



EMBRAPA - UFPA - MPEG
UNIVERSIDADE FEDERAL DO PARÁ
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS AMBIENTAIS

CÁSSIA CAMILA SILVA DA SILVA

PREVISÃO DE RAIOS ATRAVÉS DE ÍNDICES TERMODINÂMICOS
DA ATMOSFERA, EM TORNO DE BELÉM DO PARÁ

BELÉM – PA

2014

CÁSSIA CAMILA SILVA DA SILVA

**PREVISÃO DE RAIOS ATRAVÉS DE ÍNDICES TERMODINÂMICOS
DA ATMOSFERA, EM TORNO DE BELÉM DO PARÁ**

Dissertação de mestrado apresentada ao Programa de Pós-Graduação em Ciências Ambientais do Instituto de Geociências da Universidade Federal do Pará, Museu Paraense Emílio Goeldi e a EMBRAPA Amazônia Oriental, como requisito parcial para obtenção do grau de Mestre em Ciências Ambientais.

Área de concentração: Física do Clima.

Orientador: Prof^o.Dr^o José Ricardo Santos de Souza

BELÉM –PA

2014

Dados Internacionais de Catalogação de Publicação (CIP)
(Biblioteca do Instituto de Geociências/UFPA)

Silva, Cássia Camila Silva da, 1988-

Previsão de raios através de índices termodinâmicos da atmosfera, em torno de Belém do Pará / Cássia Camila Silva da Silva. – 2014.

Inclui bibliografias

Orientador: José Ricardo Santos de Souza
Dissertação (Mestrado) – Universidade Federal do Pará, Instituto de Geociências, Programa de Pós-Graduação em Ciências Ambientais, Belém, 2014.

1. Raio – Belém (PA). 2. Árvores de decisão. I. Título.

CDD 22. ed. 551.5632

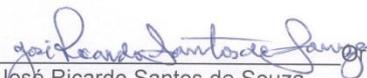
CÁSSIA CAMILA SILVA DA SILVA

PREVISÃO DE RAIOS ATRAVÉS DE ÍNDICES TERMODINÂMICOS DA
ATMOSFERA, EM TORNO DE BELÉM DO PARÁ

Dissertação apresentada para obtenção do grau de
Mestre em Ciências Ambientais do Programa de Pós-
Graduação em Ciências Ambientais, Instituto de
Geociências, Universidade Federal do Pará. Área de
Física do Clima
Orientador: Dr. José Ricardo Santos de Souza

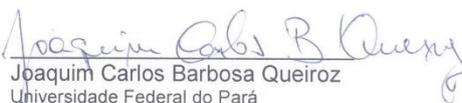
Data de aprovação: 20 / 01 / 2014

Banca Examinadora:


José Ricardo Santos de Souza Orientador
Universidade Federal do Pará


Arthur da Costa Almeida
Universidade Federal do Pará


Steel Silva Vasconcelos
EMBRAPA


Joaquim Carlos Barbosa Queiroz
Universidade Federal do Pará

Aos meus pais, Gerônimo e Marlene, ao meu orientador
José Ricardo S. de Souza e ao professor Arthur Almeida, agradeço por toda dedicação, e
apoio.

AGRADECIMENTOS

Primeiramente, a Deus pelas oportunidades, as quais me ajudaram a crescer como pessoa e alcançar meus objetivos de vida, inclusive a conclusão deste trabalho.

A meus pais, José Gerônimo e Maria Marlene, e a minha irmã Miscilene e aos meus sobrinhos Renato e Felipe, por todo carinho, paciência, e amor.

A minha Tia Célia, por me acolher de braços abertos, por sua orientação e paciência que foram de fundamental importância para meu bem estar.

Agradeço ao Prof. José Ricardo Santos de Souza pela orientação acadêmica, pelas boas conversas e pelos inúmeros ensinamentos que levo para minha vida pessoal e acadêmica e principalmente pela oportunidade de realizar o processo de Mestrado em Ciências Ambientais /UFPA, sem o que não seria possível a execução deste trabalho.

A CAPES pela concessão da bolsa, que possibilitou a minha permanência no curso e dedicação a esse estudo.

Agradeço ao Prof. Dr. Arthur da Costa Almeida pelo apoio acadêmico e o constante apoio durante todo o período de Mestrado e de desenvolvimento do trabalho. Pelos longos contatos telefônicos, inúmeras contribuições, sem as quais também não seria possível a execução dessa Dissertação.

Aos meus amigos de pesquisa: Wanda, Raimundo e Marcela.

Agradeço a Prof^a. Dr^a. Brígida Ramati Pereira da Rocha pelo apoio acadêmico e confiança depositada em mim para terminar essa Dissertação.

Agradeço aos professores Steel Vasconcelos, Edson Rocha e Joaquim Queiroz por todas as contribuições dadas na banca de qualificação deste trabalho.

Agradeço à Universidade Federal do Pará pelo curso.

À todos os Professores do Programa de Pós Graduação em Ciências Ambientais - PPGCA, pela atenção em todos os momentos em que seus conhecimentos foram transmitidos.

Aos mais que amigos: Klebson Rodrigues, Renata Nascimento, Anne Roberta e Yonara Leoni Fonseca. Seus conselhos e estímulos foram de extrema importância para concluir esta dissertação.

E a todos que direta ou indiretamente colaboraram com esta jornada de mestrado.

Como pode a Matemática, sendo produto do pensamento humano, independente da experiência, se adaptar tão admiravelmente aos objetos da realidade?

Albert Einstein

RESUMO

A Região Amazônica frequentemente tem a maior parte da sua superfície coberta por nuvens do tipo cúmulos, as quais, para se desenvolverem, precisam de uma forçante termodinâmica (fonte de calor e umidade) e uma forçante dinâmica (vento). O município de Belém se localiza numa das regiões mais chuvosas da Amazônia, devido à forte atividade convectiva da região. Um dos fenômenos naturais, além da precipitação, que provocam destruição e transtornos a população da região Amazônica são os raios, geralmente associados a nuvens de tempestades. O objetivo geral desse trabalho é criar um modelo estatístico baseado no método árvore de decisão, que compõe uma das técnicas de inteligência computacional, com o intuito de encontrar uma relação entre os índices termodinâmicos da atmosfera e a ocorrência de raios. Para tanto foram utilizados dados de sondagens da atmosfera feitas diariamente no Aeroporto de Belém e divulgados pela Universidade de Wyoming. Por outro lado foi feita uma coleta, análise e armazenamento de dados da ocorrência de raios em uma área circular de 100 km de raio, no entorno de Belém, extraídos do banco de dados disponibilizados pela Rede de Detecção de Raios STARNET, que cobre continuamente toda a Amazônia. Os resultados mostraram que as regras da árvore de decisão são significativas, pois mostraram que os valores críticos dos índices termodinâmicos do modelo são semelhantes aos descritos na literatura. O modelo de árvore de decisão estabelece regras que podem contribuir na previsão de tempo e raios para a região de Belém. O estudo de índices termodinâmicos com ocorrência de raios mostrou uma relação onde a avaliação do modelo obteve um resultado de 94 % de acerto, valores considerados bons para construção de um modelo de representação estatística, para ocorrência e severidade das descargas elétricas da atmosfera.

Palavras-chave: Previsão de Raios. Árvore de Decisão. Modelo Estatístico

ABSTRACT

The Amazon region often has most of its surface covered by cumulus type clouds which, to thrive, need a thermodynamic forcing (source of heat and humidity) and dynamic forcing (wind). The city of Belém-Pa is located in one of the rainiest regions of the Amazon, due to the strong convective activity in the region. One of the natural phenomena, besides precipitation, which causes destruction and losses to the population of the Amazon region is lightning, usually associated with the thunderstorms. The overall goal of this work is to create a statistical model based on the decision tree method, which consists of a computational intelligence technique, in order to find a relationship between thermodynamic indices of the atmosphere and the occurrence of lightning. We used data from radio soundings of the atmosphere made daily from Belém airport and published at an Internet site by the University of Wyoming. Additionally, lightning daily occurrences within a 100 km radius area around the sound launching site were drawn from the STARNET lightning monitoring network data base; which covers the entire Amazon. The results showed that the rules of the decision tree are significant because the critical values of thermodynamic indices of the model are similar to those described in the literature. The decision tree model establishes rules that can help in forecasting weather and lightning in the region of surrounding Belém. The study of atmospheric thermodynamic indices versus the occurrence of lightning showed a positive relationship where the evaluation of the model obtained a result of 94% correct values. This is considered good for building a statistical model representation for the occurrence and severity of electrical discharges from the atmosphere.

Keywords: Lightning Forecast. Decision Tree. Statistical model.

LISTA DE ILUSTRAÇÕES

Figura 1 – Danos causados por raios	16
Figura 2 – Ilustração dos diversos tipos de raios.....	18
Figura 3 – Localização dos sensores de VLF da rede STARNET	20
Figura 4 – Nuvens Cumulonimbos.....	21
Figura 5 – Lançamento de balões meteorológicos com sondas atmosféricas	22
Figura 6 – Médias climatológicas da precipitação anual (1978-2008).....	24
Figura 7 — Área de estudo.....	25
Figura 8 – Relação de distância com dissimilaridade e similaridade	40
Figura 9 – Representação do algoritmo K-Means	41
Figura 10 – Técnica K-Means na separação dos clusters.....	42
Figura 11 – Clusters de severidade.....	42
Figura 12 – Representação Heurística	45
Figura 13 – Busca pela Melhor Escolha- primeira expansão	46
Figura 14 – Busca pela Melhor Escolha- segunda expansão.....	46
Figura 15 – Busca pela Melhor Escolha- última expansão	47
Figura 16 – Mediana dos índices termodinâmicos selecionadas pelo método Best First	48
Figura 17 – Distribuição das Medianas Mensais de Raios.....	49
Figura 18 – Árvore de Decisão.....	51
Figura 19 – Imagens do programa R	53
Figura 20 – Árvore de decisão feita no software R	54
Figura 21 – Modelo da árvore de decisão com valores reais.....	55
Figura 22– Médias climatológicas da precipitação mensal (1978-2008).....	59
Figura 23 – Distribuição Espacial de Raios.....	60
Figura 24– Distribuição Horária de Raios no entorno de Belém	61
Figura 25– Temperatura da superfície do mar 2009-2010	62
Figura 26 – Comparação do Diagrama SKEW T LOG P de Belém do dia 17/02/2013 às 1200UTC com as regras do modelo da árvore de decisão	63
Figura 27 – Comparação da precipitação acumulada em Belém do dia 17/02/2013 com as imagens de satélite às 19:00UTC – 16:00 hora local	64
Figura 28– Distribuição temporal dos raios em Belém no dia 17/02/2013.....	64
Figura 29 – Comparação do Diagrama SKEW T LOG P de Belém do dia 12/01/2013 às 1200UTC com as regras do modelo da árvore de decisão	65

Figura 30 – Comparação da precipitação acumulada em Belém do dia 12/01/2013 com as imagens de satélite às 19:00UTC – 16:00 hora local	66
Figura 31 – Distribuição temporal dos raios em Belém no dia 12/01/2013	67
Figura 32 – Impactos socioambientais causados pela chuva e desligamento de energia	68
Figura 33 – Comparação do Diagrama SKEW T LOG P de Belém do dia 01/07/2013 às 1200UTC com as regras do modelo da árvore de decisão	69
Figura 34 – Comparação da precipitação acumulada em Belém do dia 01/07/2013 com as imagens de satélite às 19:00UTC, 16:00 hora local	70
Figura 35 – Distribuição temporal dos raios em Belém no dia 01/07/2013.	70
Figura 36 – Comparação do Diagrama SKEW T LOG P de Belém do dia 22/08/2013 às 1200UTC com as regras do modelo da árvore de decisão	71
Figura 37 – Comparação da precipitação acumulada em Belém do dia 22/08/2013 com as imagens de satélite às 19:00UTC, 16:00 hora local	72
Figura 38 – Distribuição temporal dos raios em Belém no dia 22/08/2013.	73

LISTA DE TABELAS

Tabela 1 – Sondagem realizada na estação SBBE (aeroporto de Belém)	26
Tabela 2 - Índices Termodinâmicos.	27
Tabela 3 - Valores críticos de show.....	29
Tabela 4 - Valores do índice K.....	30
Tabela 5 - Valores do índice Totl para convecção	31
Tabela 6 – Valores da CAPE.....	32
Tabela 7 - Exemplos de dados de raios	36
Tabela 8 - Exemplo de dados de agrupados	43
Tabela 9 - Exemplos de Dados Seleccionados para construção do modelo.....	47
Tabela 10 - Dados de Raios para a área de 100 km em torno de Belém	61

LISTA DE ABREVIATURAS E SIGLAS

BRDR - Rede de Detecção de Raios do Brasil

CAPE - Energia Potencial Disponível para convecção

Cb's - Cumulonimbus

CINE - Energia de Inibição da Convecção

CMORPH – CPC MORPHing Technique

CPTEC - Centro de Previsão do Tempo e Estudos Climáticos

ELAT - Grupo de Eletricidade Atmosférica

ENOS – El niño oscilação sul

hPa - hectopascal

LI - Linha da Instabilidade

NCAR – National Center for Atmospheric Research

NCEP – National Center for Environmental Prediction

NCE - Níveis de Convecção Espontânea

NE - Nível de Equilíbrio

RDR - Rede de Detecção de Raios

RPCH - Rede Estadual de Previsão Climática e Hidrometeorológica do Pará

RINDAT - Rede Integrada Nacional de Detecção de Descargas Atmosféricas

SIPAM - Sistema de Proteção da Amazônia

STARNET - Sferics Tracking and ranging Network

SUP - Superfície

TOA - Topo da Atmosfera

TSM – Temperatura da Superfície do Mar

Tv(amb) - Temperatura Virtual do Ambiente

Tv(par) - Temperatura Virtual da Parcela

VLF - Very Low Frequency

WWLLN - World Wide Lightning Location Network

UFPA - Universidade Federal do Pará

Z - Altura Geopotencial

ZCIT - Zona de Convergência Intertropical

P - Densidad

SUMÁRIO

1 INTRODUÇÃO	15
2 REVISÃO BIBLIOGRAFICA	18
2.1 Raios	18
2.1.1 Conceito Básico	18
2.1.2 Sistema de Detecção de Raios	19
2.1.3 Nuvens de Tempestades	21
2.1.4 Sistemas Meteorológicos atuantes na Região Amazônica	22
3 MATERIAIS E MÉTODOS	25
3.1 Localização da Área de Estudo	25
3.2 Base de Dados	26
3.2.1 Coleta de Dados da Sondagem	26
3.2.2 Coleta de Dados de Raios	36
3.2.3 Aspectos do Software R	37
3.3 Metodologia	38
3.3.1 Árvore de Decisão	38
3.3.2 Clusterização	39
3.3.3 Seleção de Índices Explicativos - Busca pela Melhor Escolha (Best First)	44
3.4 Modelo de Árvore de Decisão no R	50
3.5 Avaliação da Qualidade do Modelo	52
3.6 Classificação e Escolha do Modelo	52
4 RESULTADOS E DISCUSSÃO	55
5 ESTUDOS DE CASOS	59
5.1 Aplicação do Modelo da Árvore de Decisão	59
5.1.1 Caso1	63
5.1.2 Caso 2	68
6 CONCLUSÃO	74
REFERÊNCIAS	75
ANEXOS	82
ANEXO A - COMANDOS PARA CONSTRUÇÃO DA ÁRVORE DE DECISÃO NO R	83
ANEXO B - AVALIAÇÃO DA QUALIDADE DO MODELO	85
ANEXO C - CLASSIFICAÇÃO E ESCOLHA DO MODELO	90

1 INTRODUÇÃO

O Brasil é um dos países que mais sofrem com a incidência de raios, aproximadamente 100 milhões de ocorrências por ano, segundo o INPE - Instituto Nacional de Pesquisas Espaciais (ELAT, 2009). Também conhecemos a dimensão dos impactos negativos causados por esse fenômeno da natureza, tanto no que tange à perda de vidas humanas, na medida em que 200 pessoas morrem por ano no Brasil vítimas de acidentes com raios, quanto em prejuízos de ordem material (ELAT, 2009).

Devido a sua grande área e por estar inserida na região tropical, grande parte dos raios ocorrem na Amazônia. Essa Região em geral tem a maior parte da sua área total coberta de nuvens do tipo cumulos, as quais para se desenvolverem, precisam de uma forçante termodinâmica (fonte de calor e umidade) e uma forçante dinâmica (vento). Os fatores que governam o desenvolvimento de nuvens cumuliformes nos trópicos são: convergência horizontal de massa, profundidade e estrutura da camada subnuvens, instabilidade convectiva, orografia e características da superfície (RIEHL, 1979).

Os raios associados a nuvens de tempestades na Amazônia são os que causam maiores prejuízos. Logo, é preciso entender a associação desse fenômeno com os sistemas meteorológicos a ele associados, tais como a Zona de Convergência Intertropical (ZCIT), um dos fenômenos meteorológicos que mais influencia nas condições de tempo principalmente no leste da região. Além de outros, como ondas de leste, linhas de Instabilidade, ZCAS (Zona de Convergência do Atlântico Sul), El Niño, que influenciam, de forma significativa, as condições de tempo em toda a região.

Devido a esses prejuízos para a sociedade, mostrados na Figura 1, estudos sobre raios são de suma importância para a proteção de pessoas, animais, florestas, edificações, telecomunicações, aeronaves, instalações elétricas, linhas de transmissão de energia em altas voltagens. Muitas informações sobre esses estudos na região Amazônica já estão contribuindo para diminuição desses danos (ALMEIDA et al, 2010).

Figura 1 – Danos causados por raios, mortes de animais, incêndios e sistema elétrico



Fonte: Incêndios (2014).

A informação sobre as condições termodinâmicas na região Amazônica é muito importante para entender a ocorrência dos raios na região. Para analisar a atmosfera, ou seja, para definir se ela está instável ou estável é preciso entender o comportamento dos índices termodinâmicos. No intuito de compreender esse comportamento, diversas pesquisas já foram realizadas, uma delas é sobre técnicas de inteligência computacional, capazes de agrupar e classificar padrões atmosféricos relacionados à ocorrência de raios, usando índices termodinâmicos como: Energia Potencial Convectiva Disponível (CAPE), Nível de Convecção Livre (LFCT) e Água Precipitável (PWAT), que foram estudados para a região de Belém - PA (SÁ, ROCHA, ALMEIDA e SOUZA, 2012).

Outra pesquisa desenvolvida foi o estudo de sistemas meteorológicos de grande escala (ZCIT) e de escala menor (Linhas de Instabilidade) que encontrou influências entre a distribuição de raios e a precipitação na região de Belém – PA (POMPEU, 2012).

O estudo dos índices termodinâmicos com ocorrência de raios pode ter relação com a frequência do fenômeno na região Amazônica. Várias técnicas estatísticas podem contribuir para achar uma relação entre as variáveis termodinâmicas com a ocorrência de raios, tais quais: técnicas de aprendizagem automática, numa área da Inteligência Artificial (IA) relacionada com todo processo (não trivial) que envolve a transformação de grandes volumes de dados em conhecimento útil. A utilização de técnicas de aprendizagem automática à descoberta de conhecimento constitui nesse momento uma área de grande dinâmica, na qual, tem como expectativas obter resultados satisfatórios. Muitas áreas do conhecimento estão utilizando essas técnicas, como: medicina, finanças, agricultura, social, marketing e vendas, seguros, engenharia, militar, policiamento, ciências espaciais (PIATESKY-SHAPIRO, 1991).

A combinação da técnica de aprendizagem por máquina baseado em conhecimentos matemáticos com os conhecimentos do comportamento da atmosfera pode melhorar o entendimento do meio ambiente e dos fenômenos naturais, contribuindo para estudos futuros

sobre previsão do tempo.

No intuito de contribuir para estudos sobre previsão de raios, serão abordadas questões sobre:

- ✓ A existência de relação entre os índices termodinâmicos com a ocorrência de raios no entorno de Belém, usando o método da árvore de decisão.
- ✓ Os métodos estatísticos são capazes de explicar as condições atmosféricas mais propícias à ocorrência de raios de uma determinada região?

O objetivo geral desse trabalho é criar um modelo estatístico baseado no método árvore de decisão, que compõe uma das técnicas de inteligência computacional, com a finalidade de encontrar uma relação entre os índices derivados das variáveis termodinâmicas da atmosfera e a ocorrência de raios em uma região de teste que se situa em torno de Belém-Pará. E o objetivo específico é validar esse modelo estatístico.

Em sequência a essa breve Introdução, o tópico 2, mostra um referencial teórico sobre: os conceitos básicos de raios; sistemas de detecção de raios, nuvens de tempestades e sistemas meteorológicos atuantes na região Amazônica. No tópico 3 foram descritos os materiais e métodos utilizados na pesquisa, localização da área de estudo, base de dados e metodologia. No tópico 4 são descritos os resultados e discussão da pesquisa. No tópico 5 é feito um estudo de casos para dias que se encontram nos períodos de maior e menor precipitação e no tópico 6 foram apresentadas as conclusões dessa dissertação.

2 REVISÃO BIBLIOGRAFICA

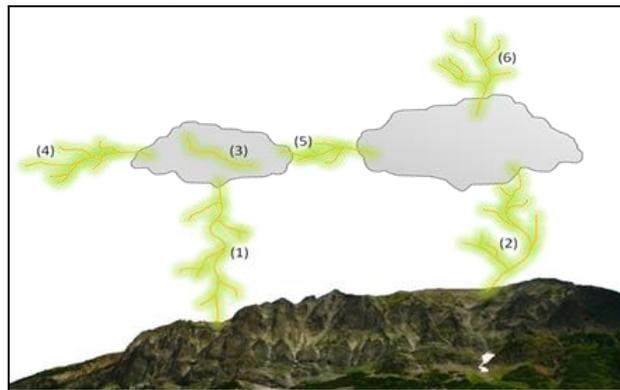
2.1 Raios

2.1.1 Conceito Básico

Os raios são intensas descargas elétricas que ocorrem na atmosfera em um curto intervalo de tempo, realizando um percurso de alguns quilômetros. Essas descargas elétricas podem ocorrer associadas às explosões nucleares, erupções vulcânicas e tornados. Contudo, geralmente, se originam em nuvens do tipo Cumulonimbus (Cb's), ou seja, nuvens de tempestades. (UMAM, 1987).

Existem diversos tipos de raios, os quais podem ser classificados em função dos locais onde ficam seus extremos, e devido ao sinal das cargas dissipadas nas nuvens. Na Figura 2 observa-se os diversos tipos de raios.

Figura 2 – Ilustração dos diversos tipos de raios.



Fonte: Silva e Coelho (2011)

Onde:

- (1) É o raio tipo nuvem-solo, ocorre quando a descarga parte da nuvem para o solo;
- (2) É o raio tipo solo-nuvem, ocorre quando a descarga parte do solo para a nuvem;
- (3) É o raio tipo intranuvem, ocorre quando o raio acontece dentro da própria nuvem;
- (4) É o raio tipo da nuvem para um ponto qualquer da atmosfera – denominados então descargas no ar;
- (5) É o raio tipo entre nuvens, ocorre quando a descarga parte de uma nuvem até outra nuvem;
- (6) É o raio tipo da nuvem para a Ionosfera, região altamente ionizada na parte externa da atmosfera terrestre.

Dentre os diversos tipos de raios, os que não atingem o solo representam cerca de 80% do total. Os raios nuvem-solo são os mais danosos. A detecção dos raios no mundo pode ser obtida por meio de satélites ou por redes de sensores localizados na superfície terrestre. (SÁ et al, 2011; MORALES; NEVES; ANSELMO,2011).

2.1.2 Sistema de Detecção de Raios

O Brasil possui cinco redes de detecção que diferem de acordo com larga cobertura e a frequência de operação dos sensores. Estas redes detectam descargas atmosféricas nuvem-solo e intra-nuvem; algumas redes distinguem a detecção. A maioria opera sob a responsabilidade de setores governamentais, e são:

1. **RINDAT** - Rede Integrada de Detecção de Descargas Atmosféricas: abrange grande parte da Área monitorada do Brasil, opera em alta e baixa frequência (VLF/VHF). É um consórcio entre companhias do setor elétrico e o INPE (Instituto Nacional de Pesquisas Espaciais).
2. **SIDDEM** - Sistema de Detecção de Descargas Atmosféricas e Eventos Meteorológicos Críticos, cobrindo o sul do Brasil.
3. **WWLLN** - World Wide Lightning Location Network: opera em baixa frequência. É um consórcio entre o INPE e a Universidade de Washington. Cobre o Brasil inteiro. Mas com baixa eficiência e precisão.
4. **RDR-SIPAM** - Rede de Detecção de Raios do Sistema de Proteção da Amazônia: cobre parte do leste da Amazônia, porém está inativa desde 2010.
5. **STARNET**- operada pelo IAG-USP, oferece dados de esféricos de baixa frequência, com ampla cobertura. A eficiência de detecção dessa rede tem melhorado desde 2007, cobrindo bem toda a Amazônia.

A STARNET - Sferics Tracking and Ranging Network, foi concebida pela Resolution Display Inc (RDI) a partir do programa de desenvolvimento inovativo de pequenas empresas da NASA. Estes receptores detectam e localizam descargas atmosféricas baseando-se na detecção dos “esféricos” (ANAGNOSTOU, et al, 2004). Os “esféricos” são ruídos de rádio emitidos pelo relâmpago em uma vasta banda do espectro eletromagnético. Na faixa do VLF, os esféricos podem se propagar a milhares de quilômetros de distância dentro do guia de onda

2.1.3 Nuvens de Tempestades

As nuvens de tempestades, Cumulonimbus (Cb's), são as responsáveis pelas alterações mais significativas do campo elétrico de tempo bom, e pela produção de raios. Devido aos gradientes térmicos no interior dessas nuvens, os processos de transporte são muito intensos (RIBEIRO, 2010). São nuvens de grande desenvolvimento vertical que produzem precipitação e descargas atmosféricas. Em situações extremas podem apresentar tornados e chuvas intensas. Em geral sua extensão vertical é de 10 a 20 km, com um diâmetro de 5 a 10 km, com duração, em média, de 30 a 90 minutos. Essas nuvens podem se apresentar tanto na forma de tempestades isoladas, ou na forma de aglomerados convectivos. Estes se apresentam, em geral, como sistemas mais severos que as células isoladas, com chuva e ventos mais intensos (PINTO JR.; PINTO, 2000).

As nuvens com topos mais altos localizam-se nos trópicos ou, também, nas latitudes médias durante o verão. Durante o processo de desenvolvimento, as velocidades verticais dentro da nuvem podem chegar a mais de 100km/h. O grande desenvolvimento das Cumulonimbus (Cb's), e conseqüentemente das tempestades locais, encontra-se associado à presença de ar quente, úmido e instável. Estas nuvens apresentam características como: descargas elétricas atmosféricas, ventos fortes, súbitas variações de temperatura e ocasionalmente, podem transformar-se em tornados (VIANELLO; ALVES, 1991).

Figura 4 – Nuvens Cumulonimbus.



Fonte: Almeida ¹

¹ Almeida [informação pessoal] arthur@ufpa.br em Agosto 2013.

A convecção profunda nos trópicos tem relação fundamental com a circulação global e a energética da atmosfera tropical, através do transporte de calor de níveis inferiores para a alta troposfera. O comportamento diurno das variáveis termodinâmicas locais varia de acordo com a quantidade de energia solar recebida e a disponibilidade de água presente na atmosfera, podendo assim contribuir para o desenvolvimento de nebulosidade (VITORINO, et al,1997).

A convecção é um dos principais meios em que calor pode ser transferido na atmosfera, além da radiação e da condução. O monitoramento das condições atmosféricas é feito entre outros através de balões meteorológicos com sondas atmosféricas como mostra a Figura 5.

Figura 5 – Lançamento de balões meteorológicos com sondas atmosféricas.



