distribuição difusa	44
Figura 20 - Distribuição bifocal.....	44
Figura 21 - Distribuição multifocal	45
Figura 22 - Distribuição com um foco.....	46
Figura 23 - Probabilidade de o veículo estar parado	55
Figura 24 - Gráficos que representam a probabilidade do veículo estar parado	56
Figura 25 - Gráficos que representam os deslocamentos	57
Figura 26 - Distribuição espacial dos eventos de monitoramento	58
Figura 27 - Estimador de Kernel (probabilidade acumulada de 99%)	59
Figura 28 – Identificação das localidades	59

Lista de Tabelas

Equação 1 - Cálculo do Kernel	26
Equação 2 - Modelo de distribuição exponencial	27
Equação 3 - Modelo de distribuição Weibull.....	28
Equação 4 - Modelo de Distribuição Log-normal:	29
Equação 5 - Função de verossimilhança com censura aleatória:	30
Equação 6 - Determinação da distância entre dois pontos.....	34
Equação 7 - Equação de regressão	38
Equação 8 - Generalização do teste T de Student	39
Equação 9 - Teste da Razão de Máxima Verossimilhança (TRV).....	40

Lista de siglas

2G - Segunda geração da tecnologia de comunicação móvel

A-GPS - Assisted GPS ou GPS assistido

AVL - Automatic Vehicle Location

BLOB - Binary Large Object

CELL_ID - Identificação de Célula - Comunicação Móvel

CESVI - Centro de Experimentação e Segurança Viária

CONTRAN - Conselho Nacional de Transito

DoD - United State Departament of Defence

GPRS - General Packet Radio Service, Padrão de transferência de dados em telefonia celular

GPS - Global Positioning System

GRISTEC - Associação Brasileira das Empresas de Gerenciamento de Risco e de Tecnologia de Rastreamento e Monitoramento

GSM - Groupe Special Mobile, Padrão para comunicação de telefonia celular

HDF - Hierarchical Data Format

ITS - Intelligent Transportation System

NetCDF - Network Common Data Form

PDA - Persona Digital Assistant

XML - Extensible Markup Language

1. INTRODUÇÃO

A utilização de sistemas de rastreamento de veículos, segundo Roth (1977) iniciou-se nas primeiras décadas do século XX, para o controle de frotas de veículos de transporte público urbano. No final da década de 60 (século XX) foi implantado o primeiro sistema automatizado para a gestão de frotas e de tráfego em Chicago. Segundo Rodrigues *et al.* (2009) estes sistemas foram denominados inicialmente de Automatic Vehicle Location (AVL) e no Brasil a utilização de sistemas de rastreamento de veículos iniciou-se em aplicações de segurança patrimonial e pessoal, com a finalidade de prevenir roubos e/ou recuperação de bens.

Existe a perspectiva de crescimento expressivo do volume de dados a ser processado pelas prestadoras de serviços de rastreamento, em decorrência do aumento natural do uso de sistemas de rastreamento e também para atender a Resolução 330 de 2009 e Deliberação 135 de 30/01/2013 do Conselho Nacional de Trânsito (CONTRAN), Esta resolução estabelece a instalação de equipamentos de rastreamento em todos os veículos novos, a partir de junho de 2013, que até o momento da finalização desta dissertação estava em vigor. Este crescimento gera a necessidade da incorporação de ferramentas analíticas nos sistemas de gerenciamento do rastreamento e monitoramento de veículos e na gestão de risco, para melhorar a sua eficiência e atender o crescimento do mercado.

Entretanto, o processo de análise dos dados provenientes do rastreamento de veículos carece de ferramentas analíticas que possam auxiliar o gerenciamento e a tomada de decisão, principalmente no gerenciamento de frotas e análise de risco, O mercado de rastreamento segundo Gonçalves (2012) é constituído principalmente por empresas de pequeno porte e não certificadas por entidade reconhecida pelo mercado, estas empresas representam 90 % do mercado. As empresas não certificadas carecem de instrumentos de análise automatizadas, utilizando processos manuais para a análise dos dados de rastreamento. Outro fator é o grande volume de dados de posicionamento gerado pelos sistemas de rastreamento, que dificulta a análise não automatizada.

Uma das análises que pode ser realizada pelas prestadoras de serviço de rastreamento é a detecção de anormalidades no comportamento da movimentação esperado de um

determinado veículo. Para tal é necessário verificar se um dado posicionamento do veículo é o esperado temporalmente e espacialmente, informando, por exemplo, se o veículo rastreado está se movimentando no local e horário previstos.

2. OBJETIVO

O presente trabalho propõe uma metodologia que permite caracterizar o comportamento de movimentação de um veículo, com a finalidade de auxiliar o processo de tomada de decisão no gerenciamento e monitoramento de veículos.

Para a caracterização do comportamento de movimentação do veículo foi desenvolvido um processo para geração de um modelo analítico do comportamento de movimentação, coletando os dados pretéritos da posição espacial e temporal. A partir dessa série histórica dos dados de rastreamento, foi construído um modelo de movimentação considerando os aspectos comportamentais espaciais e temporais, obtendo de forma independente as características do comportamento de movimentação temporal e do espacial. Com esta caracterização do comportamento identificam-se os períodos e os locais de utilização habituais do veículo.

3. REVISÃO BIBLIOGRÁFICA

3.1. Mercado das empresas de Rastreamento de veículos

O mercado de sistemas de monitoramento de veículos no Brasil está em crescimento desde a década de 1980. Apesar de não existir uma estatística oficial sobre a utilização de equipamentos de monitoramento, Buonavoglia (2011) declarou que em 2011 haviam 1,4 milhões de veículos monitorados e com uma perspectiva de crescimento de 20 % ao ano.

Segundo GRISTEC (2012), constatou-se que seus associados possuem sistemas de rastreamento que armazenam além do posicionamento, dados operacionais e de telemetria. Estes dados estão sendo utilizados para o gerenciamento de frotas, programas de manutenção de veículos, controle logístico e especialmente para a gestão de segurança. A utilização de equipamentos de posicionamento e de comunicação embarcados é comum em veículos

comerciais de transporte de carga e, com o seu barateamento, está ocorrendo expansão para o setor de veículos particulares principalmente em aplicações de gerenciamento de segurança.

Os dados de posicionamento obtidos pelo monitoramento são utilizados para o acompanhamento de progresso das viagens ou para a localização do veículo em caso de sinistros (termo utilizado pelo mercado de seguros para dano para qualquer bem segurado). Os dados na maioria dos casos não são analisados de forma automatizada, existem apenas normas e procedimentos aplicados de forma manual com o apoio de ferramentas computacionais, como gerenciadores de banco de dados e visualizadores de mapas e segundo, Rocha (2012), o fator humano ainda é fundamental no processo da análise do monitoramento.

Ainda não há consenso, mas segundo Gonçalves (2012) e GRISTEC (2012), as empresas que realizam serviços de monitoramento podem ser classificadas em duas categorias, as de gerenciamento de risco e de rastreamento e monitoramento, e são caracterizador da seguinte forma:

- Gerenciamento de risco: são procedimentos para minimizar preventivamente as perdas materiais, financeiras e físicas. Estes procedimentos envolvem, além do rastreamento, a roteirização, manutenção preventiva, escolta e treinamento.
- Rastreamento e Monitoramento: oferecem serviços de bloqueio e de rastreamento da posição do veículo, as aplicações mais comuns encontradas são a visualização da localização do veículo, controle da frota e bloqueio remoto do veículo.

Segundo Pereira (2012), nos serviços de gerenciamento de risco que envolve o acompanhamento e o gerenciamento das viagens, são utilizados procedimentos que verificam o comportamento do veículo monitorado. Esses procedimentos algumas vezes são realizados de forma não automatizada e ficam sob a responsabilidade de um operador.

3.2. Rastreamento de veículos

Os métodos e as técnicas utilizadas no monitoramento e gerenciamento da movimentação de veículos são denominados de AVL. Segundo Rodrigues *et al.* (2009) o AVL pode ser definido como sendo um conjunto de tecnologias de posicionamento e comunicação, empregado para o gerenciamento da movimentação atendendo operações de

controle logístico, de segurança, de gerenciamento de frotas, de gerenciamento de transporte público, entre outras (Figura 1).

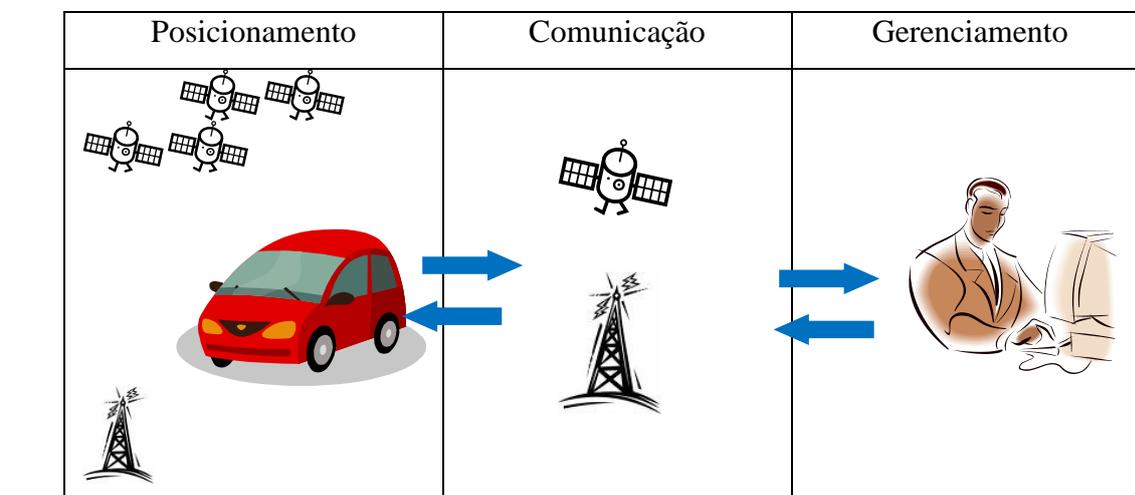


Figura 1 - Componentes do sistema de AVL (Adaptado de Rodrigues *et al.*, 2009)

A Revista CESVI n° 84 (2013), apresenta os sistemas de rastreamento aprovados na sua avaliação. Para o setor de automóveis estão certificados 37 sistemas, destes apenas 2 não utilizam o sistema Global Positioning System - GPS para determinar o posicionamento e 4 não utilizam a tecnologia de comunicação celular com GSM/GPRS. Entre as tecnologias existentes de posicionamento e comunicação, o GPS e o GSM/GPRS são as mais empregadas,

3.2.1. Posicionamento

O GPS foi concebido como um sistema de navegação, com o objetivo principal de determinar instantaneamente a posição, velocidade de um usuário, em qualquer lugar, em um mesmo referencial. Este sistema foi desenvolvido pelo Departamento de Defesa Norte Americano (DoD) e segundo Kaplan e Hegary (2006) foi considerado operacional em 1985.

O sistema é composto por 24 satélites que estão posicionados em seis planos orbitais a uma altitude aproximada de 20.200 km (Figura 2). Segundo informações oficiais divulgados pelo Governo Americano no site www.gps.gov (consultado em julho de 2013) o último lançamento foi realizado em 15 de maio de 2013. Esta configuração garante que pelo menos quatro satélites estejam visíveis em qualquer local da superfície da terra a qualquer hora do dia.

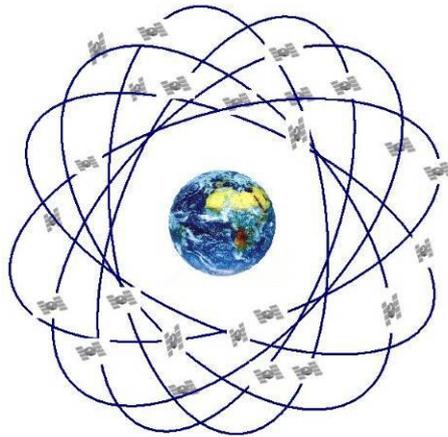


Figura 2 Esquema da Constelação de GPS (fonte: http://www.fc.up.pt/lic_eg/)

Segundo Blitzkow (2007) nos satélites estão embarcados equipamentos que geram padrões de frequência estáveis (rubídio e césio), com estes padrões obtém-se uma referência de tempo estável. Para a determinação da posição do receptor GPS é necessário inicialmente determinar a distância entre o satélite e o receptor, para isso, são utilizados os dados da posição do satélite e a correção do relógio, que é a diferença de tempo entre relógio do receptor e o relógio do satélite. Segundo Blitzkow *op.cit* para calcular a posição do receptor será necessário solucionar equações que tem como incógnitas as coordenadas cartesianas (X,Y,Z) e a correção do relógio Δt . Para este cálculo a equação da distância possui quatro incógnitas, portando para a solução desta equação devemos captar simultaneamente o sinal de quatro satélites para solucionar a equação das distâncias (Figura 3).

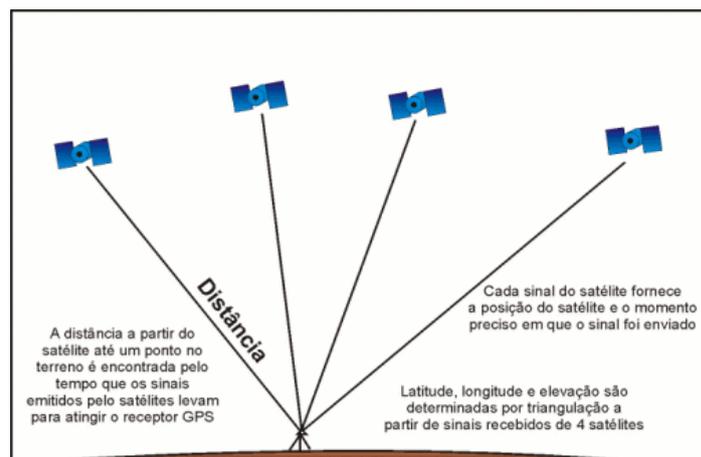


Figura 3- Esquema de Posicionamento GPS (fonte:<http://www.nara.org.br/servicos/ntp/GPS>)

Apesar da abrangência global e dos baixos preços dos receptores, estes sistemas segundo Monico (2008), possuem limitações principalmente em áreas urbanas densamente ocupadas, túneis, garagens subterrâneas e em ambientes fechados aonde os sinais dos satélites são facilmente obstruídos, impedindo a determinação da posição.

- **Sistemas Híbridos / A-GPS (*Assisted GPS*)**

Para atenuar os problemas causados pelas perdas de sinal de GPS em áreas urbanas, foi criado um método que utiliza a rede de telefonia celular para auxiliar as técnicas de posicionamento com GPS. Esta técnica segundo Djuknic *et al.* (2001) consiste no envio para o receptor GPS, através da rede de comunicação de dados, dos dados pré-processados sobre a posição dos satélites. Esta técnica reduz o tempo de cálculo para a determinação do posicionamento. O esquema de funcionamento está demonstrado na Figura 4.

O A-GPS permite utilizar outras técnicas de posicionamento como, por exemplo, o *Cell_id* ou a triangulação de rádio base, para auxiliar a determinação do posicionamento do receptor.

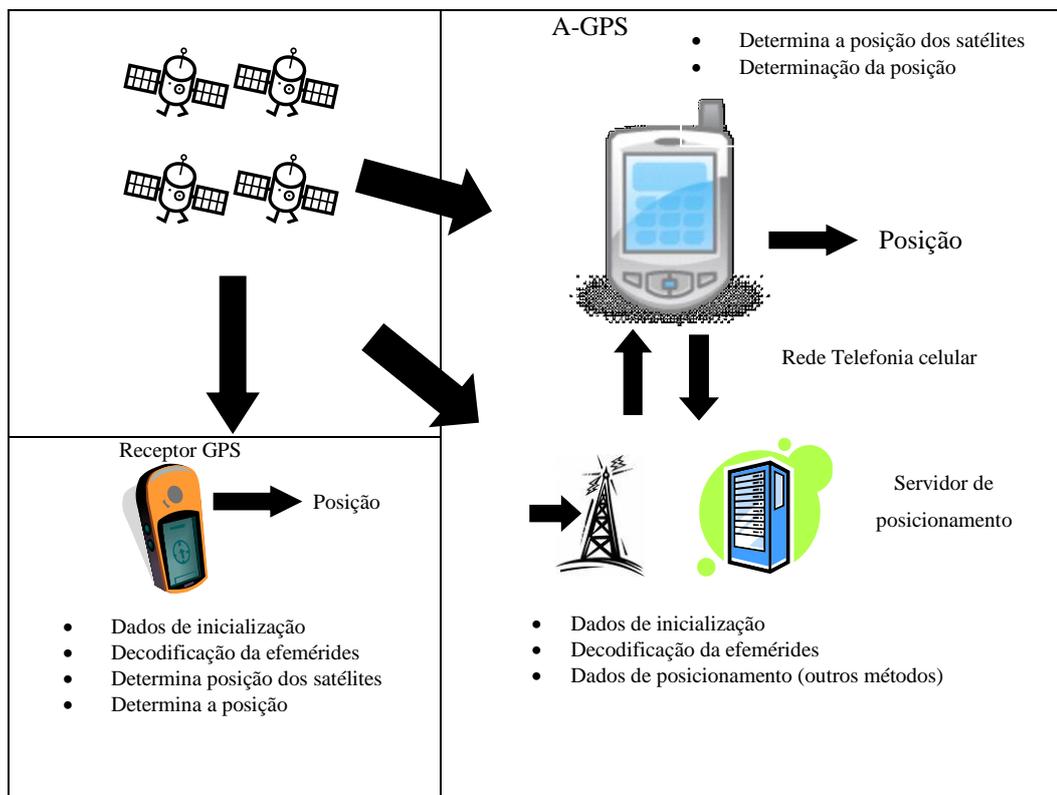


Figura 4 - Sistemas híbridos – A-GPS (Fonte: adaptado de Wang *et al* (2008))

3.2.2. Comunicação

Os sistemas de comunicação são utilizados para a troca de dados, texto ou comunicação por voz entre o veículo e a central de comunicação. A tecnologia empregada dependerá do tipo de operação realizada.

- **Via Telefonia móvel**

O sistema de telefonia móvel permite a troca de dados e a comunicação através de voz e texto. A grande área de cobertura, principalmente, nos centros urbanos e nas rodovias. Segundo dados da Anatel compilados pela TELECO em Maio de 2013, todos os municípios Brasileiros possuem cobertura celular. Por estas características este é o meio de comunicação predominante em sistemas de AVL.

A segunda geração, ou 2G, segundo Rodrigues *et al.* (2009), incorporou os sistemas digitais e a transmissão de dados. A taxa de transmissão de dados é relativamente baixa (9.600

bps). Os protocolos de comunicação utilizados para a transmissão de dados, não permitem a troca de dados de forma sincronizada e contínua. A evolução destes protocolos de comunicação viabilizou a comunicação de dados de modo síncrono e contínuo, com uma taxa de transmissão de até 384 kbps, esta evolução da segunda geração denominou de geração 2,5G.

A terceira geração das tecnologias para comunicação de celular é denominada de 3G, segundo Rodrigues *et al.* (*op. cit.*), incorporam protocolos de comunicação que permitem uma taxa de transmissão de dados maior (1 Mbit/s para upload) além de serviços adicionais como, por exemplo, o de mensagens multimídia.

3.2.3. Gestão da informação

A central de operação é responsável pela gestão da comunicação incluindo a integração dos diversos meios de comunicação, a operação de armazenamento, o processamento de dados e a gestão da informação.

A gestão da informação consiste em prover ao operador instrumentos para executar aplicações específicas para cada tipo de operação e distribuir os dados através de consultas ou pela geração de relatórios.

