



ISSN:1984-2295

# Revista Brasileira de Geografia Física

Homepage: [www.ufpe.br/rbgfe](http://www.ufpe.br/rbgfe)



## Previsão de raios através de índices termodinâmicos da atmosfera, na região de Belém do Pará

Cassia Camila Silva da Silva <sup>1</sup>; José Ricardo Santos de Souza <sup>2</sup>; Arthur da Costa Almeida <sup>2</sup>; Joaquim Carlos Barbosa Queiroz <sup>2</sup>; Marcela Machado Pompeu <sup>3</sup>; Wanda Maria do Nascimento Ribeiro <sup>4</sup>

1-Mestra em Ciências Ambientais -Professora Substituta da Universidade Estadual do Pará (UEPA) [cassia\\_css@yahoo.com.br](mailto:cassia_css@yahoo.com.br) 2-Professor da Universidade Federal do Pará [jricardo@ufpa.br](mailto:jricardo@ufpa.br) ; [arthur@ufpa.br](mailto:arthur@ufpa.br) [joaquimqz@gmail.com](mailto:joaquimqz@gmail.com) – UFPA.3-Mestra em Ciências Ambientais -Professora Colaboradora da Faculdade de Araguatins (FAIRA-PÓS) [marcelapompeu5@hotmail.com](mailto:marcelapompeu5@hotmail.com) 4-Mestra em Ciências Ambientais Física do Clima - Universidade Federal do Pará- UFPA [wandaribeiro10@hotmail.com](mailto:wandaribeiro10@hotmail.com)

Artigo recebido em 22/08/2018 e aceite em 15/02/2018

### 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. 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 significativa com o modelo criado que obteve 94% de acerto.

Palavras-chave: Previsão de raios; árvore de decisão. modelo estatístico

### Prediction of lightning thermodynamic indexes of the atmosphere, in the region of Belém do Pará

### 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. 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 surrounding Belém The study of atmospheric

1186

thermodynamic indices versus the occurrence of lightning showed a positive relationship where the evaluation model obtained a result of 94% correct values.

Keywords: Lightning forecast. decision tree. statistical model.

## Introdução

Os raios associados a nuvens de tempestades na Amazônia são os que causam maiores prejuízos. Sendo assim, 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.

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).

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.

## Material e métodos

### Localização e 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<sup>2</sup>, conforme mostrado na Figura 1. Belém é a capital do Estado do Pará e fica localizada geograficamente a 1° 27' 21" de latitude Sul e 48° 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 1).

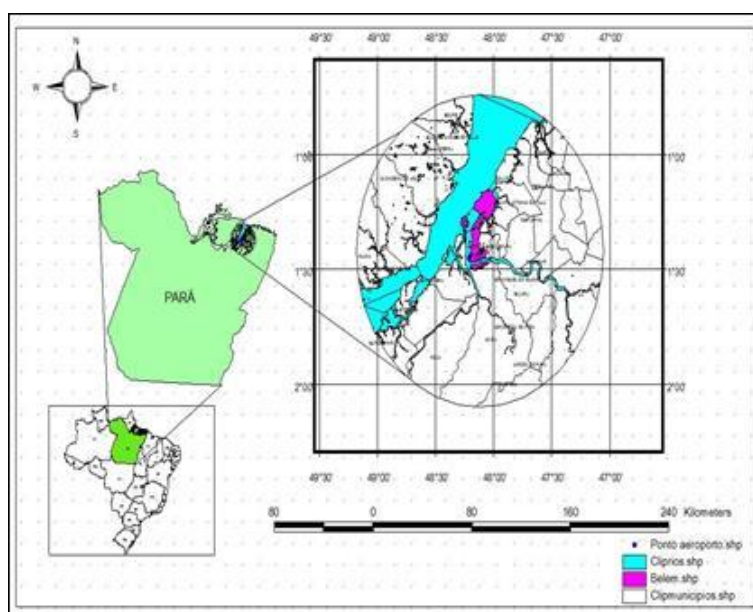


Figura 1 – Área de estudo.

A precipitação pluviométrica de Belém varia de 2000 a 3000 mm por ano, e nos 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).