Fonte: UFERSA (2012)

2.1.4 Sistemas Meteorológicos atuantes na Região Amazônica

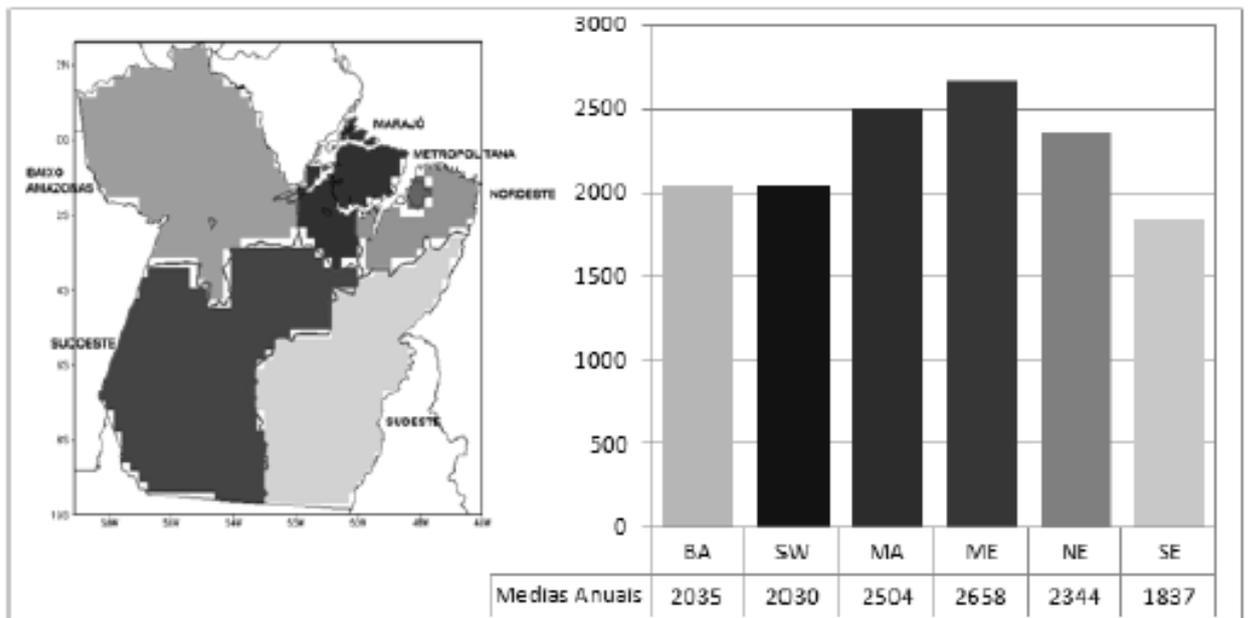
A maioria da precipitação que ocorre na região amazônica é proveniente de Sistemas Convectivos de Mesoescala (SCM). O SCM é definido como um sistema de precipitação que tem escala horizontal de 10-500 Km e inclui convecção significativa durante alguma parte de sua vida. A vida de um SCM é da ordem de 10 horas (TAVARES,2009). Os SCM podem ser classificados de acordo com suas características físicas, organização e localização de ocorrência, tais como as Linhas de Instabilidade (LI), Complexos Convectivos de Mesoescala (CCM), os Ciclones Tropicais, etc (MADDOX,1980). As linhas de Instabilidade (LI) na Amazônia são responsáveis por uma grande quantidade de precipitação na região durante o

período seco (COHEN; SILVA DIAS; NOBRE, 1995). Outro sistema de grande escala que atua sobre a região é a alta da Bolívia que se forma durante o verão do Hemisfério Sul (KOUSKY; KAYANO, 1981; MOLION, 1993). A ZCIT é caracterizada por ser um fenômeno atmosférico de grande escala e que interfere no clima planetário; com isso ela caracteriza o clima de diferentes localidades do planeta, possuindo grande importância nos índices de precipitação da região tropical (XAVIER et al., 2000; FERREIRA et al., 2005).

Nas regiões equatoriais, o gradiente horizontal da temperatura é muito pequeno, provocando a predominância de ventos fracos na superfície, levando a uma movimentação do ar basicamente na vertical, produzindo processos de convecção, conseqüentemente nuvens e chuva (FISH; VENDRAME; HANAOKA, 2007). É na Amazônia, também, que ocorrem as atividades convectivas intensas e que pode ter uma influência grande no clima de outras localidades, devido ao transporte horizontal de energia e vapor d'água (RIBEIRO et al, 2011).

A distribuição da precipitação média anual no Estado do Pará, baseado nos dados da Rede Estadual de Previsão Climática e Hidrometeorológica do Pará (RPCH), está ilustrada na Figura 6. Observa-se que os maiores índices estão concentrados no Nordeste paraense, com precipitação superior a 2000 mm, englobando as mesorregiões do Marajó, Metropolitana de Belém e Nordeste Paraense. Destaca-se que a mesorregião Metropolitana de Belém é a que apresenta o maior índice pluviométrico do estado do Pará. O fato do alto índice pluviométrico está concentrado na região nordeste se deve a influência de sistemas de grande-escala como ZCIT, forte convecção local, aglomerados de Cumulonimbus e por estar localizada às proximidades de áreas litorâneas. Por outro lado, os menores valores de precipitação, ficam reservados as mesorregiões Sudeste do estado, com valores inferiores a 1900 mm. As mesorregiões do Baixo Amazonas e Sudeste possuem precipitações anuais bem próximas, com valores de 2.035 e 2.030 mm, respectivamente (ALBUQUERQUE et al, 2010).

Figura 6 – Médias climatológicas da precipitação anual (1978-2008) representativas das mesorregiões do Pará. Unidade: mm.



Fonte: Albuquerque et al (2010)

Na Figura 6, as mesoregiões do Pará são representadas com siglas: BA: Baixo Amazonas, SW: Sudoeste, MA: Marajó, ME: Metropolitana de Belém, NE: Nordeste, SE: Sudeste.

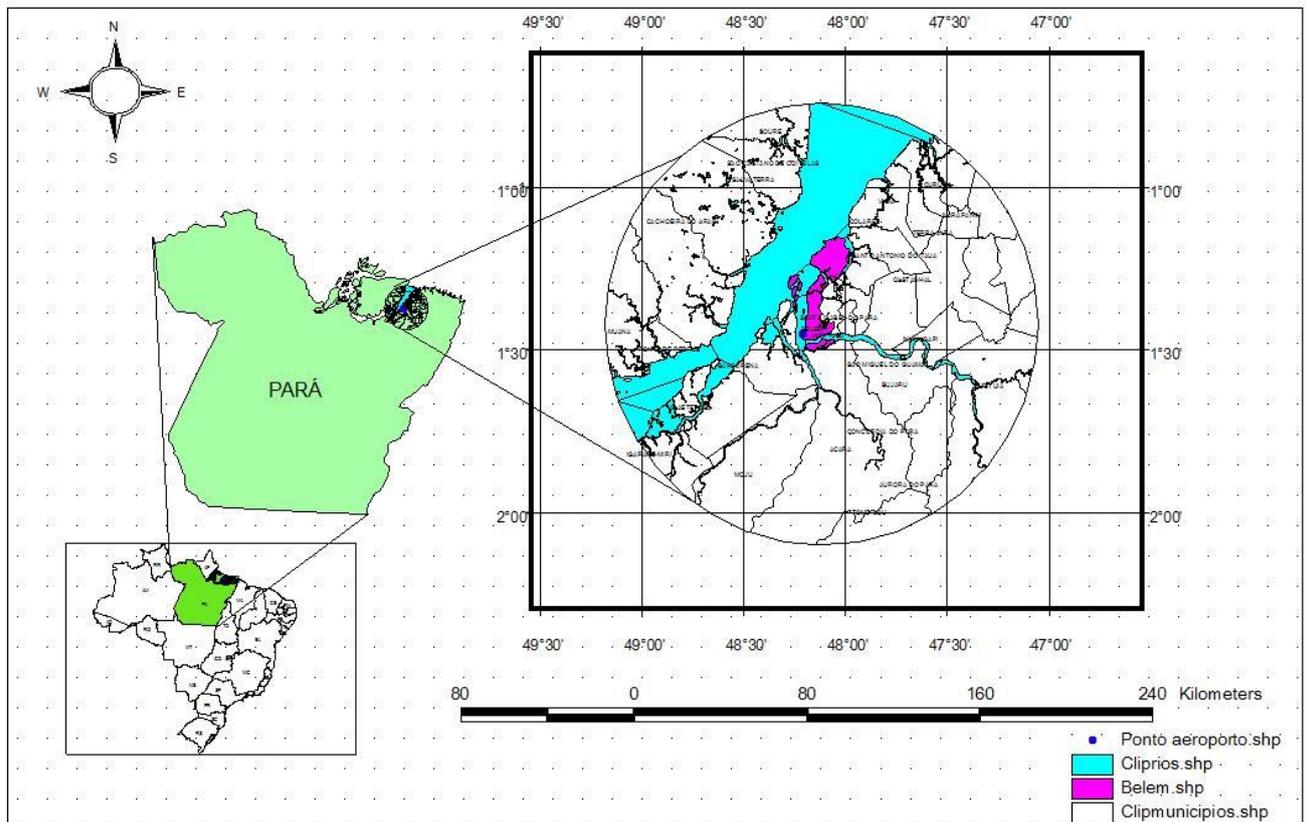
3 MATERIAIS E MÉTODOS

3.1 Localização da Área de Estudo.

A área de estudo é uma área circular com centro no Aeroporto Internacional de Belém, cujas coordenadas geográficas são latitude -1.4558 e longitude -48.5044 com raio de 100 km, totalizando uma área de 31,4 mil km^2 , conforme mostrado na Figura 7.

Belém é a capital do Estado do Pará e fica localizada geograficamente a $1^{\circ} 27' 21''$ de latitude Sul e $48^{\circ} 30' 15''$ de longitude Oeste, em média quatro metros acima do nível do mar, às margens da Baía do Guajará e do Rio Guamá, distante 120 quilômetros do mar no estuário do Rio Pará (Figura 7).

Figura 7 – Área de estudo.



Fonte: STARNET – (processados no Arc View 3.2)

A precipitação pluviométrica de Belém varia de 2000 a 3000 mm por ano, os meses de dezembro a maio são registrados os maiores volumes de chuva, associados aos padrões de circulação atmosférica quase estacionários de grande escala, a Zona de Convergência

Intertropical (ZCIT) e a Zona de Convergência do Atlântico Sul (ZCAS) (SOUZA,2003; POMPEU,2012).

3.2 Base de Dados

3.2.1 Coleta de Dados da Sondagem

Foi realizada a coleta, análise e armazenamento dos dados de sondagem feita diariamente no Aeroporto de Belém e divulgados pela Universidade de Wyoming, disponível no site <http://weather.uwyo.edu/upperair/sounding.html>. Na Tabela 1, observa-se como os dados são obtidos, de acordo com o glossário meteorológico da AMS (American Meteorological Society – Sociedade Meteorológica Americana).

Tabela 1– Variáveis termodinâmicas na sondagem realizada na estação SBBE (aeroporto de Belém).

82193 SBBE Belem (Aeroporto) Observations at 12Z 08 May 2012										
PRES	HGHT	TEMP	DWPT	RELH	MIXR	DRCT	SKNT	THTA	THTE	THTV
hPa	m	C	C	%	g/kg	deg	knot	K	K	K
1012.0	16	29.6	25.8	80	21.21	90	8	301.7	364.7	305.5
1000.0	121	27.8	24.4	82	19.69	65	12	300.9	359.2	304.5
998.0	139	27.6	24.3	82	19.60	65	12	300.9	358.9	304.4
979.0	310	25.8	23.2	86	18.68	65	18	300.8	355.9	304.1
969.0	400	26.0	22.2	80	17.80	65	21	301.8	354.6	305.1
958.0	502	26.2	21.2	74	16.85	67	19	303.0	353.2	306.1
941.0	661	29.2	9.2	29	7.82	71	16	307.6	331.8	309.1
929.0	775	27.6	19.6	62	15.71	74	14	307.1	354.8	310.0
925.0	813	27.0	20.0	66	16.19	75	13	306.9	355.9	309.9
907.0	987	24.4	20.5	79	17.05	80	10	306.0	357.4	309.1
874.0	1310	22.1	18.3	79	15.36	125	12	306.9	353.3	309.7
850.0	1553	20.4	16.6	79	14.18	115	13	307.5	350.6	310.1
824.0	1821	18.2	15.8	86	13.89	112	16	307.9	350.2	310.5
755.0	2567	13.2	10.7	85	10.81	104	25	310.3	343.6	312.3
719.0	2978	12.2	7.2	72	8.94	100	30	313.6	341.6	315.2
705.0	3143	11.8	5.8	67	8.27	96	29	314.9	341.1	316.4
700.0	3203	11.4	5.4	67	8.10	95	28	315.1	340.8	316.6
671.0	3555	9.2	4.1	71	7.71	100	25	316.4	341.0	317.9

Fonte: Universidade de Wyoming (2012).

Onde:

PRES (hPa) - Atmospheric Pressure (Pressão atmosférica);

HGTH (M) – Geopotential Height (Altura Geopotencial);

TEMP (°C) – Temperature – Temperatura;

DWPT (°C) – Dewpoint Temperature (Temperatura do Ponto de Orvalho);

RELH (%) – Relative humidity (Umidade Relativa);

MIXR (g/kg) – Mixing Ratio (Razão de Mistura);

DRCT (deg) – Wind Direction (Direção do Vento);

SKNT (knot) – Wind Speed (Velocidade do Vento);

THTA (°K) – Potential Temperature (Temperatura Potencial);

THTE (°K) – Equivalent Potential Temperature (Temperatura Potencial Equivalente);

THTV (°K) – Virtual Potential Temperature (Temperatura Potencial Virtual).

Os índices nessa tabela são calculados a partir das medidas das variáveis termodinâmicas primárias: Pressão atmosférica, Altura Geopotencial, Temperatura, Temperatura do Ponto de Orvalho, Umidade Relativa, Razão de Mistura, Direção do Vento, Velocidade do Vento, Temperatura Potencial, Temperatura Potencial Equivalente, Temperatura Potencial Virtual, proporcionados pelas radiossondagem nos diversos níveis alcançados pelo balão atmosférico.

Os objetivos dos mesmos são diversos e incluem: quantificar o estado de convecção, o conteúdo de água líquida e vapor de água, a convergência de umidade pelos ventos, o conteúdo de energia térmica por unidade de massa, e os gradientes das variáveis do ar entre dois ou mais níveis de ascensão.

Devido aos deslocamentos horizontais da sonda ao longo da atmosfera, geralmente se supõe que a representatividade da sondagem se estenda a um cilindro vertical de 100 km de raio de seção transversal da atmosfera centrada em torno do ponto de lançamento da sondagem na atmosfera.

Na Tabela 2, observam-se os exemplos de índices termodinâmicos a partir da radiossondagem acima disponíveis no endereço de Internet da Universidade de Wyoming, relativos ao aeroporto da cidade de Belém, Pará, Brasil, SBBE (indicador de localidade) e coordenadas geográficas latitude -1.4558 e longitude -48.5044.

Tabela 2 - Índices Termodinâmicos no dia 08 de maio de 2012.

STATION INFORMATION AND SOUNDING INDICES	AND
Station identifier:	SBBE

Station number:	82193
Observation time:	120508/1200
Station latitude:	-1.38
Station longitude:	-48.48
Station elevation:	16.0
Showalter index:	0.77
Lifted index:	-1.42
LIFT computed using virtual temperature:	-1.87
SWEAT index:	255.18
K index:	33.10
Cross totals index:	18.70
Vertical totals index:	22.50
Totals totals index:	41.20
Convective Available Potential Energy:	825.73
CAPE using virtual temperature:	1001.05
Convective Inhibition:	-111.62
CINS using virtual temperature:	-89.47
Equilibrium Level:	139.86
Equilibrium Level using virtual temperature:	139.80
Level of Free Convection:	822.00
LFCT using virtual temperature:	832.58
Bulk Richardson Number:	88.43
Bulk Richardson Number using CAPV	107.21
Temp [K] of the Lifted Condensation Level:	295.69
Pres [hPa] of the Lifted Condensation Level:	935.15
Mean mixed layer potential temperature:	301.42
Mean mixed layer mixing ratio:	18.84
1000 hPa to 500 hPa thickness:	5829.00
Precipitable water [mm] for entire sounding:	59.81

Onde:

STATION IDENTIFIER (Identificador da Estação); **STATION NUMBER** (Número da Estação); **OBSERVATION TIME** (Tempo de Observação); **STATION LATITUDE – SLAT** (Latitude da estação em graus); **STATION LONGITUDE – SLON** (Longitude da estação em graus, onde a longitude oeste é negativa); **STATION ELEVATION – SELV** (Elevação da estação em metros); **SHOWALTER INDEX - SHOW** (Índice Showalter), é obtido a partir da equação (1):

$$SHOW = T_{500} - T_{parcela}$$

(1)

Onde:

T_{500} = Temperatura do ar ambiente, em graus Celsius, no nível de pressão atmosférica de 500 (500 hPa). $T_{parcela}$ = Temperatura, em graus Celsius, de uma parcela de ar que sobe adiabaticamente do nível de pressão de 850 hPa até 500 hPa (considerando saturação).

Tabela 3 - Valores críticos de Show

ÍNDICES	INTERPRETAÇÃO
>4	Tempo estável
+3 a +1	Possibilidade de tempestade
+1 a -2	Grande possibilidade de tempestade
-2 a -3	Muita instabilidade/ tempestade com trovoada
< -5	Extremamente instável

Fonte: Adaptado, Loureiro (2005)

LIFTED INDEX –LIFT (Índice de Instabilidade por Levantamento), é definido pela equação 2:

$$\mathbf{LIFT = T_{500} - T_{parcela}} \quad (2)$$

Onde:

T_{500} = Temperatura do ar, em graus Celsius ($^{\circ}\text{C}$), no nível em que a atmosfera está com 500 de pressão (500 hPa).

$T_{parcela}$ = Temperatura da parcela, em graus $^{\circ}\text{C}$, de uma parcela de ar em 500 hPa após sua ascensão desde a superfície.

LIFT COMPUTED USING VIRTUAL TEMPERATURE: É calculado mediante a utilização de temperatura virtual;

SWEAT INDEX - Severe Weather Index ou Index Sweat (Índice de Ameaça severa do tempo): É um indicador utilizado para analisar o potencial de tempestades severas, calculado a partir da equação (3):

$$\mathbf{SWET = 12.(TD850) + 20.(TERM2) + 2.(SKT850) + SKT.500 + SHEAR} \quad (3)$$

Onde:

TD850 = Temperatura do ponto de orvalho, em graus $^{\circ}\text{C}$, no nível de pressão atmosférica a 850 hPa;

TERM2 = Máx (TOTL – 49), onde TOTL é o índice total- totals;

SKT850 = Velocidade do vento, em nós, no nível em que a atmosfera está com 850 hPa de pressão;

SKT500 = Velocidade do vento, em nós, no nível em que a atmosfera está com 500 hPa de pressão;

SHEAR = $125 \cdot [\sin(\text{DIR}500 - \text{DIR}850) + 0,2]$, onde DIR500 = a direção do vento no nível pressão atmosférica a 500 hPa e DIR850 = direção do vento no nível de pressão atmosférica a 850 hPa.

Não é permitido que um dos termos da equação seja negativo. Sendo que o termo SHEAR deve ser considerado como nulo se quaisquer das considerações a seguir não forem satisfeitas, como:

- ✓ A direção do vento no nível de 850 hPa se encontra no intervalo de 130 a 250°;
- ✓ A direção do vento no nível de 500 hPa se encontra no intervalo de 210 a 310°;
- ✓ A diferença entre a direção dos ventos é positiva;
- ✓ As velocidades dos ventos dos 500 e/ou 850 hPa são maiores do que 15 nós (~30Km/h = 8,3 m/s).

K INDEX –KINX (INDICE K), obtido em graus celsius, é uma medida do potencial de instabilidade baseado na variação vertical de temperatura. Este índice é muito usado para avaliar chuvas fortes, pois a presença de camadas úmidas em 850 hPa e 700 hPa implica em bastante água precipitável, é calculado pela equação (4):

$$\mathbf{KINX = (T850 - T500) + TD850 - (T700 - TD700)} \quad (4)$$

Onde:

T850, T500 e T700 = São as temperaturas do ar, em graus °C, no nível de pressão atmosférica de 850 hPa, 500 hPa e 700 hpa, respectivamente;

TD700 e TD850 = São as temperaturas do ponto de orvalho, em graus °C, no nível de pressão atmosférica de 700 hPa e 800 hpa, respectivamente.

Os valores do índice K variam com a estação do ano e localização. Quanto mais positivo for o índice, maior será a probabilidade de tempestades, sendo que o valor crítico para esse índice é de aproximadamente 30°C e, quanto maior o valor de K, maior a instabilidade atmosférica (DIAS, 2008), observados na tabela 4.

Tabela 4- Valores do índice K

ÍNDICES	INTERPRETAÇÃO
K < 20	Nenhuma

$20 < K < 25$	Tormentas isoladas
$25 < K < 30$	Tormentas amplamente dispersas
$30 < K < 35$	Tormentas dispersas
$35 < K < 40$	Muitas tormentas

Fonte: Adaptado de DIAS (2008)

CROSS TOTALS INDEX (CTOT): Significa total transversal e expressa a combinação do teor de umidade nos baixos níveis com as temperaturas superiores. É calculado pela equação (6):

$$CTOT = TD850 - T500 \quad (6)$$

VERTICAL TOTALS INDEX (VTOT): É o total vertical e expressa a taxa vertical de variação de temperatura entre dois níveis de pressão atmosférica (850 hPa e 500 hPa), é calculado pela equação (7):

$$VTOT = T850 - T500 \quad (7)$$

TOTAL TOTALS INDEX –TOTL (Índice Total) – é dado pela soma dos índices Vertical Totals e Cross Totals, é calculado pela equação (8):

$$TOTL = (T850 - T500) + (TD850 - T500) \quad (8)$$

Tabela 5 - Valores do índice Totl para convecção

Valores do Índice TOTL (C)	Interpretação
$TOTL < 44$	Potencial fraco
$44 < TOTL < 50$	Tormentas moderadas
$50 < TOTL < 55$	Tormentas fortes
$TOTL > 55$	Tormentas muito fortes

Fonte: Adaptado de DIAS (2008)

CONVECTIVE AVAILABLE POTENTIAL ENERGY (CAPE): Energia potencial convectiva disponível (índice cape), também conhecida como energia de fluabilidade, mede a energia potencial disponível para convecção, é definido pela equação (9):

$$CAPE = \int_{Z-LFCV}^{Z-EQTV} g \cdot \frac{(TP - TE)}{TE} dZ$$

(9)

Onde:

EQTV = Nível de equilíbrio; **LFCT** = Nível de convecção livre; **g** = Aceleração da gravidade terrestre ($9,8 \text{ m/s}^2$); **TP** = Temperatura de uma parcela de ar nos primeiros 500 m de altitude, elevada adiabaticamente seca até o nível de condensação por levantamento (localizado na base da nuvem), e adiabaticamente úmida depois; **TE** = Temperatura do ar.

O quadro mostra os valores de referência para CAPE, indicando o potencial para o desenvolvimento de tempestades.

Tabela 6 – Valores da CAPE

VALORES DE CAPE (J/kg)	INTERPRETAÇÃO
0 a 1000	Pouco instável
1000 a 2500	Moderadamente instável
2500 a 4000	Acentuadamente instável
>4000	Extremamente instável

Fonte: Adaptado de DIAS (2008).

CAPE USING VIRTUAL TEMPERATURE (CAPV): ÍNDICE CAPE, calculado usando a temperatura virtual, é calculado pela equação (10):

$$CAPV = \int_{Z-LFCV}^{Z-EQTV} g \cdot \frac{(TVP - TVE)}{TVE} dZ$$

(10)

Onde:

TVP = Temperatura virtual de uma parcela de ar nos primeiros 500 m de altitude, elevada adiabaticamente seca até o nível de condensação por levantamento (localizado na base da nuvem), e adiabaticamente úmida depois; **TVE** = Temperatura virtual do ar, é calculada pela equação (11):

CONVECTIVE INHIBITION (CINS): Inibição convectiva

$$CINS = g \cdot \int_{z=0}^{z=LFCT} \frac{(TP - TE)}{TE} dz$$

(11)

CINS USING VIRTUAL TEMPERATURE (CINV): CINS. Pode ser calculado mediante a utilização de temperatura virtual, é calculada pela equação (12):

$$CINS = g \cdot \int_{z=0}^{z=LFCT} \frac{(TVP - TVE)}{TVE} dz$$

(12)

EQUILIBRUM LEVEL – EQLV (Nível de Equilíbrio), obtido em hPa, é definido como o nível em que uma parcela de ar nos primeiros 500m de altitude é elevada adiabaticamente seca até o nível de condensação por levantamento (localizado na base da nuvem), e adiabaticamente úmida para um nível acima do qual a temperatura da parcela é a mesma que o ambiente.

EQUILIBRUM LEVEL USING VIRTUAL TEMPERATURE (EQTV): INDICE EQLV, calculado mediante a utilização de temperatura virtual.

LEVEL OF FREE CONVECTION (LFCT): (Nível de Convecção), obtido em hPa, é encontrado através da comparação de temperatura entre uma parcela de ar e o meio ambiente. É definido como o nível em que uma parcela de ar nos primeiros 500m de altitude, elevada adiabaticamente seca até o nível de condensação por levantamento (localizado na base da nuvem), e adiabaticamente úmida para o nível acima do qual a parcela é positivamente flutuante.

LFCT USING VIRTUAL TEMPERATURE (LFCV): Índice calculado mediante a utilização de temperatura virtual.

BULK RICHARDSON NUMBER (BRCH): é um número adimensional, ou seja, sem unidades físicas, relativo à estabilidade vertical e cisalhamento vertical do vento. Sendo que, o cisalhamento vertical do vento é a diferença vertical da velocidade do vento em dois pontos do espaço dividida pela distância vertical entre eles. O seu valor representa a proporção de turbulência térmica produzida e a turbulência gerada pelo cisalhamento vertical, determinando se a convecção é livre ou forçada. Valores elevados indicam instabilidade e/ou ambientes fracamente cisalhados; valores baixos indicam instabilidade fraca e/ou cisalhamentos verticais

do vento fortes. Geralmente, os valores no intervalo entre 10 e 45 sugerem condições ambientais favoráveis para desenvolvimento de enormes nuvens de tempestades (supercélulas). Pode ser calculado pela equação (13):

$$BRCH = \frac{CAPE}{(0,5 \cdot U^2)}$$

(13)

Onde:

CAPE = Energia Potencial Convectiva Disponível; **U** = Magnitude do cisalhamento ($u_2 - u_1$, $v_2 - v_1$); u_1, v_1 = média do vento u , v nos primeiros 500m de altitude; u_2, v_2 = média do vento u , v nos primeiros 6000m de altitude.

BULK RICHARDSON NUMBER USING CAPV (BRCV): Número de bulk Richardson , é calculado usando o CAPV, é calculada pela equação (14):

$$BRCH = \frac{CAPV}{(0,5 \cdot U^2)}$$

(14)

Onde: CAPV = Energia Potencial Convectiva Disponível, usando a temperatura virtual.

TEMP [K] OF THE LIFTED CONDENSATION LEVEL (LCLT): Temperatura (em kelvin) do nível de condensação por levantamento de uma parcela de ar elevada para os primeiros 500 metros de altitude. O nível de condensação por levantamento (Lifted Condensation Level-LCL) é formalmente definida como a altura em que umidade relativa de uma parcela de ar chega a 100% quando é resfriada pela elevação adiabática seca. A umidade relativa do ar aumenta quando ela é resfriada, já que a quantidade de vapor d'água no ar permanece constante, enquanto a saturação do vapor diminui quase exponencialmente com a diminuição da temperatura. Se a parcela de ar é elevada além do LCL, o vapor d'água na parcela de ar se condensará, formando gotículas de nuvem. O LCL é caracterizado como a altura da base da nuvem. Pode ser calculado pela equação (15):

$$LCLT = \left[\frac{1}{\frac{1}{DWPk - 56} + \frac{\ln\left(\frac{TMPk}{DWPk}\right)}{800}} \right] + 56$$

(15)

Onde:

DWPK = Temperatura do ponto de orvalho em Kelvin; **TMPK** = Temperatura da parcela de ar nos primeiros 500m de altitude, em Kelvin e **Ln** = logarítmico natural

PRESS [HPA] OF THE LIFTED CONDENSATION LEVEL (LCLP): Pressão do nível de condensação por levantamento de uma parcela de ar elevada para os primeiros 500 metros de altitude, é calculada pela equação (16):

$$LCLP = PRES. \left(\frac{LCLT}{TMPC + 273,15} \right)^{\frac{1}{KAPPA}}$$

(16)

Onde:

PRES = Pressão Atmosférica; **LCLT** = Temperatura, em Kelvin, do nível de condensação do vapor elevado em uma média nos primeiros 500 metros de altitude; **TMPC** = temperatura da parcela de ar nos primeiros 500 metros de altitude, em graus Celsius e **KAPPA** = Constante.

MEAN MIXED LAYER POTENTIAL TEMPERATURE (MLTH): Temperatura Potencial média da camada de mistura em kelvin. Desta forma, MLTH é a média da temperatura potencial dos primeiros 500 m de altitude. Sendo que a camada de mistura é uma camada que possui uma forte mistura vertical impulsionada pelo fluxo vertical de calor positivo, apresentando uma grande variação durante o dia atingindo sua altura máxima por volta do período da tarde.

MEAN MIXED LAYER MIXING RATIO (MLMR): Razão de mistura média da camada de mistura (em g/kg). é considerada como sendo a média razão de mistura dos primeiros 500 m de altitude.

1000 hPa TO 500 hPa THICKNESS (HTK): 1000 hPa a 500 hPa de espessura (em metro) é calculada pela equação (17):

$$HTK = (Z500 - Z1000)$$

(17)

Onde: Z500 = (nível atmosférico de pressão) Nível da superfície de 500 hPa; Z1000 = Nível da superfície de 1000 hPa.

PRECIPITABLE WATER [MM] FOR ENTIRE SOUNDING (PWAT): Água precipitável (em mm) para toda a sondagem: É definida como o conteúdo de vapor d'água existente na atmosfera em uma coluna de seção horizontal unitária e na vertical da superfície até o topo. Representa a quantidade de água que potencialmente pode precipitar caso todo

vapor se condense. Ressaltando que a água precipitável é considerada uma variável atmosférica e não um índice termodinâmico.

Essas definições dos índices termodinâmicos procuram determinar características da atmosfera, a partir de variáveis termodinâmicas, como descrito anteriormente: Pressão atmosférica, Altura Geopotencial, Temperatura, Temperatura do Ponto de Orvalho, Umidade Relativa, Razão de Mistura, Direção do Vento, Velocidade do Vento, Temperatura Potencial, Temperatura Potencial Equivalente, Temperatura Potencial Virtual.

3.2.2 Coleta de Dados de Raios

Por outro lado foi feita uma seleção, análise e armazenamento de dados da ocorrência de raios em uma área de 100 km no entorno de Belém, divulgados pela Rede de Detecção de Raios STARNET, uma rede de detecção de longa distancia, disponível no site, www.zeus.iag.usp.br. Os dados fornecidos são os de tempo de ocorrência e a localização geográfica do evento, esses dados são padronizados em formato ASCII, no qual os campos de dados estão separados por um espaço e cada um dos arquivos fornece dados com intervalos de tempo de 15 minutos.