Para o tratamento das informações espaciais são utilizados aplicativos que utilizam Sistemas de Informações Geográficas (SIG) e uma base de dados cartográfica. Com isso um sistema de AVL é capaz de identificar as posições espaciais de um veículo em um determinado período de tempo, sobre uma base cartográfica (Figura 5).

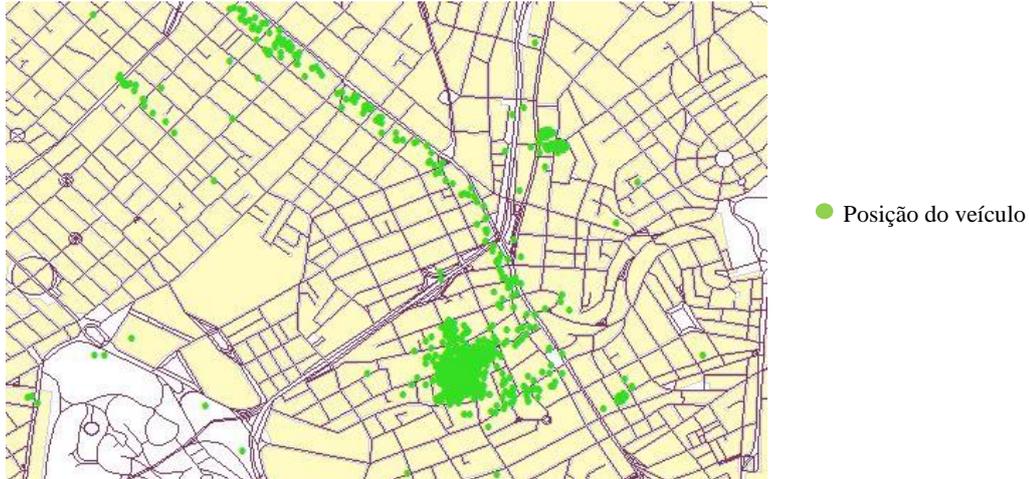


Figura 5 – Posições espaciais de um veículo

3.3. Estudo do comportamento espaço-temporal

3.3.1. Comportamento espacial - conceitos

As pesquisas sobre o comportamento espacial das populações e das pessoas tiveram início com Hägerstrand na década de 60 do século XX, os conceitos apresentados no artigo “*What about people in regional Science*” (Hägerstrand(1970)), são a base conceitual da geografia do tempo (“*Time geography*”), que estuda a movimentação dos indivíduos durante sua vida e considera que as variáveis de tempo e de espaço estão intrinsecamente relacionadas. Outra contribuição de Hägerstrand foi a elaboração de uma representação para o comportamento diário da movimentação de um indivíduo, denominada representação em prisma espaço-tempo, como pode ser observado na Figura 6. Esta representação mostra o comportamento de um indivíduo no decorrer do dia, nos eixos horizontais são representados os locais onde o indivíduo esteve, e no eixo vertical representamos o tempo, isso é, em que hora do dia o indivíduo esteve em um determinado local.

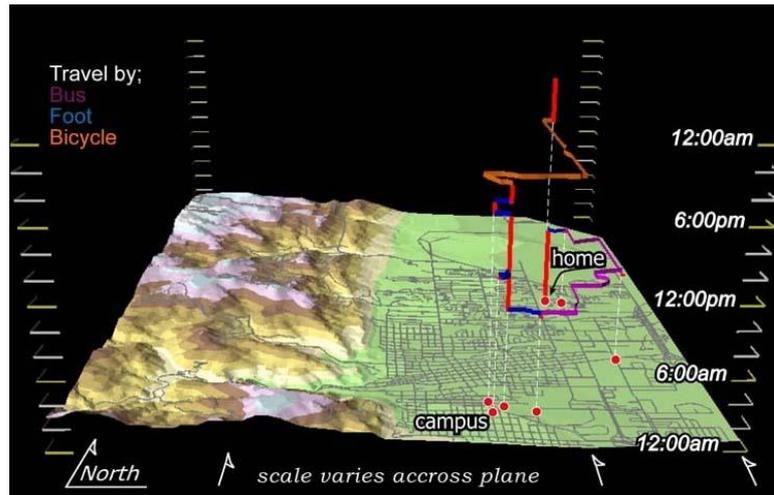


Figura 6 - Representação em prisma (temporal e espacial)

Fonte: http://www.colorado.edu/geography/cartpro/cartography2/spring2001/dettloff/time/prism_map.html (2005)

Com o desenvolvimento das ferramentas de SIG na área de transportes, Goodchild (2000) definiu o GIS-T como sendo as ferramentas e análises de SIG para a área de transportes. Aliando os conceitos de SIG com os da geografia do tempo (Time geography), Yu (2007) construiu ferramentas para visualizar e analisar as viagens e as atividades gerando assim um panorama do comportamento diário de um objeto móvel, onde este é representado pelos prismas espaço tempo (Figura 7).

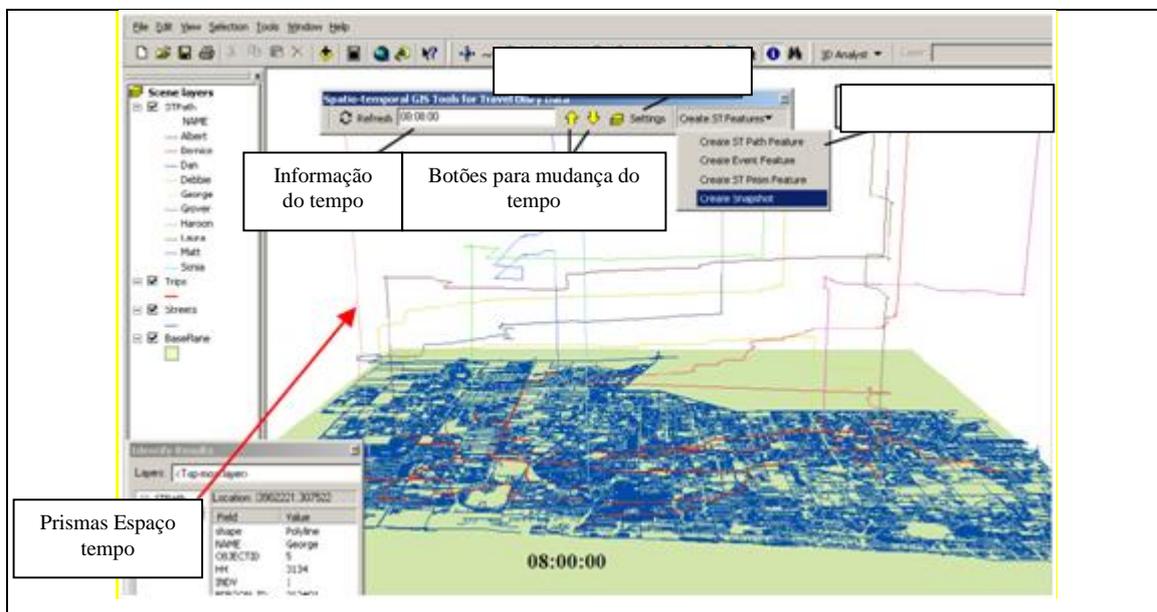


Figura 7 -- Prisma Espaço Tempo fonte:Yu(2007)

Existem diferentes aspectos que estão sendo tratados na literatura relacionada ao estudo e análise dos padrões de movimentação de objetos móveis que no caso desta Dissertação serão os veículos que possuem sistemas de rastreamento.

Inicia-se destacando a importância de se estudar objetos móveis e seus padrões de comportamento. Algumas das diferentes análises foram propostas para determinar padrões espaciais, temporais e espaço-temporais. Identifica-se grupos de bibliografias, referentes às análises individuais (um objeto móvel) e de grupos (em geral, frotas). Um ponto comum a muitas das bibliografias, tanto nas de aplicação em transportes quanto nas que relatam o desenvolvimento de algoritmos computacionais, é a busca pela segmentação das viagens e, no caso dos transportes, mais enfaticamente, a possibilidade de se associar um significado aos trechos de trajetórias. No caso específico de transportes, muito se tem desenvolvido para associar o modo de viagem utilizado em cada diferente trecho. O outro enfoque identificado é o de encontrar as similaridades entre as trajetórias e como medi-las. Há também, a constatação que o modelo espaço-temporal é muito usado e a utilização do prisma espaço tempo pelos usuários da área de transportes é frequente.

De acordo com González *et al.* (2008), apesar da importância do entendimento das leis básicas que governam o movimento humano permanecem limitadas à própria falta de ferramentas para monitorar o tempo e a localização (“the time-resolved location”) dos indivíduos. No estudo de 100.000 usuários anônimos de telefone celular cuja posição foi rastreada por seis (6) meses, mostrou que as trajetórias humanas apresentam um alto grau de regularidade espacial e temporal, sendo cada indivíduo caracterizado por uma distância própria de viagem, independentemente do tempo, e com significativa probabilidade de retornar aos lugares altamente frequentados. Após correções necessárias devido às distâncias nas viagens e as anisotropias intrínsecas a cada trajetória, os padrões de viagens recaem em distribuições de probabilidade espacial bem definida, indicando que, os seres humanos seguem padrões simples e reprodutíveis. Essa similaridade em padrões de viagem pode impactar em todos os fenômenos voltados à mobilidade urbana, planejamento urbano e modelagem baseada em agentes (“agent based modelling”).

3.3.2. Objetos móveis

Markus Schneider (<http://www.cise.ufl.edu/~mschneid>) é, na área de computação voltada aos objetos móveis, um dos maiores geradores de publicações qualificadas, adotadas na maioria dos melhores cursos de Graduação e Pós Graduação que abordam dados espaciais, além de ser também, dos mais citados. Como exemplos temos o livro *Moving Objects Database* (Güting e Schneider, 2005) pioneiro no tema de objetos móveis e vários artigos e capítulos de livros (alguns são Schneider 2009a, b, c e 2010). Em palestra (Schneider, 2012) afirma que bases de dados espaciais são bases de dados especializadas (“full-fledged databases”) que, adicionalmente, permitem o armazenamento, a recuperação, a manipulação, a busca e análise de geometrias como as representações por pontos, linhas e regiões, como por exemplo, as geometrias das cidades, estradas e estados, respectivamente. Exemplos mais complexos são as partições espaciais representando subdivisões do espaço como municípios e grafos espaciais representando redes espaciais como as de transporte. Bases de dados de objetos móveis são bases de dados espaciais que permitem a manipulação contínua da evolução das geometrias eradas pelos deslocamentos espaço-temporais do objeto móvel.

As geometrias variantes que são geradas pelos deslocamentos de um indivíduo no tempo são chamadas de objetos móveis. Exemplos são pontos móveis que representam usuários de celular, linhas móveis representando o tráfego congestionado e regiões móveis representando áreas atingidas por furacões. Güting e Schneider *op. cit.* argumentam que os tipos de dados espaciais, como comentado antes, e tipos de dados espaço-temporais, fornecem uma abstração fundamental para modelar a estrutura geométrica de objetos no espaço, a temporalidade envolvendo essa estrutura no espaço e no tempo, seus relacionamentos, propriedades e operações. Tais tipos de dados desenvolvem seu pleno poder de expressão se estão integrados como tipos de dados abstratos nas bases de dados e nas linguagens de busca (“query languages”). Os autores mencionam como tópicos de pesquisa, a compartimentação espacial e as redes espaciais como os principais em bases de dados espaciais, objetos estáticos e variando no tempo como terceira dimensão e a predição de movimentos de objetos.

Destaca que um problema importante é como implementar objetos espaciais e espaço-temporais no contexto de bases de dados. Esses objetos são altamente estruturados, podem ser grandes em tamanho e tem tamanhos de representação variáveis. Atualmente esses objetos são

armazenados ou em soluções específicas de bases de dados à venda (“spatial cartridges”, “extenders” e “data blades”), tipos de dados incorporados (“built-in data types”) às bases de dados (como XML e BLOB), ou formatos de arquivos científicos (como HDF e NetCDF). Todas essas abordagens têm seus inconvenientes: ou não fornecem o tipo de abstração de dados necessária ao usuário, ou não podem gerenciar grandes volumes de dados estruturados e simultaneamente fornecer as operações de alto nível necessárias, ou falham em funcionalidades básicas da base de dados (“lack fundamental database functionality”), ou não permitem atualizações aleatórias.

Uma simplificação e ao mesmo tempo uma estruturação do que se faz com dados espaço-temporais e mencionada em Giannotti *et al.* (2007): “Uma vez definida a potencialidade da utilidade dos padrões, os autores afirmam que é necessário prover um algoritmo de mineração de dados efetivo para extrair os padrões das trajetórias.”. Essa busca por padrões das trajetórias têm permeado a maioria dos trabalhos na literatura sobre o tema e não é diferente da abordagem deste trabalho.

3.3.3. Padrões de trajetória

Segundo Giannotti *et al.* (*op. cit.*), a crescente pervasividade¹ das tecnologias de aquisição de localização (GPS, redes GSM, etc.) está produzindo enormes bases de dados espaço-temporais e aumenta a oportunidade de se obter conhecimento útil sobre o comportamento do movimento, o qual promove o desenvolvimento de novas aplicações e serviços. Nesse sentido, o artigo mostra o desenvolvimento de uma extensão do paradigma de minerar sequências padrões para analisar trajetórias de objetos móveis. Introduzem o conceito de padrões de trajetórias (“trajectory pattern”) como uma descrição precisa de comportamentos frequentes em ambos os sentidos: o espacial (por ex.: regiões do espaço visitadas durante o movimento) e temporais (por ex.: a duração do movimento). Fornecem uma proposição formal, geral, desse novo (à época) problema de mineração e estudam várias e diferentes situações (os autores usam o termo: “instanciações”) de diferentes

¹ Capacidade de se difundir por toda parte, ou que tende a propagar-se ou estender-se totalmente por meio de diversos canais, tecnologias, sistemas, dispositivos etc.

complexidades. Ressaltam que os padrões espaço-temporais mostram, de forma sucinta, o comportamento cumulativo de uma população de objetos móveis e, são abstrações úteis para o entendimento de fenômenos relacionados à mobilidade (citam, como exemplo, os fluxos Origem/Destino-O/D). Mencionam que a “pervasiveness of ubiquitous technologies” gera a oportunidade e o desafio de se descobrir, automaticamente, a partir das trajetórias, padrões espaço-temporais.

Ainda Giannotti *et al.* (*op. cit.*), apresentam uma nova forma de padrão espaço-temporal sobre a agregação de comportamentos de movimentos. Este padrão, denominado de “padrão de trajetória”, representa um conjunto de trajetórias individuais que têm em comum, a propriedade de visitar a mesma sequência de lugares com tempos de viagem similares. Nesse contexto, duas noções são fundamentais: (i) as “regiões de interesse” em um dado espaço e, (ii) o “tempo de viagem típico” de objetos móveis de região para região. De fato, nesse enfoque, um padrão de trajetória é uma sequência de regiões espaciais que, com base numa fonte de dados de trajetória, surge como frequentemente visitada no sentido especificado pela sequência; Adicionalmente, a transição entre duas regiões consecutivas na sequência, é chamada de “tempo de viagem típico”, que emerge, novamente, das trajetórias de entrada.

Os autores fornecem três diferentes algoritmos referentes a três enfoques de crescente complexidade ao tratar a dimensão espacial. Os dois (2) primeiros enfoques diferem se as regiões de interesse são previamente definidas ou definidas à partir dos dados. Diferem, portanto, na forma de discretizar o espaço. Em ambos os casos, a dimensão temporal é usada para minerar as sequências:

- Método 1, as regiões de interesse são pré-concebidas e utiliza-se de conhecimento prévio para se identificar os locais sabidamente atratores de movimento (O/D, por exemplo).
- Método 2 as regiões são chamadas de “regiões populares”. Neste caso as regiões de interesse são identificadas como os lugares mais visitados nas trajetórias fornecidas. Os autores introduzem o conceito de “discretização espacial baseada em densidade”.
- Método 3 é um algoritmo genuinamente espaço-temporal. A região de interesse é dinamicamente identificada minerando as sequências com a informação temporal. Esse enfoque detecta mais precisamente os padrões de trajetórias e as regiões de interesse são identificadas como regiões localmente densas.

O trabalho desses autores, difere da abordagem desta Dissertação onde se busca um padrão de trajetória de um único veículo/objeto móvel sem considerar as particulares regiões

de interesse. Entretanto o método 2 acima, foi naturalmente consideradas no modelo desenvolvido nesta Dissertação, sem a necessidade de se definir tais áreas como potenciais regiões de interesse. A detecção seguida, no tempo, de regiões fora dos padrões do objeto móvel/veículo considerado, no caso desta Dissertação, podem, persistindo por um intervalo de tempo maior, deixarem de ser consideradas “outsiders²” para se tornarem regiões de interesse.

Moreno *et al.* (2010) tratam de dados de trajetórias que são gerados por amostras de pontos, os quais são difíceis de entender e de analisar, porque muitas vezes são coletados sem informação semântica. Este é o caso dos dados utilizados nesta Dissertação: os dados coletados são somente a localização do ponto e o instante (tempo) de coleta. Recentemente, dentre os muitos estudos desenvolvidos para analisar trajetórias (“trajectory data analysis”), Spaccapietra *et al.* (2008), propuseram um novo modelo que foi desenvolvido para tratar trajetórias com paradas e movimentos, onde as paradas são pontos (lugares) importantes das trajetórias. Baseado nesse trabalho, diferentes métodos têm sido desenvolvidos para instanciar esse modelo, com base em diferentes características, como velocidade e direção, ajudando a atribuir semânticas às trajetórias. Neste artigo, os autores avaliam o comportamento de uma trajetória considerando inicialmente as suas propriedades, tais como velocidade e mudança de direção e então, baseados nessa análise, propõem um domínio de conhecimento que descreve algumas características do domínio de aplicação para inferir os objetos parados. Uma avaliação é feita sobre objetos reais.

Classificam os trabalhos desenvolvidos para análises de trajetórias em dois (2) grandes grupos:

- A geração de padrões focados nas propriedades geométricas das trajetórias e definindo tipos de padrões de trajetórias como convergência, cruzamentos (“encounters”), ajuntamentos (“flocks”), direção (“leadership”), etc.;
- Foco em mineração e análise de amostras de pontos de trajetórias, considerando tempo e espaço.

Ainda Moreno et al.(2010) inclui a extração de grupos de trajetórias localizadas em regiões densas (Nanni e Pedreschi, 2006); grupos de trajetórias que se movimentam entre

² Valor atípico

regiões no mesmo intervalo de tempo (Giannotti *et al.*, 2007); padrões de trajetórias com formas (“shapes”) similares (Kruipers *et al.*, 2006) ou distâncias similares (Pelekis *et al.*, 2007).

3.3.4. Análise das trajetórias

Spaccapietra *et al.* (2008) é a base para vários trabalhos propondo modelos de dados que tratam trajetórias do pontos de vista conceitual, onde uma trajetória é um conjunto de lugares importantes chamados de paradas (“stops”).

Segundo os autores, o método proposto pode ser usado por dois motivos principais: descobrir o significado de uma parada “desconhecida” ou descobrir a atividade do objeto móvel dentro de uma parada conhecida. O proposto neste artigo pode ser é uma ideia que pode ser aplicada aos resultados derivados desta Dissertação, onde uma rota ou parada anteriormente desconhecida, podem passar a integrar rol de uma trajetória Não é o escopo desta Dissertação especular sobre as rotas e paradas, mas somente observar que rotas e paradas não conhecidas não são um problema de segurança do veículo.