#### Base de Dados

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>. Os índices 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. 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 distância, disponível no site, [www.zeus.iag.usp.br](http://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. 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.

O 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 dados e é um software livre, distribuído sob a licença pública geral ([http://fsf.org/pt\\_home.html](http://fsf.org/pt_home.html)) e pode ser livremente copiado e distribuído entre usuários, bem como pode ser instalado em diversos computadores.

O método utilizado para construção do modelo foi a Árvore de Decisão, que se baseia na 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). As Árvores de Decisões 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.

#### Resultados e discussão

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 ideia básica consiste em colocar em um mesmo grupo objetos que sejam similares de acordo com algum critério pré-determinado (Linden, 2009).

Os dados de raios foram divididos em 3 cluster : cluster 1 poucos raios ( $0\% < \text{Cluster 1} <$

33,3%), cluster 2 uma quantidade intermediária de raios(33,3% < Cluster 2 < 66,6%) e o cluster 3 representa muitos raios(66,6% < cluster 3 <100%).

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. Os 6 índices selecionados foram: Lifted, Sweat, K\_ind, Totl, CAPE e Pwat.

Os 6 Índices Termodinâmicos encontrados foram semelhantes aos que foram 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. 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.

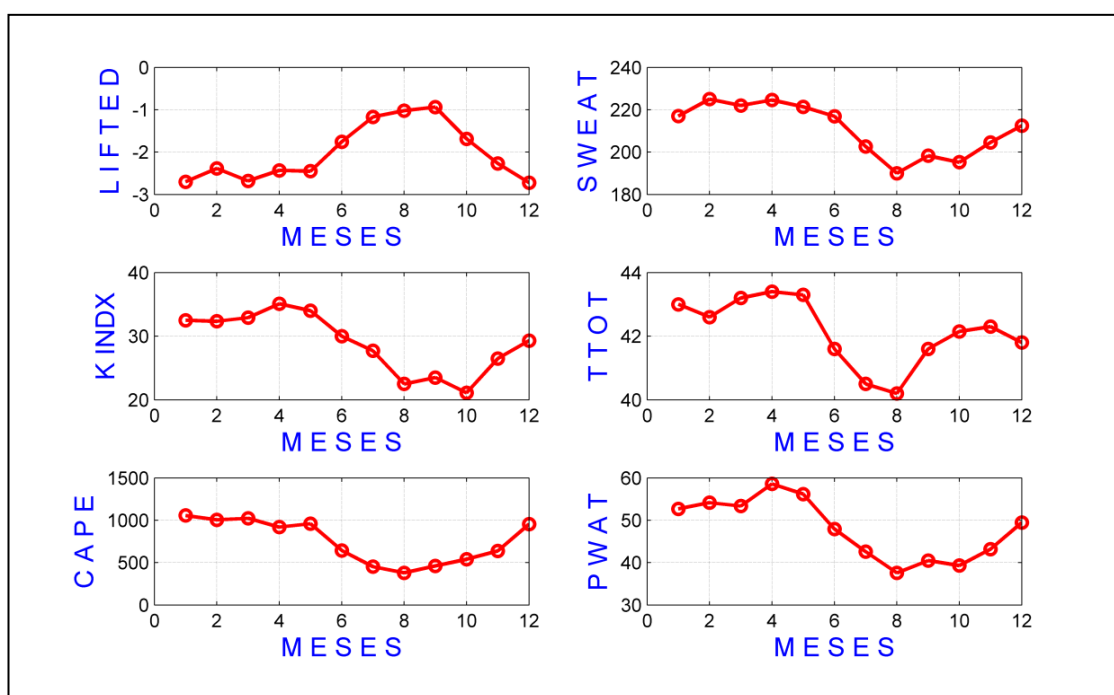


Figura 2 – Mediana mensal dos índices termodinâmicos selecionados pelo método Best First.

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.