Os dados obtidos da Rede STARNET , foram tratados utilizando o programa LEZEUS, elaborado e criado pela equipe de raios da UFPA que transforma os dados no formato “.DAT” provenientes da rede STARNET em formato “.TXT” e que funciona dentro do ambiente do software Matlab. Com os dados já no formato “.TXT” foi possível manipulá-los e trabalhá-los no software ArcView 3.2, como descrito na Tabela 7.

Tabela 7 - Exemplo de dados de incidência de 8 raios

ANO	MÊS	DIA	HORA	MINUTO	SEGUNDO	LATITUDE	LONGITUDE
2009	1	1	14	3	4	-1.0169	-49.241
2009	1	1	14	3	4	-1.0099	-49.229
2009	1	1	14	3	4	-1.0216	-49.218
2009	1	1	14	6	13	-1.0628	-49.211
2009	1	1	14	7	37	-1.0929	-49.272
2009	1	1	14	7	37	-1.0979	-49.279
2009	1	1	14	9	55	-1.0748	-49.269
2009	1	1	14	9	55	-1.0521	-49.271

Esses dados foram armazenados de forma digital, em meio magnético e ótico, e processados com programas desenvolvidos no software R, usando o método conhecido como árvore de decisão para se buscar um modelo que relacione alguns índices termodinâmicos com a ocorrência de raios na região de estudo. O período de estudo foi de 2 anos (2009 - 2010). Para entender um pouco da técnica árvore de decisão que foi utilizada e do programa onde ela foi desenvolvida, alguns conceitos são fundamentais os quais serão descritos na metodologia.

3.2.3 Aspectos do Software R

Um dos pontos fortes do R é a facilidade com que gráficos bem delineados e de alta qualidade para impressão podem ser produzidos com possibilidade de inclusão de fórmulas e símbolos matemáticos quando necessário, R é um conjunto integrado de facilidades de software para manipulação de dados, cálculos e visualização gráfica. Inclui também: Um conjunto de operadores para cálculos sobre quadros de dados, em particular as matrizes; uma grande e coerente coleção integrada de ferramentas intermediárias para análise de dado e é um software livre, distribuído sob a licença pública geral (http://.fsf.org/pt_home.html) e pode ser livremente copiado e distribuído entre usuários, bem como pode ser instalado em diversos computadores.

Com o R é possível manipular e analisar dados, visualizar gráficos e escrever desde pequenas linhas de comando até programas inteiros. O R é a versão em código aberto de uma linguagem de programação inventada nos anos 1980 no Bell Labs chamada de S. Essa linguagem tornou-se bastante popular, e vários produtos comerciais que a usam estão disponíveis, como o S-PLUS, SPSS, STATA e SAS. Um aspecto digno de nota é que a linguagem R, ao contrário de outras linguagens como Fortran e C, é uma linguagem interpretada, o que a faz ser mais fácil de programar, pois processa linhas de comando e as transforma em linguagem de máquina (código binário que o computador efetivamente lê), mas isso diminui a velocidade de processamento.

Nas linhas de comandos do R haverá um sinal de `>`, que indica o “prompt”, representado que o R está pronto para receber comandos. Se uma linha de comando não está completa, aparecerá um sinal de `+`, indicando que você poderá continuar a digitar aquela linha. Para que o “prompt” apareça novamente, pressiona-se Esc. Para que os comandos sejam executados, pressiona-se Enter. Para criar objetos, podemos utilizar os símbolos `->` ou

=. Estes símbolos representam que queremos “guardar” a informação dentro do objeto. Existem cinco classes de objetos na linguagem R: vetor, matriz, data frame, funções e lista, O R é um ambiente de programação e existem atualmente mais de 3000 pacotes que desempenham funções específicas e que precisam ser instalados e carregados independentemente. Os pacotes “stats” e “base” já vêm instalados e carregados, são estes pacotes que possuem as funções para os cálculos de modelos lineares simples, como teste t, ANOVA, χ^2 , glm e outros. A função que instala pacotes no R é a `install.packages`.

3.3 Metodologia

3.3.1 Árvore de Decisão

Árvore de Decisão é um método de classificação de dados no contexto da chamada Mineração de Dados (Data Mining). Pode ser usada em conjunto com a tecnologia de indução de regras, mas é a única a apresentar os resultados hierarquicamente (com priorização). Nelas, o atributo mais importante é apresentado na árvore como o primeiro nó, e os atributos menos relevantes são mostrados nos nós subsequentes. A vantagem principal da Árvore de Decisão é a tomada de decisões levando em consideração os atributos mais relevantes, além de compreensíveis para a maioria das pessoas. Ao escolher e apresentar os atributos em ordem de importância, as Árvores de Decisão permitem aos usuários conhecer quais fatores mais influenciam seus trabalhos.

A Árvore de Decisão é composta por nós interconectados por ramos (representados por linhas). Um nó quadrado é chamado de nó decisão porque representa uma decisão. Ramos emergindo do nó de decisão representam as diferentes alternativas para uma decisão particular (RAGSDALE, 2001). Para Garcia (2000), a Árvore de Decisão consiste de: nodos (nós), que representam os atributos, e de arcos (ramos), provenientes desses nodos e que recebem os valores possíveis para esses atributos (cada ramo descendente corresponde a um possível valor desse atributo). Nas arvores existem nodos folhas (folhas da árvore), que representam as diferentes classes de um conjunto de treinamento, ou seja, cada ramo está associado a uma classe. Cada percurso na árvore (da raiz à folha) corresponde a uma regra de classificação.

As Árvores de Decisão são representações simples do conhecimento e um meio eficiente de construir classificadores que predizem classes baseadas nos valores de atributos de um conjunto de dados (GARCIA, 2000). A construção do modelo de classificação com árvore de decisão utiliza o Software R, onde se usa uma linguagem e ambiente para

computação estatística e gráfica. É um projeto GNU similar à linguagem e ambiente S desenvolvida no *Bell Laboratories* por John Chambers e colaboradores. O software disponibiliza uma grande variedade de métodos estatísticos (modelagem linear e não-linear, testes estatísticos clássicos, séries temporais, classificação, métodos multivariados, árvore de decisão, etc.) e técnicas gráficas.

3.3.2 Clusterização

No tratamento dos dados de raios foi utilizado o método de análise de agrupamento, ou clustering, nome dado para o grupo de técnicas computacionais cujo propósito consiste em separar objetos em grupos, gerando partição da base de dados que contenha grupos (clusters), baseando-se nas características que estes objetos possuem. A idéia básica consiste em colocar em um mesmo grupo objetos que sejam similares de acordo com algum critério pré-determinado. Os grupos determinados por uma métrica de qualidade devem apresentar alta homogeneidade interna e alta separação (heterogeneidade externa). Isto quer dizer que os elementos de um determinado conjunto devem ser mutuamente similares e preferencialmente, muito diferentes dos elementos de outros conjuntos (LINDEN, 2009).

A análise de agrupamento é uma ferramenta útil para a análise de dados em muitas situações diferentes. Esta técnica pode ser usada para reduzir a dimensão de um conjunto de dados, reduzindo uma ampla gama de objetos à informação do centro do seu conjunto. Tendo em vista que clustering é uma técnica de aprendizado não supervisionado (quando o aprendizado é supervisionado, o processo é denominado de classificação), pode servir também para extrair características escondidas dos dados e desenvolver as hipóteses a respeito de sua natureza (LINDEN, 2009).

Por ser uma aplicação não supervisionada, o agrupamento pode não possuir qualquer informação prévia da base de dados da qual deve gerar partições, possivelmente nem mesmo o número de agrupamentos existente na base. Para realizar a partição de uma base de dados, um algoritmo de agrupamentos conta com heurísticas que auxiliam o particionamento dos objetos. Essas heurísticas geralmente têm relação com o conceito de similaridade entre os objetos (FONTANA e NALDI, 2009).

Para realizar a tarefa de particionamento é necessário utilizar algum critério de medida de similaridade entre objetos, ou seja, um critério que avalie o quanto dois objetos são parecidos e se devem, ou não, ser colocados em um mesmo grupo. Uma abordagem comum

para comparação de dois objetos é a associação de função de distância (calculada através dos valores dos atributos de dois objetos) ao conceito de dissimilaridade. Nesse caso, dois objetos possuem alta dissimilaridade (baixa similaridade) se a distância entre eles for alta e possuem baixa dissimilaridade (alta similaridade) se a distância entre eles for baixa. Essa relação está ilustrada na figura 8 (FONTANA e NALDI, 2009).

Figura 8 – Relação de distância com dissimilaridade e similaridade



Fonte: Fontana e Naldi, (2009).

Usando o conceito de similaridade e dissimilaridade, o algoritmo de clusterização K-means é uma técnica que busca minimizar a distância dos elementos a um conjunto de k centros dado por $x = \{x_1, x_2, \dots, x_k\}$ de forma iterativa. A distância entre um ponto p_i e um conjunto de clusters, dada por $d(p_i, x)$, é definida como sendo a distância do ponto ao centro mais próximo dele. A função a ser minimizada então, é dada pela equação (18):

$$d(P, \chi) = \frac{1}{n} \sum_{i=1}^n d(p_i, \chi)^2$$

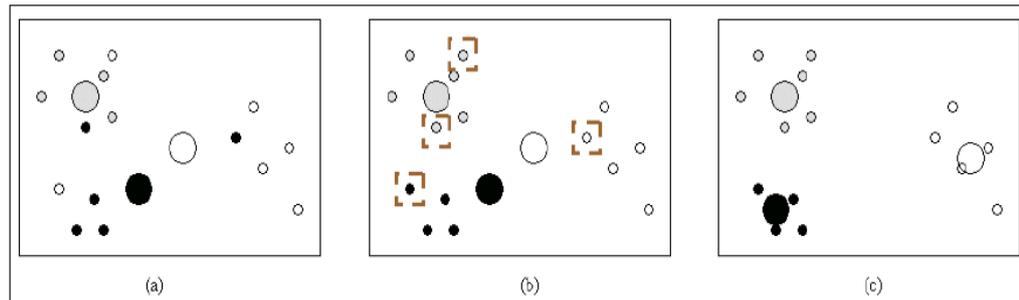
(18)

O algoritmo depende de um parâmetro ($k =$ número de clusters) definido pelo usuário. Isto costuma ser um problema, tendo em vista que normalmente não se sabe quantos clusters existem *a priori* (LINDEN, 2009). O algoritmo do K-Means pode ser descrito da seguinte maneira:

- I. Escolher k distintos valores para centros dos grupos (possivelmente, de forma aleatória)
- II. Associar cada ponto ao centro mais próximo
- III. Recalcular o centro de cada grupo
- IV. Repetir os passos II-III até nenhum elemento mudar de grupo

Este algoritmo é extremamente veloz, geralmente convergindo em poucas iterações para uma configuração estável, na qual nenhum elemento está designado para um cluster cujo centro não lhe seja o mais próximo. Um exemplo da execução do algoritmo de K-Means pode ser visto na Figura 9 (LINDEN, 2009).

Figura 9 – Representação do algoritmo K-Means.



Fonte: Linden, (op.cit.)

Na Figura 9, (a) Cada elemento foi designado para um dos três grupos aleatoriamente e os centróides (círculos maiores) de cada grupo foram calculados. (b) Os elementos foram designados agora para os grupos cujos centróides lhe estão mais próximos. (c) Os centróides foram recalculados. Os grupos já estão em sua forma final. Caso não estivessem, repetiríamos os passos (b) e (c) até que estivessem.

Os dados de raios foram divididos em 3 centróides, e os dados das variáveis foram usados para se agrupar nos centróides, usando a técnica K-Means, os 3 clusters foram formados : cluster 1 poucos raios ($0\% < \text{Cluster 1} < 33,3\%$), cluster 2 uma quantidade intermediária de raios ($33,3\% < \text{Cluster 2} < 66,6\%$) e a cluster 3 representa muitos raios ($66,6\% < \text{cluster 3} < 100\%$). Na Figura 10, os clusters estão definidos, e na Figura 11, os clusters estão representados na forma de percentis.

Figura 10 – Técnica K-Means na Separação dos Clusters.

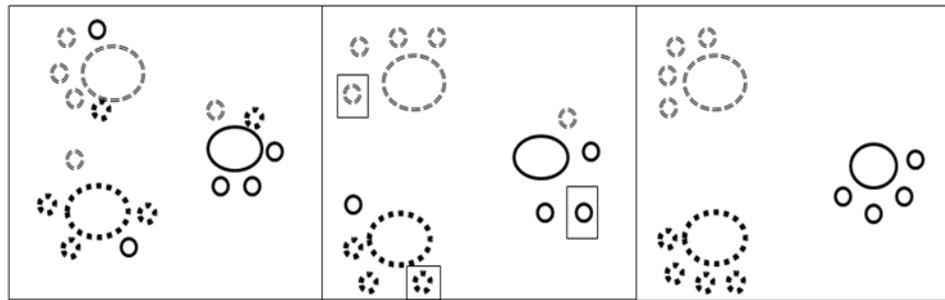
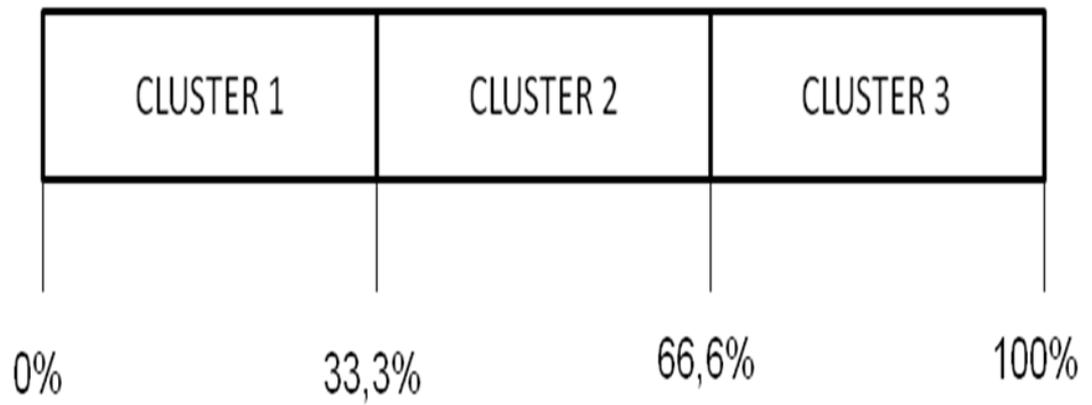


Figura 11 – Clusters de Severidade.



Os dados foram agrupados em um só arquivo, dados diários de janeiro 2009 à dezembro de 2010 de raios, as variáveis termodinâmicas, os clusters de severidade, e o total de raios, como mostra o exemplo do quadro abaixo para o mês de janeiro de 2009.

Tabela 8 – Exemplo de dados agrupados

Ano	Mês	Dia	Show	Lift	Swet	Kind	Ctot	Vtot	Totl	Cape	Cins	Eqlv	Lfct	Brch	Lclt	Lclp	Mlth	Mlmr	Thtk	Pwat	Cluster	Raios
2009	1	1	0.51	-5	218.99	32.8	20.2	23.3	43.5	2678.2	-0.43	157.93	944.47	420.05	296.72	963.89	299.87	19.43	5766	55.67	2	392
2009	1	2	1.56	-4.64	199.79	32.5	18.1	25.1	43.2	2211.1	-0.71	147.39	934.58	156.6	296.29	958.34	299.93	19.04	5771	52.7	2	86
2009	1	3	-0.79	-3.37	263.22	26.9	21.7	22.3	44	1335.7	-2.45	142.74	928.08	254.14	295.6	951	299.88	18.38	5788	43.8	2	22
2009	1	4	2.81	-3.91	169.81	21.5	17.1	25.1	42.2	741.13	-61.37	186.55	832.97	438.23	294.85	948.93	299.31	17.61	5780	39.27	2	87
2009	1	5	3.47	-2.03	187.22	20.1	16.5	23.5	40	719.37	-45.67	171.37	825.91	84.48	294.39	937.89	299.85	17.28	5780	46.28	2	45

3.3.3 Seleção de Índices Explicativos - Busca pela Melhor Escolha (Best First)

Para selecionar os índices mais representativos foi utilizado o método da busca pela melhor escolha (best first search). Alguns conceitos iniciais são importantes para entender esse método.

Algoritmo de busca são técnicas aplicadas a problemas de alta complexidade teórica e que não são resolvidos com métodos de programação convencionais, principalmente os de natureza puramente numérica. Além disso, a redução do processo de solução por meio de algum algoritmo de busca, também diminui essa complexidade consideravelmente. Do mesmo modo, a utilização de heurísticas, geralmente, reduz a explosão combinatória das possibilidades de busca. O algoritmo de busca divide-se em estratégias não-informadas e informadas (GONÇALVES, 2011).

As estratégias de busca não informada, também conhecida como busca cega ou blind search são estratégias de busca baseadas em tentativas de soluções por força bruta onde o único conhecimento que pode ser aplicado ao problema para determinar o próximo passo rumo a uma solução é dado por uma função de enfileiramento. Já na estratégia de busca informada, é acrescentada uma informação a mais na determinação da ordem de expansão dos estados durante o processo de busca. Esta informação é chamada de função de avaliação, ou heurística, e consiste em uma forma de mensurar a probabilidade de um estado convergir para uma solução baseado em seu estado corrente (SOUZA,2008)

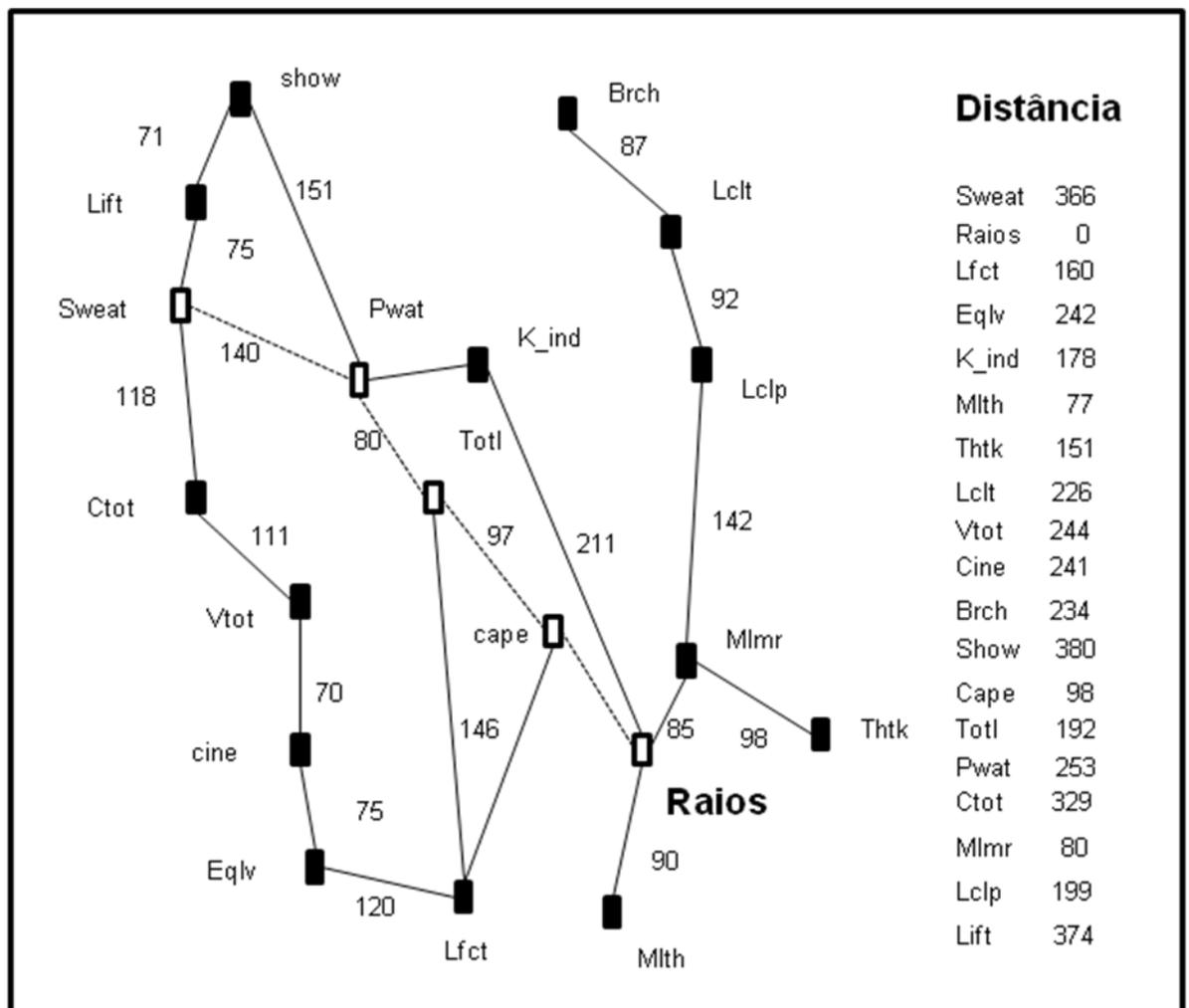
A palavra heurística tem origem da palavra grega eurisco, que significa “Eu descobro”. Portanto, heurística pode ser considerada como um desenvolvimento de métodos e regras baseadas em métodos de tentativa e erro, no qual os problemas são solucionados através de tentativas, até encontrar uma solução viável para o problema.

Métodos heurísticos são algoritmos exploratórios que buscam resolver problemas. Geralmente não envolvem a implementação computacional de um conhecimento especializado (por exemplo, um método heurístico, para resolver uma equação de segundo grau, não usaria, necessariamente, a fórmula de Báscara, mas buscaria, por outros métodos, uma solução que atendesse à equação). Por este motivo, muitas vezes, esses métodos são classificados como “busca cega”. Uma solução ótima de um problema nem sempre é alvo dos métodos heurísticos, uma vez que tendo como ponto de partida uma solução viável, baseiam-se em sucessivas aproximações direcionadas a um ponto ótimo. Logo, estes métodos costumam encontrar as melhores soluções possíveis para problemas, e não soluções exatas,

perfeitas, definitivas. Esta subjetividade, ou falta de precisão dos métodos heurísticos, não se trata de uma deficiência, mas uma particularidade análoga à inteligência humana.

A Figura 12, ilustra o caso de problema de localização de Índices Termodinâmicos com ocorrência de raios, usando a heurística de distância em linha reta, h_{DLR} . Se o objetivo é Raios, é preciso conhecer as distâncias em linha reta até Raios. Na representação heurística é possível entender melhor o método busca pela melhor escolha (Best First), na Figura 12 existem vários índices termodinâmicos, mais é determinada a distância de um índice específico para o objetivo final.

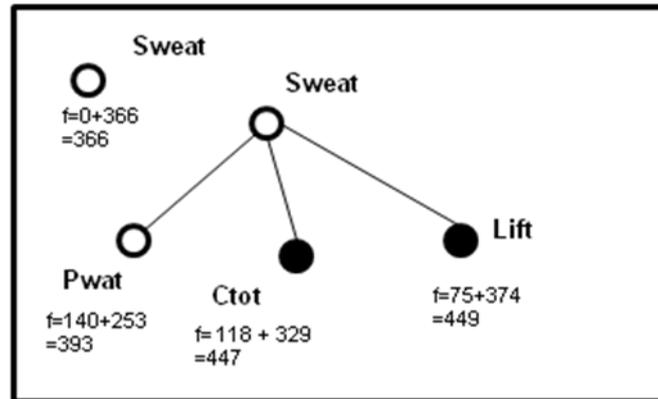
Figura 12 – Representação Heurística.



As Figuras 13, 14 e 15 demonstram o progresso da Busca pela Melhor Escolha para encontrar um caminho de Sweat até Raios. Determinado um índice específico, o método

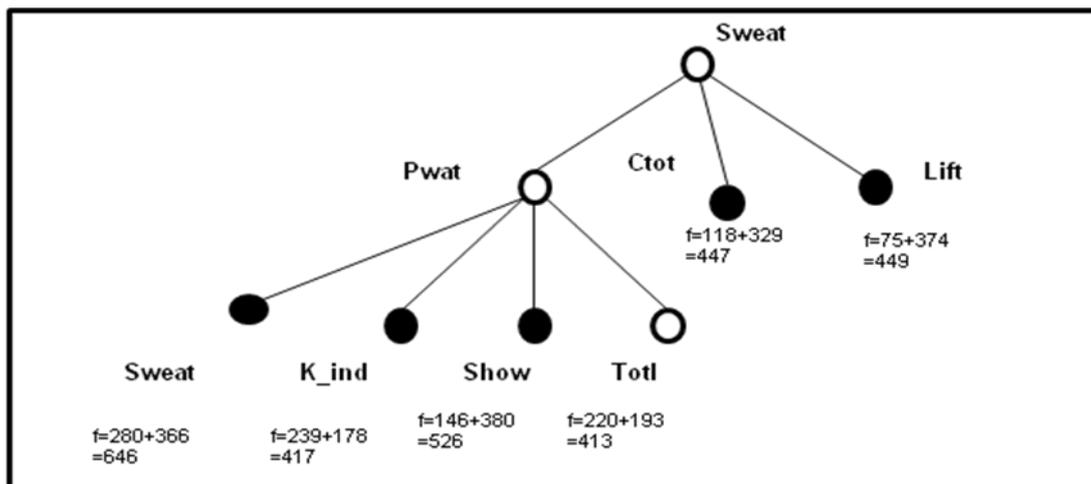
seleciona o melhor caminho, através da menor distância. Com isso ele vai fazendo várias expansões para determinar a melhor escolha. No círculo branco é destacado o melhor caminho para esse problema de localização de Índices Termodinâmicos com ocorrência de raios.

Figura 13 – Busca pela Melhor Escolha- primeira expansão.



Nota-se que na Figura 13 a primeira expansão escolhe a PWAT (água precipitável) como melhor caminho, em seguida é feita uma nova expansão para determinar o próximo Índice Termodinâmico.

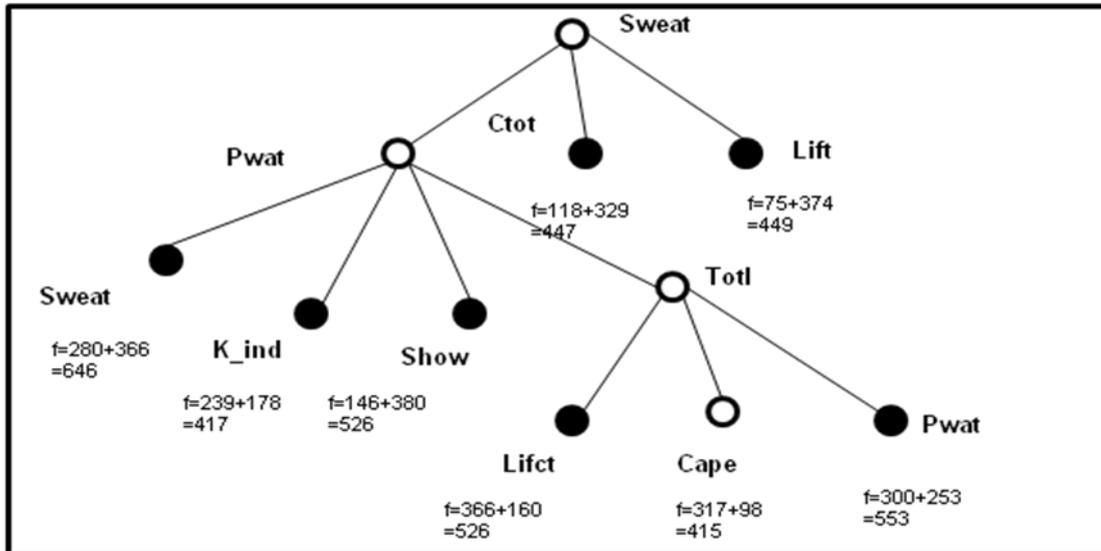
Figura 14 – Busca pela Melhor Escolha- segunda expansão.



Na Figura 14 é observada uma segunda expansão, onde o índice escolhido foi TOTL (índice total), logo o método continua fazendo expansão até encontrar o caminho.

Na Figura 15 abaixo, o método faz a última expansão, determinando o melhor caminho.

Figura 15 – Busca pela Melhor Escolha- última expansão.



A técnica da Busca pela Melhor Escolha foi usada em todos os índices termodinâmicos, para selecionar os mais significativos para ocorrência de raios. Dos 18 índices termodinâmicos foram selecionados 6 para a construção da árvore de decisão. A Tabela 9 mostra os 6 índices selecionados: Lifted, Sweat, K_ind, Totl, CAPE e Pwat.