Segundo Amorim e Campos (2012), com a evolução da tecnologia, a captura de dados em grande escala sobre o movimento de objetos móveis, tem se tornado viável técnica e economicamente. Como resultado, há um crescente número de novas aplicações auxiliando o entendimento e gerenciamento de fenômenos complexos envolvendo esses objetos. Intelligent Transportation Systems (ITS) abrange novos tipos de aplicações planejadas para incorporar tecnologias de informação e de comunicação à infraestrutura de transportes. O principal objetivo dessas aplicações é permitir ao usuário, conhecimento sobre o funcionamento do sistema e fornece novos serviços para melhorar a coordenação e a manutenção do sistema. O ITS faz uso intensivo de dados coletados por sensores colocados ao longo da rede de transportes ou embutidos em dispositivos dentro dos veículos, para analisar e monitorar as condições das vias e o fluxo de veículos e usuários de um sistema de transporte público. Os autores fazem uma revisão de artigos que tratam dos movimentos de táxis. Assumem os conceitos e definições básicas de Spaccapietra *et al.* (2008) para o tratamento de trajetórias de objetos móveis como um conceito espaço-temporal.

Em relação a esta Dissertação, à semelhança do trabalho feito pelos autores do artigo pode ser uma etapa posterior aos trabalhos desenvolvidos na Dissertação, no sentido de conceituar formalmente e relacionar o trabalho à literatura internacional da área.

Oliveira e Batista (2012) propõem um sistema para visualização e análise de dados espaço-temporais que gerencia/trata com as 6 (seis) características necessárias a um sistema com esse propósito: insumos (“resources”) para a dimensão espacial; insumos para a dimensão temporal; independência de domínio; flexibilidade; interoperabilidade e mineração de dados baseada em agrupamentos espaço-temporais. Segundo os autores, os sistemas de visualização espaço-temporais não tratam apropriadamente a dimensão temporal, uma vez que focam na visualização espacial. Essa afirmação é importante, pois valoriza o enfoque dado nessa Dissertação, onde a modelagem e a análise temporal são privilegiadas.

3.3.5. Análise das atividades diárias

Chen *et al.* (2011) afirmam que o estudo das atividades humanas no tempo e no espaço tem sido um importante tema de pesquisa na área de transportes. Os autores citam trabalhos desde 1972 até 2009 onde dados de atividades diárias são a principal fonte de informação para os estudos realizados. Uma base de dados de atividades diárias armazena atividades feitas por uma amostra de indivíduos num particular período de tempo (seja um ou múltiplos dias), incluindo informações tais como: a hora do começo e o fim de cada atividade, onde ela ocorreu, as características da atividade e seus participantes, o que fornece informações num contexto espaço-temporal. Isso possibilita identificar padrões nas atividades humanas que podem ajudar no entendimento de como indivíduos interagem com as outras pessoas e com o ambiente. Segundo os autores: “limitações dos métodos estatísticos convencionais para analisar atividades humanas ao nível de indivíduo, encorajam o desenvolvimento de análises espaço-temporais de padrões de atividades humanas no contexto espaço-tempo”. Citam publicações que usam métodos estatísticos para analisar padrões de atividade desde 1972 até 2009 e comentam que, embora muito úteis para o estudo de características agregadas das atividades individuais, são pouco úteis para analisar os padrões dessas atividades individuais e as suas interações no contexto espaço-tempo. Citam referências que falam sobre a importância dessa integração e comentam que “um ambiente de análise integrando espaço e tempo também permite aos pesquisadores investigarem atividades

individuais como processos ao invés de eventos separados. Deste modo, análises exploratórias de dados baseadas na integração espaço-tempo podem trazer importantes contribuições para os estudos de transportes baseados em atividades.” Esta Dissertação trata, com técnicas de análise exploratória de dados, atividades individuais, para estudos de segurança, utilizando dados de espaço e tempo. No caso, as duas componentes são tratadas individualmente, sem a integração e a maior contribuição está na análise temporal.

3.3.6. Análise de tempo de viagem

Muruganandhan (2010) explora os sistemas de rastreamento (“tracking”) baseados em GPS; apresenta um sistema de AVL para rastreamento de veículos em tempo real utilizando uma unidade no veículo e um servidor de rastreamento. Faz uma distinção entre sistemas passivos quando as informações não são transferidas em tempo real e sistemas ativos quando a transmissão de dados é realizada em tempo real. O conceito adotado pelo autor de AVL é de que determina a posição geográfica de um veículo e transmite essa informação a um servidor remotamente localizado.

Embora o artigo de Flamm *et al.* (2007) trate de uma combinação entre rastreamento de pessoas por GPS (“GPS-based person tracking”) e entrevistas qualitativas. Traz uma breve revisão sobre hábitos de viagem (“travel habits”). Menciona que não é um assunto novo e que teve o interesse sobre o tema revitalizado (Garling e Axhausen, 2003). No sentido de melhorar o entendimento sobre a formação de hábitos de viagens, o comportamento de cada viagem individual deve ser analisado na sua continuidade no decorrer da vida, com o foco específico no momento em que mudanças ocorrem (mudança de casa, mudança de emprego, nascimento de filhos, etc.). Numa outra escala o problema é similar ao tratado nesta Dissertação: as mudanças observadas ocorrem em momentos específicos (não no modo de vida), mas sugerem mudanças de hábito ou ocorrências a serem investigadas. Mencionam que a localização de atividades pode ser baseada ou na velocidade ou em algoritmos de análise de densidade espacial. Em ambos os casos, um tempo mínimo de permanência define se a posição da sequência de dados (“data stream”) pode ser considerada uma atividade (ausência de movimento em um determinado instante).

As viagens são então definidas como porções da sequência de dados que conectam duas atividades. Esse conceito difere do adotado nesta Dissertação, onde uma viagem tem origem e destino definidos (esperados) e não são esperadas outras atividades no decorrer da viagem (o que é considerado um evento a ser investigado e que, eventualmente, após múltiplas constatações, será incluído na viagem).

Para Moghaddam *et al.* (2011), características do tempo de viagem de veículos em trânsito como a média e o desvio-padrão são de importância crítica tanto para o planejamento quanto para a operação de transportes. A predição dessas medidas não somente ajuda as agências de trânsito a programar e realocar recursos de forma adequada como também facilita o desenvolvimento de modos de escolha mais robustos (“more robust mode choice”) e modelos de tempo de partida. Usam dados de AVL para prever a média e o desvio-padrão dos tempos de viagem para rotas de trânsito propostas. Fazem essas estimativas de duas formas:

- Estimam a média e o desvio-padrão do tempo de viagem regredindo os valores observados contra características operacionais e da via;
- Estimam o desvio-padrão entre origens e destinos considerando os desvios-padrões dos segmentos individuais e a correlação entre esses segmentos.

Ainda segundo esses autores, muitas agências de trânsito têm suas frotas instrumentadas com AVL ou APC (“automática passenger counting”). Esses dados fornecem às agências, uma bem definida distribuição do tempo de viagem tanto para as rotas como um todo, como para os segmentos individuais de cada rota. Para valorizar os dados de AVL, os autores mencionam que é desejável gerar um modelo generalizado do tempo de viagem como uma função do “alignment” (alinhamento: entendido como consistência com a rota) e parâmetros operacionais de tal modo que os tempos de viagem (e suas distribuições) possam ser acuradamente estimados para uma rota proposta e para a qual não existem dados de AVL. Segundo esses autores, é necessário predizer a locação de recursos para um sistema não somente baseado no tempo médio de viagem, mas também na variabilidade desse tempo médio de viagem.

Nesta Dissertação, à semelhança do trabalho supracitado, considera em cada viagem feita por um veículo (pois, após algumas viagens, deduzimos sua origem e seu destino), considerando a variação no tempo de viagem além de possíveis mudanças de rota.

3.3.7. Ferramentas de análise

O artigo de Chen *et al.* (2011) apresenta uma extensão desenvolvida para o software ArcGIS denominada Activity Pattern Analyst (APA) para facilitar a análise exploratória de dados de atividades diárias, baseada no conceito de geografia do tempo (“time geography”) de Hägerstrand, que representa atividades individuais num ambiente integrado espaço-tempo. É uma representação gráfica ortogonal 3D, onde o plano é o plano do espaço e a terceira dimensão o tempo, mostrando o comportamento individual no tempo, de um movimento no espaço. Esse conceito considera que “não fazer nada” é uma atividade, ou seja, ficar estacionado em um ponto fixo é uma atividade.

Citam várias referências de 1991 a 2009, de esforços para incorporar os conceitos de geografia do tempo em softwares SIG. A extensão desenvolvida (APA) tem 6 (seis) componentes: os 3 (três) primeiros relacionados à caracterização espaço-tempo; os 3 (três) últimos dedicados à detecção de padrões. O primeiro componente é o “gerador de caminhos espaço-tempo”, usado para construir os caminhos espaço-tempo individuais; é derivado de dados coletados via GPS ou via questionários ou via codificação de endereços. O segundo componente é um “filtro de caminhos espaço-tempo”, usado para extrair subconjuntos de caminhos individuais espaço-tempo, baseados nos seus atributos não espaciais (ex.: sexo, idade, educação, ocupação, salário, etc.), bem como suas características espaciais, tal como o local da residência. O terceiro componente é um “segmentador de caminhos espaço-tempo”, baseado conceito de segmentação dinâmica temporal (para qualquer período de tempo, a função dinamicamente extrai os dados espaciais e temporais). O quarto componente é a “busca por atividade (activity query)” que, dinamicamente, associa dados de atividades de cada indivíduo à correspondente posição do indivíduo no caminho espaço-tempo. O quinto componente é a “análise da distribuição/densidade da atividade”, que recupera, dinamicamente, a distribuição espacial de todos os indivíduos a qualquer tempo especificado pelo usuário. O sexto componente é a “análise de agrupamentos (cluster analysis) dos caminhos espaço-tempo”, que agrega num mesmo grupo, caminhos com a geometria similar de caminhos espaço-tempo.

O trabalho mostrado nesta Dissertação (2011) foi desenvolvido antes da publicação do artigo ora revisado e, uma possível atividade futura, seria aplicar o modelo (ou a forma de

modelar os dados) desenvolvido nesta Dissertação, aos grupos finais, caracterizando assim, uma análise espaço-temporal, de acordo com esses autores.

3.3.8. Padrões de distribuição espacial – transporte publico

Mesbah *et al.* (2012) comentam que, apesar da importância, poucas pesquisas abordam padrões espaciais sobre o desempenho operacional do transporte público urbano. No artigo é analisado o desempenho operacional do transporte público “on-road” usando análise espacial e histórica de uma grande rede. Usam dados advindos de AVL (10 anos de monitoramento sobre as paradas de um veículo leve sobre trilhos (VLT) em Melbourne) e SIG para a visualização das mudanças dos padrões no tempo e no espaço. Se utilizam de estatísticas simples, o desvio-padrão do tempo de viagem e o coeficiente de variação (medidas, segundo os autores, bem estabelecidas para medir o desempenho operacional do trânsito) para posteriormente num SIG explorar a variação espacial, usando o inverso das distâncias ponderadas (IDW: Inverse Distance Weighted). Apresentam uma revisão de artigos que se utilizam tanto dessas estatísticas quanto de SIG. Ainda segundo os autores, o artigo demonstra o crescente valor das análises espaciais e temporais de grandes montantes de dados de uma rede extensa, no caso o VLT de Melbourne. A metodologia é uma ferramenta de visualização para criar medidas concisas do desempenho a partir de grandes bases de dados de AVL. Chamam o processo de Automatic Vehicles Monitoring (AVM).

O artigo de Fang *et al.* (2012) apresenta uma análise espaço-temporal do que os autores chamam de elos críticos de transporte (“critical transportation links”) baseada no conceito de geografia do tempo. Utilizam dados de rastreamento, por GPS, de táxis pra analisar espaço-temporalmente padrões de viagens O/D que se utilizam de 3 (três) pontes críticas, e o conceito do prisma espaço tempo para identificar caminhos alternativos no espaço e no tempo. O estudo de caso, segundo os autores, ilustra que é viável e há benefícios em se usar o enfoque da geografia do tempo para se analisar padrões espaço-temporais de viagens que demandam elos críticos e seus caminhos alternativos dentro do sistema de transportes. Elos críticos de transporte podem, algumas vezes, serem parcialmente interrompidos ou completamente incapacitados (“disabled”) devido ao excesso de veículos, acidentes de tráfego, clima adverso ou falhas técnicas.

Downs e Hoener (2012) afirmam que dados de rastreamento de veículos (“vehicle tracking data”) são frequentemente usados para explicar o comportamento humano e padrões de atividades (“activity patterns”) em viagens. A geografia do tempo é um enfoque útil para se analisar esses conjuntos de dados e provê um meio para se identificar possíveis rotas e paradas de um veículo entre localizações conhecidas, que são denominadas “árvore de caminhos possíveis” (“potential path tree”).

Embora este seja o artigo, até agora, com abordagem mais próxima desta Dissertação, nesta tudo é conhecido ou esperado (origem, destino e caminho) a partir de um determinado número de viagens observadas, ou seja, as origens, destinos e caminhos são extraídos a partir da amostra de viagens fornecida.

Os autores afirmam também, que um enfoque para modelar possíveis movimentos de um objeto entre localizações conhecidas é a geografia do tempo, como apresentada por Miller (2005). O mesmo pode ser usado para calcular um conjunto de potenciais localizações de um objeto num espaço Euclidiano, dadas a sua capacidade de movimento e as restrições espaço-tempo impostas pelas localizações consideradas/observadas. Para quaisquer dois (2) pontos consecutivos, essa área de caminhos potenciais (“potential path area”) assume a forma de uma elipse, quando mapeado num espaço contínuo bidirecional. Ainda segundo os autores, para veículos viajando em estradas, a formulação clássica de geografia do tempo não é válida (Raubal *et al.*, 2007) e é preciso adotar o dispositivo baseado em redes (“network-based framework” descrito em Miller (1991)).

A geografia do tempo se mostra útil para analisar os movimentos de um veículo e a principal crítica é que as árvores de caminho potencial (“potential path trees”) somente delinea localizações onde um objeto possivelmente viajará, mas não fornece nenhuma medida de quais rotas ou paradas são as mais prováveis (Pred, 1977). Em Andrienko *et al.* (2010), os autores citam que desenvolvem artefatos de geografia do tempo probabilísticos ou estatísticos, sendo um deles a formulação de estimação por densidade de Kernel tradicional, usada com padrões de pontos gerados por objetos móveis. Segundo os autores, essa técnica ainda não é compatível com dados de rastreamento de veículos ou o caso mais geral, de “discrete network space”. No artigo, os autores apresentam um estimador de densidade de

geografia do tempo baseado em redes e como ele pode-se analisar dados de rastreamento de veículos.

No artigo, os autores usam a técnica para construir árvores de caminhos potenciais probabilísticos que podem ser usados para inferir a verossimilhança de que um veículo viaje em uma particular rota e/ou particular parada em determinadas localidades.

Horner *et al.* (2012), afirma que a partir do uso de bases de dados de pesquisa de viagem ao nível de indivíduo, descrevendo detalhes de atividades residenciais (“activities of household”), é possível analisar os movimentos humanos com alto grau de precisão. Entretanto, dados de pesquisas de viagens não estão livres de aspectos de qualidade. Existe a possibilidade de origens e destinos que não estão georreferenciados, talvez devido à falta de documentação/registo da informação, ou devida às inconsistências de bases de dados de endereços espaciais, o que pode limitar o seu uso. Quando o interesse é pela sequência de localização de atividades, essa ausência de dados é um problema sério. O artigo apresenta um dispositivo (“framework”) e um enfoque computacional básico para explorar a localização de atividades não localizadas inerentes às viagens pesquisadas (“travel surveys”). O sistema estima as possíveis localizações de origens e destinos perdidos. Não é o enfoque desta Dissertação, pois nesta, a ausência de informação é motivo para tomada de decisão e não especular a possível localização do veículo. Pode ser usada para minimizar as ações de tentar localizar o veículo se as ocorrências (dada à falta de informação a localização estimada for correta) se repetirem com frequência no tempo.

3.3.9. Equipamentos para coleta de dados

Xia e Arrosmith (2008) discutem quais dados são necessários para descrever o movimento espaço-temporal de visitantes (“visitors”). Apresentam uma revisão das técnicas para rastrear e contabilizar o movimento de visitantes. Segundo um dos autores (Arrosmith *et al.*, 2006), os dados necessários para simular o comportamento de um visitante podem ser categorizados em parâmetros espaciais e temporais:

- Identidade (“identity”): identificação do objeto
- Posição (“position”): localização geográfica do objeto
- Distância (“distance”): medida de distância linear no plano

- Tempo (“time”): “time point” e duração
- Duração: período de tempo gasto numa atividade ou viagem
- Velocidade: relaciona distância e tempo
- Sequência e itinerário: sequência é a ordem na qual um objeto móvel individual visita um certo número de atrações. Das sequências se deduz o itinerário.
- Fluxo (“flow”): é utilizada para medir a intensidade do tráfego. É o número de objetos que passa por um ponto por unidade de tempo.

Os autores apresentam métodos de contar e rastrear dados. Quando abordam o rastreamento, fazem uma breve revisão sobre GPS, PDA (integração de GPS com SIG), “mobile phone tracking” e “closed circuit television monitoring (CCTV)”.

Xiao *et al.* (2012) afirmam que pesquisas de atividades de transporte investigam quando, onde e como pessoas viajam em áreas urbanas para fornecer as informações necessárias ao planejamento de transporte urbano. Os autores estão desenvolvendo um sistema de levantamento de atividades de transportes baseado em smartphones para substituir as pesquisas domiciliares (“households surveys”) tradicionais.

Smartphones equipados com vários sensores podem ser usados para coletar informações provenientes de GPS, GSM, acelerômetros e outras informações contextuais para detectar lugares, rastrear trajetórias (“trajectory tracking”) e detectar o modo de transportes. O algoritmo desenvolvido processa os dados coletados pelos smartphones para detectar o modo de transporte e as paradas. A parte do algoritmo para a detecção do modo de transporte, usa estatísticas derivadas de informações provenientes de GPS e da rede celular e o desvio-padrão da magnitude da força de amostras de acelerômetros. A parte referente à detecção de parada (“stop detection”) pode identificar quando e onde o usuário para e também inferir se as paradas são transferências de transporte público. O algoritmo é baseado na distância do movimento geográfico numa janela de tempo (“geographical moving distance in time windows”), no padrão de troca de informação do celular, e o contexto do modo de transporte. O artigo se assemelha à Dissertação quando se refere à identificação de paradas: no artigo usam informações de contexto para caracterizá-la e na Dissertação é objeto de investigação e eventual alarme com consequente chamada do usuário.

Para Nitsche *et al.* (2012), a aquisição de dados de viagens é, baseada em questionários caros e demorados e que fornecem uma imagem incompleta devido à limitada cobertura e atualização inadequada. O artigo introduz um novo enfoque, segundo os autores, desenvolvendo o uso inteligente de smartphones para pesquisas de viagens de grande escala. É baseado nos sinais de acelerômetros (“embedded accelerometers”) e receptores GPS e usa um conjunto de classificadores probabilísticos treinados para reconstruir, automaticamente, as viagens individuais que compõem uma viagem (“tour”), incluindo a escolha modal. Os modelos foram treinados e avaliados para detectar caminhadas e viagens por bicicletas em Viena, na Áustria. O acelerômetro é usado quando há perda de sinal do GPS.

3.3.10. Técnicas e métodos selecionados

O método denominado de Kernel que será utilizado nesta Dissertação, calcula a densidade de pontos detectados em torno de um centroide, com uma determinada vizinhança. Para o estudo da movimentação de veículos a técnica de análise de Kernel ou também conhecida análise de densidade de pontos, delimita e explica a movimentação do veículo.