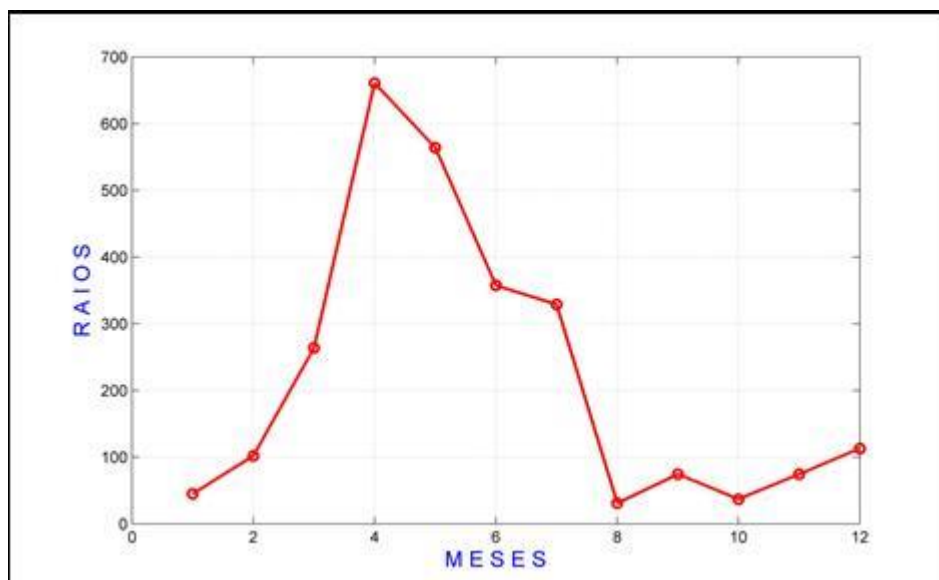


Figura 3 – Distribuição de medianas mensais de raio

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 3.

Para construção do modelo, foram utilizados os 6 índices termodinâmicos considerando a radiossondagem das 12:00 UTC (9:00 Hora Local de Belém –Pa).

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 1. Os valores ficaram no intervalo de [0,1].

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

O R possui um pacote chamado “rpart”, que 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.

É possível avaliar a um modelo através de uma matriz de confusão.



Previsão modelo

Matriz de confusão

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

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.

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. Dos 10 casos, foi escolhido o modelo com percentual de 94 % de acerto.

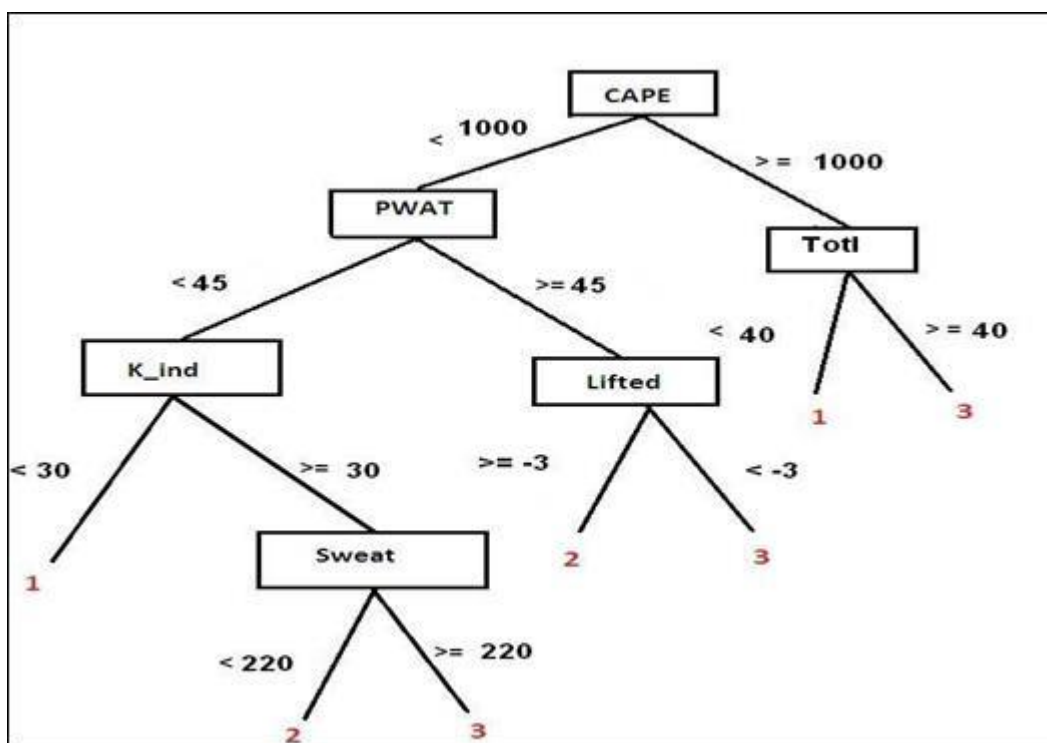


Figura 4 – Modelo da árvore de decisão com valores reais.