Tabela 9– Exemplo de dados selecionados para construção do modelo

Lifted	Sweat	K_ind	Totl	CAPE	Pwat	Cluster
-4.64	199.79	32.50	43.20	2211.10	52.70	3
-3.37	263.22	26.90	44.00	1335.70	43.80	3
-3.91	169.81	21.50	42.20	741.13	39.27	1
-2.03	187.22	20.10	40.00	719.37	46.28	1
-4.82	232.82	34.10	42.50	2600.70	51.16	3
-4.49	195.60	34.50	43.00	1698.20	54.31	3

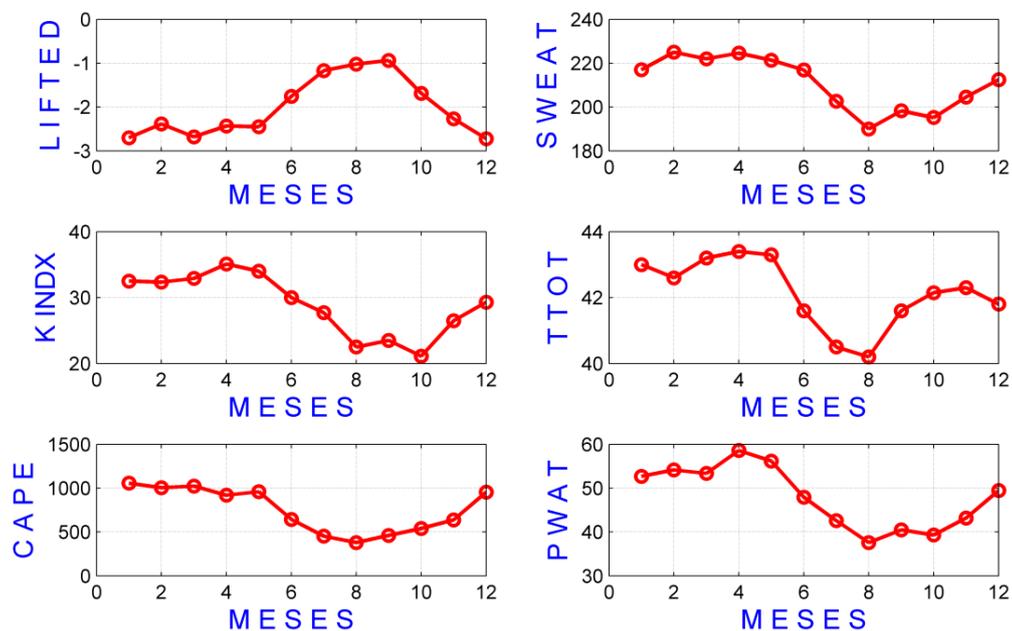
Os dados da Tabela 9 foram organizados e classificados com os clusters na última coluna, demonstrando como foi a entrada do conjunto de dados para a construção do modelo de árvore de decisão no software R.

Os 6 Índices Termodinâmicos encontrados foram semelhantes os que são estudados por (NASCIMENTO, 2005) no intuito de encontrar parâmetros de previsão convectiva para a América do Sul. Os índices de instabilidade termodinâmica, conteúdo de água precipitável e fluxos de energia estática associados ao ciclo anual de precipitação em uma região do Rio Grande do Norte são estudados por (SANTOS; SILVA; LEME, 2011).

O método de Best First (Busca pela Melhor escolha) utilizou informação de raios para encontrar os índices termodinâmicos mais propícios a ocorrências de raios. Os resultados encontrados descritos na Tabela 10 tornam a solução do método em uma solução viável, por apresentar valores descritos na literatura.

Utilizando métodos estatísticos é possível acompanhar o comportamento dos índices termodinâmicos selecionados pelo método Best First, para os anos de 2009- 2010. Foi utilizada a mediana, por ser menos sensível a valores extremos de dados.

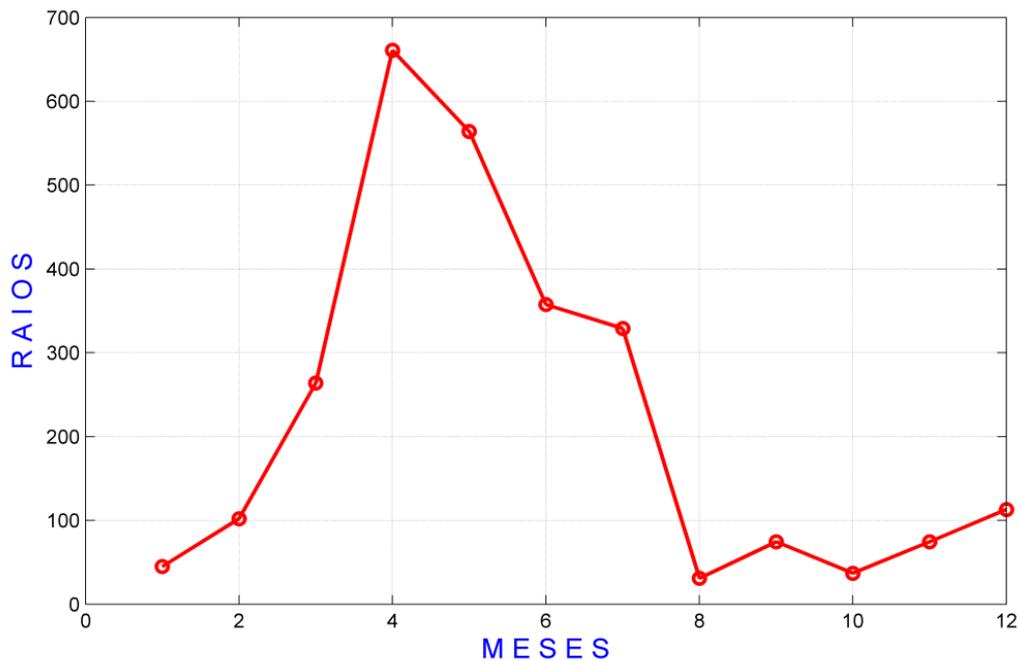
Figura 16 – Mediana mensal dos índices termodinâmicos selecionados pelo método Best First.



O comportamento dos índices termodinâmicos durante os anos 2009-2010 foi de muita instabilidade atmosférica para um período de maior precipitação entre (Dezembro, Janeiro, Fevereiro, Março, Abril, Maio) de acordo com a variável LIFTED, quanto menor o valor de LIFTED, maior a instabilidade. O índice SWEAT é a avaliação de potencial para tempestade severa, onde maiores valores são no período de maior precipitação. Os valores do índice K

variam com a estação do ano e localização. Quanto mais positivo for o índice, maior será a probabilidade de tempestades. Na região estudada os maiores valores estão no período de maior precipitação. Os maiores valores de CAPE, TOTL ou TOTT e PWAT também são encontrados no período de maior precipitação, de acordo com a Figura 16.

Figura 17 – Distribuição das medianas mensais de raios.



Os maiores valores dos índices termodinâmicos são semelhantes aos picos de maior valor de raios distribuídos ao longo do ano para o período estudado (2009-2010), no período de maior precipitação entre (Dezembro, Janeiro, Fevereiro, Março, Abril, Maio), representado na Figura 17.

Para construção do modelo, foram utilizados os 6 índices termodinâmicos considerando a radiossondagem das 12:00 UTC (9:00 Hora Local de Belém –Pa). Os comandos para construção do modelo no software R estão disponíveis no anexo A. No anexo B estão disponíveis os comandos para avaliação do modelo da árvore de decisão.

3.4 Modelo de Árvore de Decisão no R

Uma árvore de decisão consiste numa hierarquia de testes a alguns dos índices envolvidos no problema de decisão. Para criação do modelo são as variáveis Lifted, Sweat, K_ind, Totl, CAPE e Pwat. A árvore pode ser “lida” a partir do teste encontrado na parte superior da mesma, normalmente chamado nó raiz da árvore. Ou seja, para usar uma árvore deste tipo para tomar decisões, começamos por fazer o teste do nó raiz. Se for verdadeiro, seguimos para o ramo esquerdo da árvore e fazemos o teste que aí se encontra, se for falso seguimos o ramo direito, e assim sucessivamente até chegarmos a um nó terminal (normalmente conhecidos por folhas da árvore), onde temos a decisão do modelo para o caso em análise. É possível traduzir uma árvore de decisão para um conjunto de regras de decisão, para implementar como uma função que toma decisões para novos casos.

Após termos um conjunto de dados numa planilha (data frame, na linguagem do R) estamos em condições de obter uma árvore de decisão. Para criar o modelo o conjunto de dados com as variáveis foi normalizado, descrito na equação 19.

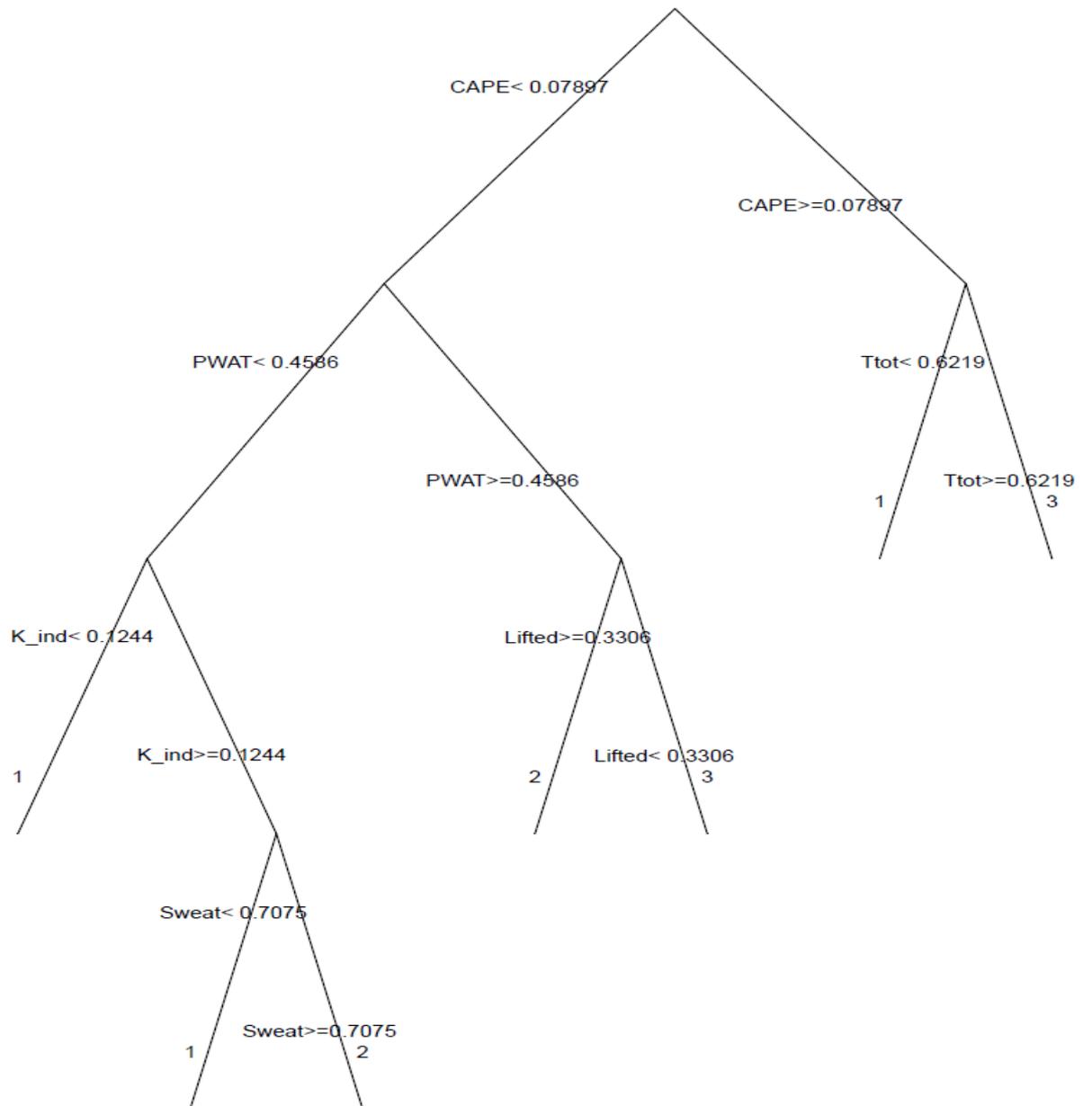
$$X_{Norm} = \frac{X_i - X_{min}}{X_{máx} - X_{min}}$$

(19)

Os valores ficaram no intervalo de [0,1]

O R possui um pacote chamado “rpart”, contém as funções para obter árvores de decisão. O rpart pode construir modelos de classificação ou regressão de uma estrutura muito geral usando um processo em duas fases; os modelos resultantes podem ser representados como árvores binárias. Os comandos para construção do modelo estão no anexo A. Na Figura 18 é demonstrado o resultado do modelo na forma gráfica, em que os resultados estão com os dados normalizados.

Figura 18 – Árvore de Decisão (valores normalizados)



3.5 Avaliação da Qualidade do Modelo

Os comandos para avaliação do modelo da árvore de decisão estão no anexo B. É possível avaliar um modelo através de uma matriz de confusão.

`previsões.modelo`

	1	2	3
1	25	5	6
2	9	30	8
3	0	2	15

Matriz de confusão

Somando todos os valores da Matriz de Confusão, ela representa 100%, e a soma da diagonal principal é o resultado de acerto em porcentagem da previsão. A matriz acima representa 25 % de acerto para o cluster 1, 30% para o cluster 2 e 15% para o cluster 3, totalizando 70% na soma da diagonal principal. A soma do restante da matriz de confusão representa o erro da matriz que é de 30 % de um total de 100 casos.

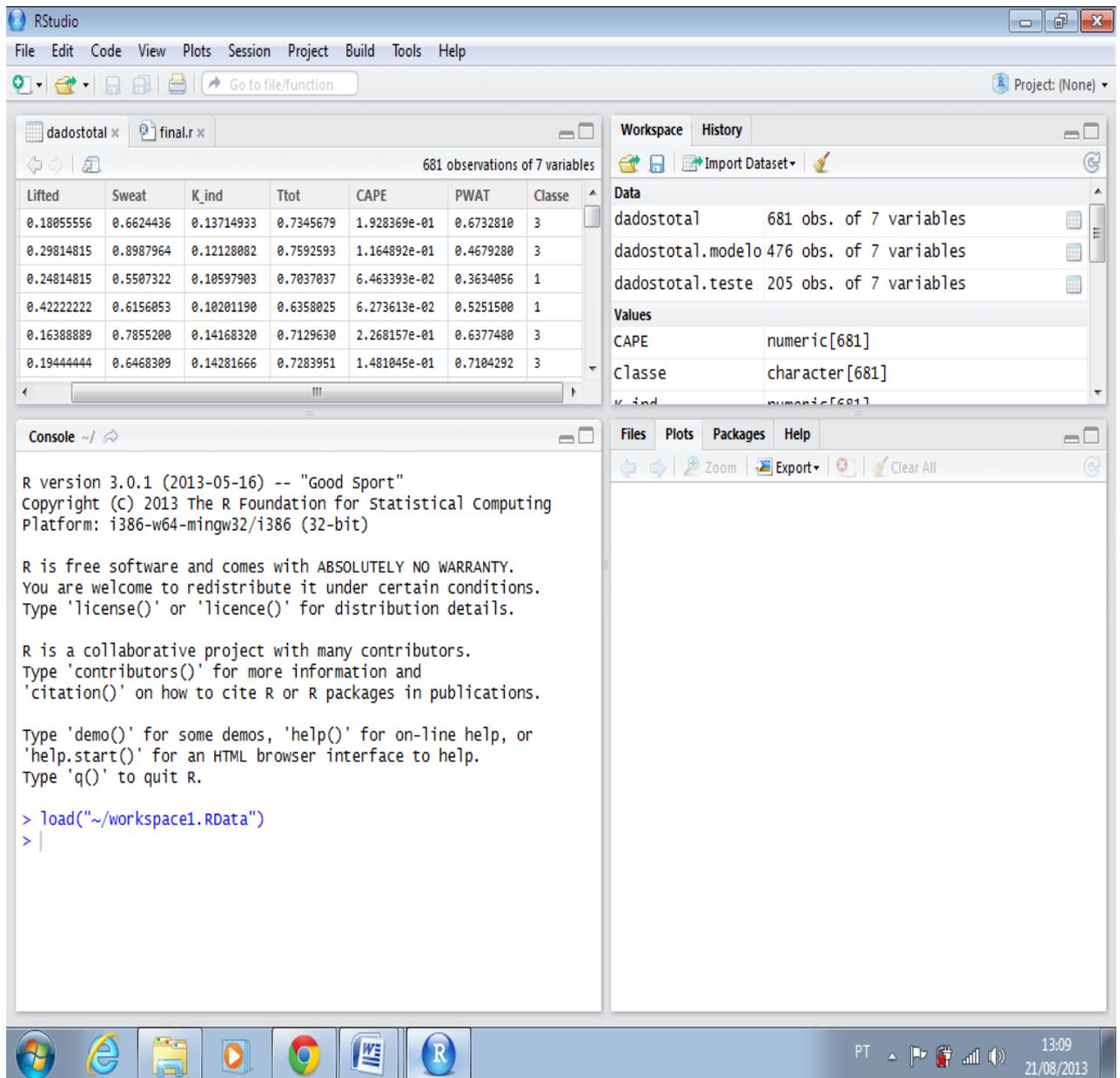
3.6 Classificação e Escolha do Modelo

Foi utilizado o método de validação cruzada para classificar o modelo. A validação cruzada é uma técnica para avaliar a capacidade de generalização de um modelo, a partir de um conjunto de dados. Esta técnica é amplamente empregada em problemas onde o objetivo da modelagem é a previsão. Busca-se então estimar o quão acurado é este modelo na prática, ou seja, o seu desempenho para um novo conjunto de dados.

O conceito central das técnicas de validação cruzada é o particionamento do conjunto de dados em subconjuntos mutuamente exclusivos, e posteriormente, utilizam-se alguns destes subconjuntos para a estimação dos parâmetros do modelo (dados de treinamento) e o restante dos subconjuntos (dados de validação ou de teste) são empregados na validação do modelo.

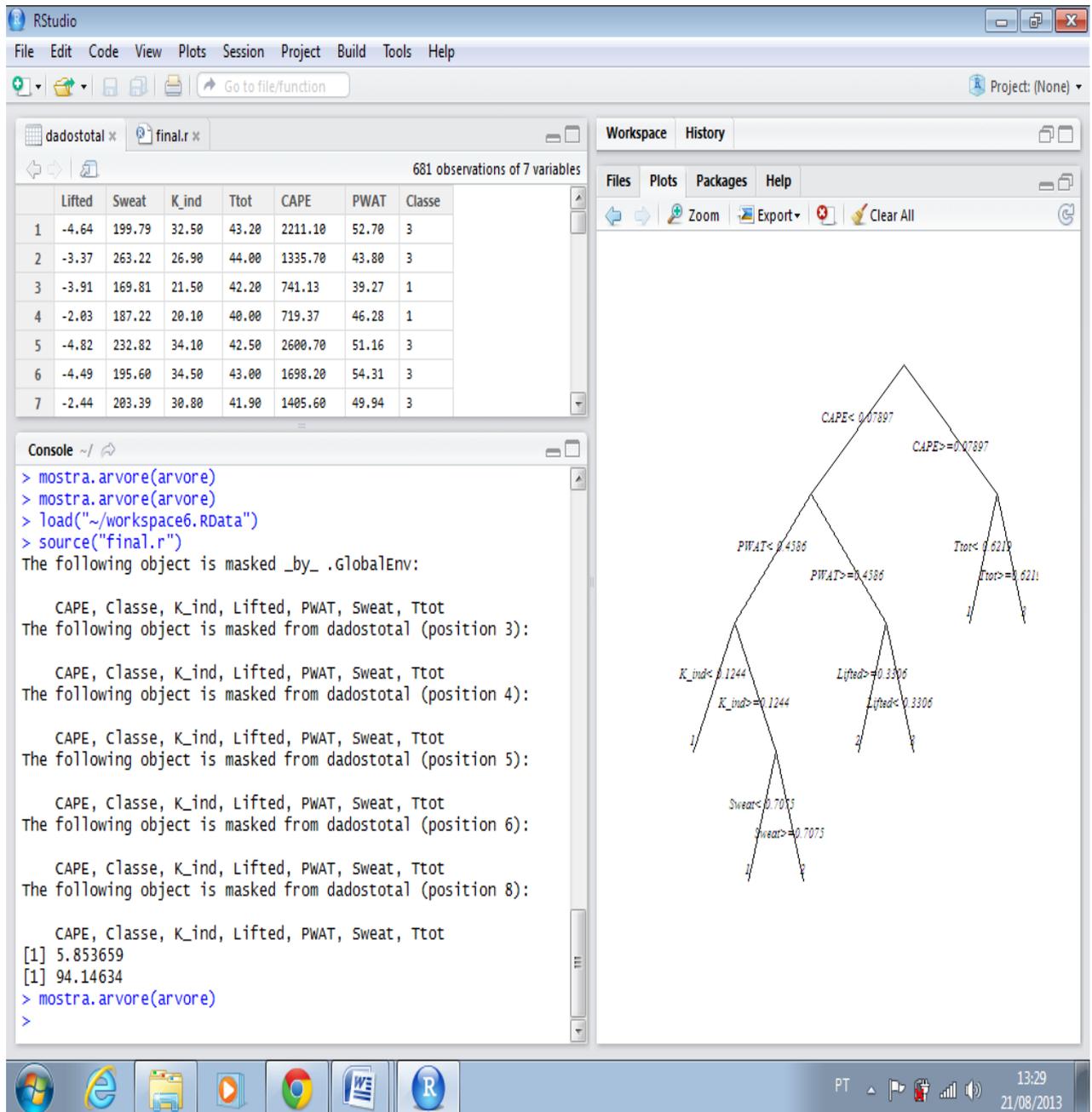
Foram separadas 10 amostras para escolher a melhor representação do modelo da árvore de decisão. Cada classificação testa a árvore com subconjuntos diferentes do total de dados, e essa escolha é feita de forma aleatória pelo software R, onde estão disponíveis no anexo C.

Figura 19 – Imagens do Software R



A Figura 19 mostra uma imagem da área de trabalho do software R, onde se observa o conjunto de dados, o local dos comandos, e onde o modelo será definido na forma gráfica.

Figura 20 – Árvore de decisão feita no software R

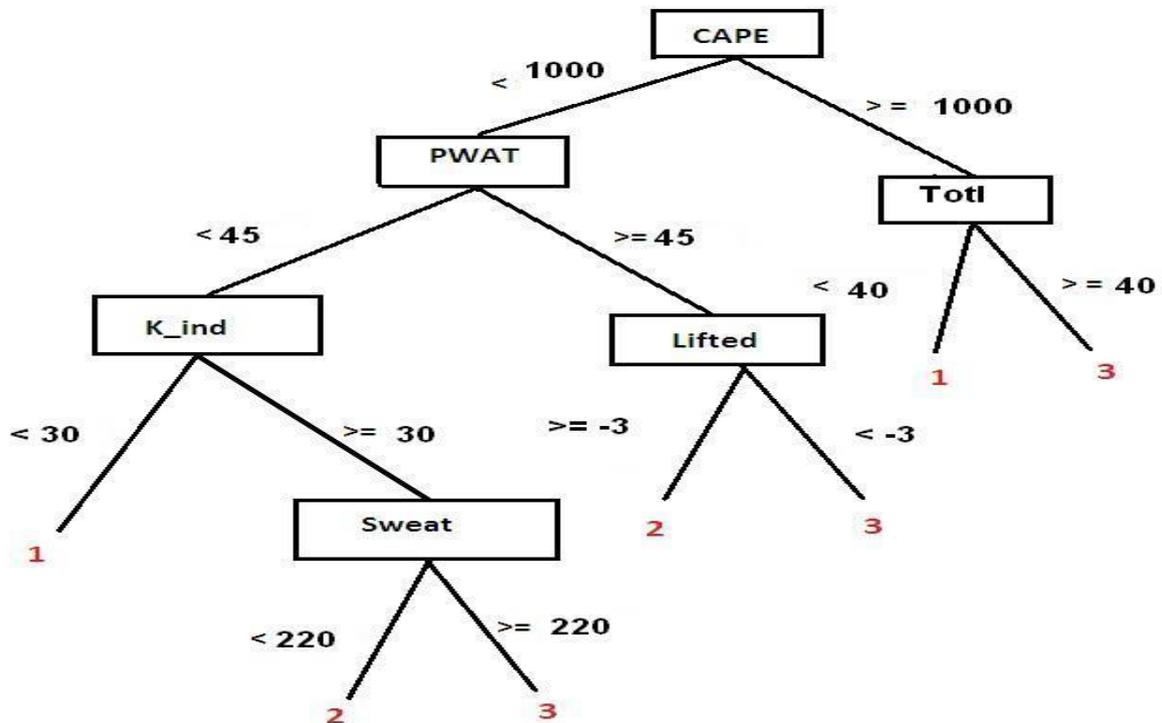


Dos 10 casos, foi escolhido o modelo com maior percentual de acerto. A Figura 20 mostra esse percentual de 94 %.

4 RESULTADOS E DISCUSSÃO

A escolha foi do modelo da Figura 20, porque o percentual de acerto foi de 94%, os valores encontram-se normalizados, e na Figura 21 os dados já estão transformados nos valores reais.

Figura 21 – Modelo da árvore de decisão com valores reais



A Figura 21 representa a forma gráfica da árvore decisão, que pode ser descrita através de 7 regras como :

1. CAPE ≥ 1000 , e Totl ≥ 40 , temos cluster 3 de raios.
2. CAPE ≥ 1000 , e Totl < 40 , temos cluster 1 de raio.
3. CAPE < 1000 , Pwat ≥ 45 , e Lifted < -3 , temos cluster 3 de raios.
4. CAPE < 1000 , Pwat ≥ 45 , e Lifted ≥ -3 , temos cluster 2 de raios.
5. CAPE < 1000 , Pwat < 45 , K_ind ≥ 30 , e Sweat ≥ 220 , temos cluster 3 de raios.
6. CAPE < 1000 , Pwat < 45 , K_ind ≥ 30 , e Sweat < 220 , temos cluster 2 de raios.
7. CAPE < 1000 , Pwat < 45 , K_ind < 30 , temos cluster 1 de raio.

Parâmetros de instabilidade para previsão convectiva: Índice de Instabilidade por levantamento (Lifted):

$$\text{LIFT} = T_{500} - T_{\text{parcela}}$$

Originalmente, T_{500} é a temperatura do ar ambiente em 500 hPa e T_{parcela} é a temperatura do ar ambiente em 500 hPa (ambas em graus Celsius) após ascender – via curva adiabática seca, e úmida após saturação a partir da superfície com a temperatura máxima prevista para o dia em questão e com a razão de mistura média para os primeiros 900m da sondagem. Valores negativos de Lifted indicam condições instáveis, valores Lifted abaixo de -5°C indicam situação muito instável, e extremamente instável se abaixo de -10°C . O Lifted é uma informação importante para analisar o grau de instabilidade atmosférica, especialmente quando combinada com a CAPE (NASCIMENTO, 2005).

Um dos parâmetros mais utilizados em previsão convectiva (severa ou não) é a energia potencial convectiva disponível (CAPE) que pode ser definida como:

$$\text{CAPE} = \int_{Z-\text{LFCV}}^{Z-\text{EQTIV}} g \cdot \frac{(TP - TE)}{TE} dz$$

EQTIV = Nível de equilíbrio; **LFCT** = Nível de convecção livre; g = Aceleração da gravidade terrestre ($9,8 \text{ m/s}^2$); **TP** = Temperatura de uma parcela de ar nos primeiros 500 m de altitude, elevada adiabaticamente seca até o nível de condensação por levantamento (localizado na base da nuvem), e adiabaticamente úmida depois; **TE** = Temperatura do ar.

De uma forma geral, valores de CAPE de 100 a 2500 J kg^{-1} são considerados altos; valores de 2500 J kg^{-1} indicam instabilidade acentuada, e acima de 4000 J kg^{-1} indicam instabilidade extrema (NASCIMENTO, 2005).

Outro parâmetro estudado para previsão de tempo severo é o índice SWEAT (Índice de ameaça severa do tempo) definido como:

$$\text{SWET} = 12 \cdot (\text{TD850}) + 20 \cdot (\text{TERM2}) + 2 \cdot (\text{SKT850}) + \text{SKT} \cdot 500 + \text{SHEAR}$$

Onde:

TD850 = Temperatura do ponto de orvalho, em graus $^{\circ}\text{C}$, no nível de pressão atmosférica a 850 hPa;

TERM2 = Máx (TOTL – 49), onde TOTL é o índice total- totals;

SKT850 = Velocidade do vento, em nós, no nível em que a atmosfera está com 850 (850 hPa) de pressão;

SKT500 = Velocidade do vento, em nós, no nível em que a atmosfera está com 500 hPa de pressão;

$SHEAR = 125.[\text{sen}(\text{DIR}500 - \text{DIR}850) + 0,2]$, onde DIR500 = a direção do vento no nível pressão atmosférica a 500 hPa e DIR850 = direção do vento no nível de pressão atmosférica a 850 hPa.

Não é permitido que um dos termos da equação seja negativo. Sendo que o termo SHEAR deve ser considerado como nulo se quaisquer das considerações a seguir não forem satisfeitas, como:

- ✓ A direção do vento no nível de 850 hPa se encontra no intervalo de 130 a 250°;
- ✓ A direção do vento no nível de 500 hPa se encontra no intervalo de 210 a 310°;
- ✓ A diferença entre a direção dos ventos é positiva;
- ✓ As velocidades dos ventos dos 500 e/ou 850 hPa são maiores do que 15 nós (~30Km/h = 8,3 m/s).

O Sweat combina informação sobre as estruturas termodinâmica e cinemática da atmosfera, sendo, portanto desenvolvido para a previsão específica de potencial para tempestades severas. Valores de Sweat acima de 300 são considerados indicativos de situações propícias a tempestades severas, enquanto que valores acima de 400 indicam maior probabilidade de eventos tornádicos (NASCIMENTO, 2005).