A escolha do método de análise considerou o sucesso no uso do estimador de Kernel por Kharoufeh e Goulias (2002), para o estudo das atividades diárias dos indivíduos.

Para esta Dissertação foi definido como valor da largura da faixa (vizinhança) como sendo igual ao do erro do equipamento de posicionamento embarcado no veículo. Como resultado do cálculo de Kernel, foram identificadas as regiões nas quais o veículo pode estar se movimentado com uma maior probabilidade. Na Figura 8 estas regiões são evidenciadas. Como resultado do cálculo de Kernel é possível observar em amarelo as regiões com probabilidade acumulada de 50 % para o posicionamento do veículo, em rosa são as regiões com 75% de probabilidade de ocorrência, em azul as regiões onde o veículo pode estar com uma probabilidade de 95%.

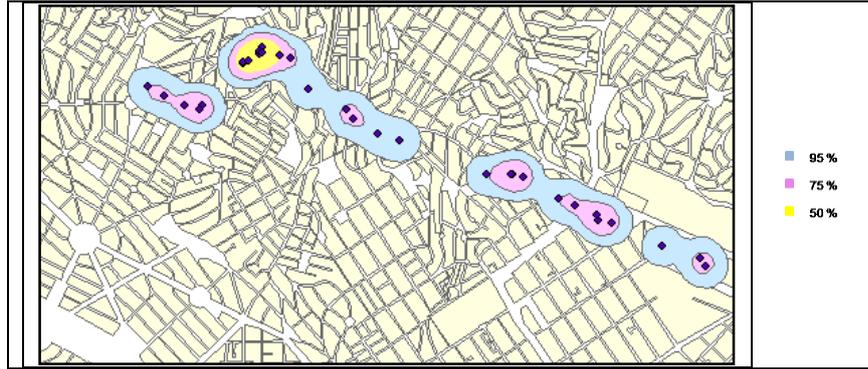


Figura 8 - Probabilidade de ocorrência do evento

O cálculo do Kernel gera uma representação da distribuição, que é caracterizada pela densidade de pontos que ocorrem na vizinhança de um centroide (Figura 9) e pode ser calculada utilizando a seguinte fórmula:

$$\hat{\lambda}_\tau(s) = \sum_{i=1}^n \frac{1}{\tau^2} I\left(\frac{(s-s_i)}{\tau}\right)$$

Equação 1

Onde :

$\hat{\lambda}_\tau(s)$: Kernel na área inserida num raio de interesse t em relação ao ponto S

τ : raio da área de interesse, definindo a área que estará sendo analisada.

S : localização qualquer na área de estudo

S_i : localizações detectadas dos objetos móveis.

h_i Distância entre S e S_i

n representa o número de detecções.

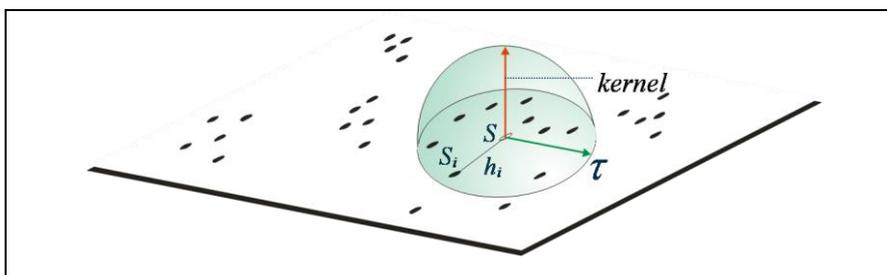


Figura 9 - Estimador de Kernel (adaptado de Câmara e Carvalho, 2004)

Para a análise do comportamento temporal do veículo como mostrado anteriormente, existem outras técnicas, nesta Dissertação será utilizado a análise de sobrevivência que analisa

as distancias percorridas até a ocorrência do evento de interesse, no caso, a detecção da parada do veículo. Essas distâncias, segundo Cox (1972) são denominadas de “tempo de falha”.

Uma característica dos dados de rastreamento de objetos móveis é a presença de censura que segundo Colosiomo e Giolo (2006) é a observação parcial da resposta é relevante. Por exemplo, na área médica, podem ocorrer interrupções no acompanhamento de gravidez, mesmo que não seja registrado o parto, ele deverá ocorrer em até 42 semanas, desta forma temos uma censura em registros de tempo maiores que 42 semanas. No caso de rastreamento de veículo, pode ocorrer a interrupção na coleta de dados por falha de comunicação por diversos motivos como a falta de sinal ou erros no equipamento utilizado para o monitoramento, por exemplo, quando se registra grandes deslocamentos em períodos pequenos.

Os modelos probabilísticos que melhor descrevem os fenômenos analisados nesta dissertação são semelhantes as utilizadas em análise de confiabilidade de processos (tempo de falha) e segundo Colosiomo e Giolo *op. cit.*, são o modelo de distribuição exponencial, o Weibull e o Log-normal.

Os modelos exponencial, o Weibull e o Lognormal estão descritos abaixo:

- Modelo de distribuição exponencial

Este modelo é o mais simples que é utilizado para descrever uma situação aonde consideramos a taxa de falha constante. Isso significa, por exemplo, em um estudo de produto ou dispositivo considera-se que a taxa de falha, não muda com o tempo, sendo produto novo ou velho.

A função densidade de probabilidade é dada por:

$$f(d) = \frac{1}{\alpha} \exp\left\{-\left(\frac{d}{\alpha}\right)\right\}, d \geq 0 \quad \text{Equação 2}$$

Onde:

α : parâmetro de escala (média dos deslocamentos)

d : deslocamento até a parada

Esta função pode ser representada pelo gráfico mostrado na figura 10,

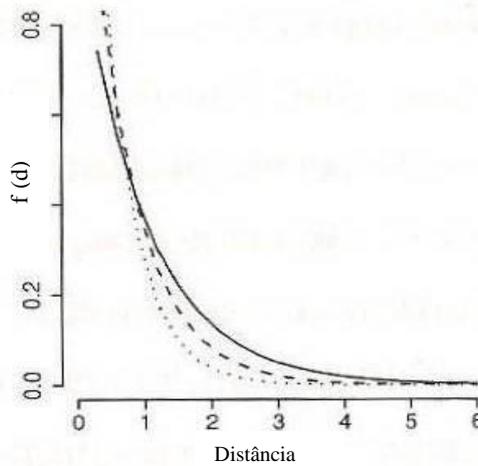


Figura 10 – Função densidade de probabilidade, para valores de α 1 (—); 0,7 (--) e 0,5 (---)
(adaptado de (Colosimo e Giolo, 2006))

- Modelo de distribuição Weibull

Este modelo segundo Montgomery (2004) é flexível, e com uma seleção adequada dos valores de α e γ , a distribuição pode assumir diversas formas, esta característica permite que possa ser utilizado em diversas áreas como a engenharia de confiabilidade e em estudos biomédicos.

A função densidade de probabilidade é dada por:

$$f(d) = \frac{\gamma}{\alpha^\gamma} d^{\gamma-1} \exp\left\{-\left(\frac{d}{\alpha}\right)^\gamma\right\}, d \geq 0 \quad \text{Equação 3}$$

Onde:

α : parâmetro de escala (média dos deslocamentos)

γ : parâmetro de forma (desvio padrão dos deslocamentos)

d: deslocamento até a parada

A função pode ser representada pelo gráfico mostrado na figura 11

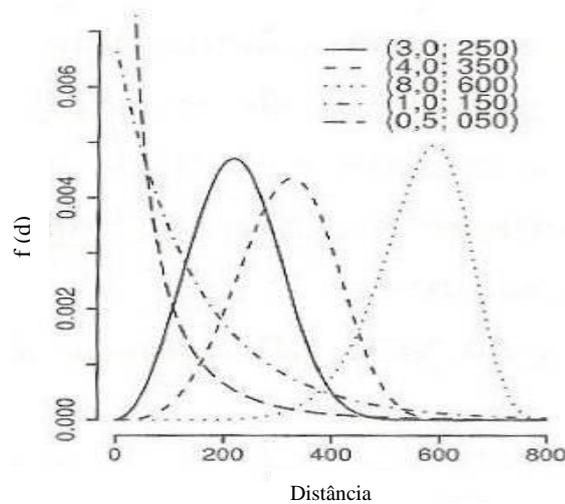


Figura 11 - Função distribuição de probabilidade – distribuição Weibull para valores de α e γ (fonte: adaptado de Colosiomo e Giolo (2006))

- Modelo de Distribuição Log-normal

Esta distribuição é muito utilizada para a caracterização de tempo de vida de pessoas e produtos. Este modelo é o logaritmo de uma variável com distribuição log-normal de parâmetros m e s : Isto significa que os dados provenientes desta distribuição podem ser analisados segundo uma distribuição normal de média m e desvio padrão s .

Esta distribuição não possui uma taxa de falha monótona, isso significa que a taxa é variável. Desta forma como no caso do modelo de distribuição Weibull, a forma da distribuição pode ser ajustada, para explicar melhor a o fenômeno estudado.

A função densidade de probabilidade é dada por:

$$f(d) = \frac{1}{\sqrt{2\pi}d\sigma} \exp\left\{-\frac{1}{2}\left(\frac{\log(d) - \mu}{\sigma}\right)^2\right\}, d > 0$$

Equação 4:

Onde:

μ : média dos deslocamentos
 σ : desvio padrão dos deslocamentos
 d : deslocamento

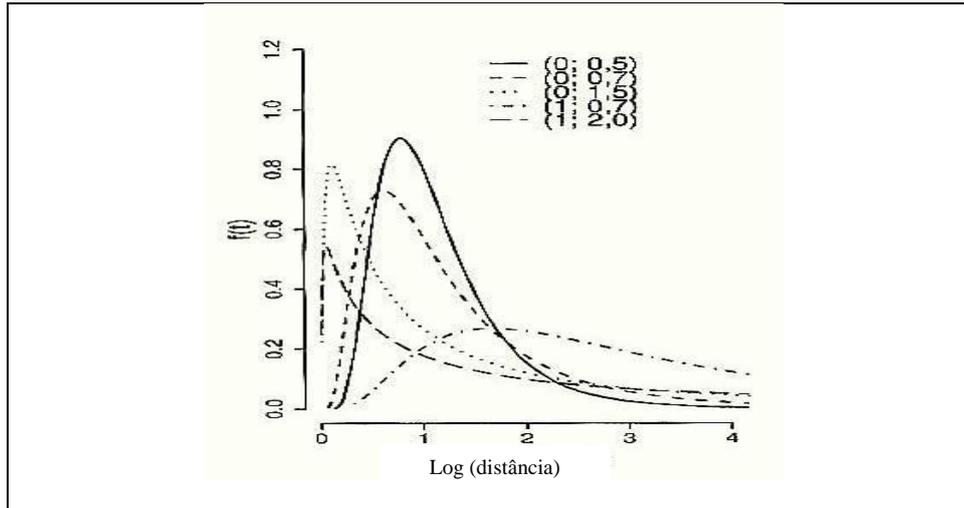


Figura 12 - Função distribuição de probabilidade Log normal (fonte: adaptado de Colosimo e Giolo (2006))

Os modelos de distribuição são definidos por quantidades desconhecidas, chamadas de parâmetros. Para conhecer e avaliar os parâmetros, são utilizados os estimadores estatísticos. O método mais adequado para a estimação dos parâmetros do modelo é o de Máxima Verossimilhança, por incorporar o tratamento de dados censurados e por ser conceitualmente simples e de fácil aplicação, principalmente em grandes amostras.

A partir deste princípio, segundo Bussab e Morettin (2006), devemos selecionar o valor do parâmetro desconhecido que maximiza a probabilidade de obter a amostra observada, ou seja, que torna a amostra a “mais provável”. Matematicamente a função de verossimilhança com censura aleatória é definida como:

$$L(\theta) = \prod_{i=1}^r f(d_i; \theta) \prod_{i=r+1}^n (d_i; \theta)$$

Equação 5:

Onde:

n = número de observações (Detecções do veículo)

r = número de falhas (Detecção da parada do veículo)

n-r = número de censuras (Perda de sinal de GPS ou deslocamento menor que a precisão do equipamento)

d = Deslocamento.

Analisando a equação obtida na estimação dos parâmetros pode-se realizar uma readequação no modelo de distribuição, testando e excluindo elementos do estimador de

parâmetros (preditora + coeficiente) que não possuam significância para o modelo, ou seja, excluir os elementos que não contribuem na definição do modelo.

A partir do modelo de distribuição e dos seus parâmetros será gerada a equação que caracteriza o comportamento movimentação do veículo ou objeto móvel analisado.

4. METODOLOGIA

4.1. Obtenção dos dados de rastreamento dos dados

Para a realização deste estudo foram utilizados dados de monitoramento contendo apenas a posição e horário de coleta do dado, sem identificação do veículo ou proprietário, dados estes fornecidas por uma empresa de rastreamento, que solicitou sigilo.

Estes dados foram obtidos em formato texto e se referem à coleta de dados de um (1) mês de movimentação de trinta (30) veículos, estes foram selecionados aleatoriamente de seu banco de dados, o intervalo nominal entre as coletas de 3 minutos. Esta frequência para a coleta de dados é normalmente utilizada em sistemas de rastreamento.

Para este conjunto de dados foram selecionados para a análise aqueles que os dados dos veículos que apresentavam registro em um maior número de dias para o período selecionado e que os dados de posição estivessem consistentes, apresentando deslocamentos e velocidades praticáveis.

4.2. Ferramentas utilizadas para a análise

As ferramentas de softwares utilizados no desenvolvimento do modelo foram:

- Minitab Versão 15 – Software estatístico.
- SciLab – Software de computação numérica.
- Arcview versão 3.1, com os módulos:
 - Spatial Analyst;
 - Modulo Animal Movement – USGS – Alaska Science Center.
- ArcMap/ArcInfo versão 10, com o módulo Spatial Analyst.

Para a análise temporal, foram estudados métodos analíticos para a caracterização dos deslocamentos do veículo e verificada a existência de padrões de deslocamento nos diversos dias da semana e horas do dia. Para a análise espacial foram avaliadas as posições espaciais do veículo detectadas durante o processo de monitoramento. O modelo de movimentação será a composição do modelo temporal com o modelo espacial do veículo estudado.

4.3. Análise dos dados de movimentação de um veículo.

Para a obtenção de informações sobre o comportamento de movimentação, desenvolveu-se uma metodologia para determinar o padrão de movimentação de um determinado veículo por meio da modelagem do comportamento temporal e espacial.

O estudo sobre o comportamento de movimentação do veículo baseou-se na análise de uma série histórica dos dados de rastreamento contendo os deslocamentos e a sua posição espacial. Estes deslocamentos caracterizam os padrões diários de movimentação que dependem do estilo de vida e das atividades realizadas.

A caracterização do padrão de movimentação do veículo ou objeto móvel foi realizada analisando separadamente o comportamento espacial e o comportamento temporal. O esquema apresentado na figura 13 apresenta as etapas que serão realizadas para a construção do modelo de movimentação do veículo.

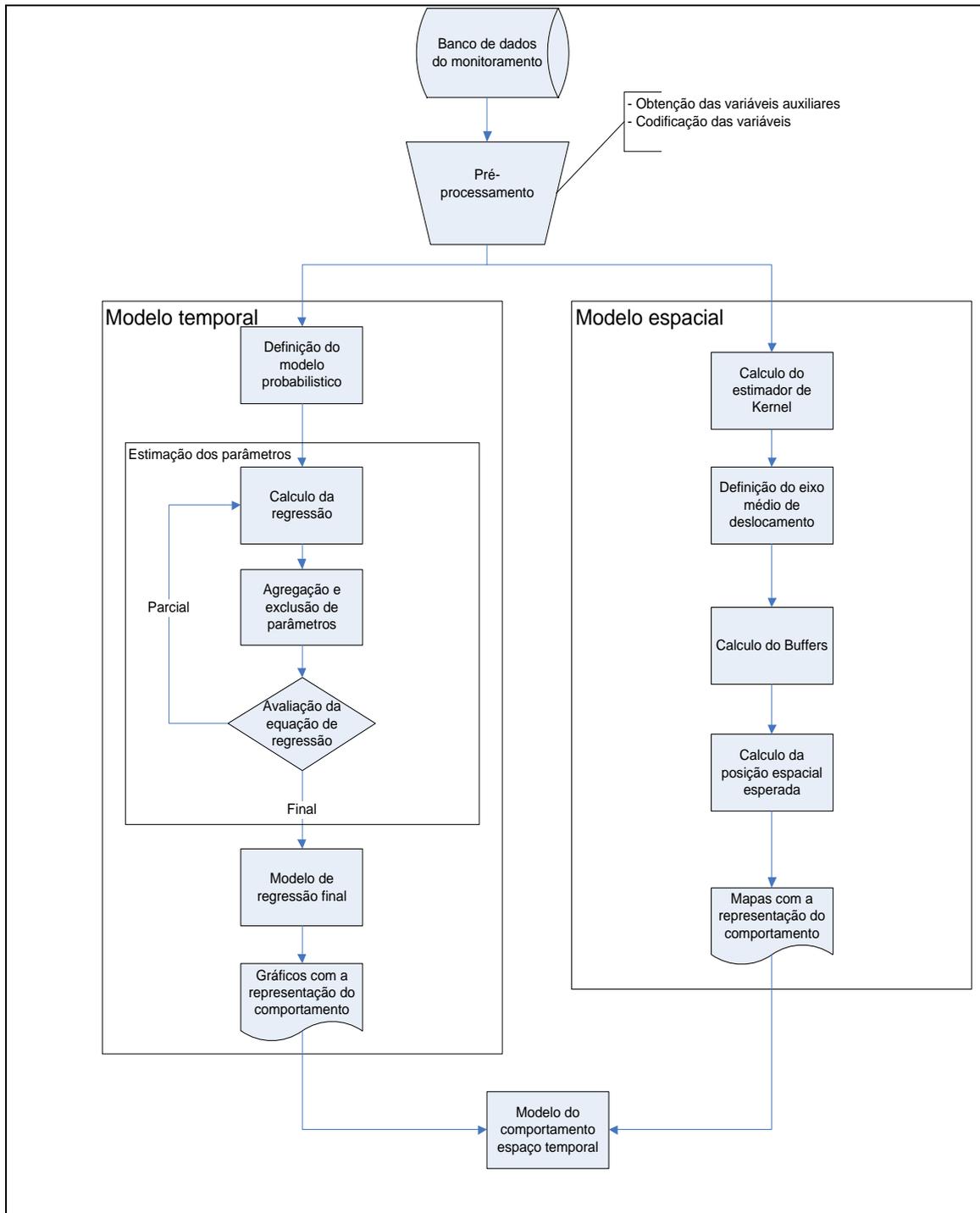


Figura 13 - Esquema para análise da movimentação do veículo

4.4. Pré-processamento dos dados de rastreamento

Os dados foram obtidos em uma empresa de rastreamento de veículos, que no momento não deseja ser citada, estes dados não possuem identificação do veículo ou do proprietário, os dados produzidos pelo processo de rastreamento dos veículos possuem as seguintes características, conforme exemplo mostrado na Tabela 1:

- Identificação do veículo
- Posição espacial (em coordenadas UTM - Este, Norte)
- Data e hora do registro da posição espacial (Data - Dia/mês/ano, hora – hora: minuto)

Tabela 1 - Exemplo de dados obtidos pelo rastreamento

ID_veic	Este	Norte	data_hora
Veic01	485552	7419651	26/2/2008 01:32
Veic01	485543	7419657	26/2/2008 01:35
Veic01	485545	7419655	26/2/2008 02:05

Os dados são coletados em uma frequência nominal bem definida, dependendo do tipo de aplicação que está associado. A frequência de coleta pode variar também conforme as condições de transmissão dos dados e de coleta do dado de posicionamento do veículo. Os dados foram coletados durante um mês.