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

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

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  $-5^{\circ} C$  indicam situação muito instável, e extremamente instável se abaixo de  $-10^{\circ} C$ . 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:

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

**QTV** = 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  $2500 \text{ J kg}^{-1}$  são considerados altos; valores de  $2500 \text{ J kg}^{-1}$  indicam instabilidade acentuada, e acima de  $4000 \text{ J kg}^{-1}$  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:

$$SWET = 12.(TD850) + 20.(TERM2) + 2.(SKT850) + SKT.500 + SHEAR$$

Onde:

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

$TERM2 = \text{Máx} (TOTL - 49)$ , 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  $\text{DIR}500$  = a direção do vento no nível pressão atmosférica a 500 hPa e  $\text{DIR}850$  = 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  $250^{\circ}$ ;

✓ A direção do vento no nível de 500 hPa se encontra no intervalo de 210 a  $310^{\circ}$ ;

✓ 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 ( $\sim 30 \text{ Km/h} = 8,3 \text{ m/s}$ ).

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 extremos (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  $^{\circ}C$ , 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  $^{\circ}C$ , 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 = 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 conforme equação abaixo :

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

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.

O estudo de caso começa 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 analisados.

## Estudo De Caso

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, (Albuquerque et al, 2010) .

Foi escolhido o dia 01-07-2013 no período com menor precipitação correspondentes aos trimestres (Junho, Julho, Agosto) e (Setembro, Outubro, Novembro). Esse dia foi selecionado 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 Almeida<sup>1</sup>. A Figura 5 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.

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.



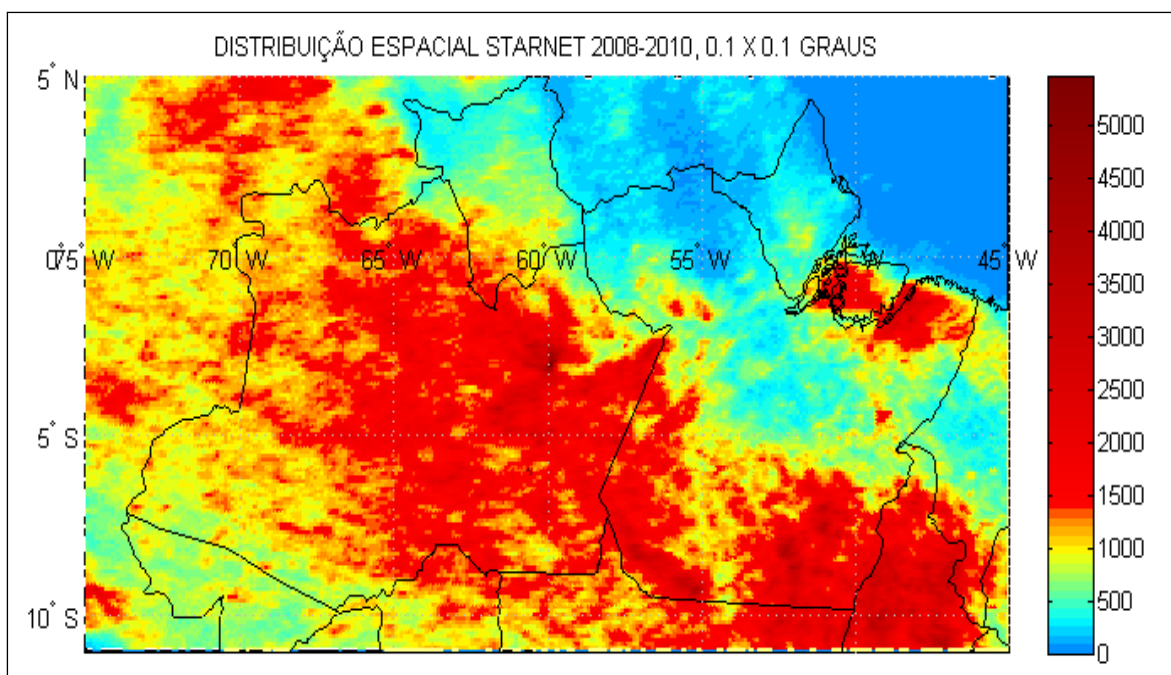


Figura 5- Distribuição Espacial de Raio

A Figura 6 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).