Índice K é obtido em graus celsius, é uma medida do potencial de instabilidade baseado na variação vertical de temperatura. Este índice é muito usado para avaliar chuvas fortes, pois a presença de camadas úmidas em 850 hPa e 700 hPa implica em bastante água precipitável

$$KINX = (T850 - T500) + TD850 - (T700 - TD700)$$

Onde:

T850, T500 e T700 = São as temperaturas do ar, em graus °C, no nível de pressão atmosférica de 850 hPa, 500 hPa e 700 hPa, respectivamente;

TD700 e TD850 = São as temperaturas do ponto de orvalho, em graus °C, no nível de pressão atmosférica de 700 hPa e 800 hPa, respectivamente.

Os valores do índice K variam com a estação do ano e localização. Quanto mais positivo for o índice, maior será a probabilidade de tempestades, sendo que o valor crítico para esse índice é de aproximadamente 30°C e, quanto maior o valor de K, maior a instabilidade atmosférica (DIAS, 2008).

De uma forma geral, valores de TOTL a partir de 40° C indicam situações favoráveis a tempestades, e acima de 50 °C indicam chances de tempestades severas (NASCIMENTO, 2005).

Água precipitável ($PW = \text{Precipitable Water}$) é a integral vertical da razão de mistura do vapor d' água. Expressa a quantidade de água (em altura ou em massa), que poderia ser obtida se todo o vapor de água contido numa coluna de atmosfera de secção transversal horizontal unitária se condensasse e precipitasse:

$$PW = \frac{1}{g} \int_{z_1}^{z_2} r \, dz$$

Em que g (m s^{-2}) é a aceleração da gravidade; r (g kg^{-1}) é a razão de mistura do vapor; z_1 e z_2 são as alturas correspondentes aos níveis de pressão 1000 e 300 hPa, respectivamente. Sendo calculada pela Equação acima, a PW pode ser expressa em mm , são bem elevados, usa-se por praticidade a unidade cm . Ressaltando que a água precipitável é considerada uma variável atmosférica e não um índice termodinâmico (SANTOS; SILVA; LEME, 2011).

Todos os índices termodinâmicos encontrados na árvore de decisão são semelhantes aos descritos na literatura para determinar a estabilidade ou instabilidade da atmosfera. Com intuito de prever raios, no capítulo 5 foram feitos 2 estudos de casos, com 4 dias determinados de forma aleatória, mas levando em consideração 2 dias no período de maior precipitação e 2 dias no período de menor precipitação para região de Belém, e considerando os horários de maior incidência de raios para região de 100 km de raio no entorno de Belém para validação do modelo.

Os estudos de casos no capítulo 5 começam com uma breve informação da distribuição de raios para toda a região Amazônica, e seu horário de maior intensidade. Foi analisada a distribuição horária de raios para região de estudo no entorno de Belém, para entender a ocorrência de raios na área alvo

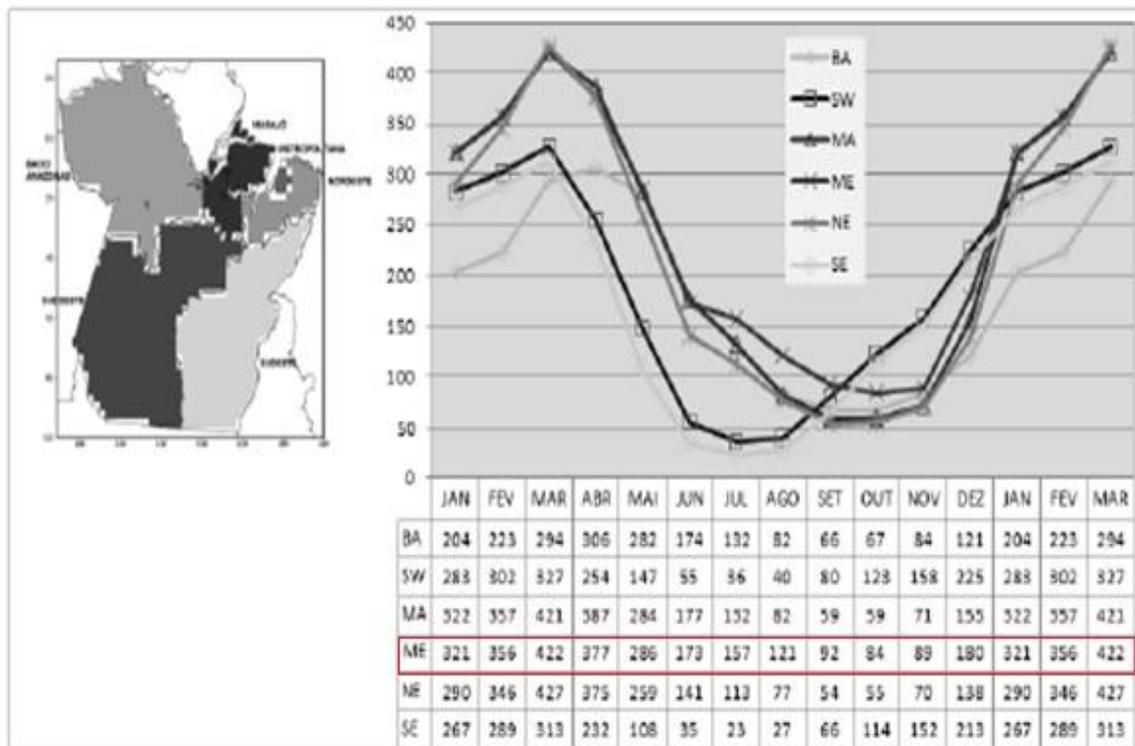
Foram separados os dados de raios no período de estudo (2009-2010), e foram feitas as médias mensais desse período, imagens da temperatura da superfície do mar também foram utilizadas para análise dos Casos 1 e 2.

5 ESTUDOS DE CASOS

5.1 Utilizando o Modelo da Árvore de Decisão

A análise do modelo da árvore de decisão levará em consideração as médias climatológicas mensais de precipitação (1978-2008) para a região metropolitana de Belém, descrito na Figura 22 como ME (ALBUQUERQUE et al, 2010)

Figura 22 – Médias climatológicas da precipitação mensal (1978-2008) representativas das mesorregiões do Pará para os meses de janeiro a dezembro. Unidade: mm.

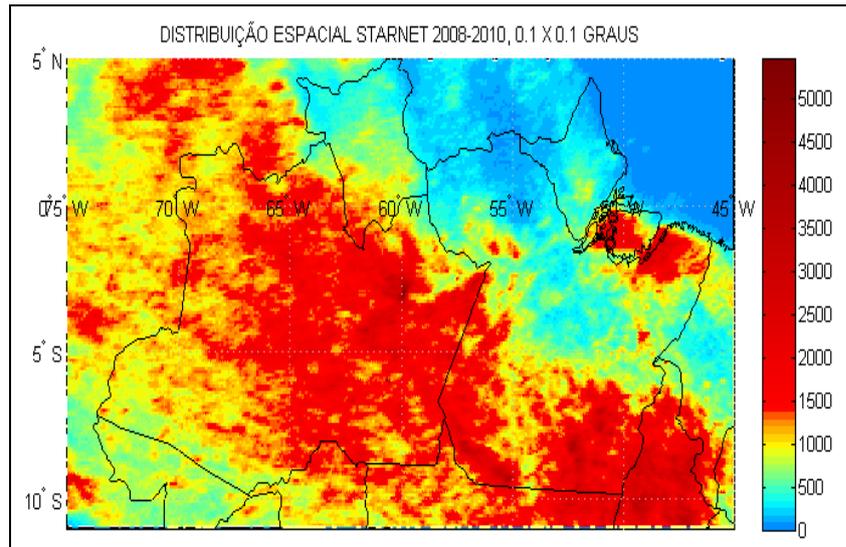


Fonte: Albuquerque et al. (2010)

Foram escolhidos 2 dias no período de maior precipitação nos trimestres (Dezembro, Janeiro, Fevereiro) e (Março Abril, Maio), considerando a Figura 22, os dias escolhido foi 17-02-2013 e 12-01-13, para analisar o Caso 1, e para o Caso 2, foram escolhidos 2 dias no período com menor precipitação correspondentes aos trimestres (Junho, Julho, Agosto) e (Setembro, Outubro, Novembro) Os dias foram 01-07-2013 e 22-08-2013. Esses 4 dias foram selecionados com o objetivo de prever raios numa área de 100 km no entorno de Belém a partir das condições termodinâmicas.

O estudo da distribuição espacial e temporal dos raios na Amazônia com os dados da STARNET, que já havia sido previamente realizado por Almeida¹. A Figura 23 mostra a distribuição espacial, que foi calculada em uma grade de 0.1 x 0.1 grau na extensão do retângulo que cobre a Amazônia: Longitude: -45 a -75 e Latitude: 5 a -11, no período: junho de 2008 a dezembro de 2010.

Figura 23 - Distribuição Espacial de Raios.



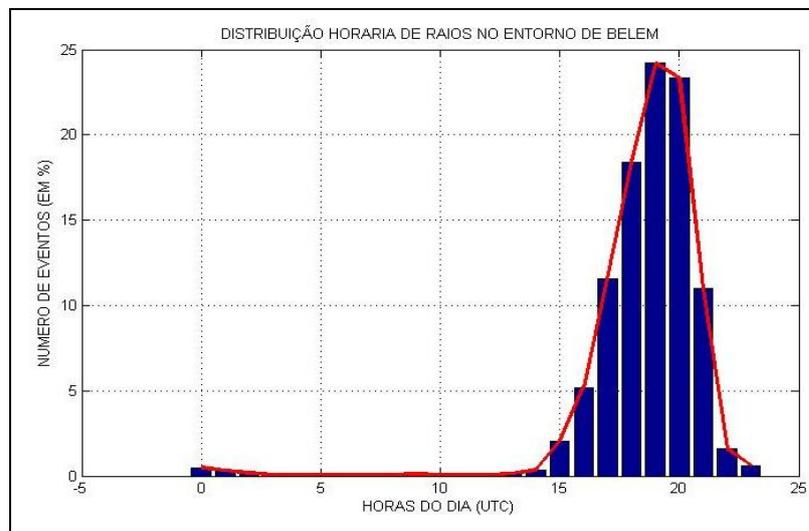
Fonte: Almeida²

Nota-se que na região do Pará, a maior parte dos raios concentra-se principalmente nos extremos do estado (nordeste e sul). Na região da cidade de Belém, localizada na região nordeste do estado, o quantitativo de raios chega a 5000 ocorrências entre 2008-2010. Esse padrão pode estar relacionado com o clima local, uma vez que, a capital Belém além de estar na região tropical também é uma cidade litorânea, ou seja, possui uma grande disponibilidade de elevadas temperaturas e umidade durante todo o ano. Essas características somadas aos efeitos locais provocados pela urbanização (CORRÊA, 2011), geram uma dinâmica atmosférica extremamente favorável à formação de nuvens cumulonimbus.

A Figura 24 mostra a distribuição dos raios na região de estudo, nota-se uma maior intensidade às 19:00 UTC (16:00 horário de Belém). Para prever raios que se intensificam a tarde, a radiossondagem utilizada das 12:00 UTC (9:00 horário local).

² Almeida [informação pessoal] arthur@ufpa.br em Agosto 2013.

Figura 24 - Distribuição Horária de Raios.



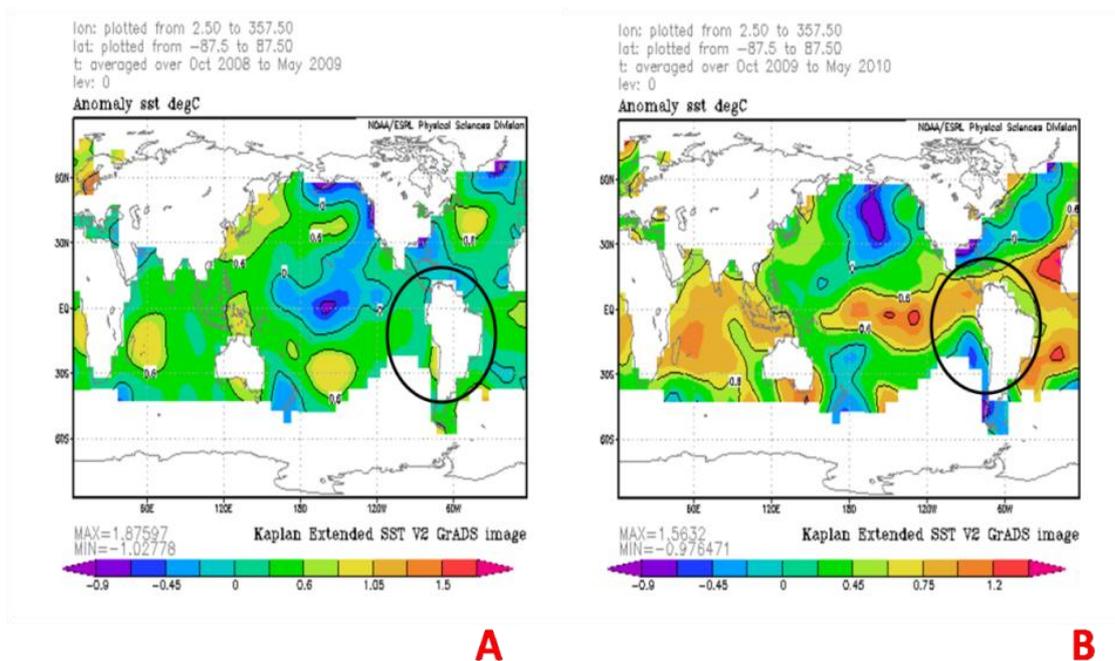
A Tabela 10 mostra a mediana diária de raios para o período de 2009-2010, buscando-se analisar os casos futuros de 2013, sendo feita uma comparação com os meses da Tabela 10.

Tabela 10 – Dados de Raios na área de estudo anos de 2009 e 2010.

Meses 2009-2010	Total	Mediana mensal	Mediana diária
Jan09-Jan10	19340	289	4,82
Fev09-Fev10	19554	277,5	4,63
Mar09-Mar10	32380	752	12,53
Abr09-Abr10	84278	1480,5	24,67
Mai09-Mai10	66588	1541,5	25,69
Jun09-Jun10	58020	1176	19,6
Jul09-Jul10	52738	951	15,85
Ago09-Ago10	17248	162	2,7
Set09-Set10	19494	269	4,48
Out09-Out10	10646	120	2
Nov09-Nov10	7872	156,5	2,6
Dez09-Dez10	19795	222	3,7

O fenômeno de grande escala interanual : o ENOS (El niño Oscilação sul), influenciam na precipitação da região Amazônica principalmente no primeiro semestre de cada ano. A figura 25 mostra a temperatura da superfície do mar (TSM) do NCEP/NCAR (National Center for Environmental Prediction/ National Center for Atmospheric Research) disponibilizados pelo NOAA (National Oceanic and Atmospheric Administration).

Figura 25 – Temperatura da superfície do mar. A) Ocorrência de um La Niña fraco entre 2008-2009. B) Ocorrência de um El Niño entre 2009-2010.



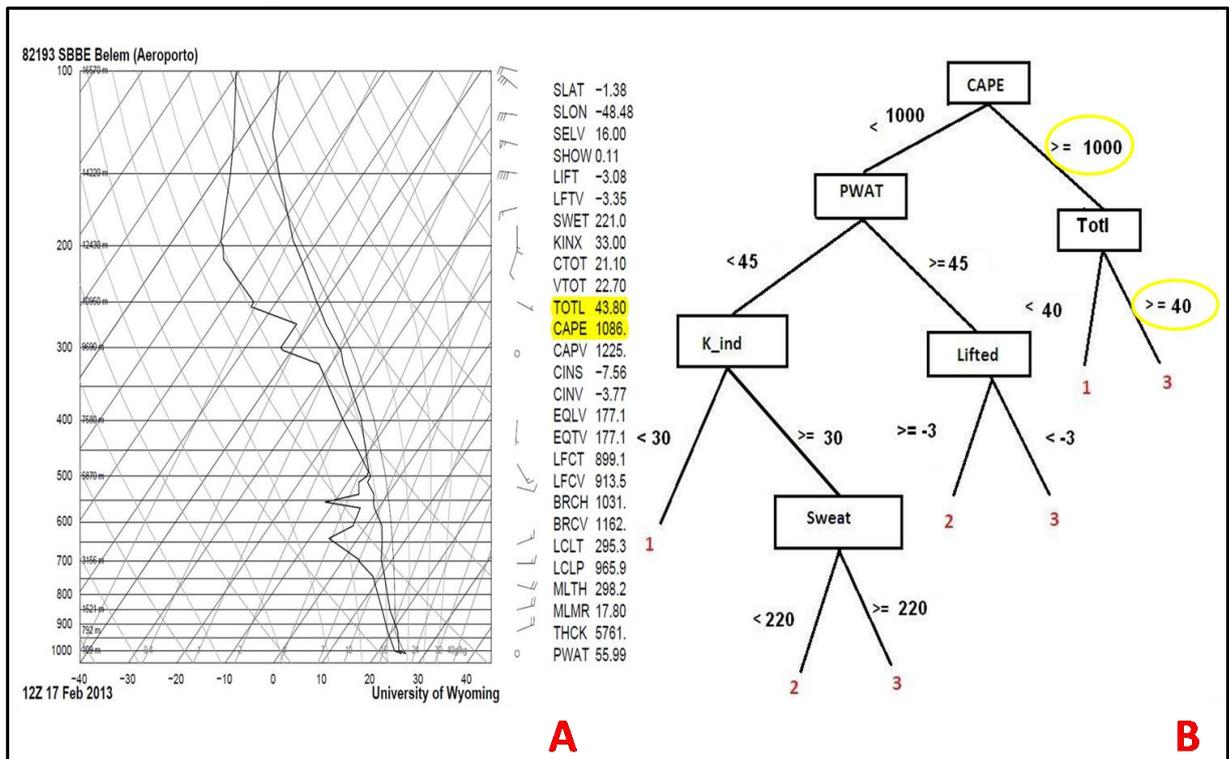
Fonte: NCEP/NCAR

Analisando imagens do diagrama SKEW T LOG P, do satélite GOES 13, precipitação acumulada utilizando dados do CMORPH, TSM e imagens de raios com dados da rede Starnet é possível analisar as regras geradas do modelo da árvore de decisão. A idéia é analisar os casos de ocorrência de raios em dias escolhidos de forma aleatória e em meses diferentes para o ano de 2013 para validar o modelo de árvore de decisão. A análise poderá auxiliar na previsão de tempo para a cidade de Belém- Pa.

5.1.1 Caso1

O modelo de árvore de decisão, será comparado com o Diagrama SKEW T LOG P para uma região que abrange 100 km no entorno de Belém, foi feita essa comparação para 2 dias considerando o período de maior precipitação (Dezembro, Janeiro, Fevereiro) e (Março Abril, Maio); o primeiro dia é o 17/02/2013, e o segundo dia é o 12/01/2013; foi feito um estudo com as imagens de satélite e as imagens da quantidade de raios na região, com o intuito de compreender o comportamento dos índices termodinâmicos e o resultado do modelo.

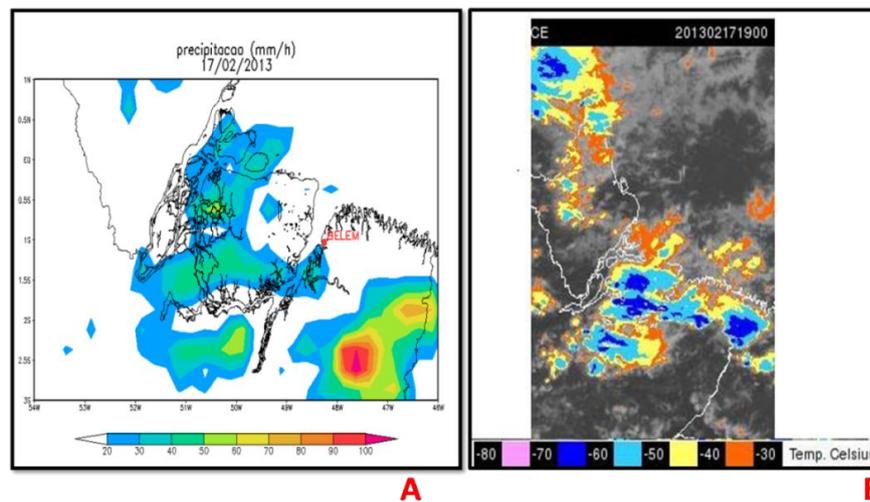
Figura 26 – Comparação do Diagrama SKEW T LOG P de Belém do dia 17/02/2013 às 1200UTC com as regras do modelo da árvore de decisão.



Fonte : Universidade de Wyoming, (2013)

As regras geradas pelo modelo na figura **26B**, diz que a CAPE é o índice termodinâmico a ser considerado com o valor mínimo de 1000, sendo que na figura **26A**, a CAPE foi 1.086, sendo, portanto maior que 1000; logo, o índice seguinte a ser considerado foi a TOTL, que no modelo na Figura **26B** assume um valor limite de 40. Observando-se a Figura **26A**, o TOTL foi 43.80, maior que o valor limite do modelo. Logo, a regra finaliza com o cluster 3, o que representa muitos raios.

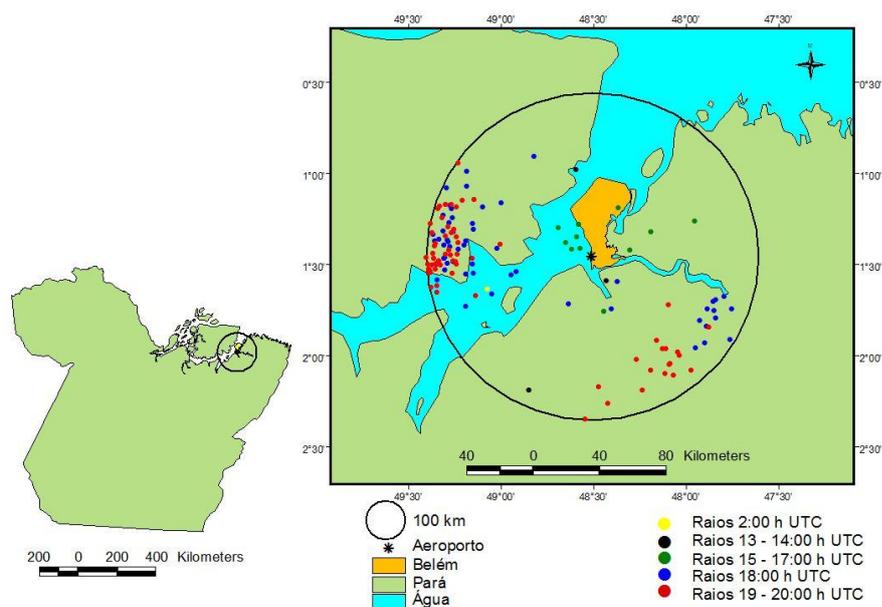
Figura 27– Comparação da precipitação acumulada em Belém do dia 17/02/2013 com as imagens de satélite às 19:00UTC – 16:00 hora local.



Fonte : Satelite GOES 13- Canal – 4 “infravermelho realçado”- CPTEC e precipitação acumulada do CMORPH

A Figura 27A, mostra pouca chuva de 20 - 30 mm, e no horário das 19:00 UTC (16:00 Hora local Belém), está bem nublado de acordo com a figura 27B. É esperado um cluster 3 (muitos raios). Na Figura 28 é observada a distribuição de raios feita no Software Arc View.

Figura 28 – Distribuição temporal dos raios em Belém no dia 17/02/2013.



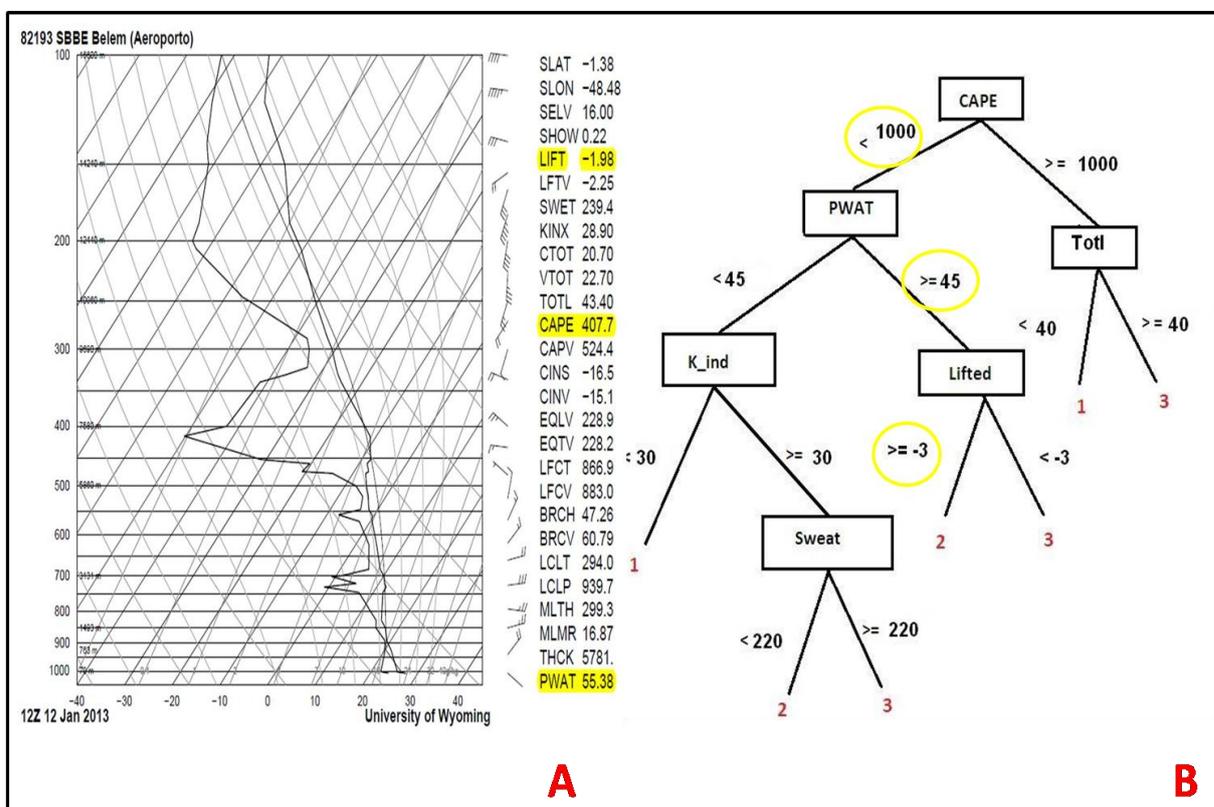
Fonte : Software Arc View, (2013)

A mediana diária de raios considerando a Tabela 10 para o mês de fevereiro foi de aproximadamente 5 raios. O modelo de árvore de decisão indica cluster 3 de raios, o que significa muitos raios, acima da mediana. O total de raios que caíram na região para o dia 17/02/2013 de acordo com a rede da Starnet foi de 137 raios mostrados na figura 28, o que representa cluster 3, acima da mediana.

Valores associados ao índice TOTL, indicam que entre 41 - 43 uma instabilidade moderada, e 44 - 46 formação de tempestades esparsas, e valores de CAPE, indicam que entre 0 - 1000 Limite para convecção profunda, e 1000 - 2500 Convecção profunda moderada. Como os valores dos índices termodinâmicos estão muito próximos dos valores descritos na literatura.

Para o dia 12/01/2013, o modelo prevê cluster 2 de raios, em entorno da mediana como mostra a figura 29.

Figura 29 – Comparação do Diagrama SKEW T LOG P de Belém do dia 12/01/2013 às 1200UTC com as regras do modelo da árvore de decisão.

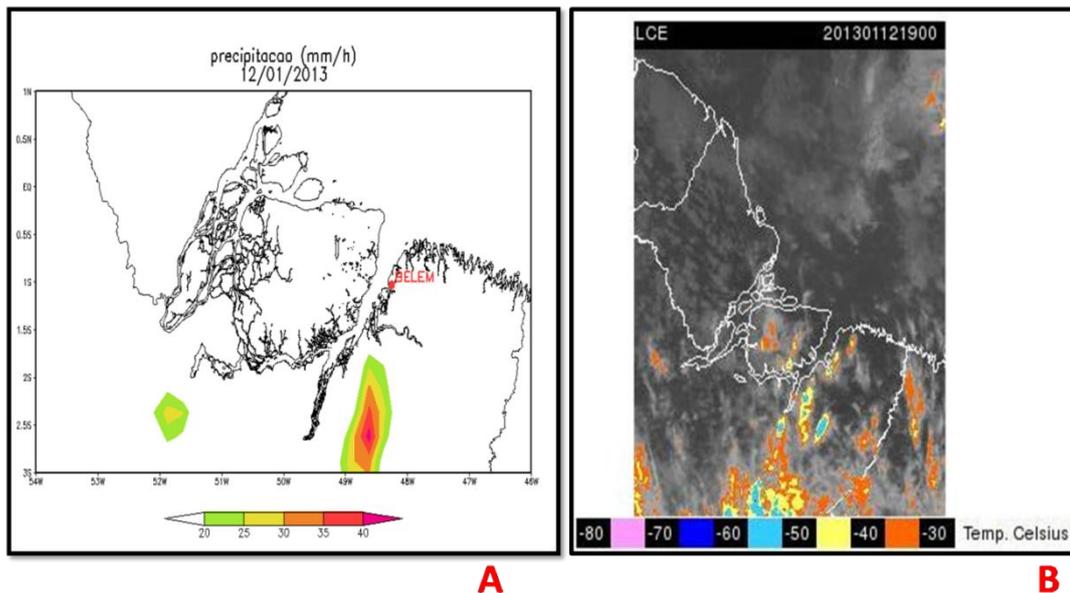


Fonte : Universidade de Wyoming, (2013).