O campo data_hora foi desmembrado em dois novos campos: o dia da semana e a hora do dia. Além destes definiu-se um campo para o deslocamento. O deslocamento é definido como sendo a distância, em metros. Entre a posição espacial detectada e a última posição coletada (Tabela 2). A posição espacial é registrada em coordenadas no sistema UTM, este sistema utiliza o metro como unidade e para medir a distância entre dois pontos neste sistema podemos utilizar a seguinte formula:

$$d = \sqrt{((E^1 - E^2)^2 + (N^1 - N^2)^2)}$$

Equação 6

onde:

E^x : Coordenada na direção Este

N^x : Coordenada na direção Norte

d: Distância entre os dois pontos.

Tabela 2 - Exemplo de dados utilizados para o tratamento

ID_veic	Este	Norte	Data_hora	Dia da semana	Deslocamento (m)
Veic01	485552	7419651	26/2/2008 01:32	Terça-feira	20
Veic01	485543	7419657	26/2/2008 01:35	Terça-feira	10
Veic01	485545	7419655	26/2/2008 02:05	Terça-feira	4

Para a caracterização da movimentação foram empregadas técnicas estatísticas aplicadas em análise de confiabilidade. Estas técnicas são utilizadas para avaliar as preditoras que influenciam no caso desta Dissertação as distancias percorridas e a confiabilidade de um processo.

Neste estudo definimos como preditoras, os dias da semana e as horas do dia. Na análise das características dos dados foram detectadas limitações para a utilização dos valores de deslocamento obtidos. A principal é a precisão do equipamento utilizado para a determinação do posicionamento, pela especificação do fornecedor do dado, os equipamentos utilizados possuem uma precisão nominal para o posicionamento de quinze (15) metros. A detecção de deslocamento menor que a precisão do equipamento será desconsiderada, neste caso não sabe se realmente ocorreu movimentação ou o veículo está parado. Para evitar conclusões viciadas pela exclusão destas observações, a análise de confiabilidade permite a inclusão de um evento censurado, o evento é considerado, mas o valor associado a ele não, para este tipo de evento é arbitrado um valor pré-definido que é denominado de valores censurados. Desta forma são incluídos todos os eventos para os cálculos estatísticos, pois as observações realizadas, mesmo incompletas, fornecem informações importantes para a caracterização dos deslocamentos.

A fim de registrar a movimentação do veículo e definir os valores censurados foram criados mais dois campos auxiliares:

- “Deslocamento – início”; onde é registrado se o valor medido é menor que o valor censurado. Esta forma de censura é conhecida como “censura a esquerda”
- “Deslocamento – fim”; onde é registrado se o valor medido é maior que o valor máximo. Esta forma de censura é conhecida como “censura a direita”

O valor mínimo para esta dissertação foi definido como o valor do erro nominal do equipamento de posicionamento, que no caso é o GPS e para este equipamento o erro

atribuído é de 15 metros. E o valor máximo atribuído é de 6000 metros que equivale a 120 km/h, considerando que a taxa de coleta nominal de dados seja de 3 minutos, esta velocidade é o limite máximo das principais Rodovias paulistas.

Na tabela 3 temos um exemplo do registro com censura à esquerda para um valor de deslocamento menor que 15 metros.

Tabela 3 - Exemplo de registro de dados censurados

ID_veic	Este	Norte	Data_hora	Dia da semana	Deslocamento (m)	Início (m)	Fim (m)
Veic01	485552	7419651	26/2/2008 01:32	Terça-feira	20	20	20
Veic01	485543	7419657	26/2/2008 01:35	Terça-feira	10	*	15
Veic01	485545	7419655	26/2/2008 02:05	Terça-feira	4	*	15

Para descrever o comportamento temporal de movimentação do veículo, será avaliada a variação dos valores de deslocamento e realiza-se uma avaliação para determinar o modelo de distribuição de probabilidade mais adequado, que descreve o comportamento de movimentação, assim como a equação de regressão que será utilizada para modelar a movimentação do veículo. Para a análise do comportamento espacial será realizada análise de distribuição espacial de pontos. (Kernel).

4.5. Comportamento Temporal

A seleção do modelo de distribuição estatística mais adequado para explicar a movimentação do objeto móvel, foi realizada por método gráfico, comparando os valores obtidos utilizando o modelo de distribuição proposto com os valores obtidos com a utilização do modelo gerado utilizando o estimador de máxima verossimilhança. As comparações foram entre os modelos de distribuição Weibull, Exponencial e Log-normal, estes são os modelos mais frequentemente utilizados. O modelo mais adequado é aquele em que os valores obtidos pelo estimador de máxima verossimilhança estiverem mais próximos aos pontos obtidos na amostra. A Figura 14 ilustra um exemplo do método gráfico, as linhas em azul delimitam a região aonde o modelo de distribuição Weibull é significativo estatisticamente, dentro de um intervalo de confiança arbitrado para esta Dissertação em 95% e os pontos vermelhos representam aos pontos obtidos pela amostra. Define-se que o modelo ideal é aquele no qual a

distribuição dos pontos calculados utilizando o modelo estão dentro dos limites delimitados pelas linhas azuis, que representa os limites do intervalo de confiança.

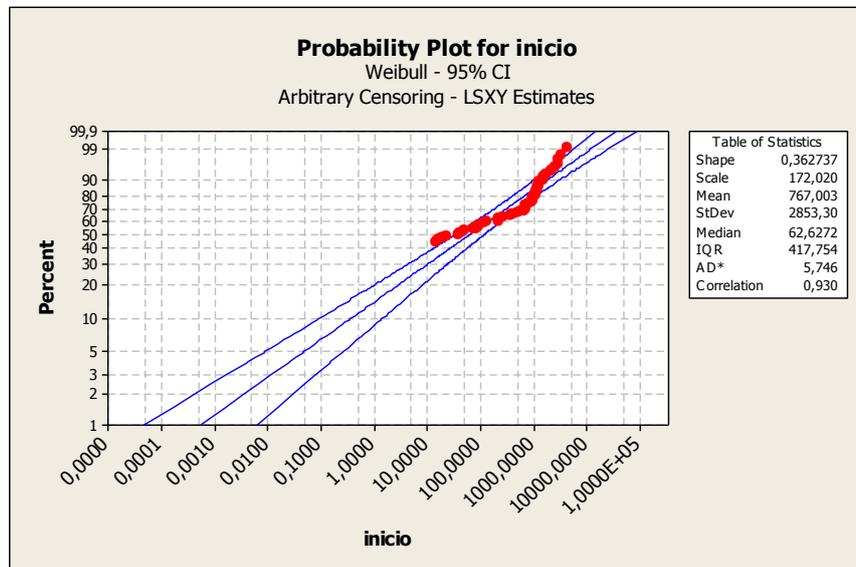


Figura 14 - Exemplo do método gráfico – gerado no software minitab

A avaliação realizada nos dados de rastreamento do veículo selecionado para o estudo indicou que o modelo distribuição Weibull é adequado para a modelagem dos valores de deslocamento. Desta forma utilizaremos este modelo para construir as equações de regressão que serão utilizadas para caracterizar o comportamento temporal do veículo.

4.5.1. Estimação dos coeficientes das preditoras do modelo

Apesar de definido o modelo de distribuição de probabilidade, as preditoras e os seus coeficientes que caracterizam esta distribuição ainda são desconhecidas. As preditoras e seus coeficientes serão estimados por meio da avaliação das observações de movimentação realizadas. A partir deste modelo será possível caracterizar a movimentação do veículo e seu comportamento.

As preditoras que serão utilizadas referem-se ao tempo, para este estudo será adotado inicialmente como preditoras as horas do dia e os dias da semana. Codificou-se esta preditora com o auxílio de variáveis dummy ou binária. Os seus coeficientes assumem o valor 0 ou 1, indicando a presença ou ausência da preditora. No presente estudo os dias da semana e as

horas do dia foram codificadas segundo este tipo de técnica por exemplo, para codificar a terça-feira e adotou-se para esta notação os códigos apresentados no quadro 1, A partir do qual obteve-se a seguinte expressão:

$$\text{Terça Feira} = 0.ds1 + 0.ds2 + 1.ds3 + 0.ds4 + 0.ds5 + 0.ds6 + 0.ds7$$

Código	Dia da Semana
ds1	Domingo
ds2	Segunda-feira
ds3	Terça-feira
ds4	Quarta-feira
ds5	Quinta-feira
ds6	Sexta-feira
ds7	Sábado

Quadro 1 – Codificação para os dias da semana

Para simplificar os cálculos e a codificação das horas do dia foi definido que foi considerado um intervalo de uma hora. Quando há necessidade de identificar o comportamento com um intervalo de tempo menor, a preditora poderá ser recodificada.

A partir da definição do modelo de distribuição de probabilidades para os deslocamentos, e determinadas as preditoras e seus coeficientes para os modelos, foram criadas equações de regressão capazes de prever o deslocamento do veículo a um determinado nível de significância.

A equação de regressão seguirá a seguinte estrutura:

$$d = \theta_0 + \theta_1 X_1 + \theta_2 X_2 + \dots + \theta_k X_k + \varepsilon \quad \text{Equação 7}$$

onde:

- d = Deslocamento
- θ_0 = Intercepto (constante)
- $\theta_1 \dots \theta_k$ = Coeficiente de regressão
- X = Preditora (dias da semana e horas do dia)
- ε = Termo de erro aleatório

A partir destas equações elaborou-se um conjunto de gráficos representando as probabilidades de o veículo estar parado em função dos parâmetros definidos pelo modelo,

além daqueles que representam o deslocamento do veículo em função das horas do dia, com uma probabilidade acumulada de P%. Estes gráficos mostram o comportamento de movimentação esperado para o veículo, em função das horas do dia.

Pode haver similaridade do comportamento de movimentação do veículo, em determinados dias da semana e horas do dia. Para avaliar esta similaridade, foram testadas as preditoras e os seus coeficientes, aplicando-se o Teste de Wald. Para executar este teste será adotado como preditoras de referência o sábado as 0:00hs como referência.

Este teste consiste numa generalização do teste T de Student e é utilizado para testar hipóteses relativas a um único parâmetro de $\hat{\theta}$. É calculado utilizando-se a seguinte equação:

$$W = \theta C' [C' \text{Var}(\theta) C]^{-1} C \theta \quad \text{Equação 8}$$

Onde:

θ = vetor com os coeficientes da regressão,

C = vetor que contém o subconjunto dos coeficientes da regressão a serem testados

$\text{Var}(\theta)$ = matriz de variância e covariância

Se o valor de W é menor ou igual ao valor crítico obtido da distribuição teórica com os mesmos coeficientes da regressão e a um determinado nível de significância.

Este teste permite construir uma nova equação de regressão utilizando somente as preditoras significativas. A nova equação de regressão é testada para verificar a equivalência entre as equações, o teste utilizado é o Teste da Razão de Máxima Verossimilhança (TRV) que compara os valores dos logaritmos de máxima verossimilhança, e é dado como;

Equação A: Equação que utiliza todas as preditoras,

Equação B: Equação que contém o subconjunto de preditoras.

Hipóteses:

H0: Equação A = Equação B $TRV < \chi_c^2$

H1 Equação A \neq Equação B $TRV > \chi_c^2$

$$TRV = -2 \log \left[\frac{L(\hat{\theta}_0)}{L(\hat{\theta})} \right] = 2 [\log L(\hat{\theta}) - \log L(\hat{\theta}_0)] \quad \text{Equação 9}$$

A avaliação entre a equação de regressão será realizada até que no teste da razão da máxima verossimilhança apresente que a hipótese H1 seja aceita, isto é, até quando as equações testadas sejam significativamente diferentes.

A equação de regressão obtida deste processo e os gráficos de probabilidade permitirão a definição do padrão de comportamento temporal do veículo.

4.6. Comportamento Espacial

Para a caracterização espacial foram utilizadas técnicas de análise espacial de eventos, que segundo Câmara (2004) estuda fenômenos expressos através de ocorrências identificadas como pontos localizados no espaço, denominados processos pontuais. O objetivo da análise é estudar a distribuição espacial desses pontos, testando hipóteses sobre o padrão observado. São exemplos: localização de crimes, ocorrências de doenças e localização de espécies vegetais. O objeto de interesse é a própria localização espacial dos eventos em estudo. O termo evento, no contexto de Câmara (*op. cit.*), refere-se a qualquer tipo de fenômeno localizável no espaço. Neste caso, é a localização de um veículo,

Ao analisar o padrão de distribuição espacial de pontos, procuramos identificar se os eventos observados formam algum padrão sistemático, observando a existência de conglomerados espaciais (clusters). Uma das estratégias de análise do comportamento de padrão de pontos é utilizar um estimador de intensidade (Kernel Estimation); para isso ajustamos uma função sobre os eventos considerados, compondo superfícies que são proporcionais à densidade da amostra (número de detecções) por unidade de área. Esta função realiza a contagem de todos os pontos de uma região de influência, ponderada pela a distância de cada um à localização de interesse.

4.6.1. Análise das distribuições espaciais

Utilizando os mesmos dados de movimentação dos veículos obtidos na empresa de rastreamento, para explorar as potencialidades da análise espacial, representamos as distribuições espaciais de 30 veículos. Para construir estas representações utilizamos o software ArcView versão 3.2 da ESRI com os módulos Spatial Analyst e Home Range Analysis. No entanto, a representação dos pontos não é suficiente compreender as distribuições espaciais como mostra a Figura 15, mostrando apenas os locais aonde foram coletados os dados do rastreamento, dificultando a caracterização do comportamento de movimentação espacial do veículo,



Figura 15 - Representação dos pontos amostrados

Existem diversas técnicas utilizadas para a definição da distribuição espacial. Cada técnica possui características específicas que explicam melhor o padrão de movimentação de um determinado indivíduo.

Exemplos de análise de distribuição espacial foram analisados com o auxílio do módulo *Home Range Analysis* concebido por Hooge *et.al* (2000) que possui diversos algoritmos para demonstrar o padrão de movimentação. Estes exemplos estão representados na Figura 16. Utilizando-se o método do mínimo polígono convexo – MPC (Mohr, 1947), construímos o menor polígono onde estão contidos todos os pontos que foram detectados. O método Jennrich-Turner (Jennrich & Turner, 1969) define um elipsóide que represente, com certa confiabilidade, o espalhamento espacial dos pontos detectados. O terceiro método denominado de Kernel (Worton, 1989) é calculado pela densidade de pontos detectados em

torno de um centróide, com uma determinada vizinhança. Para o estudo da movimentação de veículos a técnica de análise de Kernel ou também conhecida análise de densidade de pontos, delimita e explica melhor a movimentação do veículo.

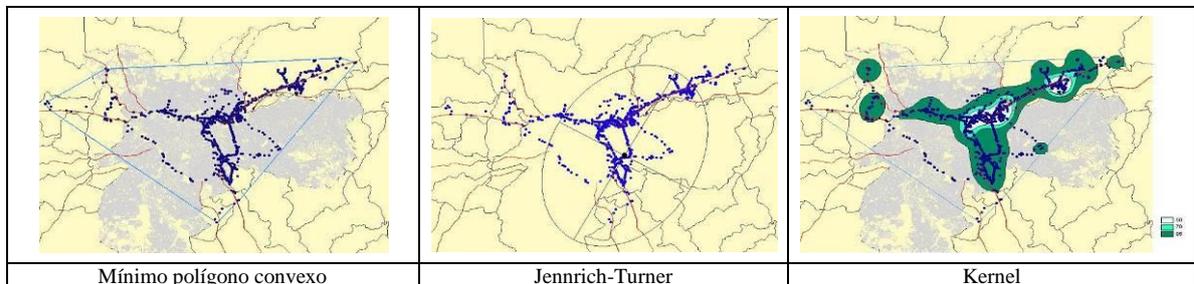


Figura 16 - Exemplos de análise de envoltórias

Como resultado do cálculo de Kernel aplicado aos veículo da amostra, identificamos as regiões nas quais há uma maior probabilidade do veículo estar (Figura 17). Em amarelo estão delimitadas as regiões com probabilidade acumulada de 50 % para o posicionamento do veículo, em rosa, regiões com 75% de probabilidade de ocorrência, em azul as regiões onde o veículo pode estar com uma probabilidade de 95%.

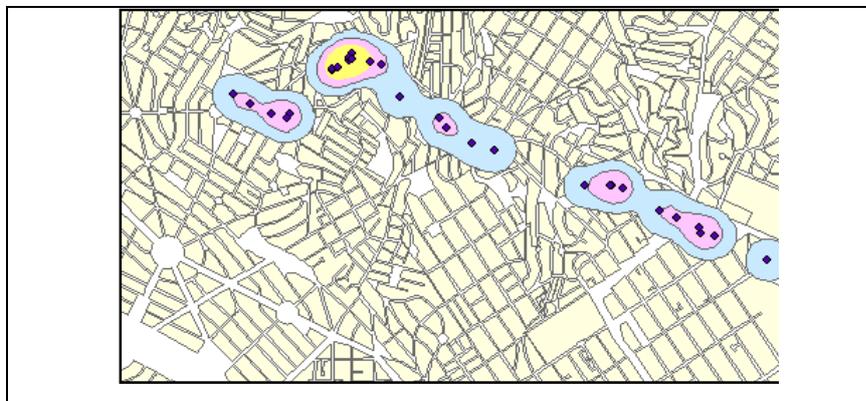


Figura 17 – Mapa de isoprobabilidade de ocorrência do evento

Dependendo das características de utilização do veículo, temos uma distribuição espacial do estimador de Kernel distintas. Avaliando a distribuição espacial de 30 veículos, identificamos alguns padrões de movimentação espacial, que podem fornecer informações sobre o comportamento espacial, subsidiando a modelagem espacial. Para estes padrões que não estão formalizados na bibliografia consultada, sugerimos a seguinte terminologia (Figura 18).