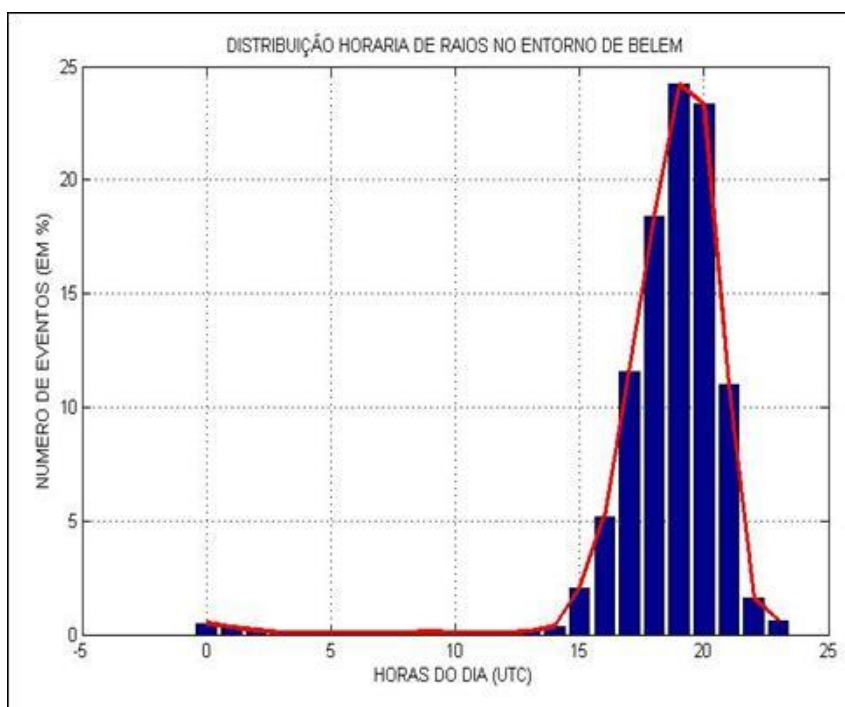


Figura 6 - Distribuição Horária de Raios.

A Tabela 1 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 1.

Tabela 1 – Dados de Raios na área de estudo anos de 2009 e 2010.

<b>Meses 2009-2010</b>	<b>Total</b>	<b>Mediana mensal</b>	<b>Mediana diária</b>
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

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 ideia 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.

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 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.

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 com problemas de congestionamento em diversos pontos da cidade conforme observamos na figura 7. 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 7 – Impactos socioambientais causados pela chuva e desligamento de energia.

Na Figura 8 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 9.

As regras geradas pelo modelo na Figura **8B**, demonstram que a CAPE é o índice termodinâmico a ser considerado com o valor limite de 1000, sendo que na Figura **8A**, 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 **8B** assume um valor limite de 40. Observando a Figura **8A**, 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 9A, de precipitação acumulado em torno de 30-70 para toda região próximo a Belém e na Figura 9B, nuvens esparsas sobre a região no horário das 19:00 UTC (16:00 hora local de Belém).