As regras geradas pelo modelo na Figura 29B, diz que a CAPE é o índice termodinâmico a ser considerado com o valor limite de 1000, sendo que na Figura 29A, a CAPE foi 407.7, portanto, menor que 1000. Desta maneira, o índice seguinte a ser considerado foi a PWAT, que no modelo na Figura 29B assume um valor limite de 45; observando a Figura 29A, o PWAT foi 55.38, maior que o valor limite do modelo. Em seguida é considerada outra variável Lifted que no modelo na Figura 29B assumem o valor limite de - 3, sendo que na Figura 29A, o Lifted foi de - 1.98, maior que o valor limite do modelo. Logo, a regra finaliza com o cluster 2, o que representa uma quantidade em torno da mediana de raios.

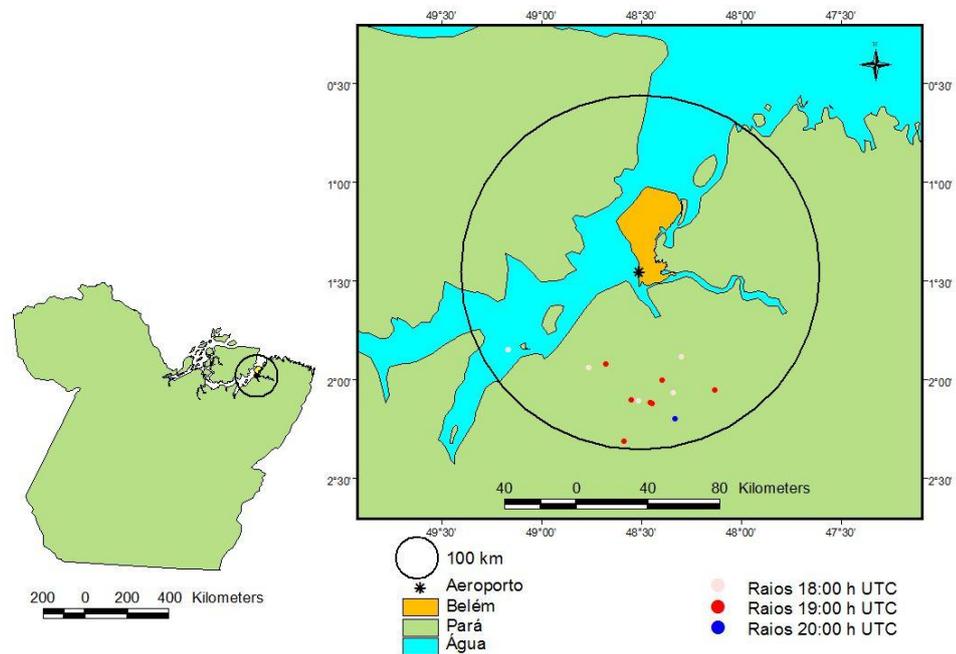
Valor associado ao índice CAPE, indicam que entre 0 - 1000 Limite para convecção profunda, e valores de LIFTED ou LI, indicam que entre -2 <LI<0 Possibilidade de ocorrência de convecção profunda. Na Figura 30A, a precipitação acumulada foi abaixo de 20mm e na figura 30B, no horário de maior ocorrência de raios, as 19:00h (16:00h Hora Local de Belém) não foi possível ver muitas nuvens. A distribuição de raios para o dia 12/01/2013 é observada na Figura 31.

Figura 30 – Comparação da precipitação acumulada em Belém do dia 12/01/2013 com as imagens de satélite às 19:00UTC – 16:00 hora local.



Fonte : Satelite GOES 13- Canal – 4 “infravermelho realçado”- CPTEC e precipitação acumulada do CMORPH

Figura 31 – Distribuição temporal dos raios em Belém no dia 12/01/2013.



Fonte : Software Arc View, (2013).

Na Figura 31 observa-se a quantidade de raios que caíram no dia 12/01/2013. Foi feita uma comparação com a mediana diária de raios, considerando a Tabela 10 para o mês de janeiro foi de aproximadamente 5 raios, o modelo de árvore de decisão indica cluster 2, o que significa raios em torno da mediana. Mas o total de raios que caíram na região para o dia 12/01/2013 de acordo com a rede da Starnet foi de 14 raios, o que representa cluster 2, entorno da mediana.

A quantidade de raios que caíram nos dia 17/02/2013 e 12/01/2013, ficaram na faixa de severidade dos clusters 3 e 2.

5.1.2 Caso 2

Os dias escolhidos para analisar o caso 2 consideram o período de menor precipitação (Junho, Julho, Agosto) e (Setembro, Outubro, Novembro). Os dias 01-07-2013 e 22-08-2013.

Notícia no jornal ORM da cidade de Belém para o dia 01/07/2103: “Chuva deixa trânsito de Belém complicado nesta segunda-feira. Chuva torrencial começou no início da noite e durou mais de duas horas. Várias avenidas da capital ficaram congestionadas com o fluxo de veículos. A forte chuva que caiu em Belém a partir das 18h desta segunda-feira (1º) interrompeu o fornecimento de energia em alguns bairros, o que deixou o trânsito complicado em diversos pontos da cidade. De acordo com o Instituto Nacional de Meteorologia (INMET) na capital paraense, o dia 01/07/2013 esteve parcialmente nublado com pancadas de chuvas.”

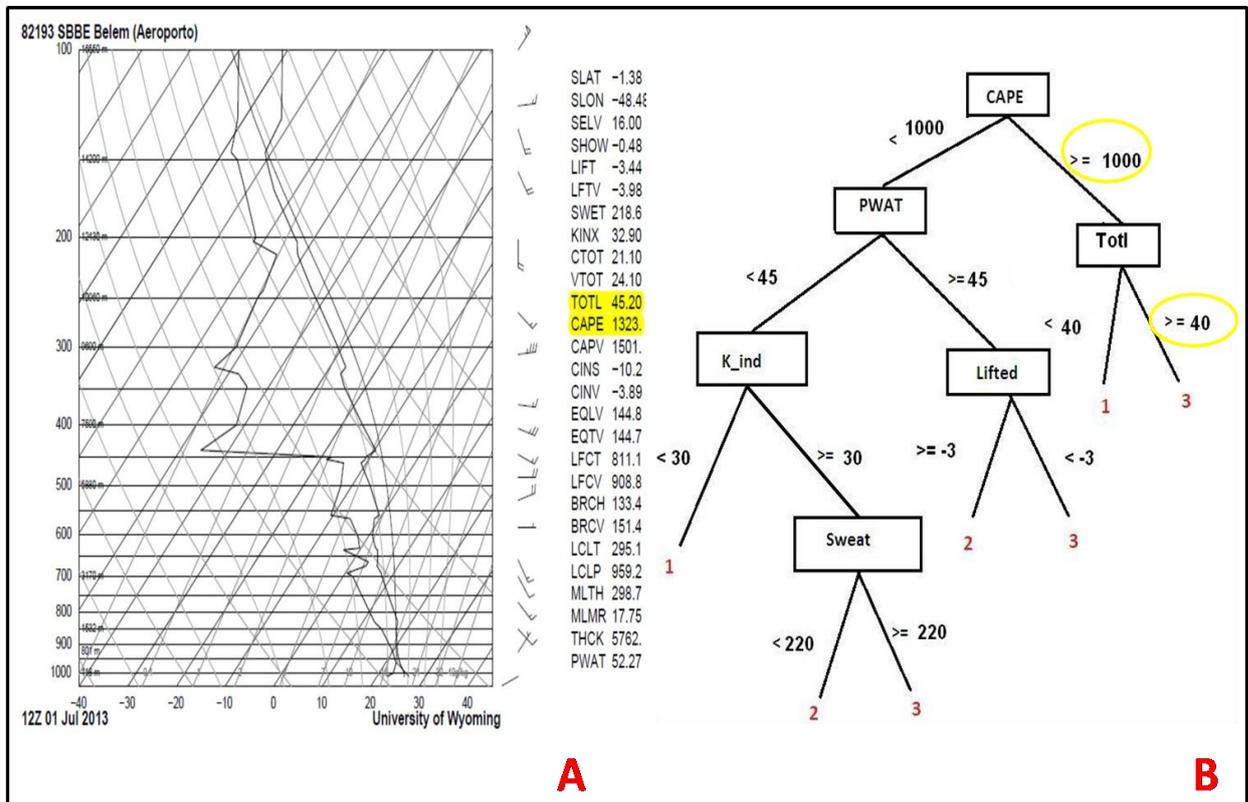
Figura 32 – Impactos socioambientais causados pela chuva e desligamento de energia.



Fonte : Portal ORM (2013)

Na Figura 33 foi feita uma comparação da radiossondagem do dia 01/07/2013 e dados do modelo da árvore de decisão, depois foram observadas as imagens de precipitação acumulada e do satélite GOES 13 na Figura 34.

Figura 33 – Comparação do Diagrama SKEW T LOG P de Belém do dia 01/07/2013 às 1200UTC com as regras do modelo da árvore de decisão.



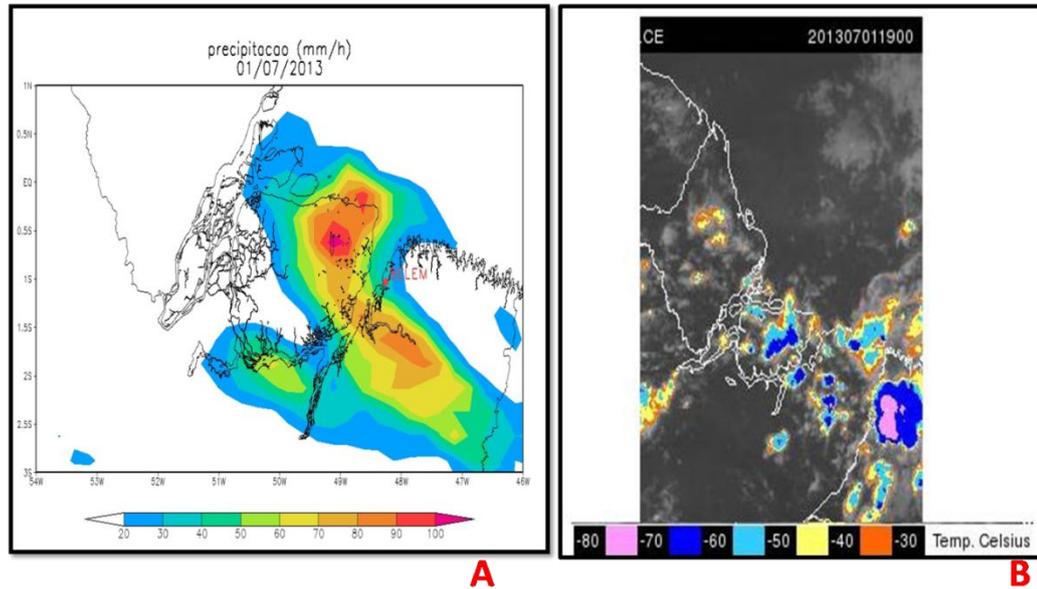
Fonte : Universidade de Wyoming, (2013).

As regras geradas pelo modelo na Figura 33B, diz que a CAPE é o índice termodinâmico a ser considerado com o valor limite de 1000, sendo que na Figura 33A, a CAPE foi 1.323, sendo, portanto maior que 1000, logo a variável seguinte a ser considerada foi a TOTL, que no modelo na Figura 33B assume um valor limite de 40. Observando a Figura 33A, a TOTL foi 45.20 maior que o valor limite do modelo. O que finaliza com um cluster de raios 3, o que representa muitos raios.

Valores associados ao índice TOTL, indicam que entre 41 - 43 uma instabilidade moderada, e 44 - 46 Formação de tempestades esparsas, e valores de CAPE, indicam que entre 0 - 1000 Limite para convecção profunda, e 1000 - 2500 Convecção profunda moderada.

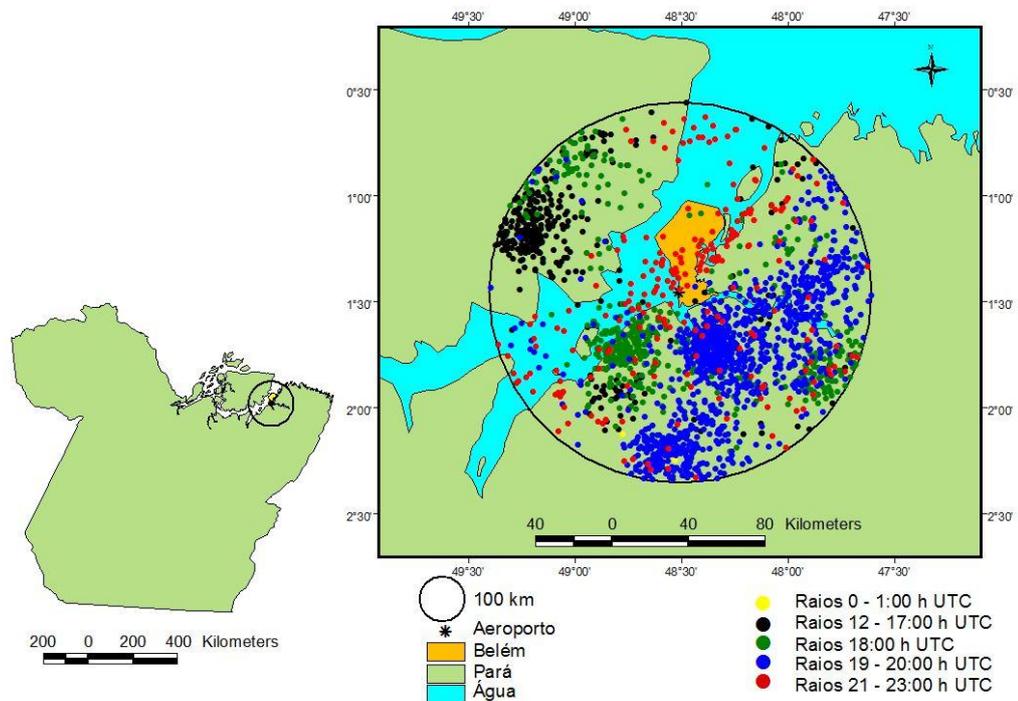
Os valores observados na Figura 34A, de precipitação acumulado em torno de 30-70 para toda região próximo a Belém e na Figura 34B, nuvens esparsas sobre a região no horário das 19:00 UTC (16:00 hora local de Belém). A Figura 35 mostra a distribuição de raios no dia 01/07/2013.

Figura 34 – Comparação da precipitação acumulada em Belém do dia 01/07/2013 com as imagens de satélite às 19:00UTC, sendo 16 hora local.



Fonte : Satelite GOES 13- Canal – 4 “infravermelho realçado”- CPTEC e precipitação acumulada do CMORPH

Figura 35 – Distribuição temporal dos raios em Belém no dia 01/07/2013.

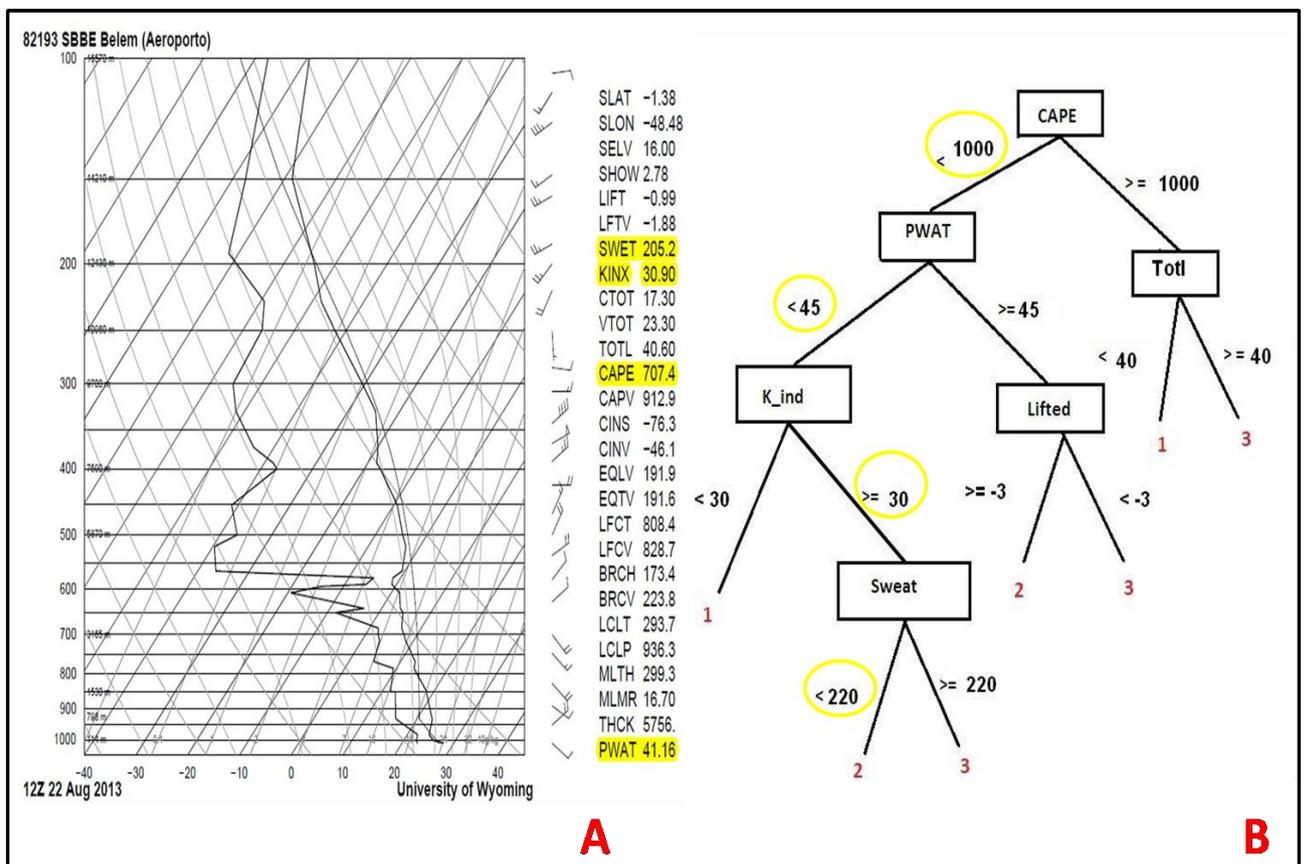


Fonte : Software Arc View (2013).

A mediana diária de raios considerando a Tabela 10 para o mês de julho foi de aproximadamente 16 raios, o modelo de árvore de decisão indica cluster 3 de raios o que significa muitos raios, acima da mediana. O total de raios que caíram na região para o dia 01/07/2013 de acordo com a rede da Sarnet foi de 2271 raios, o que coincide com o modelo, cluster 3.

Para o dia 22/08/2013, o modelo prevê cluster 2 de raios, em torno da mediana como mostra a figura 36.

Figura 36 – Comparação do Diagrama SKEW T LOG P de Belém do dia 22/08/2013 às 1200UTC com as regras do modelo da árvore de decisão.



Fonte : Universidade de Wyoming, (2013).

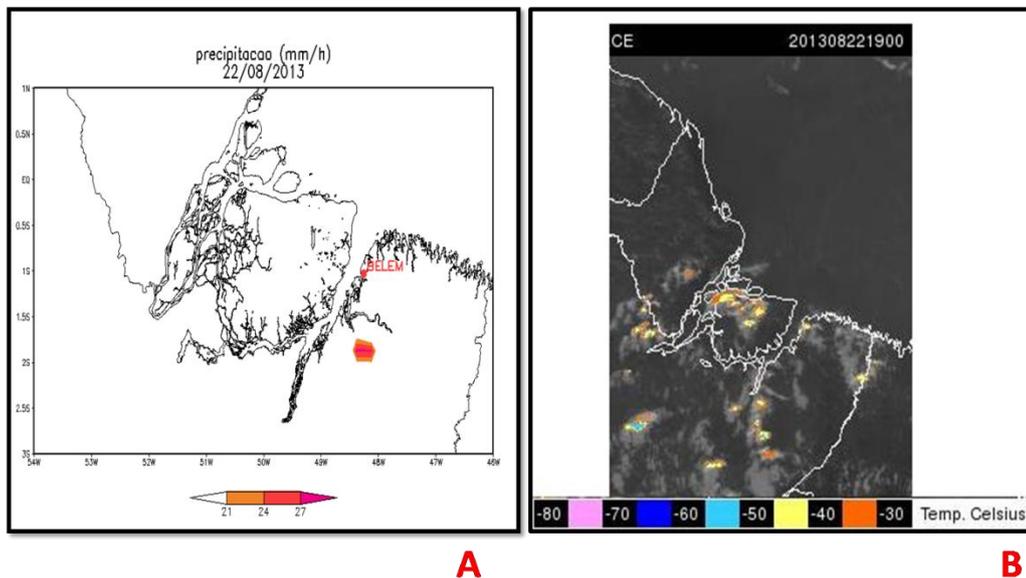
As regras geradas pelo modelo na figura **36B**, dizem que a CAPE é a variável a ser considerada com o valor limite de 1000, sendo que na figura **36A**, a CAPE foi 707.4, sendo, portanto, menor que 1000. Assim, a variável seguinte a ser considerada foi a PWAT, que no modelo na figura **36B** assume um valor limite de 45, observando a figura **36A**, o PWAT foi 41.16, menor que o valor limite do modelo, em seguida é considerada outra variável K_ind

que no modelo na Figura **36B** assumem o valor limite de 30, sendo que na Figura **36A**, o K_{ind} foi de 30.9, maior que o valor limite do modelo. Sendo maior é considerada a variável SWEAT, que na Figura **36B**, assume um valor limite de 220, observando a Figura **36A**, o SWEAT foi de 205.2, menor que o valor do modelo. Logo a regra finaliza com o cluster 2, o que representa uma quantidade em torno da mediana de raios.

Valores associados ao índice CAPE, indicam que entre 0 - 1000 Limite para convecção profunda, e valores de K_{IND} , indicam que entre 26 – 29, 50% Possibilidade de formação de tempestade, 30 – 35, 85% Possibilidade de formação de tempestade severa. Os valores de SWEAT acima de 300 são considerados indicativos de situações propícias a tempestades severas, enquanto que valores acima de 400 indicam maior probabilidade de eventos tornádicos (NASCIMENTO, 2005).

A Figura 37A, mostra precipitação acumulada do dia 22/08/2013 abaixo de 20mm, e no horário das 19:00 UTC (16:00 hora local de Belém), não é possível ver nuvens em cima da região de Belém.

Figura 37 – Comparação da precipitação acumulada em Belém do dia 22/08/2013 com as imagens de satélite às 19:00UTC, 16:00 hora local.

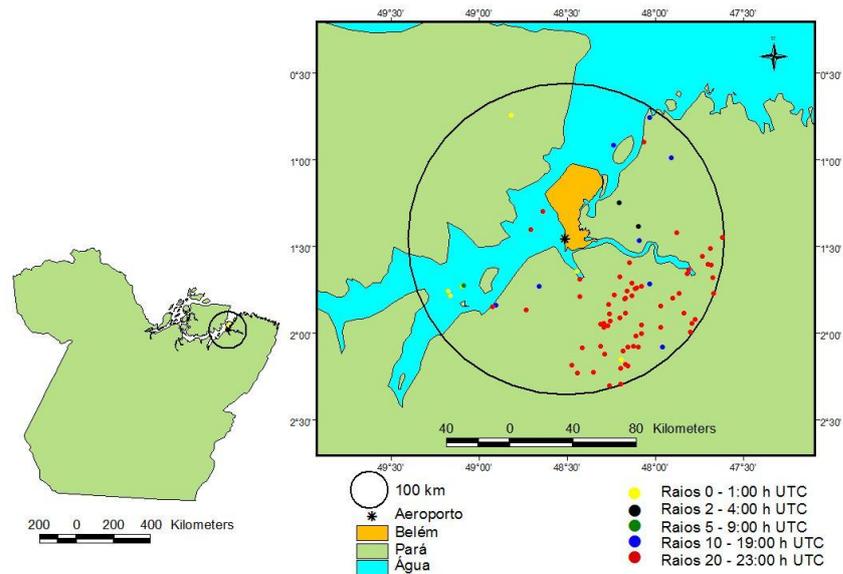


Fonte: Satélite GOES- Canal – 4 “infravermelho realçado”- CPTEC e precipitação acumulada do CMORPH

Na Figura 38 observa-se a distribuição de raios feita no software Arc View, para verificar a quantidade de ocorrência para o dia 22/08/2013. A mediana diária de raios considerando-se a Tabela 10 para o mês de agosto, foi de aproximadamente 3 raios; o modelo

de árvore de decisão indica cluster 2 de raios o que significa quantidade intermediária de raios, em torno da mediana. O total de raios que caíram na região para o dia 22/08/2013 de acordo com a rede da Sarnet foi de 80 raios. O que representa cluster 3 e não 2.

Figura 38 – Distribuição temporal dos raios em Belém no dia 22/08/2013.



Fonte : Software Arc View (2013).

A estação chuvosa é influenciada pela forçante dinâmica de grande escala (ZCIT) que favorece a formação da convecção profunda precipitante; então as nuvens não têm grande desenvolvimento vertical (não alcançam grandes alturas). Todavia, as chuvas são mais abundantes. Já na estação seca, os eventos extremos estão em sua maioria associados a linhas de instabilidade (COHEN; SILVA DIAS; NOBRE, 1995; TAVARES, 2009).

6 CONCLUSÃO

Este estudo teve como foco o uso de técnicas estatísticas para construção de um modelo de árvore de decisão, que fosse capaz de encontrar regras baseadas em aprendizado de máquina supervisionado, para relacionar os Índices Termodinâmicos da atmosfera com a ocorrência de raios na região de 100 km de raio, em torno de Belém- Pa.

O modelo baseado em árvore de decisão foi criado com um acerto de 94% e um erro de 6%. Com a finalidade de validar esse modelo foram analisados os resultados do modelo com os dados reais de raios divididos em dois 2 Casos para dias de maior e menor precipitação na área, respectivamente. A avaliação do modelo com a previsão da mediana diária de raios para os 2 casos em 2013, obteve um resultado de 75 % de acerto para o (Caso 1 e um dia no Caso 2) e 25% de erro para (1 dia do Caso 2).

Em comparação com os dados reais em 2013, o modelo respondeu de forma positiva, com uma margem de acerto de 75 %, considerado bom em uma análise estatística. Em geral o modelo obtido não é determinístico, pois avaliar a quantidade exata de raios depende das condições termodinâmicas e das influências dos sistemas meteorológicos atuantes na região de estudo. Uma das vantagens do modelo de árvore de decisão criado, é a sua adaptação para outras regiões, onde existam radiossondagens rotineiras (como as realizadas em aeroportos), ou deliberadas. Assim sendo esse método de previsão de faixas ou níveis de ocorrências de raios não se limita à área estudada no entorno Belém- Pa.

Os resultados obtidos sugerem novas pesquisas a serem desenvolvidas: a) Com relação a eventuais melhorias no nível de detecção de raios da STARNET; b) Aplicações a outras regiões urbanas que contam com aeroportos e radiossondagens rotineiras; c) Aumentar o número de estudos de casos, para refinar a avaliação estatística do desempenho do modelo na prática.

REFERÊNCIAS

ALBUQUERQUE, M. F.; SOUZA, E. B.; OLIVEIRA, M. C. F.; SOUZA, J. A. J. Precipitação nas mesorregiões do estado do Pará: climatologia, variabilidade e tendências nas últimas décadas (1978-2008). **Revista Brasileira de Climatologia**, v. 6. p. 151-168, 2010.

ALMEIDA, A. C.; ROCHA, B. R. P.; SOUZA, J. R. S.; SÁ, J. A.; PISSOLATO FILHO, J. A. Cloud-to-ground lightning observations over the Eastern Amazon region. **Atmospheric Research**. v. 117, p. 86-90. 2012.

ALMEIDA, A.C.; ROCHA, B. R. P.; MONTEIRO, J.H.A.; SOUZA, J. R. S.; VIEIRA, G.R.; SOUZA, E.B.; CUNHA, H.T.S.; SOUZA, W.A.P. Lightning characteristics associated to severe rainfall events, around Belém PA Brazil. In: INTERNATIONAL SYMPOSIUM ON LIGHTNING PROTECTION, 9, 2007, Foz do Iguaçu- SP [**Anais...**]. Foz do Iguaçu- SP: Institute of Electrotechnics and Energy of the University of São Paulo, v. 1. p. 57-60. Disponível em: <http://hidroweb.ana.gov.br/2007>. Acesso em 2012.

ALMEIDA, A. C.; SOUZA, J.R. S.; SÁ, J.A. S.; MONTEIRO, J.H. A.; ROCHA, B. R. P.; PISSOLATO FILHO, J. Características dos raios nuvem-solo na Amazônia Oriental: subsídios para proteção de sistemas elétricos. In: SIMPÓSIO BRASILEIRO DE SISTEMAS ELÉTRICOS, 3., 2010, Belém-PA. **Anais...** Belém-PA: UFPA, 2010.

ANAGNOSTOU, E.N.; CHRONIS.T. **The worth of long-range lightning observations on overland satellite rainfall estimation**. [S.l.]: Kluwer Academic Publishers, 2004.(Book Chapter).

CAREY, L. D., MURPHY, M. J., MCCORMICK, T. L., DEMETRIADES, N. W. S.. “Lightning location relative to storm structure in a leading-line, trailing-stratiform mesoscale convective system”, **J. Geophys. Res.**, v. 110, D03105, doi:10.1029/2004JD004371, 2005

CARVALHO, D. R. **Árvore de decisão/Algoritmo Genético para tratar o problema de pequenos disjuntos em classificação de dados**. 2005. 173f. Tese (Doutorado em Ciências em Engenharia Civil) – Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2005.