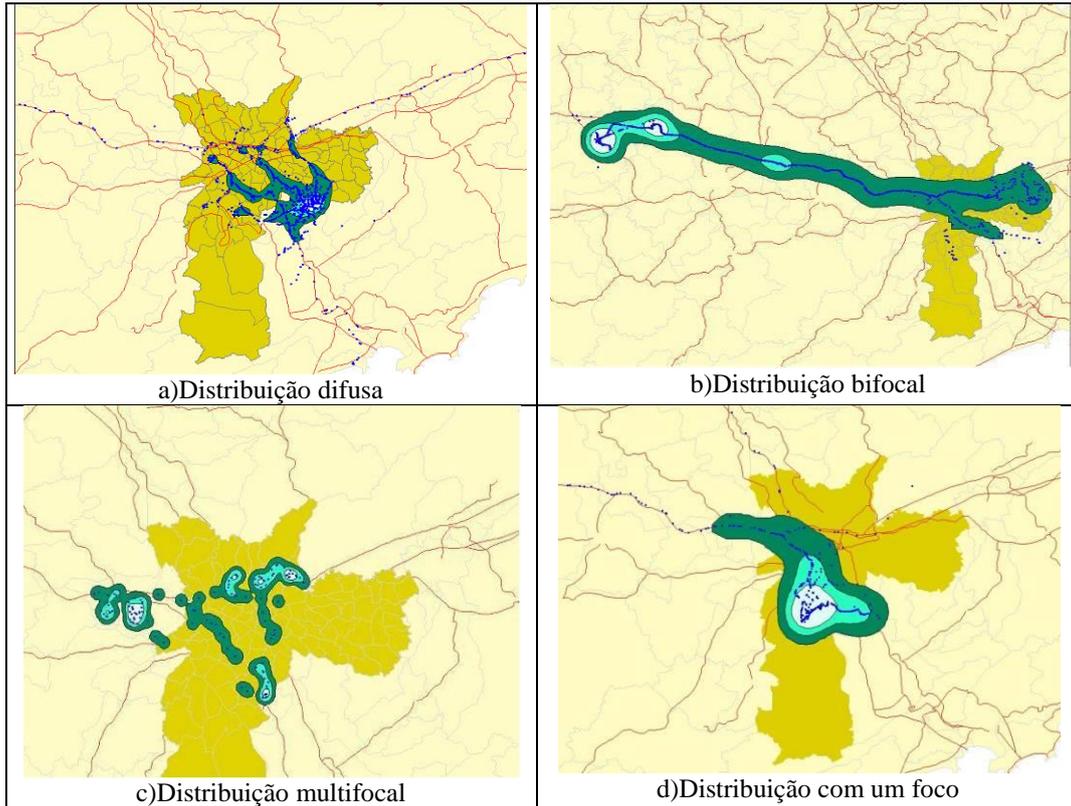


Figura 18 - Padrões de distribuição espacial da movimentação de veículos

Para a obtenção desses tipos além dos padrões de distribuição espacial, a forma de utilização também foi levada em consideração, e são descritas e representadas a seguir:

Distribuição difusa

Neste padrão (Figura 19) observamos um centro bem marcado e a ocorrência de outras concentrações de menor significância em várias regiões no entorno. Este padrão observado sugere a movimentação de veículos comerciais de entrega, que executam as tarefas de distribuição de mercadorias a partir de um centro de distribuição, sem preferência na seleção da área atendida.

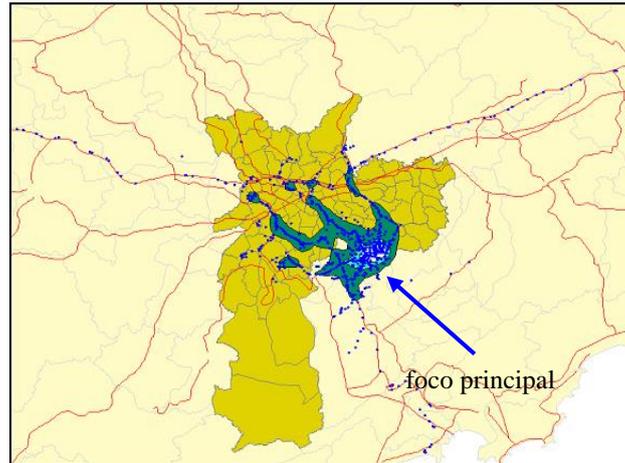


Figura 19 - Distribuição difusa

Distribuição bifocal

Neste padrão são observadas duas concentrações de pontos e uma região alongada entre estas duas concentrações, conforme observado na Figura 20

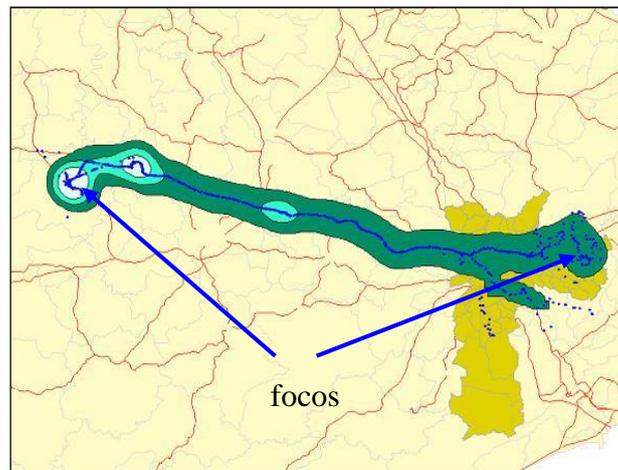


Figura 20 - Distribuição bifocal

Este tipo foi identificado em veículos que realizam um movimento rotineiro entre duas localidades, este movimento pode ser chamado de pendular. Veículos que realizam transporte entre dois (2) centros de distribuição ou entre empresas do mesmo grupo, possuem este tipo de movimento.

Um veículo de transporte pessoal que é utilizado apenas para a movimentação da residência para o local de trabalho gera também este padrão de movimentação. Nesta

distribuição de pontos as alterações dos padrões de movimentação são raras e detectadas facilmente.

Distribuição multifocal

A distribuição multifocal apresenta várias áreas de concentração de ocorrências dos pontos, e uma distribuição espacial ampla, sendo observadas múltiplas regiões para a ocorrência dos pontos.

O tipo de distribuição espacial multifocal (Figura 21) foi observado na movimentação de veículos de uso pessoal utilizados para executar diversas atividades, como por exemplo, ir ao trabalho, levar os filhos para a escola, ir ao clube, ir para academia de ginástica e as compras no shopping ou pode significar que existem diversos condutores. Outro tipo de veículo que apresenta este mesmo padrão é o de veículos de serviços ou socorro mecânico que tem várias bases de apoio e atendem em diversos locais.

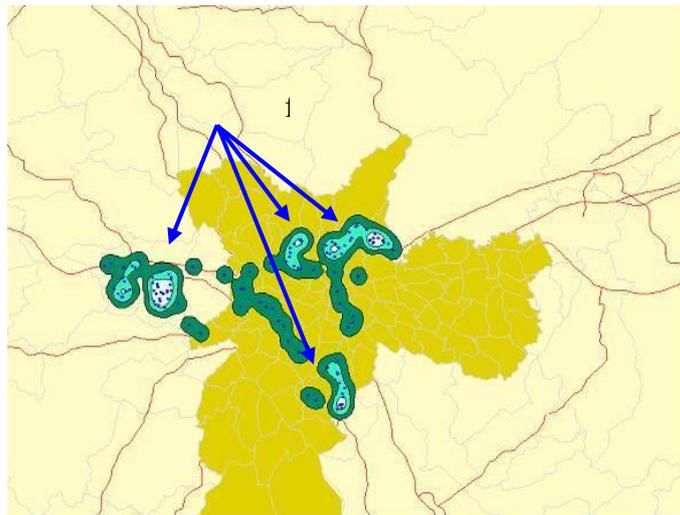


Figura 21 - Distribuição multifocal

Distribuição com um foco

Esta distribuição é caracterizada pela concentração dos pontos em apenas um local e a distribuição espacial é restrita a uma porção do espaço, como pode ser observado na Figura 22.

Este tipo de distribuição é característico para veículos de entregas ou prestação de serviços que tem um centro de distribuição ou apoio e uma região bem definida para a realização das entregas ou serviços.

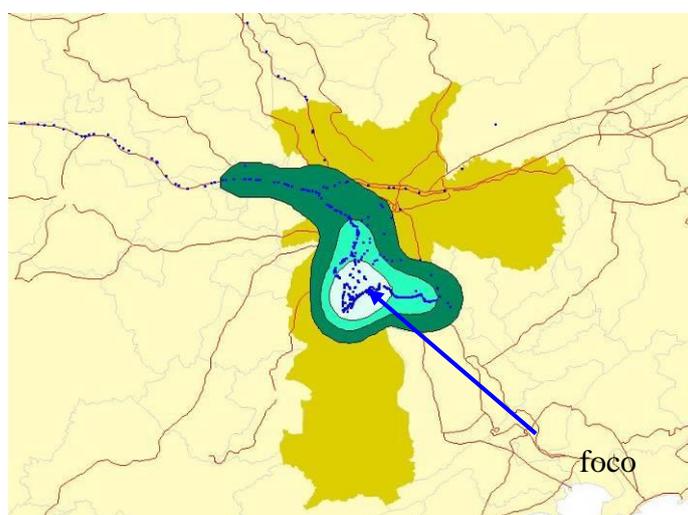


Figura 22 - Distribuição com um foco

A forma da distribuição espacial que representa a movimentação do veículo pode ser utilizada como indicador do tipo de utilização do veículo, como exemplo, a distribuição bifocal pode apontar o uso pessoal, ou mesmo o uso de transporte de carga entre dois (2) centros de distribuição.

5. EXPERIMENTO

5.1. Caracterização dos dados coletados

Para a modelagem temporal e análise espacial consideramos os dados de rastreamento referentes a utilização de um veículo por um (1) mês, com uma frequência nominal de coleta de dados de posição de 3 minutos. Com estas características de coleta foram detectados

16.382 eventos. O método de determinação da posição do veículo foi o GPS, com uma precisão nominal de 15 metros e a forma de comunicação utilizada foi a rede de telefonia móvel. Para esta fase da dissertação, arbitrou-se inicialmente a resolução temporal de 1 hora, desta forma consideramos que o comportamento de movimentação é homogêneo neste período de tempo.

5.2. Modelagem temporal

A partir das informações obtidas do monitoramento, foi construída uma tabela com os dados de posição espacial (UTM – E e N) e a data (hora e dia da coleta do dado), para facilitar a análise colunas auxiliares foram construídas:

- A coluna dia da semana - calculada a partir da data,
- A coluna hora - extraída da coluna hora e dia.
- A coluna deslocamento - calculada a partir da posição espacial detectada e da última posição coletada.

Como primeiro passo calculamos a equação de regressão utilizando como parâmetros todos os dias da semana e todas as horas do dia, seguindo a seguinte estrutura:

$$d = \theta_0 + \theta_1 X_1 + \theta_2 X_2 + \dots + \theta_k X_k + \varepsilon$$

d = Deslocamento

θ_0 = Intercepto (constante)

$\theta_1 \dots \theta_k$ = Coeficiente de regressão

X = Predictoras (dias da semana e horas do dia)

ε = Erro aleatório

Os cálculos de regressão foram realizados com o auxílio da software Minitab versão 16, disponibilizado pelo Centro de Computação eletrônica – CCE da Universidade de São Paulo. Os testes de hipótese (TRV e Wald) foram desenvolvidos no Software SciLab versão 4.1.2.

Na primeira etapa foram testadas as predictoras a um nível de 5% , determinando se estas são significativas e influenciam o modelo de regressão. Na segunda etapa foram agregadas as predictoras que possuíam similaridade, com a finalidade de diminuir a dimensionalidade da equação, simplificando desta forma o modelo de regressão. A seguir testamos a similaridade entre as equações de regressão, até encontrar uma equação com

melhor número de preditoras, que consiga explicar a movimentação do veículo e que seja equivalente à equação inicial que possui todas as preditoras identificadas.

Para a codificação das preditoras foi adotada a notação apresentada na tabela 4 e 5 :

Tabela 4 - Codificação das preditoras de dia da semana

Preditora	Dia da semana
ds1	Domingo
ds2	Segunda-feira
ds3	Terça-feira
ds4	Quarta-feira
ds5	Quinta-feira
ds6	Sexta-feira
ds7	Sábado

Tabela 5 - Codificação das preditoras de hora do dia

Preditora	Hora do dia
h0	0:00 as 0:59
h1	1:00 as 1:59
h2	2:00 as 2:59
h3	3:00 as 3:59
h4	4:00 as 4:59
h5	5:00 as 5:59
h6	6:00 as 6:59
h7	7:00 as 7:59
h8	8:00 as 8:59
H9	9:00 as 9:59
h10	10:00 as 10:59
h11	11:00 as 11:59
h12	12:00 as 12:59
h13	13:00 as 13:59
h14	14:00 as 14:59
h15	15:00 as 15:59
h16	16:00 as 16:59
h17	17:00 as 17:59
h18	18:00 as 18:59
h19	19:00 as 19:59
h20	20:00 as 20:59
h21	21:00 as 21:59
h22	22:00 as 22:59
h23	23:00 as 23:59

5.2.1. Avaliação da significância das preditoras

Para o cálculo da regressão foram adotados arbitrariamente, como referência as preditoras ds7 e h0, isto é o período de referência da equação de regressão são os sábados no

horário entre as 0:00 e as 0:59. Para melhor visualização da equação de regressão ela será apresentada em forma de tabela.

Tabela 6 - Equação de regressão

Preditora	Coefficiente	Probabilidade
Intercepto	-11.75	0
ds1	-3.60	0
ds2	0.14	0.686
ds3	0.06	0.856
ds4	0.69	0.035
ds5	1.58	0
ds6	-0.40	0.253
h1	3.43	0.002
h2	3.31	0.004
h3	1.07	0.4
h4	1.90	0.112
h5	1.67	0.166
h6	-1.06	0.396
h7	7.83	0
h8	7.11	0
h9	4.80	0
h10	5.76	0
h11	7.98	0
h12	8.14	0
h13	9.31	0
h14	6.92	0
h15	7.09	0
h16	6.95	0
h17	5.84	0
h18	8.61	0
h19	8.71	0
h20	8.57	0
h21	6.56	0
h22	4.98	0
h23	1.99	0.096
Erro	0.12	

Para esta equação obtemos -11026.066 como valor do logaritmo da Máxima verossimilhança.

A coluna probabilidade indica se a preditora é estatisticamente significativa. Os valores reduzidos de probabilidade indicam que existe uma diferença significativa entre as preditoras, ou seja, é improvável que a preditora tenha ocorrido por acaso. Para este estudo foi adotado uma significância de 5%, isto é, foram excluídas as preditoras que possuem uma probabilidade maior que 0,05. A adoção deste valor indica um limiar para avaliar o modelo

temporal, estudos adicionais deverão ser realizados para determinar se o valor adotado é adequado.

Para o primeiro refinamento da equação de regressão, as predictoras não utilizadas são o ds2, ds3, ds6, h3, h4, h5, h6, h23. A equação de regressão está representada na tabela 7

Tabela 7 - Equação de regressão (primeiro refinamento)

Preditora	Coefficiente	Probabilidade
Intercepto	-10.42	0
ds1	-3.56	0
ds4	0.73	0.003
ds5	1.62	0
h1	2.04	0.001
h2	1.89	0.004
h7	6.46	0
h8	5.72	0
h9	3.42	0
h10	4.39	0
h11	6.62	0
h12	6.78	0
h13	7.93	0
h14	5.55	0
h15	5.72	0
h16	5.58	0
h17	4.46	0
h18	7.21	0
h19	7.35	0
h20	7.20	0
h21	5.17	0
h22	3.60	0
Erro	0.12	

Para esta equação obtemos -11029.618 como valor do logaritmo da Máxima verossimilhança.

Para identificar se as duas equações são iguais para um nível de significância de 5 % utilizamos o Teste da Razão da Máxima Verossimilhança (TRV):

$$H_0: \text{Equação 1} = \text{Equação 2} \quad TRV < \chi_c^2$$

$$H_1: \text{Equação 1} \neq \text{Equação 2} \quad TRV > \chi_c^2$$

GL = Preditoras equação 1 – Preditoras equação 2 = 8

$$\text{TRV} = 2 [\log L(\hat{\theta}) - \log L(\hat{\theta}_0)] = 2 (-11026.066 + 11029.618)$$

$$\text{TRV} = 7.104$$

Para um nível de significância de 5% e um grau de liberdade = 8 , temos um $\chi_c^2 = 15.507$, para este valor aceitamos H0, isto é as equações são iguais.

5.2.2. Avaliação para agregação das preditoras

Para avaliar a possibilidade de agregar as preditoras, convencionamos que se a diferença entre os coeficientes era menor que 0,7 a sua igualdade deverá ser testada. Utilizando-se o teste de Wald é possível determinar a existência da igualdade entre as preditoras, quando o teste da igualdade for aceito, os dados são agrupados em uma nova preditora. A equação resultante deste processo de agregação foi comparada à primeira equação de regressão utilizando o teste da máxima verossimilhança.

- **Teste para as preditoras h1 e h2**

Testamos a seguinte hipótese

$$H_0: h_1 = h_2 \quad W < T_c^2$$

$$H_1: h_1 \neq h_2 \quad W > T_c^2$$

$$\text{Teste de Wald : } W = 0.0359249$$

Para $\alpha = 5\%$ e um GL = 1 temos um T de 3,84 então aceitamos H0, isto é os deslocamentos entre as 1:00 e as 2:59 são iguais.

Para esta situação calculamos a terceira equação de regressão e realizamos o teste da razão da máxima verossimilhança.

Teste da Razão de Máxima Verossimilhança (TRV)
Tabela 8 - Equação de Regressão (segundo refinamento)

Preditora	Coefficiente	Probabilidade
Intercepto	-10.42	0,000
ds1	-3.56	0,000
ds4	0.74	0,003
ds5	1,63	0,000

Tabela 9 - Equação de Regressão (segundo refinamento)(continuação)

Preditora	Coefficiente	Probabilidade
h1-h2	1,97	0,000
h7	6,46	0,000
h8	5,72	0,000
h9	3,43	0,000
h10	4,39	0,000
h11	6,62	0,000
h12	6,78	0,000
h13	7,93	0,000
h14	5,55	0,000
h15	5,73	0,000
h16	5,58	0,000
h17	4,46	0,000
h18	7,22	0,000
h19	7,36	0,000
h20	7,21	0,000
h21	5,17	0,000
h22	3,60	0,000
Erro	0,12	0,000

Para esta equação obtemos -11029.636 como valor do logaritmo da Máxima verossimilhança.

$$H0: \text{Equação 1} = \text{Equação 3} \quad TRV < \chi_c^2$$

$$H1 \text{ Equação 1} \neq \text{Equação 3} \quad TRV > \chi_c^2$$

$$GL = \text{Preditoras equação 1} - \text{Preditoras equação 2} = 8$$

$$TRV = 2(-11029.636 + 11026.066) = 7.14$$

Para um nível de significância de 5% e um grau de liberdade = 8 temos um $\chi_c^2 = 15.507$, para este valor aceitamos H0, isto é as equações são iguais.

- **Teste para as preditoras h7 e h8**

Testamos a seguinte hipótese

$$H0: h7 = h8 \quad W < \chi_c^2$$

$$H1: h7 \neq h8 \quad W > \chi_c^2$$

Teste de Wald :W = 1.9330528

Para $\alpha = 5\%$ e um GL = 1 temos um χ_c^2 de 3,84 então aceitamos H0, isto é os deslocamentos entre as 7:00 e as 8:59 são iguais.

Tabela 10 - Equação de regressão (terceiro refinamento)

Preditora	Coefficiente	Probabilidade
Intercepto	-10,43	0,000
ds1	-3,56	0,000
ds4	0,73	0,004
ds5	1,63	0,000
h1-h2	1,97	0,000
h7-h8	6,13	0,000
h9	3,43	0,000
h10	4,40	0,000
h11	6,62	0,000
h12	6,79	0,000
h13	7,94	0,000
h14	5,55	0,000
h15	5,72	0,000
h16	5,58	0,000
h17	4,47	0,000
h18	7,22	0,000
h19	7,35	0,000
h20	7,21	0,000
h21	5,18	0,000
h22	3,61	0,000
Erro	0,12	0,000

Para esta equação obtemos -11030.602 como valor do logaritmo da Máxima verossimilhança

H0: Equação 1 = Equação 4 $TRV < \chi_c^2$

H1: Equação 1 \neq Equação 4 $TRV > \chi_c^2$

GL = Preditoras equação 1 – Preditoras equação 4 = 8

$TRV = 2(-11030.602 + 11026.066) = 9.072$

gl = 8 15.507 ,

Para um nível de significância de 5% e um grau de liberdade = 8 temos um $\chi_c^2 = 15.507$, para este valor aceitamos H_0 , isto é as equações são iguais.

Os refinamentos serão realizados até encontrar a equação refinada que possua o menor número possível de preditoras, e que seja significativamente diferente da equação que contém todas as preditoras. Os cálculos estão descritos no Anexo I, e foram feitos até o sétimo refinamento.