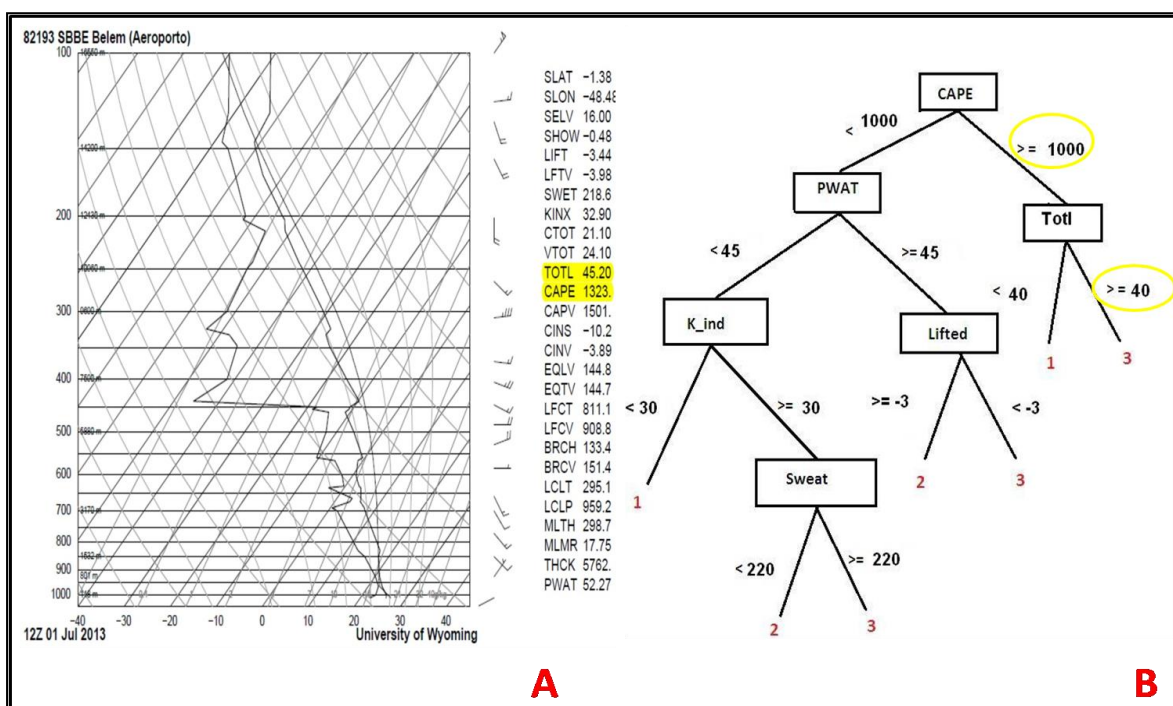


Figura 8 – Comparação do Diagrama SKEW T LOG P de Belém do dia 01/07/2013 às 1200UTC com as regras do modelo da árvore de decisão.

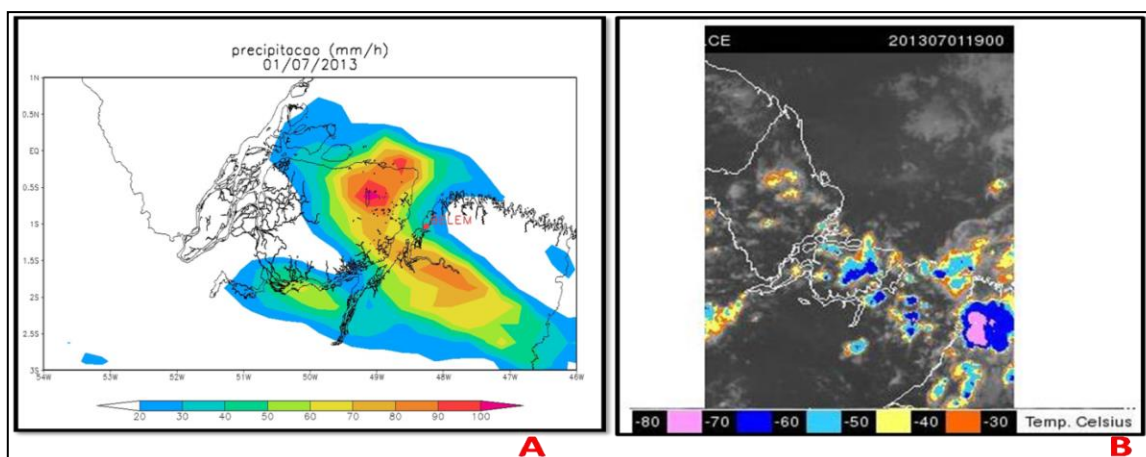


Figura 9 – Comparação da precipitação acumulada em Belém do dia 01/07/2013 com as imagens de satélite às 19:00UTC, sendo 16 hora local.