COHEN, J.C.P., SILVA DIAS, M.A.F., NOBRE, C.A. Environmental conditions associated with Amazonian squall lines: a case study. **Monthly Weather Review**, v.123, n. 11, p.3163-3174. 1995.

CORRÊA, L. V. **Análise dos efeitos térmicos da cidade de Belém utilizando imagens de satélite**. Belém, 2011. Dissertação (Mestrado em Ciências Ambientais) - Programa de Pós-Graduação. Universidade Federal do Pará, Belém, 2011.

DIAS, V. S. **Análise dos índices de instabilidade para previsão de formações convectivas severas para a região do Aeródromo do Galeão, Aeroporto Tom Jobim**. Rio de Janeiro, 2008. Instituto de Controle do Espaço Aéreo, Divisão de Ensino, Especialização em Meteorologia Aeronáutica, 2008. (Relatório de Pesquisa). Disponível em: <http://www.redemet.aer.mil.br/Artigos/inst_form_conect_sbgl.pdf> Acesso em: 26/06/2011.

FERREIRA, E.B. OLIVEIRA, M.S. **Introdução à estatística básica com R**. Lavras, 2008. 124f. Especialização (Gestão de empresas com Ênfase em Qualidade) - Universidade Federal de Lavras- MG, Lavras, 2008.

FERREIRA, N. J., REPELLI, A. C., ALVES, J. M. B., SOUZA, E. B. LOCZCIT – Um procedimento numérico para localização do eixo central da Zona de Convergência Intertropical no Atlântico Tropical. **Revista Brasileira de Meteorologia**, v. 20. n. 2, p. 159-164. 2005.

FISCH, G.; VENDRAME, I. F; HANAOKA, P. C. M. Variabilidade espacial da chuva durante o experimento LBA/TRMM 1999 na Amazônia. **Revista Acta Amazonica**. v. 37, n.4, p. 583-590. 2007.

FONSECA, J.M.M.R. **Indução de árvore de decisão**: histclss- proposta de um algoritmo não paramétrico. Lisboa, 1994. 151f. Dissertação (Mestrado) – Universidade Nova de Lisboa. Lisboa, 1994.

FONTANA, A; NALDI, M.C. **Estudo e comparação de métodos para estimação de números de grupos em problemas de agrupamento de dados**. Instituto de Ciências Matemáticas e de computação, São Carlos, 2009. (Relatório Técnico, n.340).

GARCIA, S.C. O uso de árvores de decisão na descoberta de conhecimento na área da saúde. In: SEMANA ACADÊMICA, 2000. Rio Grande do Sul. Atas... Universidade do Rio Grande do Sul, Rio Grande do Sul.2000.

GONÇALVES, A. R. **Otimização em ambientes dinâmicos com variáveis contínuas empregando algoritmos de estimação de distribuição**. 2011. 111f. Dissertação (Mestrado em Engenharia Elétrica) – Faculdade de Engenharia Elétrica e de Computação, Campinas, 2011.

GOMES, M. A. S. S. **Estudo dos relâmpagos na região Sudeste do Brasil em função das características geográficas**. São José dos Campos, 2002. 149. Dissertação (Mestrado em Geofísica Espacial) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2002.

GUNGLE, B.; KRIDER, E. P. (2006). Cloud-to-ground Lightning and Surface Rainfall in warm-season Florida Thunderstorms. **J. Geophys. Res.**, 111, D19203. 2006.

GRUPO DE ELETRICIDADE ATMOSFÉRICA (ELAT). **Descargas atmosféricas**. Disponível em : <<http://www.inpe.br/webelat/homepage/>>. Acessado em 12/12/2011.

HISER, H.W. Sferics and radar studies of South Florida thunder storms. **J.Applied. Met.** v.12, p. 4479-483. 1973.

HOREL, J.D.; HAHMANN, A.N.; GEISLER, J.E. **An investigation of the Annual Cycle of Convective Activity over the Tropical Americas**. Journal of Climate, 2(11), 1388 - 1403, 1989;

IAG/USP/STARNET. Disponível em:< <http://www.zeus.iag.usp.br/>> Acesso em Jan, 2012.

INCÊNDIOS. Disponível em: <<http://g1.globo.com/rio-de-janeiro/noticia/2012/12/raio-provoca-incendio-na-pedra-da-gavea-no-rio.html>>

KOUSKY, V.E.; KAYANO, M.T. A climatological study of the tropospheric circulation over the Amazon region. **Acta Amazonica**, v. 11, p. 743-758. 1981.

LEE, A.C.L.. The Limiting accuracy of long wavelength lightning flash location. **American Meteorological Society**, 6, p. 43 – 49. 1989.

LESSA, L.A.S., ROCHA, B.R.P., SOUZA, J.R.S., ALMEIDA, A.C., MONTEIRO, J. H.A. Lightning strikes along an electric energy transmission line in Belem-PA-Brazil. In: GROUND & 3D INTERNATIONAL CONFERENCE ON LIGHTNING PHYSICS AND EFFECTS, 2008, Florianopolis, Brasil. **Conference Proceedings**. Florianopolis, Brasil: [s,n.],2008. p. 512-514.2008.

LUDLAM, F.H. Cumulus and cumulonimbus convection. **Tellus**, v. 18, n. 4. p.687-698. 1966.

LIMA,W. F. A; MACHADO, L. A. T; MORALES, C. A; PINTO JR . A Estimativa de probabilidade de ocorrências de relâmpagos. 2006. In: CONGRESSO BRASILEIRO DE METEOROLOGIA, 14, Fortaleza, **Anais**, SBMET, CD ROM, 2006.

LINDEN, R. Técnicas de agrupamento. **Revista de Sistemas de Informação da FSMA**. n.4.p.18-36. 2009.

LOUREIRO, R. S. **Análise de uma linha de instabilidade costeira na região Leste do estado do Pará**. 2005. 111f. Dissertação (Mestrado em Ciências). Programa de Pós-Graduação em Meteorologia, Universidade Federal de Pelotas, Pelotas-RS, 2005.

MACGORMAN, D. R.; RUST, W. D.; KREHBIEL, P.; RISON, W.; BRUNING, E.; WIENS, K. The electrical structure of two supercell storms during STEPS. **Mon. Wea. Rev.**, 133, p. 2583–2607. 2005.

MADDOX, R. A. Mesoscale convective complexes. **Bulletin of American Meteorological Society**, v. 61, p. 1374-1387. 1980.

MOLION, L. C. B. **Amazonian rainfall and its variability**. In: HYDROLOGY and Water Management in the Humid Tropics, Cambridge University, 1993. p. 99-111.

MORALES,C.A; NEVES,J.R, ANSELMO.E; Sferics Timing and Ranging Network- STARNET: evaluation over South America. Rio de Janeiro, 2011. In: INTERNATIONAL CONFERENCE ON ATMOSPHERIC ELECTRICITY,14., Rio de Janeiro, Brazil, 2011.

MCMURDIE, L; HOUZE, R. A. Weather systems, chapter 3. In: WALLACE, J. M.;HOBBS, P. V. **Atmospheric Science**: an introductory survey. Second Edition, Boston: AcademicPress, 2006. p. 313-373.

NACCARATO, K. P. **Estudo de relâmpagos no Brasil com base na análise de desempenho do sistema de localização de tempestades**. 2001. 165f. São José dos Campos, 165. Dissertação (Mestrado em Geofísica Espacial) – Instituto Nacional de Pesquisas Espaciais, 2001. (INPE-8380-TDI/770).

NASCIMENTO, E. L. Previsão de tempestades severas utilizando-se parâmetros convectivos e modelos de mesoescala: Uma estratégia operacional adotável no Brasil? **Revista Brasileira de Meteorologia**, v.20, n.1, 121 – 140. 2005.

NECHET, D. Dias de trovoadas na Amazônia. In: CONGRESSO BRASILEIRO DE METEOROLOGIA, 8. E CONGRESSO LATINO-AMERICANO E IBÉRICO DE METEOROLOGIA, 2., 1994, Belo Horizonte. **Anais...** Belo Horizonte – MG: SBMET, 1994. v. 1, p. 184-187,

PEREIRA, C.S. **Elementos de sensores de eletricidade atmosférica**. 2010, Dissertação (Mestrado) - Universidade Federal do Pará, Instituto de Tecnologia, Programa de Pós-Graduação em Engenharia Elétrica, Belém, 2010.

PIATETSKY-SHAPIRO, G. Knowledge discovery in real databases: a report on the IJCAI-89 Workshop. **AI Magazine**, v. 11, N. 5, p. 68-70, Jan. 1991, Special issue.

PINTO JR. O. ; PINTO, I. R. C. A.; GIN, R.B.B.; MENDES JR., O. A coordinated study of a storm system over the South American continent 1. Weather information and quasi-dc stratospheric electric field data. **Journal of Geophysical Research**, v. 97, n. D16, p. 18195-18204, Nov.1992.

PINTO JR,O. Lighting in the tropics: a review. In: INTERNATIONAL SYMPOSIUM ON LIGHTING PROTECTION- SIPDA, 10., 2009, Curitiba, Brazil. **Anais...** Curitiba, Brazil: SIPDA, 2009.

PINTO JR,O; PINTO, I.R.C.A. **Tempestades e relâmpagos no Brasil**. Instituto Nacional de Pesquisas Espaciais- INPE- São José dos Campos- SP ,2000.

POMPEU, M. M; SOUZA,J.R.S; MOTA.M.A.S; RIBEIRO,W.M.N; ALMEIDA.A.C. Lighting observations during CAPE Experiment BARCA. In: INTERNATIONAL CONFERENCE ON ATMOSPHERIC ELECTRICITY- ICAE, 15.,2011, Rio de Janeiro, Brazil. **Proceedings...**2011.

POMPEU, M. M. **Relações entre raios e chuvas na Amazônia Oriental**. 2012. 124f. Dissertação (Mestrado em Ciências Ambientais) - Universidade Federal do Pará, Instituto de Geociências, Museu Paraense Emilio Goeldi e EMBRAPA, Programa de Pós-Graduação em Ciências Ambientais, Belém, 2012.

PORTAL ORM: Impactos socioambientais causados pela chuva e desligamento de energia Disponível em: <<http://noticias.orm.com.br/noticia.asp>> Acesso em 2013.

RAGSDALE. Disponível em: <<http://archive.ite.journal.informs.org/Vol3No3/Ragsdale/>> Acesso em setembro de 2011.

RADIOSSONDA. Disponível em: <<http://www2.ufersa.edu.br/portal/noticias/3385/>> Acesso em janeiro de 2012.

R DEVELOPMENT CORE TEAM. **R: A Language and Environment for Statistical Computing**. R Foundation for Statistical Computing, Vienna, Austria. 2005. Disponível em <<http://www.R-project.org>>. Acesso em janeiro 2012.

RENNÓ, N. O.; INGERSOL, A. P. Natural convection as a heat engine: a theory for CAPE. **J. Atmos. Sci.**, v. 53, p. 572-585. 1996.

RIBEIRO, W.M.N. Incidência de raios próximo à uma linha de transmissão da CELPA. In: SIMPÓSIO BRASILEIRO DE SISTEMAS ELÉTRICOS, 2010, **Anais....** SBSE, 2010.

RIBEIRO, W.M.N; SOUZA,J.R.S; ALMEIDA,A.C; SANTOS F.A.A; ROCHA, B.R.P. **Lighting and precipitations over different surface types in Eastern Amazonia**. In: XIV INTERNATIONAL CONFERENCE ON ATMOSPHERIC ELECTRICITY, Rio de Janeiro, Brazil(2011).

RIBEIRO, W.M.N.; SOUZA, J. R. S.; SILVA,R.M.B.; ALMEIDA,A.C.; SOUSA,J.R.A.; MONTEIRO,J.H.A.; ROCHA B. R. P.; CUNHA,H,T,S. Análise das distribuições horárias de Raios e Chuvas em Belém-PA. In: CONGRESSO BRASILEIRO DE METEOROLOGIA, 15., São Paulo. **Anais...** São Paulo: CBMET, 2008.

RIEHL, H. **Climate and weather in the Tropics**. London: Academic Press,1979. 611p

ROCHA, B.R.P.; ANDRADE JR, J.G.; RIBEIRO, W.M.N.; SANTOS, M.M.N.; SOUZA, J.R.S. Efeitos de descargas elétricas nuvem-solo no sistema telefônico de Belém no período 95-96. In: International conference on grounding and earthing, 1., 1988. Belo Horizonte-MG **Proceedings** v. 1, p. 89-92, Belo Horizonte - MG, 1998.

ROCHA, B. R. P.; SOUZA, J.R.S .;CONDE, C. L. R., NEVES, A. J. A.; ROCHA, A. O. **Efeitos de descargas elétricas no sistema de distribuição de Belém** – In: IV SIMOPA SEMINÁRIO INTEGRADO DE MANUTENÇÃO, OPERAÇÃO E ADMINISTRAÇÃO, 20-24 de out/96.

ROCHA, B. R. P.; SOUZA, J. R. S.; SILVA, I. M. O.; COSTA, J. M. Precipitation estimates from an electromagnetic lightning detector data in radiofrequencies. In: International Microwave and Optoelectronics Conference, 1997, Natal. **Proceedings...** Natal: [s.n.], 1997b. v.2, p. 655-659.

ROCHA, B. R. P.; SOUZA, J. R. S.; PEREIRA, C. S.; MOTA,G.V.; ROCHA, E. J. P; LOBATO, M. P.; LESSA, L.A.S.; MONTEIRO , J.H.A. Lightning characteristics time distribution over four locations in eastern Amazonia. In: International Symposium on Lightning Protection, 9., 2007, São Paulo. **Proceedings...** São Paulo: Institute of Electrotechnics and Energy of the University of Foz do Iguaçu - São Paulo, 2007. v. 1. p. 31-34. 2007.

- ROOHR, P.B., VONDER HAAR, T.H. A Comparative Analysis of the Temporal Variability of Lightning Observations and GOES Imagery. **Journal of Applied Meteorology.**, v. 33, p. 1271-1290. 1994.
- RPART. Pacote do Programa. Disponível: <<http://cran.r-project.org/web/packages/rpart/.pdf>>. Acesso em: jan. de 2012.
- RUTLEDGE, S. Middle latitude and tropical mesoscale convective systems. **Review of Geophysics.** Supplement, p. 88-97. 1991.
- SÁ, J.A.S; ALMEIDA,A.C; ROCHA, M.A.S; MOTA, SOUZA,J.R.S. Lighting forecast based on the hourly evolution of the convective available potential energy (cape).In: XIV INTERNATIONAL CONFERENCE ON ATMOSPHERIC ELECTRICITY, Rio de Janeiro, Brazil, 2011.
- SÁ, J.A.S; ROCHA, B; ALMEIDA, A; SOUZA, J.R. Recurrent self-organizing map for severe weather patterns recognition. **Recurrent Neural Networks and Soft Computing**, p 151- 174, Croatia, 2012.
- SANTOS, C.M.S; SILVA, F.R.; LEME, N.P. Ciclo anual de precipitação como função de índices de instabilidade termodinâmica e fluxos de energia estática: análises em maxaranguape-rn. **Revista Brasileira de Meteorologia**, v.26, n.2, 235 – 242. 2011.
- SANTOS, A. A. S. **Influência de sistemas meteorológicos de grande escala sobre a distribuição de raios na Amazônia Oriental, com aplicação no setor elétrico.**116 p. Dissertação (Mestrado em Ciências Ambientais), Instituto de Geociências, Universidade Federal do Pará, Museu Paraense Emilio Goeldi e EMBRAPA, Belém, 2010.
- SANTOS.A.P.P; SOUZA,J.R.S; SOUZA,E.B; CARMO,A.M.C; RIBEIRO.W.M.N; OLIVEIRA.R.A.J. **The climatic context of lightning storms, associated to electric systems power outages in Eastern Amazonia.** In: XIV INTERNATIONAL CONFERENCE ON ATMOSPHERIC ELECTRICITY, Rio de Janeiro, Brazil, 2011.
- SATELITE GOES, Disponível em < <http://satelite.cptec.inpe.br/> > acessado em 2013.
- SERRA, A. Trovoada e névoa seca. **Boletim geográfico.** V. 35, N. 255, p. 142-163, Rio de Janeiro - RJ, 1977.
- SILVA, C.C.S; COELHO,V.F. **Localização geográfica de raios usando o algoritmo de diferença do tempo de chegada.** In: I COLÓQUIO DE MATEMÁTICA DA REGIÃO NORDESTE,2011.
- SOUZA, S. C.B. **Planejamento de trajetória para um robô móvel com duas rodas utilizando um algoritmo A-Estrela modificado.** Dissertação (Mestrado em Engenharia Elétrica, Universidade Federal do Rio de Janeiro, (COPPE), Rio de Janeiro 2008.
- SORIANO, L. R., PABLO, F., DÍEZ, E. G. Relationship between convective precipitation and cloudto- ground lightning in the Iberian Peninsula. **Monthly Weather Review**, V.129, p. 2998-3003. 2001

SOUZA, J.R.S.; ROCHA, B.R.P.; CARRERA, G. T Lightning observation (and applications) around Belém, during the 1995-1998 period. In: INTERNATIONAL SYMPOSIUM ON LIGHTNING PROTECTION, 1999, São Paulo. **Proceedings...** São Paulo: [s.n.], 1999. v. 1, p. 17-21.

SOUZA, J.R.S.; ROCHA, B.R.P.; COSTA, J.M.; DUTRA, S.L.G. Características dos eventos de raios nuvem-solo e sua distribuição em área no Leste do Pará. In: CONGRESSO INTERNACIONAL DA SOCIEDADE BRASILEIRA DE GEOFÍSICA, 5., 1997, São Paulo. **Anais...** São Paulo - SP, 1997. v. 2, p. 1152-1155.

SOUZA, J.R.S.; ROCHA, E.J.P.; COHEN, J.C.P. **Avaliação dos Impactos Antropogênicos no Ciclo da Água na Amazônia.** Capítulo 4 In: Aragon L.E.; Clüsener-Godt, M. . Problemática do uso local e global da água da Amazônia. Org. EDIÇÕES UNESCO BRASIL, Brasília - DF, p. 69-94, 2003.

TAVARES, J. P. N. **Tempestades severas na região metropolitana de Belém:** Avaliação das condições termodinâmicas e impactos socioeconômicos. Dissertação (Mestrado em Ciências Ambientais). Universidade do Pará, Instituto de Geociências, Museu Paraense Emílio Goeldi e EMBRAPA, Belém, 2009.

TEIXEIRA, V.A; MOTA, G.V; SOUZA, J.R.S; RIBEIRO, W.M.N. **Lighting density distribution over eastern Amazonia, from 10-years data of observations by TRMM-LIS.** In: XIV International Conference on Atmospheric electricity, Rio de Janeiro, Brazil, 2011.

UMAN, M. A. **The lightning discharge.** [S.l.]: Academic Press, 1987. 377 p.

UMAN, M. A.; KRIDER; PHILIP. Natural and Artificially Initiated Lightning. **Science**, USA, V. 246, n. 4929, p. 457-464, USA, 1989.

VIANELLO, R. L.; ALVES, A. R. Meteorologia básica e aplicações. Viçosa: UFV – **Imprensa Universitária.** 337 p. 1991.

VISACRO FILHO, S. **Descargas Atmosféricas:** Uma abordagem de engenharia. São Paulo: Artliber, 2005, 272p.

VITORINO, M.A. SILVA, M.E.S.; ALVES, J.M.B. Classificação de sistemas convectivos de Mesoescala no setor Norte do Nordeste Brasileiro. **Revista Brasileira de Meteorologia**, v.12, p. 21-32. 1997.

XAVIER, T. M. B., XAVIER, A. F. S., SILVA DIAS, P. L., SILVA DIAS, M. A. F. A Zona de Convergência Intertropical – ZCIT e suas relações com a chuva no Ceará (1964-98). **Revista Brasileira de Meteorologia**, v. 15 n. 1, p. 27-43. 2000.

WALLACE, J. M.; HOBBS, P.V. **Atmospheric science:** An introductory survey. San Diego: Academic Press, 1977. 467p.

WILLIMAS, E.; RENNÓ, N. An analysis of the conditional instability of the tropical atmosphere. **Monthly Weather Review**, Boston, v.121, p.21-36. 1993.

ANEXOS

ANEXO A – COMANDOS PARA CONSTRUÇÃO DA ÁRVORE DE DECISÃO NO R

```
dadostotal <- read.table("C:/Users/Camila/Desktop/tese/dadostotal.txt", header=T,
quote="\")
```

```
View(dadostotal)
```

```
attach(dadostotal)
```

```
Lifted<-(Lifted- min(Lifted))/(max(Lifted)-min(Lifted))
```

```
Sweat<-(Sweat- min(Sweat))/(max(Sweat)-min(Sweat))
```

```
K_ind<-(K_ind- min(K_ind))/(max(K_ind)-min(K_ind))
```

```
Ttot<-(Ttot- min(Ttot))/(max(Ttot)-min(Ttot))
```

```
CAPE<-(CAPE- min(CAPE))/(max(CAPE)-min(CAPE))
```

```
PWAT<-(PWAT- min(PWAT))/(max(PWAT)-min(PWAT))
```

```
Classe <- as.character(Classe)
```

Variáveis
normalizadas

Leitura dos dados

```
dadostotal <- data.frame(Lifted,Sweat,K_ind,Ttot,CAPE,PWAT,Classe)
```

```
library(rpart) (2012)
```

Pacote que contém as funções para obter árvores de decisão.

```
arvore<-rpart(Classe ~ .,dadostotal)
```

Fórmula que indica a forma funcional do modelo que o R deve ter. Nesse caso é um modelo pra prever a variável “classe” usando todas as outras.

```
arvore
```

```
n= 476
```

```
node), split, n, loss, yval, (yprob)
* denotes terminal node
```

```
1) root 476 310 3 (0.315126050 0.336134454 0.348739496)
 2) CAPE< 0.07896504 293 142 2 (0.460750853 0.515358362 0.023890785)
   4) PWAT< 0.4585833 142 13 1 (0.908450704 0.091549296 0.000000000)
     8) K_ind< 0.1243978 120 1 1 (0.991666667 0.008333333 0.000000000)
    *
     9) K_ind>=0.1243978 22 10 2 (0.454545455 0.545454545 0.000000000)
     18) Sweat< 0.7075493 12 3 1 (0.750000000 0.250000000 0.000000000)
    *
```

```

19) Sweat>=0.7075493 10 1 2 (0.100000000 0.900000000 0.000000000)
*
5) PWAT>=0.4585833 151 13 2 (0.039735099 0.913907285 0.046357616)
10) Lifted>=0.3305556 143 7 2 (0.041958042 0.951048951 0.006993007)
*
11) Lifted< 0.3305556 8 2 3 (0.000000000 0.250000000 0.750000000) *
3) CAPE>=0.07896504 183 24 3 (0.081967213 0.049180328 0.868852459)
6) Ttot< 0.6219136 14 6 1 (0.571428571 0.285714286 0.142857143) *
7) Ttot>=0.6219136 169 12 3 (0.041420118 0.029585799 0.928994083) *

```

```
plot(arvore)
```

```
text(arvore)
```

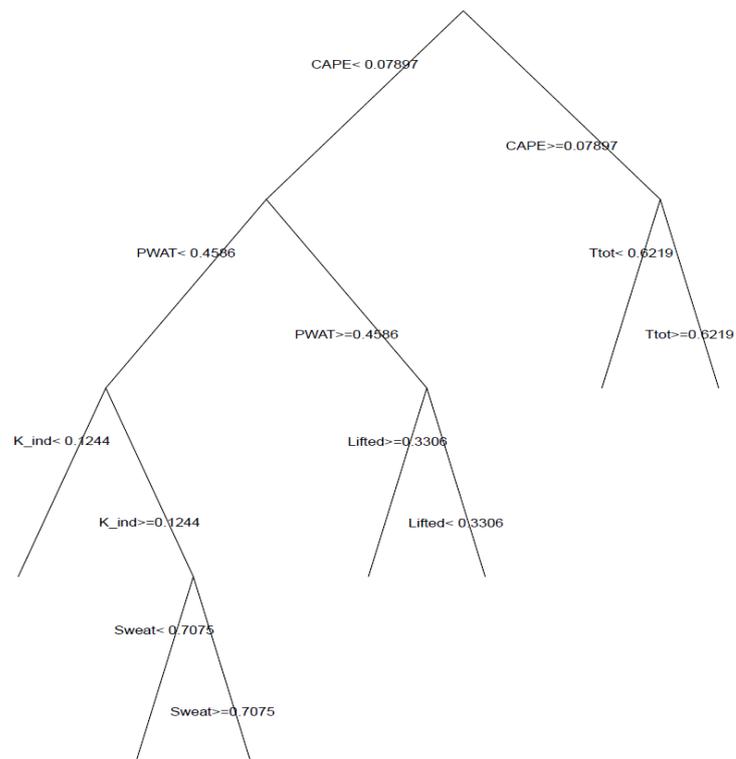
```
mostra.arvore <- function(arvore){
```

```
plot(arvore,uniform=T,branch=0)
```

```
text(arvore,digits=3,cex=0.65, font=8,pretty=0,fancy=T,fwidth=0,fheight=0)}
```

```
mostra.arvore(arvore)
```

Forma gráfica de
uma árvore de
decisão.



ANEXO B – AVALIAÇÃO DA QUALIDADE DO MODELO

```
amostra.modelo <- sample(1:nrow(dadostotal),as.integer(0.7*nrow(dadostotal)))
```

```
dadostotal.modelo <- dadostotal[amostra.modelo,]
```

```
dadostotal.teste <- dadostotal[-amostra.modelo,]
```

```
arvore <- rpart(Classe ~.,dadostotal.modelo)
```

```
arvore
```

Testa fazendo
comparação do
modelo com as
decisões das
classes.

Cria o
modelo de
decisão a
partir de
70% dos

```
n= 476
```

```
node), split, n, loss, yval, (yprob)  
* denotes terminal node
```

```
1) root 476 310 3 (0.315126050 0.336134454 0.348739496)  
 2) CAPE< 0.07896504 293 142 2 (0.460750853 0.515358362 0.023890785)  
   4) PWAT< 0.4585833 142 13 1 (0.908450704 0.091549296 0.000000000)  
    8) K_ind< 0.1243978 120 1 1 (0.991666667 0.008333333 0.000000000)  
*  
   9) K_ind>=0.1243978 22 10 2 (0.454545455 0.545454545 0.000000000)  
    18) Sweat< 0.7075493 12 3 1 (0.750000000 0.250000000 0.000000000)  
*  
   19) Sweat>=0.7075493 10 1 2 (0.100000000 0.900000000 0.000000000)  
*  
   5) PWAT>=0.4585833 151 13 2 (0.039735099 0.913907285 0.046357616)  
    10) Lifted>=0.3305556 143 7 2 (0.041958042 0.951048951 0.006993007)  
*  
   11) Lifted< 0.3305556 8 2 3 (0.000000000 0.250000000 0.750000000) *  
  3) CAPE>=0.07896504 183 24 3 (0.081967213 0.049180328 0.868852459)  
   6) Ttot< 0.6219136 14 6 1 (0.571428571 0.285714286 0.142857143) *  
   7) Ttot>=0.6219136 169 12 3 (0.041420118 0.029585799 0.928994083) *
```

```
previsões.modelo <- predict(arvore,dadostotal.teste)
```

```
previsões.modelo
```

```
      1      2      3  
2  0.04142012 0.029585799 0.928994083  
5  0.04142012 0.029585799 0.928994083  
7  0.04142012 0.029585799 0.928994083  
8  0.04142012 0.029585799 0.928994083  
10 0.04142012 0.029585799 0.928994083  
11 0.04142012 0.029585799 0.928994083  
12 0.04142012 0.029585799 0.928994083  
18 0.04195804 0.951048951 0.006993007  
20 0.04142012 0.029585799 0.928994083  
23 0.04195804 0.951048951 0.006993007  
27 0.04142012 0.029585799 0.928994083  
31 0.04142012 0.029585799 0.928994083  
32 0.04142012 0.029585799 0.928994083  
36 0.04142012 0.029585799 0.928994083  
37 0.04195804 0.951048951 0.006993007  
42 0.04195804 0.951048951 0.006993007  
44 0.04195804 0.951048951 0.006993007  
46 0.04142012 0.029585799 0.928994083
```

A árvore prevê que existe 93% de probabilidade de ser um caso de classe 3, 3% para classe 2 e 4% para classe 1.