5.2.3. Constatações da modelagem temporal

Definimos como a equação final de regressão a equação 7 (Anexo I), esta equação representa o modelo de comportamento temporal do veículo e a partir dela foram construídos gráficos de probabilidade do veículo estar parado (deslocamento menor que 15 metros) em função das horas do dia (figuras 23 e 24), e um conjunto de gráficos que representam o deslocamento em metros dado que temos uma probabilidade acumulada (figura 25), que foram utilizados para caracterizar a movimentação do veículo.

Deste processo de modelagem concluímos que temos um comportamento temporal com as seguintes características:

- Aos sábados, segundas, terças e sextas feiras temos um comportamento temporal similar.
- O comportamento temporal aos domingos, quartas e quintas feiras é distinto.
- Das 3:00hs às 6:00 e das 23:00 às 0:00, temos o comportamento similar.
- Os deslocamentos entre as 7:00 e as 8:59 são iguais.
- Os deslocamentos entre as 11:00 e as 12:59 são iguais.
- Os deslocamentos entre as 14:00 e as 16:59 são iguais
- Os deslocamentos entre as 18:00 e as 20:59 são iguais

Utilizando-se da equação final de regressão calculamos a probabilidade de o veículo estar parado e quantificamos o seu deslocamento. Com os resultados destes cálculos construímos os gráficos representados nas figuras 23, 24 e 25. A partir destes gráficos foi possível identificar o estado esperado para do veículo em determinada posição temporal.

A Figura 23 indica a probabilidade que o veículo esteja parado, por hora do dia, cada linha do gráfico, representa um dia ou um conjunto de dias da semana em que o

comportamento é similar. Para melhor observar este comportamento dividimos por dias da semana (Figura 24) e acrescentamos a representação do intervalo de confiança. Nestes gráficos temos então a probabilidade do veículo estar parado, em um intervalo de confiança de 5%.

Na figura 25, representamos o deslocamento esperado para uma probabilidade acumulada de 50 %, 90 % e 95%, definimos diversas faixas para uma melhor caracterização dos deslocamentos.

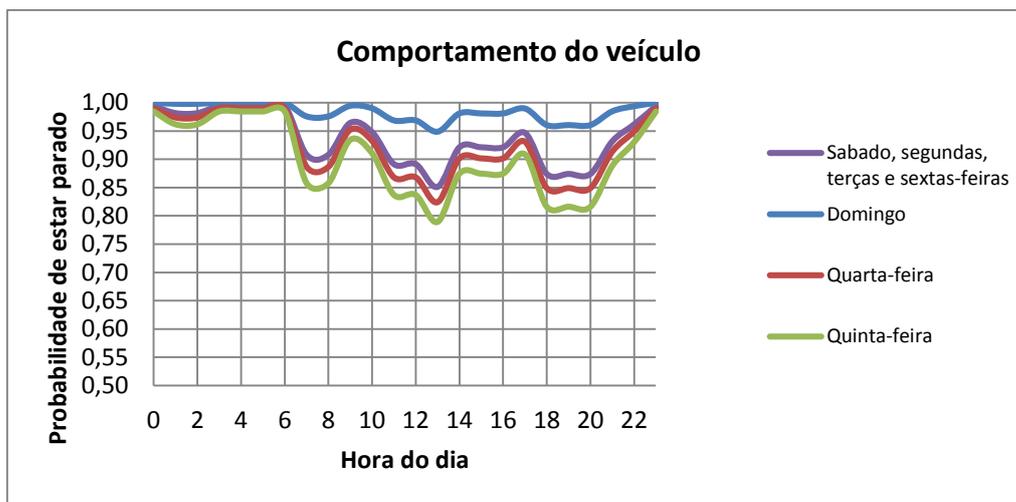


Figura 23- Probabilidade de o veículo estar parado

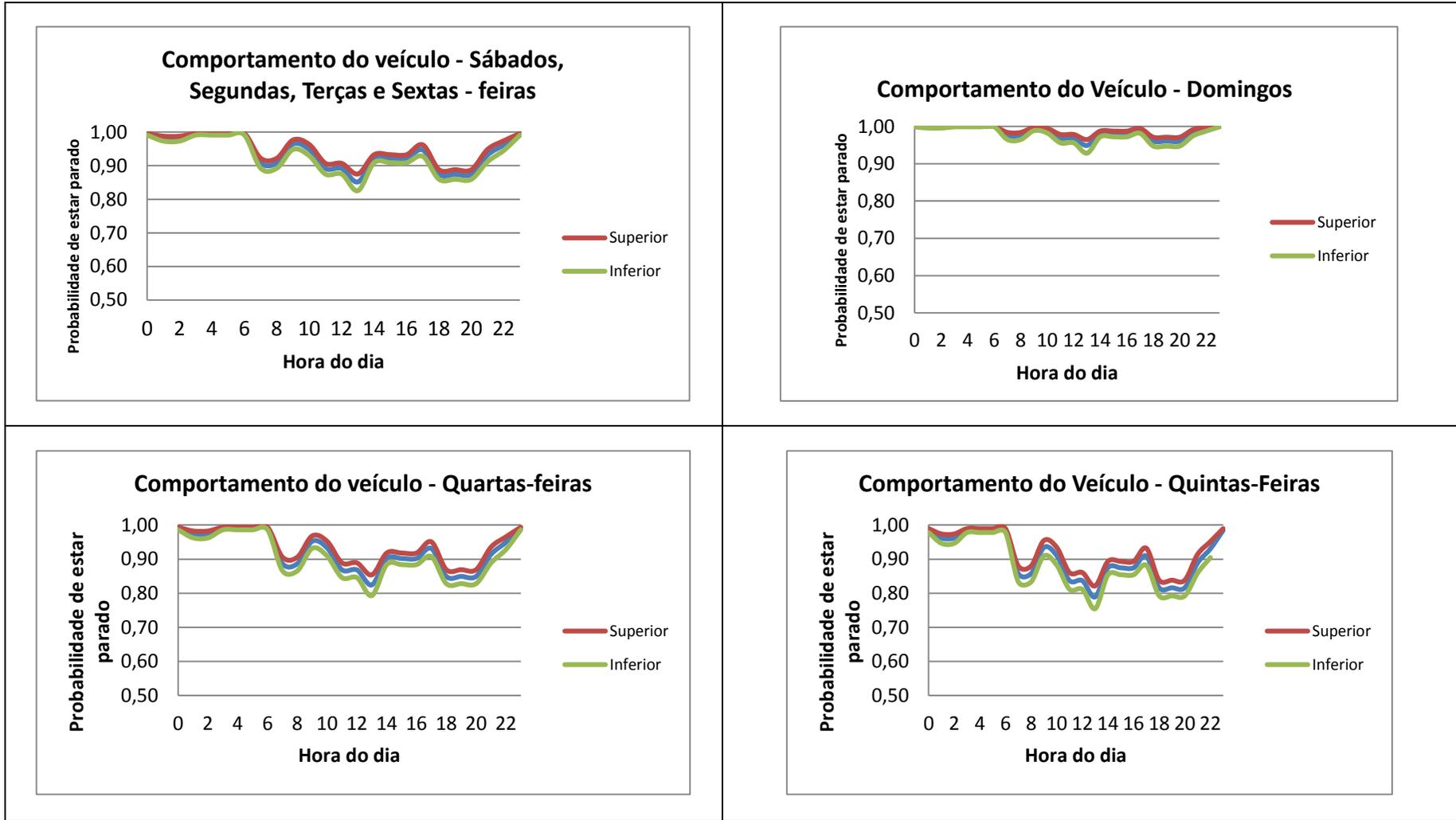


Figura 24 – Gráficos que representam a probabilidade do veículo estar parado

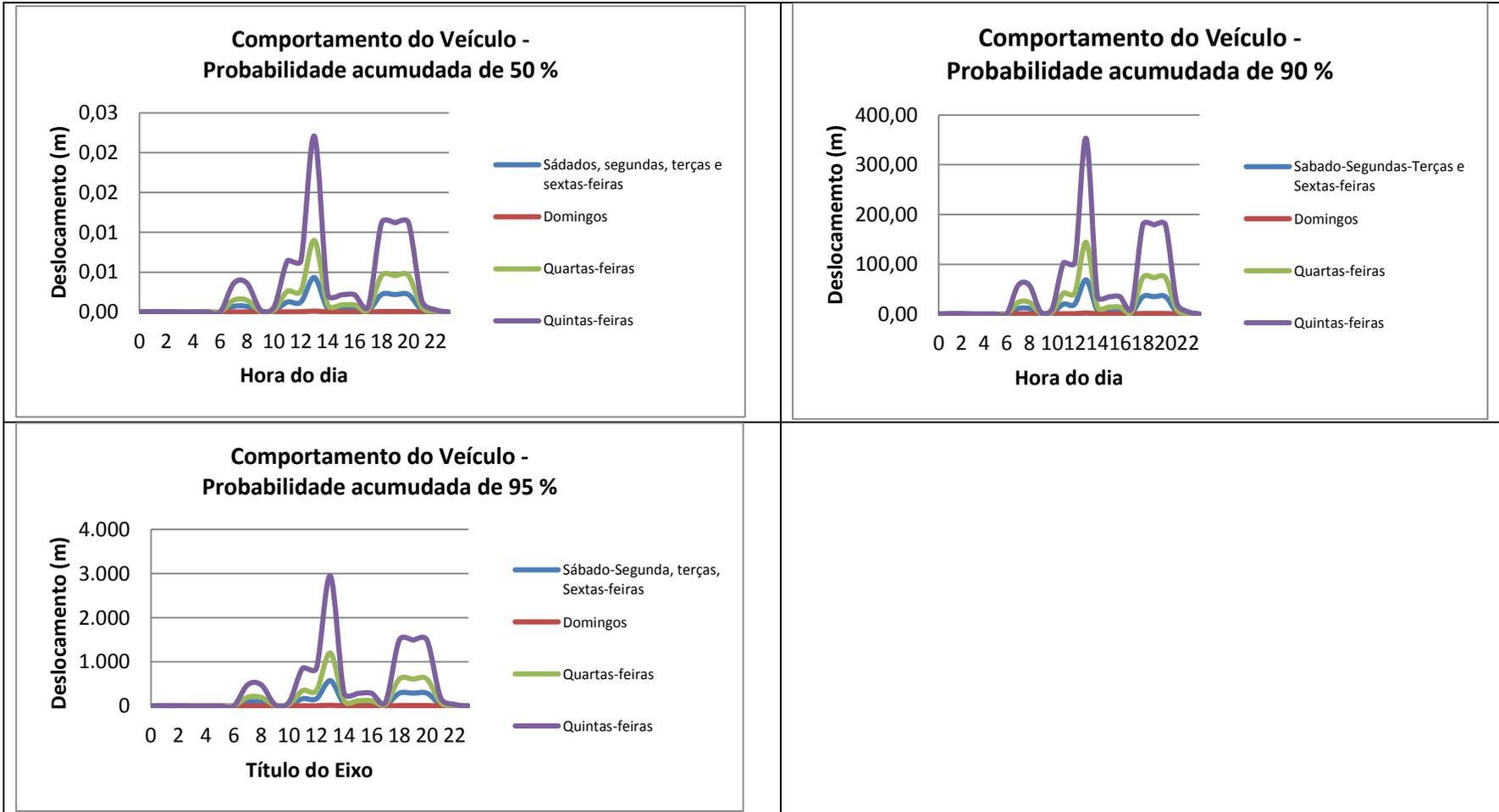


Figura 25 - Gráficos que representam os deslocamentos

5.3. Modelagem espacial.

Utilizando-se do mesmo conjunto de dados da modelagem temporal, geramos inicialmente uma modelagem com as posições espaciais dos eventos de monitoramento. Estes dados estão representados na figura 26. Constatamos que durante o tempo de observação o veículo deslocou-se apenas nos municípios de Santana do Parnaíba, Barueri e Carapicuíba.

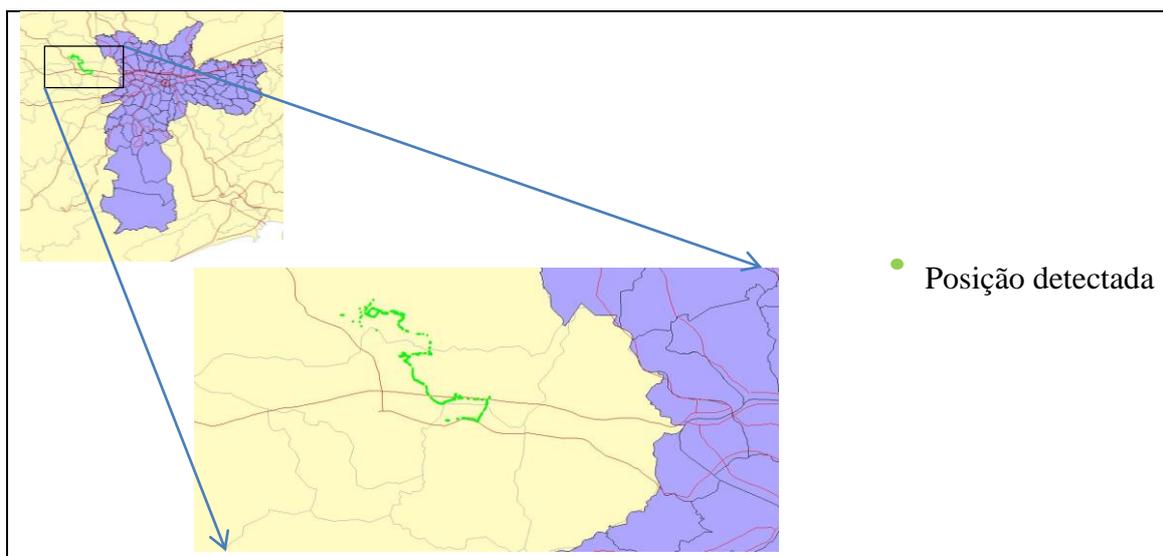


Figura 26 - Distribuição espacial dos eventos de monitoramento

Para estas ocorrências calculamos os valores do estimador de Kernel para determinar quais são as localidades aonde o veículo pode ser encontrado, utilizamos 15 metros para o parâmetro largura de faixa. O resultado desta análise está representado no mapa da Figura 27.



Figura 27 - Estimador de Kernel (probabilidade acumulada de 99%)

Para aplicação do método gráfico para a quantificação e delimitação do posicionamento espacial do veículo, foi construído um eixo de trajeto médio, utilizando como referência o mapa do estimador de Kernel e a delimitação de uma região ao redor do eixo da trajetória de 15 metros (Buffer) apresentado na figura 28.

As localidades definidas pelo estimador de Kernel e pelo Buffer do eixo do trajeto médio identificam as regiões mais prováveis onde o veículo ou objeto móvel pode ser .

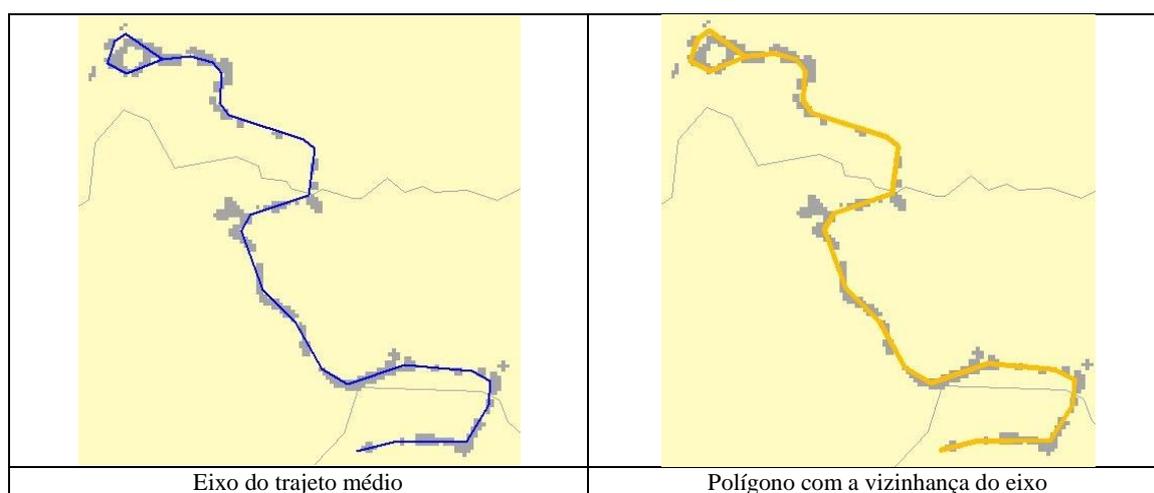


Figura 28 – Identificação das localidades

6. CONCLUSÃO

A metodologia aplicada para a modelagem temporal e espacial forneceram informações suficientes para identificar o comportamento espacial e temporal do veículo monitorado para um nível de confiabilidade de 5%, definindo os horários de utilização do veículo, as localidades mais prováveis e predizendo a movimentação esperada para o veículo. A avaliação e a obtenção do modelo espacial e temporal, assim como, os resultados obtidos com as equações de regressão, referem-se apenas ao veículo estudado. Para a caracterização de outros veículos, devemos aplicar novamente o processo de modelagem temporal e espacial.

Destaca-se que os resultados obtidos nesta dissertação, possui uma caracterização do comportamento temporal, mais detalhada em comparação com Chen *et al.* (2011) entretanto a caracterização espacial obtida neste trabalho não possui a sofisticação observada nos trabalhos de Moreno *et al.* (2010) e de Mesbah *et al.* (2012).

Em continuidade a este trabalho, a elaboração de uma modelagem espacial mais refinada e um integração entre os modelos temporais e espaciais trará uma importante contribuição para a geração de modelos espaço-temporal para objetos móveis.

7. REFERENCIAS BIBLIOGRAFIAS

AMORIM, A. M.; CAMPOS, J. **A conceptual model for representation of taxi trajectories**. Proc. of XIII Brazilian Symposium on GeoInformatics - Geoinfo, p. 96-107. 2012.

ANDRIENKO, G.; ANDRIENKO, N.; DEMSAR, U.; DRANSCH, D.; DYKES, J.; FABRIKANT, S.I.; JERN, M.; KRAAK, M.; SCHUMANN, H.; TOMINSKI, C. Space, time and visual analytics, **International Journal of Geographical Information Science**, ed. Taylor & Francis, vol. 24, Issue 10, 2010.

BATTY, M.; COLE, S. Time and space: geographic perspectives on the future. **Future**, vol 29 n 3-4, Londres, p 277-289, maio, 1997 ISSN: 0016-3287.

BLITZKOW, D.; CINTRA, J.P.;NETTO, N.P.; FONSECA JUNIOR, E.S. **Informações Espaciais II** . São Paulo: EPUSP, 2007, 153p., Departamento de Engenharia de Transportes, Notas de Aula da disciplina PTR2202.

BRASIL. Conselho Nacional de Transito – CONATRAN, **Altera a Resolução nº 330, de 14 de agosto de 2009, que estabelece o cronograma para a instalação do equipamento obrigatório definido na Resolução nº 245/07**. Deliberação 135 de 30 de Janeiro de 2013.

BRASIL. Conselho Nacional de Transito – CONATRAN, **Estabelece o cronograma para a instalação do equipamento obrigatório definido na Resolução nº 245/2007, denominado antifurto nos veículos novos, nacionais e importados**. Resolução 330 de 14 de agosto de 2009.

BUCKLAND, S.T.;ELSTON, D.A. Empirical models for the spatial distribution of wildlife. **The journal of applied ecology**, vol30, n 3, Londres, p. 478-495, 1993.