A Figura 10 mostra a distribuição de raios no dia 01/07/2013. A mediana diária de raios considerando a Tabela 1 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 Starnet foi de 2271 raios, o que coincide com o modelo, cluster 3. 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).



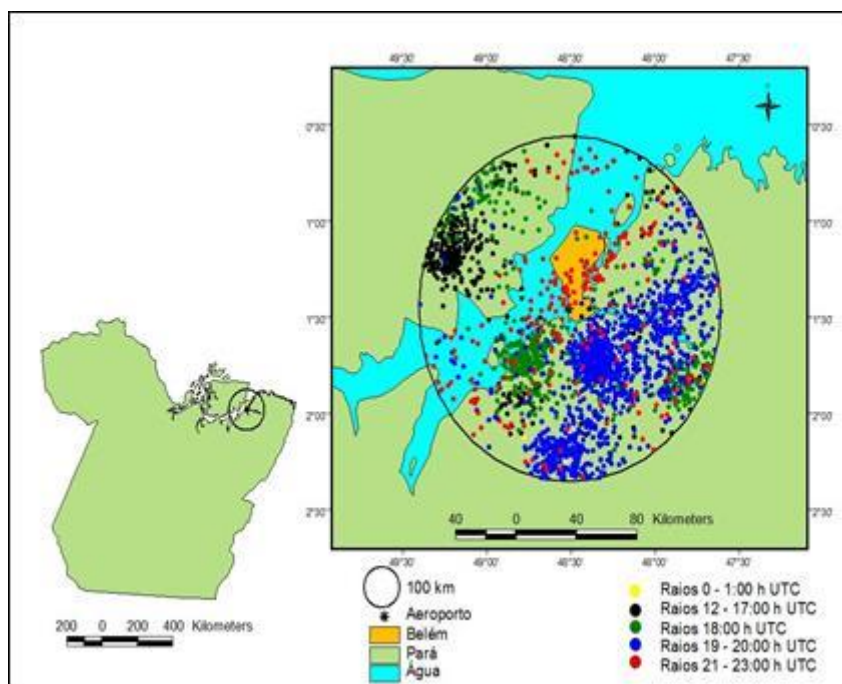


Figura 10 – Distribuição temporal dos raios em Belém no dia 01/07/2013.

## 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 para um dia de menor precipitação na área. A avaliação do modelo com a previsão da mediana de raios foi significativa, o modelo coincide com os dados reais, considerado bom em uma análise estatística. 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.

## Agradecimentos

Os autores agradecem ao Programa de pós-graduação em Ciências Ambientais da Universidade Federal do Pará, À USP / IAG pelos dados da STARNET liberados pelo Dr. Carlos A. Morales, Agradecem também a CAPES pelas bolsas concedidas a alguns dos autores.

## Referências

Albuquerque, M.F.; Souza, e. B.; Oliveira, M. C. F.; Souza, j. A. J. 2010. Precipitação nas mesorregiões do estado do Pará: climatologia, variabilidade e tendências nas últimas décadas (1978-2008). Revista Brasileira de Climatologia 6, 151-168.