48	0.04195804	0.951048951	0.006993007
51	0.04195804	0.951048951	0.006993007
54	0.04195804	0.951048951	0.006993007
56	0.04142012	0.029585799	0.928994083
57	0.04142012	0.029585799	0.928994083
61	0.04195804	0.951048951	0.006993007
62	0.04195804	0.951048951	0.006993007
65	0.00000000	0.250000000	0.750000000
69	0.04142012	0.029585799	0.928994083
76	0.04195804	0.951048951	0.006993007
78	0.04195804	0.951048951	0.006993007
79	0.04195804	0.951048951	0.006993007
82	0.04195804	0.951048951	0.006993007
92	0.04142012	0.029585799	0.928994083
94	0.04195804	0.951048951	0.006993007
95	0.04142012	0.029585799	0.928994083
96	0.04142012	0.029585799	0.928994083
103	0.04195804	0.951048951	0.006993007
104	0.04142012	0.029585799	0.928994083
105	0.57142857	0.285714286	0.142857143
106	0.04142012	0.029585799	0.928994083
112	0.04195804	0.951048951	0.006993007
124	0.04195804	0.951048951	0.006993007
126	0.04195804	0.951048951	0.006993007
136	0.04195804	0.951048951	0.006993007
139	0.04195804	0.951048951	0.006993007
140	0.04142012	0.029585799	0.928994083
141	0.04142012	0.029585799	0.928994083
142	0.04195804	0.951048951	0.006993007
148	0.04195804	0.951048951	0.006993007
153	0.99166667	0.008333333	0.000000000
155	0.04142012	0.029585799	0.928994083
158	0.04142012	0.029585799	0.928994083
159	0.04195804	0.951048951	0.006993007
174	0.04195804	0.951048951	0.006993007
176	0.99166667	0.008333333	0.000000000
177	0.99166667	0.008333333	0.000000000
179	0.75000000	0.250000000	0.000000000
183	0.99166667	0.008333333	0.000000000
188	0.10000000	0.900000000	0.000000000
189	0.99166667	0.008333333	0.000000000
194	0.99166667	0.008333333	0.000000000
198	0.99166667	0.008333333	0.000000000
204	0.99166667	0.008333333	0.000000000
209	0.99166667	0.008333333	0.000000000
212	0.99166667	0.008333333	0.000000000
213	0.99166667	0.008333333	0.000000000
215	0.99166667	0.008333333	0.000000000
220	0.04195804	0.951048951	0.006993007
221	0.75000000	0.250000000	0.000000000
227	0.99166667	0.008333333	0.000000000
229	0.99166667	0.008333333	0.000000000
235	0.99166667	0.008333333	0.000000000
239	0.04195804	0.951048951	0.006993007
240	0.04195804	0.951048951	0.006993007
241	0.04195804	0.951048951	0.006993007
244	0.99166667	0.008333333	0.000000000
247	0.10000000	0.900000000	0.000000000
275	0.99166667	0.008333333	0.000000000
278	0.99166667	0.008333333	0.000000000
279	0.99166667	0.008333333	0.000000000
282	0.99166667	0.008333333	0.000000000
283	0.04142012	0.029585799	0.928994083
288	0.04195804	0.951048951	0.006993007
294	0.99166667	0.008333333	0.000000000
295	0.99166667	0.008333333	0.000000000
299	0.99166667	0.008333333	0.000000000
301	0.04195804	0.951048951	0.006993007
302	0.04195804	0.951048951	0.006993007
317	0.04142012	0.029585799	0.928994083

321	0.04142012	0.029585799	0.928994083
322	0.04195804	0.951048951	0.006993007
326	0.00000000	0.250000000	0.750000000
328	0.04142012	0.029585799	0.928994083
332	0.04142012	0.029585799	0.928994083
335	0.04142012	0.029585799	0.928994083
336	0.04142012	0.029585799	0.928994083
338	0.04142012	0.029585799	0.928994083
339	0.04195804	0.951048951	0.006993007
353	0.04142012	0.029585799	0.928994083
354	0.04142012	0.029585799	0.928994083
355	0.04142012	0.029585799	0.928994083
356	0.04195804	0.951048951	0.006993007
357	0.04195804	0.951048951	0.006993007
361	0.04142012	0.029585799	0.928994083
364	0.04195804	0.951048951	0.006993007
373	0.04142012	0.029585799	0.928994083
374	0.04142012	0.029585799	0.928994083
376	0.04142012	0.029585799	0.928994083
377	0.04195804	0.951048951	0.006993007
378	0.04195804	0.951048951	0.006993007
379	0.04142012	0.029585799	0.928994083
380	0.04195804	0.951048951	0.006993007
381	0.04195804	0.951048951	0.006993007
382	0.04142012	0.029585799	0.928994083
383	0.04195804	0.951048951	0.006993007
384	0.57142857	0.285714286	0.142857143
394	0.04142012	0.029585799	0.928994083
397	0.04142012	0.029585799	0.928994083
401	0.04195804	0.951048951	0.006993007
403	0.99166667	0.008333333	0.000000000
405	0.04142012	0.029585799	0.928994083
406	0.04142012	0.029585799	0.928994083
410	0.04195804	0.951048951	0.006993007
412	0.04195804	0.951048951	0.006993007
417	0.04142012	0.029585799	0.928994083
421	0.04195804	0.951048951	0.006993007
433	0.04142012	0.029585799	0.928994083
436	0.04142012	0.029585799	0.928994083
440	0.04195804	0.951048951	0.006993007
441	0.04195804	0.951048951	0.006993007
448	0.04142012	0.029585799	0.928994083
449	0.04195804	0.951048951	0.006993007
453	0.04195804	0.951048951	0.006993007
457	0.04195804	0.951048951	0.006993007
467	0.04195804	0.951048951	0.006993007
469	0.04195804	0.951048951	0.006993007
474	0.04142012	0.029585799	0.928994083
476	0.04142012	0.029585799	0.928994083
479	0.04142012	0.029585799	0.928994083
480	0.04142012	0.029585799	0.928994083
484	0.04142012	0.029585799	0.928994083
485	0.04142012	0.029585799	0.928994083
488	0.04195804	0.951048951	0.006993007
489	0.99166667	0.008333333	0.000000000
491	0.04142012	0.029585799	0.928994083
494	0.99166667	0.008333333	0.000000000
495	0.00000000	0.250000000	0.750000000
498	0.99166667	0.008333333	0.000000000
500	0.99166667	0.008333333	0.000000000
501	0.04195804	0.951048951	0.006993007
504	0.04195804	0.951048951	0.006993007
505	0.04142012	0.029585799	0.928994083
508	0.04142012	0.029585799	0.928994083
511	0.99166667	0.008333333	0.000000000
512	0.04142012	0.029585799	0.928994083
513	0.04195804	0.951048951	0.006993007
514	0.04195804	0.951048951	0.006993007
518	0.04195804	0.951048951	0.006993007
522	0.75000000	0.250000000	0.000000000

```

523 0.04142012 0.029585799 0.928994083
524 0.04195804 0.951048951 0.006993007
527 0.04195804 0.951048951 0.006993007
533 0.57142857 0.285714286 0.142857143
534 0.75000000 0.250000000 0.000000000
538 0.04195804 0.951048951 0.006993007
539 0.04195804 0.951048951 0.006993007
540 0.99166667 0.008333333 0.000000000
547 0.99166667 0.008333333 0.000000000
550 0.99166667 0.008333333 0.000000000
551 0.99166667 0.008333333 0.000000000
552 0.04195804 0.951048951 0.006993007
555 0.04142012 0.029585799 0.928994083
558 0.99166667 0.008333333 0.000000000
563 0.99166667 0.008333333 0.000000000
567 0.75000000 0.250000000 0.000000000
569 0.99166667 0.008333333 0.000000000
571 0.99166667 0.008333333 0.000000000
572 0.99166667 0.008333333 0.000000000
577 0.99166667 0.008333333 0.000000000
580 0.04142012 0.029585799 0.928994083
586 0.75000000 0.250000000 0.000000000
588 0.04142012 0.029585799 0.928994083
597 0.57142857 0.285714286 0.142857143
600 0.04142012 0.029585799 0.928994083
602 0.99166667 0.008333333 0.000000000
606 0.04195804 0.951048951 0.006993007
614 0.04142012 0.029585799 0.928994083
615 0.04142012 0.029585799 0.928994083
616 0.10000000 0.900000000 0.000000000
617 0.57142857 0.285714286 0.142857143
619 0.04142012 0.029585799 0.928994083
626 0.99166667 0.008333333 0.000000000
627 0.99166667 0.008333333 0.000000000
639 0.04142012 0.029585799 0.928994083
645 0.04142012 0.029585799 0.928994083
646 0.04142012 0.029585799 0.928994083
656 0.04142012 0.029585799 0.928994083
661 0.04142012 0.029585799 0.928994083
663 0.04142012 0.029585799 0.928994083
665 0.04195804 0.951048951 0.006993007
668 0.99166667 0.008333333 0.000000000
670 0.04142012 0.029585799 0.928994083
671 0.04142012 0.029585799 0.928994083
673 0.04142012 0.029585799 0.928994083
677 0.04142012 0.029585799 0.928994083
678 0.04142012 0.029585799 0.928994083

```

Uma vez que usamos um conjunto de dados diferente para obter o modelo, a árvore obtida é diferente da anterior.

```
previsões.modelo <- predict(arvore,dadostotal.teste,type='class')
```

```
previsões.modelo
```

```

> previsões.modelo
 2  5  7  8 10 11 12 18 20 23 27 31 32 36 37
 3  3  3  3  3  3  3  2  3  2  3  3  3  3  2
42 44 46 48 51 54 56 57 61 62 65 69 76 78 79
 2  2  3  2  2  2  3  3  2  2  3  3  2  2  2
82 92 94 95 96 103 104 105 106 112 124 126 136 139 140
 2  3  2  3  3  2  3  1  3  2  2  2  2  2  3
141 142 148 153 155 158 159 174 176 177 179 183 188 189 194

```

3	2	2	1	3	3	2	2	1	1	1	1	2	1	1
198	204	209	212	213	215	220	221	227	229	235	239	240	241	244
1	1	1	1	1	1	2	1	1	1	1	2	2	2	1
247	275	278	279	282	283	288	294	295	299	301	302	317	321	322
2	1	1	1	1	3	2	1	1	1	2	2	3	3	2
326	328	332	335	336	338	339	353	354	355	356	357	361	364	373
3	3	3	3	3	3	2	3	3	3	2	2	3	2	3
374	376	377	378	379	380	381	382	383	384	394	397	401	403	405
3	3	2	2	3	2	2	3	2	1	3	3	2	1	3
406	410	412	417	421	433	436	440	441	448	449	453	457	467	469
3	2	2	3	2	3	3	2	2	3	2	2	2	2	2
474	476	479	480	484	485	488	489	491	494	495	498	500	501	504
3	3	3	3	3	3	2	1	3	1	3	1	1	2	2
505	508	511	512	513	514	518	522	523	524	527	533	534	538	539
3	3	1	3	2	2	2	1	3	2	2	1	1	2	2
540	547	550	551	552	555	558	563	567	569	571	572	577	580	586
1	1	1	1	2	3	1	1	1	1	1	1	1	3	1
588	597	600	602	606	614	615	616	617	619	626	627	639	645	646
3	1	3	1	2	3	3	2	1	3	1	1	3	3	3
656	661	663	665	668	670	671	673	677	678					
3	3	3	2	1	3	3	3	3	3					

```
table(dadostotal.teste$Classe,previsões.modelo)
```

```
previsões.modelo
  1  2  3
1 51  3  2
2  3 65  3
3  0  1 77
```

Matrix de confusão

```
m.conf <- table(dadostotal.teste$Classe,previsões.modelo)
```

```
perc.erro<100*(m.conf[1,2]+m.conf[1,3]+m.conf[2,1]+m.conf[2,3]+m.conf[3,1]+m.conf[3,2]
)/ sum(m.conf)
```

```
print(perc.erro)
```

```
[1] 5.853659
```

```
print(100-perc.erro)
```

```
[1] 94.14634
```

Nos 205 casos de Teste, A árvore acerta 94% dos casos.

ANEXO C – CLASSIFICAÇÃO E ESCOLHA DO MODELO

The screenshot displays the RStudio interface. The console on the left shows the execution of R commands and the resulting decision tree structure. The plot on the right visualizes this tree with nodes and splits.

Console Output:

```
R version 3.0.1 (2013-05-16) -- "Good Sport"
Copyright (C) 2013 The R Foundation for Statistical Computing
Platform: i386-w64-mingw32/i386 (32-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> load("~/workspace1.RData")
> source("final.r")
The following object is masked _by_ '.GlobalEnv':

  CAPE, Classe, K_ind, Lifted, PWAT, Sweat, Ttot
[1] 12.68293
[1] 87.31707
> mostra.arvore(arvore)
> |
```

Decision Tree Structure:

- Root Node: Split on $CAPE < 0.07838$
 - Left Branch: Split on $PWAT < 0.4341$
 - Left Leaf: 1
 - Right Branch: Split on $PWAT >= 0.4341$
 - Left Branch: Split on $K_ind < 0.1098$
 - Left Leaf: 1
 - Right Branch: Split on $K_ind >= 0.1098$
 - Right Leaf: 2
 - Right Leaf: 3
 - Right Leaf: 3

The plot area on the right shows this decision tree with nodes and splits labeled as above. The leaves are numbered 1, 2, and 3.

The taskbar at the bottom shows the system tray with the date 21/08/2013 and time 13:25.

RStudio

File Edit Code View Plots Session Project Build Tools Help

Go to file/function Project: (None)

Source Workspace History

Console ~/ ↶

```
R version 3.0.1 (2013-05-16) -- "Good Sport"
Copyright (C) 2013 The R Foundation for Statistical Computing
Platform: i386-w64-mingw32/i386 (32-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> load("~/workspace1.RData")
> source("final.r")
The following object is masked by_ .GlobalEnv:

  CAPE, Classe, K_ind, Lifted, PWAT, Sweat, Ttot
[1] 12.68293
[1] 87.31707
> mostra.arvore(arvore)
> load("~/workspace2.RData")
> source("final.r")
The following object is masked by_ .GlobalEnv:

  CAPE, Classe, K_ind, Lifted, PWAT, Sweat, Ttot
The following object is masked from dadostotal (position 4):

  CAPE, Classe, K_ind, Lifted, PWAT, Sweat, Ttot
[1] 6.829268
[1] 93.17073
> mostra.arvore(arvore)
>
```

Files Plots Packages Help

Zoom Export Clear All

```

graph TD
    Root(( )) --- L1[CAPE < 0.07838]
    Root --- R1[CAPE >= 0.07838]
    L1 --- L2[PWAT < 0.4587]
    L1 --- L3[PWAT >= 0.4587]
    R1 --- R2[Ttot < 0.6219]
    R1 --- R3[Ttot >= 0.6219]
    L2 --- L4[K_ind < 0.1246]
    L2 --- L5[K_ind >= 0.1246]
    R2 --- R4[Sweat < 0.7073]
    R2 --- R5[Sweat >= 0.7073]
    L4 --- L6[1]
    L5 --- L7[2]
    R4 --- R6[1]
    R5 --- R7[2]
  
```

Windows taskbar: PT 13:25 21/08/2013

RStudio

File Edit Code View Plots Session Project Build Tools Help

Go to file/function Project: (None)

Source Workspace History

Files Plots Packages Help

Zoom Export Clear All

Console ~/

```
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> load("~/workspace1.RData")
> source("final.r")
The following object is masked _by_ .GlobalEnv:

  CAPE, Classe, K_ind, Lifted, PWAT, Sweat, Ttot
[1] 12.68293
[1] 87.31707
> mostra.arvore(arvore)
> load("~/workspace2.RData")
> source("final.r")
The following object is masked _by_ .GlobalEnv:

  CAPE, Classe, K_ind, Lifted, PWAT, Sweat, Ttot
The following object is masked from dadostotal (position 4):

  CAPE, Classe, K_ind, Lifted, PWAT, Sweat, Ttot
[1] 6.829268
[1] 93.17073
> mostra.arvore(arvore)
> load("~/workspace3.RData")
> source("final.r")
The following object is masked _by_ .GlobalEnv:

  CAPE, Classe, K_ind, Lifted, PWAT, Sweat, Ttot
The following object is masked from dadostotal (position 3):

  CAPE, Classe, K_ind, Lifted, PWAT, Sweat, Ttot
The following object is masked from dadostotal (position 5):

  CAPE, Classe, K_ind, Lifted, PWAT, Sweat, Ttot
[1] 13.17073
[1] 86.82927
> mostra.arvore(arvore)
> |
```

Decision tree structure:

- Root Node: CAPE < 0.07763
 - Left Branch: PWAT < 0.464
 - Left-Left Branch: K_ind < 8.1237
 - Left-Left-Left Branch: Sweat < 0.7073
 - Left-Left-Right Branch: K_ind >= 8.1237
 - Right Branch: PWAT >= 0.464
- Right Branch: PWAT < 0.413
 - Right-Right Branch: PWAT >= 0.413

Windows Taskbar: PT, 13:26, 21/08/2013

RStudio

File Edit Code View Plots Session Project Build Tools Help

Go to file/function Project: (None)

Source Workspace History

Console

```

The following object is masked from dadostotal (position 4):
  CAPE, Classe, K_ind, Lifted, PWAT, Sweat, Ttot
[1] 6.829268
[1] 93.17073
> mostra.arvore(arvore)
> load("~/workspace3.RData")
> source("final.r")
The following object is masked by_ .GlobalEnv:
  CAPE, Classe, K_ind, Lifted, PWAT, Sweat, Ttot
The following object is masked from dadostotal (position 3):
  CAPE, Classe, K_ind, Lifted, PWAT, Sweat, Ttot
The following object is masked from dadostotal (position 5):
  CAPE, Classe, K_ind, Lifted, PWAT, Sweat, Ttot
[1] 13.17073
[1] 86.82927
> mostra.arvore(arvore)
> load("~/workspace4.RData")
> source("final.r")
The following object is masked by_ .GlobalEnv:
  CAPE, Classe, K_ind, Lifted, PWAT, Sweat, Ttot
The following object is masked from dadostotal (position 3):
  CAPE, Classe, K_ind, Lifted, PWAT, Sweat, Ttot
The following object is masked from dadostotal (position 4):
  CAPE, Classe, K_ind, Lifted, PWAT, Sweat, Ttot
The following object is masked from dadostotal (position 6):
  CAPE, Classe, K_ind, Lifted, PWAT, Sweat, Ttot
[1] 9.756098
[1] 90.2439
> mostra.arvore(arvore)
>

```

Files Plots Packages Help

Zoom Export Clear All

```

graph TD
    Root(( )) --- Node1[CAPE < 0.00003]
    Node1 --- Node2[PWAT < 0.4586]
    Node1 --- Node3[Ttot < 0.6173]
    Node2 --- Node4[K_ind < 0.1233]
    Node2 --- Node5[PWAT >= 0.4586]
    Node3 --- Node6[Ttot < 0.6173]
    Node3 --- Node7[Ttot >= 0.6173]
    Node4 --- Node8[Sweat < 0.7443]
    Node4 --- Node9[K_ind >= 0.1233]
    Node8 --- Node10[Sweat >= 0.7443]
    Node8 --- Node11[Sweat < 0.7443]

```

PT 13:27 21/08/2013

RStudio

File Edit Code View Plots Session Project Build Tools Help

Go to file/function Project: (None)

Source Workspace History

Files Plots Packages Help

Zoom Export Clear All

```

> load("~/workspace4.RData")
> source("final.r")
The following object is masked by_ .GlobalEnv:

  CAPE, Classe, K_ind, Lifted, PWAT, Sweat, Ttot
The following object is masked from dadostotal (position 3):

  CAPE, Classe, K_ind, Lifted, PWAT, Sweat, Ttot
The following object is masked from dadostotal (position 4):

  CAPE, Classe, K_ind, Lifted, PWAT, Sweat, Ttot
The following object is masked from dadostotal (position 6):

  CAPE, Classe, K_ind, Lifted, PWAT, Sweat, Ttot
[1] 9.756098
[1] 90.2439
> mostra.arvore(arvore)
> load("~/workspace5.RData")
> source("final.r")
The following object is masked by_ .GlobalEnv:

  CAPE, Classe, K_ind, Lifted, PWAT, Sweat, Ttot
The following object is masked from dadostotal (position 3):

  CAPE, Classe, K_ind, Lifted, PWAT, Sweat, Ttot
The following object is masked from dadostotal (position 4):

  CAPE, Classe, K_ind, Lifted, PWAT, Sweat, Ttot
The following object is masked from dadostotal (position 5):

  CAPE, Classe, K_ind, Lifted, PWAT, Sweat, Ttot
The following object is masked from dadostotal (position 7):

  CAPE, Classe, K_ind, Lifted, PWAT, Sweat, Ttot
[1] 12.68293
[1] 87.31707
> mostra.arvore(arvore)
> |

```

Decision tree structure:

- Root Node: CAPE < 0.00007
 - Left Branch: PWAT < 4.506
 - Left Sub-branch: K_ind < 1255
 - Left: 1
 - Right: K_ind >= 1255
 - Left: Ttot < 0.6451
 - Left: 1
 - Right: Ttot >= 0.6451
 - Left: 1
 - Right: 2
 - Right Sub-branch: PWAT >= 4.506
 - Left: 2
 - Right: 1
 - Right Branch: Ttot < 0.6173
 - Left: 1
 - Right: Ttot >= 0.6173
 - Left: 1
 - Right: 2

RStudio

File Edit Code View Plots Session Project Build Tools Help

Go to file/function Project: (None)

dadostotal * final.r * 681 observations of 7 variables

	Lifted	Sweat	K_ind	Ttot	CAPE	PWAT	Classe
1	-4.64	199.79	32.50	43.20	2211.10	52.70	3
2	-3.37	263.22	26.90	44.00	1335.70	43.80	3
3	-3.91	169.81	21.50	42.20	741.13	39.27	1
4	-2.03	187.22	20.10	40.00	719.37	46.28	1
5	-4.82	232.82	34.10	42.50	2600.70	51.16	3
6	-4.49	195.60	34.50	43.00	1698.20	54.31	3
7	-2.44	203.39	30.80	41.90	1405.60	49.94	3

Workspace History

Files Plots Packages Help

Zoom Export Clear All

```

graph TD
    Root[ ] --- Node1[CAPE < 0.07897]
    Node1 --- Node2[PWAT < 0.4586]
    Node1 --- Node3[Ttot < 0.6219]
    Node2 --- Node4[K_ind < 0.1244]
    Node2 --- Node5[Lifted < 0.3306]
    Node3 --- Node6[Ttot >= 0.6219]
    Node4 --- Node7[Sweat < 0.7073]
    Node5 --- Node8[Lifted < 0.3306]
    Node7 --- Node9[Sweat >= 0.7073]
  
```

Console ~/

```

> mostra.arvore(arvore)
> mostra.arvore(arvore)
> load("~/workspace6.RData")
> source("final.r")
The following object is masked _by_ .GlobalEnv:
  CAPE, Classe, K_ind, Lifted, PWAT, Sweat, Ttot
The following object is masked from dadostotal (position 3):
  CAPE, Classe, K_ind, Lifted, PWAT, Sweat, Ttot
The following object is masked from dadostotal (position 4):
  CAPE, Classe, K_ind, Lifted, PWAT, Sweat, Ttot
The following object is masked from dadostotal (position 5):
  CAPE, Classe, K_ind, Lifted, PWAT, Sweat, Ttot
The following object is masked from dadostotal (position 6):
  CAPE, Classe, K_ind, Lifted, PWAT, Sweat, Ttot
The following object is masked from dadostotal (position 8):
  CAPE, Classe, K_ind, Lifted, PWAT, Sweat, Ttot
[1] 5.853659
[1] 94.14634
> mostra.arvore(arvore)
>
  
```

Windows taskbar: PT 13:29 21/08/2013

RStudio

File Edit Code View Plots Session Project Build Tools Help

Go to file/function Project: (None)

Source Workspace History

Console ~/\

```

CAPE, Classe, K_ind, Lifted, PWAT, Sweat, Ttot
The following object is masked from dadostotal (position 6):

CAPE, Classe, K_ind, Lifted, PWAT, Sweat, Ttot
The following object is masked from dadostotal (position 8):

CAPE, Classe, K_ind, Lifted, PWAT, Sweat, Ttot
[1] 5.853659
[1] 94.14634
> mostra.arvore(arvore)
> load("~/workspace7.RData")
> source("final.r")
The following object is masked by_ .globalEnv:

CAPE, Classe, K_ind, Lifted, PWAT, Sweat, Ttot
The following object is masked from dadostotal (position 3):

CAPE, Classe, K_ind, Lifted, PWAT, Sweat, Ttot
The following object is masked from dadostotal (position 4):

CAPE, Classe, K_ind, Lifted, PWAT, Sweat, Ttot
The following object is masked from dadostotal (position 5):

CAPE, Classe, K_ind, Lifted, PWAT, Sweat, Ttot
The following object is masked from dadostotal (position 6):

CAPE, Classe, K_ind, Lifted, PWAT, Sweat, Ttot
The following object is masked from dadostotal (position 7):

CAPE, Classe, K_ind, Lifted, PWAT, Sweat, Ttot
The following object is masked from dadostotal (position 9):

CAPE, Classe, K_ind, Lifted, PWAT, Sweat, Ttot
[1] 5.853659
[1] 94.14634
> mostra.arvore(arvore)
>

```

Files Plots Packages Help

Zoom Export Clear All

```

graph TD
    Root[ ] --- Node1[CAPE < 0.07838]
    Node1 --- Node2[PWAT < 0.4582]
    Node1 --- Node3[Ttot < 0.6219]
    Node2 --- Node4[K_ind < 0.1238]
    Node2 --- Node5[PWAT >= 0.4582]
    Node4 --- Node6[K_ind >= 0.1238]
    Node5 --- Node7[Lifted < 0.3301]
    Node5 --- Node8[Lifted >= 0.3301]
    Node7 --- Node9[Sweat < 0.7075]
    Node7 --- Node10[K_ind < 0.1087]
    Node9 --- Node11[Sweat >= 0.7075]
    Node10 --- Node12[K_ind >= 0.1087]

```

PT 13:31 21/08/2013

RStudio

File Edit Code View Plots Session Project Build Tools Help

Go to file/function Project: (None)

Source Workspace History

Files Plots Packages Help

Zoom Export Clear All

```

CAPE, Classe, K_ind, Lifted, PWAT, Sweat, Ttot
[1] 8.780488
[1] 91.21951
> mostra.arvore(arvore)
> load("~/workspace9.RData")
> source("final.r")
The following object is masked by_ .GlobalEnv:

CAPE, Classe, K_ind, Lifted, PWAT, Sweat, Ttot
The following object is masked from dadostotal (position 3):

CAPE, Classe, K_ind, Lifted, PWAT, Sweat, Ttot
The following object is masked from dadostotal (position 4):

CAPE, Classe, K_ind, Lifted, PWAT, Sweat, Ttot
The following object is masked from dadostotal (position 5):

CAPE, Classe, K_ind, Lifted, PWAT, Sweat, Ttot
The following object is masked from dadostotal (position 6):

CAPE, Classe, K_ind, Lifted, PWAT, Sweat, Ttot
The following object is masked from dadostotal (position 7):

CAPE, Classe, K_ind, Lifted, PWAT, Sweat, Ttot
The following object is masked from dadostotal (position 8):

CAPE, Classe, K_ind, Lifted, PWAT, Sweat, Ttot
The following object is masked from dadostotal (position 9):

CAPE, Classe, K_ind, Lifted, PWAT, Sweat, Ttot
The following object is masked from dadostotal (position 11):

CAPE, Classe, K_ind, Lifted, PWAT, Sweat, Ttot
[1] 7.317073
[1] 92.68293
> mostra.arvore(arvore)
>

```

Decision tree structure:

- Root Node: CAPE < 0.07763
 - Left Branch: PWAT < 0.4586
 - Left Sub-branch: K_ind < 0.1244
 - Left Leaf: 1
 - Right Sub-branch: K_ind >= 0.1244
 - Left Leaf: 1
 - Right Leaf: 2
 - Right Sub-branch: PWAT >= 0.4586
 - Leaf: 2
 - Right Branch: Ttot < 0.6173
 - Left Leaf: 1
 - Right Leaf: 2



RStudio

File Edit Code View Plots Session Project Build Tools Help

Go to file/function Project: (None)

Source Workspace History

Console ~/

The following object is masked from dadostotal (position 11):

```

CAPE, Classe, K_ind, Lifted, PWAT, Sweat, Ttot
[1] 7.317073
[1] 92.68293
> mostra.arvore(arvore)
> load("~/workspace10.RData")
> source("final.r")

```

The following object is masked by_ .GlobalEnv:

```

CAPE, Classe, K_ind, Lifted, PWAT, Sweat, Ttot

```

The following object is masked from dadostotal (position 3):

```

CAPE, Classe, K_ind, Lifted, PWAT, Sweat, Ttot

```

The following object is masked from dadostotal (position 4):

```

CAPE, Classe, K_ind, Lifted, PWAT, Sweat, Ttot

```

The following object is masked from dadostotal (position 5):

```

CAPE, Classe, K_ind, Lifted, PWAT, Sweat, Ttot

```

The following object is masked from dadostotal (position 6):

```

CAPE, Classe, K_ind, Lifted, PWAT, Sweat, Ttot

```

The following object is masked from dadostotal (position 7):

```

CAPE, Classe, K_ind, Lifted, PWAT, Sweat, Ttot

```

The following object is masked from dadostotal (position 8):

```

CAPE, Classe, K_ind, Lifted, PWAT, Sweat, Ttot

```

The following object is masked from dadostotal (position 9):

```

CAPE, Classe, K_ind, Lifted, PWAT, Sweat, Ttot

```

The following object is masked from dadostotal (position 10):

```

CAPE, Classe, K_ind, Lifted, PWAT, Sweat, Ttot

```

The following object is masked from dadostotal (position 12):

```

CAPE, Classe, K_ind, Lifted, PWAT, Sweat, Ttot
[1] 12.19512

```

Files Plots Packages Help

Zoom Export Clear All

```

graph TD
    Root[ ] --- Node1[CAPE < 0.0823]
    Node1 --- Node2[PWAT < 0.4586]
    Node1 --- Node3[K_ind < 0.1041]
    Node2 --- Node4[K_ind < 0.1244]
    Node2 --- Node5[PWAT >= 0.4586]
    Node4 --- Node6[ ]
    Node4 --- Node7[K_ind >= 0.1244]
    Node7 --- Node8[Sweat < 0.7441]
    Node7 --- Node9[Sweat >= 0.7441]
    Node3 --- Node10[ ]
    Node3 --- Node11[K_ind >= 0.1041]
    
```

PT 13:33 21/08/2013