



Simpósio de Métodos Numéricos em Engenharia

25 a 27 de outubro, 2017

Identificação de Evento de Tempo Severo com dados 3D de radar meteorológico e Support Vector Machine

Tulipa G. G. J. da Silva
PPGMNE
Universidade Federal do Paraná
Curitiba, Brasil
tulipa@simepar.br

Paulo H. Siqueira
Departamento de Expressão gráfica
Universidade Federal do Paraná
Curitiba, Brasil

Cesar A. A. Beneti
Departamento de Monitoramento e Previsão
SIMEPAR
Curitiba, Brasil

Maiko F. Buzzi
Departamento de Matemática
Universidade Tecnológica Federal do Paraná
Curitiba, Brasil

Leonardo Calvetti
Departamento de Meteorologia
Universidade Federal de Pelotas
Pelotas, Brasil

Resumo—Análises e estudos meteorológicos permitem prever o clima e eventos severos, como fortes chuvas, tempestades elétricas e tornados. Este artigo apresenta uma aplicação da técnica Support Vector Machine (SVM) para a previsão de Eventos de Tempo Severo (SWE - Severe Weather Event). Os dados utilizados como entrada do modelo, consistem em dados coletados de um radar de polarização dupla no sul do Brasil, em Cascavel, Paraná. Estes dados são inseridos de forma volumétrica (3D) ou através do produto de radar CAPPI (2D). Após o treinamento, obteve-se um modelo que pode servir de suporte à decisão sobre alertas do SWE no estado do Paraná, o modelo obtido pelo treinamento como dado volumétrico. Os resultados indicam uma detecção de 93,02% para os SWEs estudados e um acordo de 67,84% para casos identificados por uma rede de detecção de raios até 240 km de distância do radar. O modelo obtido a partir do treinamento do produto CAPPI, apresentou acurácia muito inferior ao modelo com entrada da dados na forma volumétrica. Este estudo preliminar mostrou que, com a utilização da SVM o produto CAPPI não é suficiente para identificação de SWE, sendo sugerido o treino do modelo com mais elevações.

Palavras-chave—Evento de tempo severo; Support Vector Machine; Radar Meteorológico; Descargas Elétricas Atmosféricas.

I. INTRODUÇÃO

No período de janeiro de 2015 a julho de 2016 a Defesa Civil efetuou 106 registros de eventos de tempo severo, no estado do Paraná, causados por tempestades convectivas. Segundo a Defesa Civil, ao menos 39753 pessoas foram diretamente afetadas por estas tempestades [1], [2]. Os eventos registrados são do tipo vendaval, chuvas intensas, granizo ou tornados.

Para o monitoramento e previsão meteorológica, o radar é uma das melhores ferramentas disponíveis, pois oferece monitoramento em tempo real, com boa resolução espacial e possibilita uma visão tridimensional dos dados [3].

As técnicas de aprendizagem de máquina têm sido utilizadas na meteorologia, quer seja para classificação como mostram [4] e [5] ou para previsão como mostra o trabalho de [6].

Com o intuito de identificar a ocorrência de SWEs, será aplicada a técnica de aprendizagem de máquina Support Vector Machine (SVM) em dados de radar polarimétrico, obtendo modelos de identificação que estimam se uma tempestade convectiva pode conter um SWE em muito curto prazo (30 minutos ou menos). Os resultados obtidos são também comparados aos dados de descargas elétricas atmosféricas, para melhor avaliação dos modelos obtidos.

Esta pesquisa foi financiada pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e pelo Sistema Meteorológico do Paraná (SIMEPAR). Os dados de descargas elétricas atmosféricas foi gentilmente cedido pela Earth Networks Total Lightning (ENTL).

II. CONCEITOS METEOROLÓGICOS

A. Evento de tempo severo (SWE)

Uma tempestade convectiva severa é uma tempestade capaz de gerar vendaval, chuva intensa, granizo, tornados e ou descargas elétricas atmosféricas [7]. Cada um dos eventos do tipo vendaval, chuvas intensas, granizo, tornados e tempestades elétricas, decorrentes de tempestade convectiva é chamado Evento de Tempo Severo (SWE - Severe Weather Event) [7].

Os limiares para velocidade do vento, tamanho do granizo, volume de precipitação e outros fenômenos associados a SWEs são baseados na região de estudo, sendo adequado que os limiares sejam baseados no conceito de que um SWE é caracterizado por ser capaz de gerar fenômenos de impacto social e econômico [8].