- Almeida, A. C.; Rocha, B. R. P.; Souza, J. R. S.; Sá, J. A.; Pissolato Filho, J. A. 2012. Cloud-to-ground lightning observations over the Eastern Amazon region. *Atmospheric Research* 117, 86-90.
- 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. 2007. Lightning characteristics associated to severe rainfall events, around Belém PA Brazil. In: *International Symposium On Lightning Protection*, 9, 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. 2010. 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.
- Cohen, J.C.P., Silva Dias, M.A.F., Nobre, C.A. 1995. Environmental conditions associated with Amazonian squall lines: a case study. *Monthly Weather Review* 11, 3163-3174.
- Corrêa, L. V. 2011. 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.
- Dias, V. S. 2008. 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. (Relatório de Pesquisa). Disponível em: [http://www.redemet.aer.mil.br/Artigos/inst\\_form\\_conect\\_sbgl.pdf](http://www.redemet.aer.mil.br/Artigos/inst_form_conect_sbgl.pdf) Acesso em: 26/06/2011.
- Ferreira, E.B. Oliveira, M.S. 2008. Introdução à estatística básica com R. Lavras, 2008. 124f. Especialização (Gestão de empresas com Ênfase em Qualidade) - Universidade Federal de Lavras- MG, Lavras.
- Fonseca, J.M.M.R. 1994. Indução de árvore de decisão: histclss- proposta de um algoritmo não paramétrico. Lisboa, 1994. 151f. Dissertação (Mestrado) - Universidade Nova de Lisboa. Lisboa.
- Fontana, A; Naldi, M.C. 2009. 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. (Relatório Técnico, n.340).
- Garcia, S.C. 2000. O uso de árvores de decisão na descoberta de conhecimento na área da saúde. In: *Semana Acadêmica*, Rio Grande do Sul. Atas... Universidade do Rio Grande do Sul, Rio Grande do Sul.2000.
- Grupo De Eletricidade Atmosférica (Elat). 2011. Descargas Atmosféricas. Disponível em: <http://www.inpe.br/webelat/homepage/>. Acessado em 12/12/2011.
- 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>
- Lessa, L.A.S., Rocha, B.R.P., Souza, J.R.S., Almeida, A.C., Monteiro, J. H.A. 2008. 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.], 512-514.
- Linden, R. 2009. Técnicas de agrupamento. *Revista de Sistemas de Informação da FSMA*. 4. 18-36.

- 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.
- 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.
- Nascimento, E.L. 2005. Previsão de tempestades severas utilizando-se parâmetros convectivos e modelos de mesoescala: uma estratégia operacional adotável no Brasil?. Revista Brasileira de Meteorologia 20, 121 – 140.
- Piatetsky-Shapiro, G. 1991. Knowledge discovery in real databases: a report on the IJCAI-89 Workshop. AI Magazine 11, 68-70, Jan. Special issue.
- Pompeu, M. M; Souza,J.R.S; Mota.M.A.S; Ribeiro,W.M.N; Almeida.A.C. 2011. Lighting observations during Cape Experiment Barca. In: International Conference On Atmospheric Electricity-Icae, 15. Rio de Janeiro, Brazil. Proceedings...
- 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.
- Portal Orm: 2013. 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.itejournal.informs.org/Vol3No3/Ragsdale/>> Acesso em setembro de 2011.
- Radiossonda. 2012. Disponível em: <<http://www2.ufersa.edu.br/portal/noticias/3385/>> Acesso em janeiro de 2012.
- R 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.
- Ribeiro, W.M.N; Souza,J.R.S; Almeida,A.C; Santos F.A.A; Rocha, B.R.P. 2011. Lighting and precipitations over different surface types in Eastern Amazonia. In: Xiv International Conference On Atmospheric Electricity, Rio de Janeiro, Brazil(2011).
- RPART. Pacote do Programa. Disponível: <<http://cran.r-project.org/web/packages/rpart/.pdf>>. Acesso em: jan. de 2012.
- Rutledge, S. 1991. Middle latitude and tropical mesoscale convective systems. Review of Geophysics. Supplement 88-97.
- Sá, J.A.S; Rocha, B; Almeida, A; Souza, J.R. 2012. Recurrent self-organizing map for severe weather patterns recognition. Recurrent Neural Networks And Soft Computing, p 151- 174, Croatia.
- Santos, C.M.S; Silva, F.R.; Leme, N.P. 2011. Ciclo anual de precipitação como função de índices de instabilidade termodinâmica e fluxos de energia estática: análises em maxaranguape-rn. Revista Brasileira de Meteorologia 26, 235 – 242.
- Santos, A. A. S. 2010. 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.
- Satelite Goes, 2013. Disponível em <<http://satelite.cptec.inpe.br/>> acessado em 2013.

Silva, C.C.S; Coelho,V.F. 2011. Localização geográfica de raios usando o algoritmo de diferença do tempo de chegada. In: I Colóquio De Matemática Da Região Nordeste.

Souza, J.R.S.; Rocha, E.J.P.; Cohen, J.C.P. 2003. 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, 69-94.

Tavares, J. P. N. 2009. 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 Emilio Goeldi e EMBRAPA, Belém.