BUONAVOGLIA, C., Grupo José Alves investe em rastreamento de veículos [Depoimento a Dubes Sônego e Ana Paula Machado], **Revista Brasil Econômico**, São Paulo, 2011 disponível em http://brasileconomico.ig.com.br/noticias/grupo-jose-alves-investe-em-rastreamento-de-veiculos_101207.html.

BUSSAB, W. O.; MORETTIN, P. A. **Estatística Básica**. 5.ed. Rio de Janeiro: Editora Saraiva, 2006.

CÂMARA, G; CARVALHO, M.S. Análise de eventos pontuais In: DRUCK, S.; CARVALHO, M.S.; CÂMARA, G.; MONTEIRO, A.V.M. (eds) **Análise Espacial de Dados Geográficos**. Brasília, EMBRAPA, 2004. c 2 p. 1-15 ISBN: 85-7383-260-6. Disponível em: <<http://www.dpi.inpe.br/gilberto/livro/analise/>> acesso em 18 julho 2005.

CHEN, J.; SHAW, S.-L.; YU, H.; LU, F.; CHAI, Y.; JIA, Q. Exploratory data analysis of activity diary data: a space-time GIS approach. **Journal of Transport Geography**, 19, p. 394-404. 2011.

COLOSIMO, E.A.; GIOLO, S.R. **Análise de Sobrevivência Aplicada**, 1 ed. Editora Edgard Blucher, 2006.

COX, D.R.. Regression Models and Life-Tables. **Journal of the Royal Statistical Society. Serie B (Metodological)**, Vol 20 n 2, Londres, p. 187-220, 1972.

DJUKNIC, G.M.; RICHTON, R.E. Geolocation and Assisted GPS. **COMUNICATION**, p.123 – 125, IEEE Computer Society, 2002 ISSN 0018-9162.

DOWNS, J.A.; HOENER, M. Probabilistic potential path tree for visualizing and analyzing vehicle tracking data. **Journal of Transport Geography**, 23, p. 72-80. 2012.

FANG, Z.; SHAW, S.L.; TU, W.; LI, Q.; LI, Y. Spatiotemporal analysis of critical transportation links based on time geographic concepts: a case study of critical bridges in Wuhan, China. **Journal of Transport Geography**, 21, p. 44-59. 2012.

FLAMM, M.; JEMELIM, C.; KAUFMANN, V. **Combining person based GPS tracking and prompted recall interviews for a comprehensive investigation of travel behavior adaptation processes during life course transitions**. 7th Swiss Transport Research Conference, Monte Verità/Ascona, p. 1 – 17. 2007.

GÄRLING, T.; AXHAUSEN, K. W. Introduction: Habitual Travel Choice **Transportation**, ed. Springer, Vol. 30, Issue 1, p.1-11 February 2003.

GIANNOTTI, F.; NANNI, M.; PEDRESCHI, D.; PINELLI, F. **Trajectory pattern mining**. In: Berkhin, P., Caruana, R., Wu, X., eds: KDD, ACM, p. 330-339. San José, California, USA. 2007.

GONÇALVES, W.G. Rastreamento e monitoramento: dois setores distintos, que ainda provocam confusão. **Revista LogWEB**, n 128, São Paulo, p.49-52, 2012. Disponível em <http://www.logweb.com.br/novo/revista/edicoes/ver/eid/MTI4>, Acesso em 03/02/2013.

GONZÁLEZ, M. C.; HIDALGO, C. A.; BARABÁSI, A.-L. Understanding individual human mobility patterns. **Nature**, 453 (7196):779-82. 2008.

GOODCHILD, M. GIS and Transportation: Status and Challenges, **Geoinformatica**, vol 4 n 2, Springer, 2000.

GRISTEC - **Associação Brasileira das Empresas de Gerenciamento de Riscos e de Tecnologia**, São Paulo. Disponível em <http://www.gristec.com.br/index.php>, Acesso em 07/02/2012.

GÜTING, R. H.; SCHNEIDER, M. **Moving Objects Databases**. Morgan Kaufmann Publishers, San Francisco, USA, 2005.

HÄGERTRAND, T. What About People in Regional Science, **European Congress Regional Science Association**. Vol. 24, Regional Science Association, 1970.

HOOGE. P. N.; EICHERNLAUB, W.M.; SALOMON, E.K. **Using GIS to analyze animal movements in the marine environment**, Alaska: USGS, 2000 disponível em <http://www.absc.usgs.gov/glba/gistools/anim_mov_useme.pdf> acesso em: 18 julho 2005.

HORNER, M. W.; ZOOK, B.; DOWNS, J. A. Where were you? Development of a time-geographic approach activity destination re-construction. **Computers, Environment and Urban Systems**, Ed. Elsevier, Vol. 36 (2012), p. 488–499.

JENNERICH, D.L ; TURNER F.B, **Measurement of non-circular home range**, Theoretical Biology, Elsevier, Vol 22, p. 227-237, 1969.

KAPLAN, E.D.; HEGARY, C. **Understanding GPS: Principles and Applications**, Ed. Artech House, LONDON, 2 ed., 2006.

MESBAH, M.; CURRIE, G.; LENNON, C.; NORTHCOTT, T. Spatial and temporal visualization of transit operations performance data a network level. **Journal of Transport Geography** 225, p. 15-25, 2012.

MILLER, H. J., Modelling Accessibility Using Space-Time Prism Concepts within Geographic Information Systems. **International Journal of Geographical Information Systems** vol. 5(3), 287-301. 1991.

MILLER, H.J., Activities in Space And Time, **Handbook of transport geography and spatial systems**, Hensher, D., Button, K. (Eds), Elsevier, Cap 63 p 647-659, 2004.

MILLER, H.J., Necessary space time conditions for human interaction **Environment and Planning B: Planning and Design** vol.32, p. 381 - 401, 2005.

MOGHADDAM, S. S.; NOROOZI, R.; CASELLO, J. M.; HELLINGA, B. Predicting the mean and variance of transit segment and route travel times. Transportation Research Record. **Journal of Transportation Research Board**, no. 2217, Washington, DC, p. 30 – 37. 2011.

MONICO, J.F.G, **Posicionamento pelo GNSS - Descrição, Fundamentos e Aplicações**, ed. UNESP, 2ª ed. 2008 480p.

MONTGOMERY, D. C. **Introdução ao Controle Estatístico de Qualidade**. LTC, 4.ed., 2004.

MORENO B. N.; TIMES, V. C.; RENSO. C.; BOGORNY, V. **Looking inside the stops of trajectories of moving objects**. XI Brazilian Symposium on GeoInformatics - Geoinfo 2010, Campos do Jordão, Brazil, INPE, p. 9-20. 2010.

MURUGANANDHAN, P. R. M. Real time Web based vehicle tracking using GPS. World Academy of Science, **Engineering and technology**, 61, p. 91 - 99. 2010.

NANNI, M. PEDRESCHI, D. Time-focused clustering of trajectories of moving objects, **Journal of Intelligent Information Systems**. ed. Springer ,vol.27, Issue 3, p 267-289, November 2006.

NITSCHKE, P.; WIDHRALM, P.; BREUSS, P. M. A strategy on how to utilize smartphones for automatically reconstructing trips in travel surveys. *Procedia - Social and Behavioral Sciences*, Vol. 48, p. 1033-1046. **Transport Research Arena**. 2012.

OLIVEIRA, M. G.; BATISTA, C. S. **GeoSTAT- A system for visualization, analysis, and clustering of distributed spatiotemporal data**. XIII Brazilian Symposium on GeoInformatics Geoinfo 2012, Campos do Jordão, Brazil, INPE, p. 108-119. 2012.

PELEKIS N.; KOPANAKIS, I.; MARKETOS, G.; NTOUTSI I.; ANDRIENKO,G,; THEODORIDIS, Y. Similarity Search in Trajectory Databases . **14th International Symposium on Temporal Representation and Reasoning**, ed. IEEE, p 2007.

PEREIRA, H.R., Rastreamento e monitoramento: dois setores distintos, que ainda provocam confusão [Depoimento para Wanderley Gonelli Gonçalves]. **Revista LogWEB**, n 128, São Paulo, p.49-52, 2012. Disponível em <http://www.logweb.com.br/novo/revista/edicoes/ver/eid/MTI4>, Acesso em 03/02/2013.

PRED, A., The Choreography of Existence: Comments on Hägerstrand's Time-Geography and Its Usefulness, **Economic Geography**, ed. Clark University vol. 53, n°. 2, Abr. 1977.

RAUBAL, M. WINTER, S.; TEBMANN, S. GAISBAUER, C. Time geography for *ad-hoc* shared-ride trip planning in mobile geosensor networks, **ISPRS Journal of Photogrammetry and Remote Sensing**, ed. Elsevier vol. 62, Issue 5, p. 366–381 October 2007.

REVISTA CESVI. **Sistemas de Rastreamento e Bloqueio**, São Paulo: Ed. CESVI, n° 84, Março-Abril, 2013 p 34-35.

ROCHA, A., CESVI Promove debate sobre o mercado de rastreamento, **Revista Apólice**, São Paulo, Dez/2012, disponível em <http://revistaapolice.com.br/2012/12/cesvi-promove-debate-sobre-o-mercado-de-rastreamento/>. Acesso em 23/02/2013.

RODRIGUES, M.; CUGNASCA; C.E.; QUEIROS FILHO, A, **Rastreamento de Veículos**, Oficina dos Textos, 1ª Ed., 2009.

ROTH, S H. History of Automatic Vehicle Monitoring (AVM), **IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY**, VOL. VT-26, N°. 1 , IEEE,1977.

SCHNEIDER, M. Moving Objects in Databases and GIS: State-of-the-Art and Open Problems. **Research Trends in Geographic Information Science**. Springer-Verlag, 169-188, 2009c.

SCHNEIDER, M. Soft Computing Techniques in Spatial Databases. Handbook of Research on Soft Computing Applications for Database Technologies: **Techniques and Issues**. IGI Global, 49-71, 2010.

SCHNEIDER, M. **Spatial and moving objects database: state of the art and future research challenges**. XIII Brazilian Symposium on GeoInformatics – GeoInfo 2012 – Final Program – Abstracts. Campos do Jordão, Brazil, INPE,2012.

SCHNEIDER, M. Spatial and Spatio-Temporal Data Models and Languages. **Encyclopedia of Database Systems**. Springer-Verlag, 2009a.

SCHNEIDER, M. Spatial Data Types. **Encyclopedia of Database Systems**. Springer-Verlag, 2009b.

SPACCAPIETRA, S.; Parent, C.; Damiani, M. L.; Macedo J. A., Porto, F., Vangeno, C. A conceptual view on trajectories. **Data & Knowledge Engineering**, ed. Elsevier vol. 65, Issue 1, Pages 126–146, April 2008.

TELECO – Inteligência em telecomunicações - **Cobertura das Operadoras e População Atendida**, disponível em <<http://www.teleco.com.br/cobertura.asp>> acesso em 01/07/2013.

WANG, S, MIN, J. YI, B.K Location Bases Service for Mobiles: Technologies and Standards **IEEE International Conference on Communications**, Beijing, IEEE 2008.

WORTON, B.J Kernel methods of estimating the utilization distribution in home range studies. **Ecology**, Tempe, v.70 p 164-168, 1989.

XIA, J.; ARROSMITH, C. A. Techniques for counting and tracking the spatial and temporal movement of visitors. **Monitoring, Simulation, and Management of Visitors Landscape**. Randy Gimblett e Hans Skov-Petersom (eds.). 85-105. 2008.

XIAO, Y.; LOW, D.; BANDARA, T.; PATHAK, P.; LIM, A. B.; GOYAL, D.; SANTOS, J.; COTTRILL, C.; PEREIRA, P.; ZEGRAS, C.; BEM-AKIVA, M. Transportation activity analysis using smartphones. **Proc. of IEEE Consumer Communication & Networking Conference - CCNC'12**, p. 60-61. 2012.

YU, H. Visualizing and analyzing activities in an integrated space-time environment: temporal GIS design and implementation, **Transportation Research Record: Journal of the Transportation Research Board**, Volume 2024, p. 54-62.2007.

Anexo I

Memória de Cálculo do Refinamento das Equações de Regressão

- **Quarto Refinamento**
- **Teste para as preditoras h11 e h12**

Testamos a seguinte hipótese

$$H_0: H_{11} = h_{12} \quad W < \chi_c^2$$

$$H_1: H_{11} \neq h_{12} \quad W > \chi_c^2$$

Teste de Wald: $W = 0.1026486$

Para $\alpha = 5\%$ e um GL = 1 temos um χ_c^2 de 3,84 então aceitamos H_0 , isto é os deslocamentos entre as 11:00 e as 12:59 são iguais

Equação de regressão (quarto refinamento)

Preditoras	Coefficiente	Probabilidade
Intercepto	-10,43	0,000
ds1	-3,58	0,000
ds4	0,73	0,004
ds5	1,63	0,000
h1-h2	1,97	0,000
h7-h8	6,13	0,000
h9	3,43	0,000
h10	4,40	0,000
h11-h12	6,71	0,000
h13	7,94	0,000
h14	5,55	0,000
h15	5,72	0,000
h16	5,58	0,000
h17	4,47	0,000
h18	7,22	0,000
h19	7,35	0,000
h20	7,21	0,000
h21	5,18	0,000
h22	3,61	0,000
Erro	0,12	

Para esta equação obtemos -11030.653 como valor do logaritmo da Máxima verossimilhança

H0: Equação 1 = Equação 5 $TRV < \chi_c^2$

H1 Equação 1 \neq Equação 5 $TRV > \chi_c^2$

GL = Preditora equação 1 – Preditora equação 5 = 8

$TRV = 2(-11030.653+11026.066) = 9.174$

Para um nível de significância de 5% e um grau de liberdade = 8, temos um $\chi_c^2 = 15.507$, para este valor aceitamos H0, isto é as equações são iguais.

- **Quinto Refinamento**
- **Teste para os preditoras h14, h15 e h16**

Testamos a seguinte hipótese:

H0: h14 = h15 = h16 $W < \chi_c^2$

H1: h14 \neq h15 \neq h16 $W > \chi_c^2$

Teste de Wald: $W = 0.1049474$

Para $\alpha = 5\%$ e um GL = 2 temos um χ_c^2 de 3,84 então aceitamos H0, isto é os deslocamentos entre as 14:00 e as 16:59 são iguais

Equação de regressão (quinta refinamento)

preditora	Coefficiente	Probabilidade
Intercepto	-10,43	0,000
ds1	-3,58	0,000
ds4	0,74	0,003
ds5	1,64	0,000
h1-h2	1,97	0,000
h7-h8	6,13	0,000
h9	3,43	0,000
h10	4,40	0,000
h11-h12	6,71	0,000
h13	7,94	0,000
h14-h15-h16	5,62	0,000
h17	4,46	0,000
h18	7,22	0,000
h19	7,35	0,000
h20	7,21	0,000
h21	5,18	0,000
h22	3,61	0,000
Erro	0,12	0,000

Para esta equação obtemos -11030.705 como valor do logaritmo da Máxima verossimilhança

$$H_0: \text{Equação 1} = \text{Equação 6} \quad TRV < \chi_c^2$$

$$H_1 \text{ Equação 1} \neq \text{Equação 6} \quad TRV > \chi_c^2$$

$$GL = \text{Preditora 1} - \text{Preditora equação 6} = 8$$

$$TRV = 2(-11030.705 + 11026.066) = 9.278$$

Para um nível de significância de 5% e um grau de liberdade = 8, temos um $\chi_c^2 = 15.507$, para este valor aceitamos H_0 , isto é as equações são iguais.

- **Sexto refinamento**
- **Teste para as preditoras h18, h19 e h20**

Testamos a seguinte hipótese

$$H_0: h_{18} = h_{19} = h_{20} \quad w < \chi_c^2$$

$$H_1: h_{18} \neq h_{19} \neq h_{20} \quad w > \chi_c^2$$

$$\text{Teste de Wald : } W = 0.0961239$$

Para $\alpha = 5\%$ e um $GL = 2$ temos um χ_c^2 de 3,84 então aceitamos H_0 , isto é os deslocamentos entre as 18:00 e as 20:59 são iguais

Equação de regressão (sexto refinamento)

Preditora	Coefficiente	Probabilidade
Intercepto	-10,43	0,000
ds1	-3,58	0,000
ds4	0,74	0,003
ds5	1,63	0,000
h1-h2	1,97	0,000
h7-h8	6,12	0,000
h9	3,43	0,000
h10	4,40	0,000
h11-12	6,71	0,000
h13	7,94	0,000
h14-h15-h16	5,62	0,000
h17	4,47	0,000
h18-h19-h20	7,26	0,000
h21	5,18	0,000
h22	3,61	0,000
Erro	0,12	0,000

Para esta equação obtemos -11030.753 como valor do logaritmo da Máxima verossimilhança

$$H_0: \text{Equação 1} = \text{Equação 7} \quad TRV < \chi^2_c$$

$$H_1: \text{Equação 1} \neq \text{Equação 7} \quad TRV > \chi^2_c$$

$$GL = \text{Preditoras equação 1} - \text{Preditores equação 6} = 8$$

$$TRV = 2(-11030.755 + 11026.066) = 9.378$$

Para um nível de significância de 5% e um grau de liberdade = 8, temos um $\chi^2_c = 15.507$, para este valor aceitamos H_0 , isto é as equações são iguais.

Com este teste encerramos a avaliação dos preditoras que possuem uma diferença de até 0,7 entre os coeficientes das preditoras vizinhas.

Para o próximo passo testamos as preditoras que possuem um coeficiente com variação maior que 0,7 e menor que 1,5.

- **Sétimo Refinamento**
- **Teste para as preditoras h9 e h10**

Testamos a seguinte hipótese

$$H_0: h_9 = h_{10} \quad w < \chi_c^2$$

$$H_1: h_9 = h_{10} \quad w > \chi_c^2$$

$$\text{Teste de Wald: } W = 2.7148218$$

Para $\alpha = 5\%$ e um $GL = 2$ temos um χ_c^2 de 3,84 então aceitamos H_0 , isto é os deslocamentos entre as 9:00 e as 10:59 são iguais

Equação de regressão (sétimo refinamento)

Preditora	Coefficiente	Probabilidade
Intercepto	-10,44	0,000
ds1	-3,56	0,000
ds4	0,73	0,003
ds5	1,62	0,000
h1-h2	1,97	0,000
h7-h8	6,13	0,000
h9-h10	3,95	0,000
h11-12	6,71	0,000
h13	5,62	0,000
h14-h15-h16	7,26	0,000
h17	7,94	0,000
h18-h19-h20	4,47	0,000
h21	5,18	0,000
h22	3,61	0,000
Erro	0,12	0,000

Para esta equação obtemos -11031.868 como valor do logaritmo da Máxima verossimilhança.

$$H_0: \text{Equação 1} = \text{Equação 8} \quad TRV < \chi_c^2$$

$$H_1: \text{Equação 1} \neq \text{Equação 8} \quad TRV > \chi_c^2$$

$$GL = \text{Preditoras equação 1} - \text{Preditoras equação 6} = 8$$

$$TRV = 2(-11031.868 + 11026.066) = 13.604$$

Para um nível de significância de 5% e um grau de liberdade = 8, temos um $\chi_c^2 = 15.507$, para este valor aceitamos H0, isto é as equações são iguais.

Apesar de aceitar a hipótese h0, o valor da TRV está próximo ao valor crítico, isto significa que a equação resultante deste agrupamento de preditoras está se tornando significativamente diferente a equação de regressão formulada com todas as preditoras observadas. Por esta razão encerraremos os testes de agrupamento das preditoras, desconsiderando a equação 8.