B. Radar

O radar é uma das melhores ferramentas disponíveis para o monitoramento e previsão meteorológica, pois oferece monitoramento em tempo real, com boa resolução espacial e possibilita uma visão tridimensional dos dados [9].

A coleta de dados ocorre quando o radar transmite um feixe de radiação eletromagnética e capta a energia refletida pelas partículas presentes na atmosfera. Essa energia é amplificada e, conhecido o intervalo de tempo Δt entre emissão e retorno do eco, é possível determinar a distância do alvo ao radar. Essa distância r pode ser calculada através da Equação 1, onde c_m é a velocidade de propagação das ondas eletromagnéticas no meio [11].

$$r = \frac{c_m \Delta t}{2} \quad (1)$$

Os dados coletados pelo radar são obtidos em coordenadas esféricas, onde a posição de cada dado é descrita em termos de (θ, ϕ, r) , onde θ é o ângulo de elevação da antena, ϕ é o ângulo de rotação, em sentido horário, da antena em relação ao Norte geográfico, chamado de ângulo de azimute, e r a distância do dado ao radar.

A coleta de sucessivos giros da antena do radar a ângulos fixos, é conhecida como varredura volumétrica, pois representa o dado em sucessivos cones [3]. Cada um destes cones também é chamado de PPI - Plan Position Indicator.

Uma outra forma de visualização do dado de radar é o CAPPI (Constant altitude plan position indicator). O CAPPI possibilita a obtenção do dado de radar a uma determinada altitude, de forma que todas as elevações possam contribuir com o produto gerado [3]. A Fig. 1 ilustra a coleta de CAPPI para altitude de 3 km em relação a elevação dos PPIs.

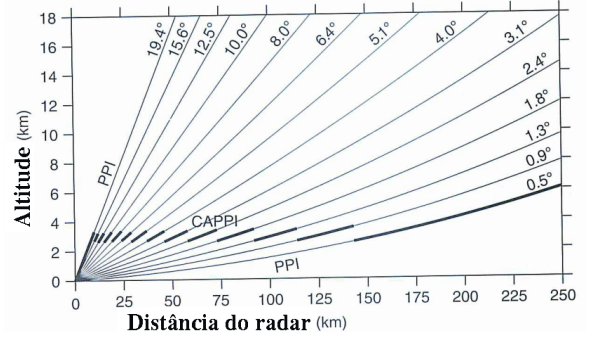


Figura 1: Exemplo de obtenção do CAPPI para 3 km de altitude. As linhas mais finas representam os PPIs em um corte vertical e as linhas mais grossas os dados que são usados para o CAPPI. Fonte: [3].

C. Momentos Doppler

Radares do tipo Doppler operam comparando a frequência do sinal transmitido com o sinal recebido, a fim de calcular a velocidade da partícula captada pelo sinal [12]. Para um radar meteorológico de tipo Doppler, são medidos três momentos: Refletividade (Z), Velocidade Radial (V) e Largura Espectral (W) [3].

Refletividade (Z). A medida de eficiência de um alvo em interceptar e retornar a energia emitida pelo radar é chamada de refletividade. Dessa forma, refletividade é dependente dos tamanhos, formatos, quantidade e propriedades dielétricas dos alvos, relação expressa pela Equação 2, onde c_r é chamada constante do radar, definida por características do radar e suas configurações, e p_r é a energia refletida para o radar [3].

$$z = c_r r^2 p_r. \quad (2)$$

A refletividade é expressa pela unidade mm^6/m^3 . Porém, os valores que a refletividade pode assumir com essa unidade variam entre 0 até 36000, sendo convencional fazer a compressão desses valores, adotando a escala logarítmica. Nesse caso, a refletividade pode ser escrita como:

$$Z = 10 \log_{10} \left(\frac{z}{1mm^6/m^3} \right). \quad (3)$$

Assim, na nova escala, a refletividade é denotada pela unidade dBZ e possui valores típicos entre -24 dBZ e 72 dBZ [3].

A refletividade é uma das variáveis de maior utilização na previsão meteorológica. Valores elevados de refletividade estão, em geral, associados à chuva intensa e altos valores de refletividade, 45 dBZ ou mais, entre 5 e 12 km de altitude podem estar associados a tempestade severa [3].

Velocidade Radial (V). Utilizando o efeito Doppler, o radar estima

a velocidade com que o alvo se afasta ou se aproxima do radar na direção do feixe emitido [3].

A velocidade é a derivada da distância r da partícula com relação ao tempo t , sendo descrita como:

$$V = \frac{dr}{dt}. \quad (4)$$

A velocidade radial é a velocidade com que o alvo detectado está se distanciando ou se aproximando do radar. Por convenção, valores negativos representam partículas se aproximando do radar e valores positivos, partículas se afastando do radar [3].

Largura Espectral (W). A largura espectral (W) representa o desvio padrão das medidas de velocidade obtidas. Em geral, está associado a turbulência no volume e sua unidade de medida é expressa em metros por segundo (m/s). Altos valores de largura espectral podem indicar convecção [3].

D. Variáveis Polarimétricas

No radar convencional, polarização simples, o sinal eletromagnético emitido e recebido possui apenas uma orientação (vertical ou horizontal), sendo capaz de obter informações dos alvos em apenas nesta orientação. O radar polarimétrico dual ou radar de dupla polarização, tem a propagação de sinal eletromagnético com orientações tanto horizontal como vertical, sendo capaz de obter informações nas duas orientações [3].

Alguns estudos para a previsão meteorológica utilizando as variáveis polarimétricas são listados por [13], pois as variáveis polarimétricas fornecem informações adicionais sobre o alvo, haja vista que, com as novas orientações é possível obter informações sobre a forma e até mesmo tamanho do alvo [13], [3].

Devido às combinações de orientações possíveis, além dos momentos Doppler, o radar de dupla polarização gera novas variáveis tais como ZDR , $RHOHV$, $PHIDP$ e KDP , denominadas variáveis polarimétricas.

Refletividade Diferencial (ZDR). A Refletividade Diferencial (ZDR), é a diferença entre Z_{HH} (refletividade com sinal emitido e captado horizontalmente) e Z_{VV} (refletividade com sinal emitido e captado verticalmente). Quando as partículas são quase esféricas, ZDR é próximo de 0 dB, no entanto, gotas de chuva são achatadas verticalmente (formato oblato) e quanto maiores e mais pesadas forem as gotas, mais oblato se tornam. Valores típicos de ZDR variam de -4 dB a 7 dB, com valores acima de zero, em geral, associados a chuva [3]. A ZDR é muito utilizada para diferenciar o que representa chuva, granizo, neve, e até mesmo alvos não meteorológicos [13].

Correlação Copolar ($RHOHV$). A Correlação Copolar ($RHOHV$), não possui unidade de medida e representa a correlação entre os sinais horizontais e verticais polarizados de Z , em um determinado ponto no espaço [9].

Os valores de $RHOHV$ variam entre 0 e 1.0 e podem ser vistos como medida da similaridade entre os sinais horizontal e vertical de polarização [13]. Esta variável descreve características físicas do alvo, sendo mais próximo de 1.0 quanto mais uniforme for o alvo, pois os sinais vertical e horizontal tendem a serem iguais.

Por ser calculado por meio da correlação dos sinais, o $RHOHV$ possui independência em relação à concentração de hidrometeoros na atmosfera e a efeitos de propagação, possuindo apenas dependência em relação às formas do hidrometeoros, suas inclinações e as misturas dos mesmos [14].

Diferencial de Fase ($PHIDP$). Seja Φ a fase em que o pulso foi emitido e recebido, o diferencial de fase ($PHIDP$), medido em graus, representa a diferença entre as fases Φ_{HH} e Φ_{VV} para o trajeto completo do pulso [3].

A variável $PHIDP$ possui variações radiais, pois apresenta mudanças cumulativas na diferença de fase para a viagem completa do pulso. Assim, $PHIDP$ não raramente é substituída pela sua derivada espacial, a KDP [3].

Diferencial de Fase Específica (KDP). A variável Diferencial de Fase Específica (KDP) é a derivada espacial de $PHIDP$ [3].

KDP tem como unidade de medida $^{\circ}/km$ e é um excelente estimador para precipitação, como apresentado por [15], com valores acima de $2^{\circ}/km$ indicando quantidade significativa de água líquida e ou em formato oblato.

E. Cisalhamento

Devido ao desenvolvimento vertical, as tempestades convectivas estão associadas a cisalhamento, especialmente a baixo e médio níveis [3].

Tradicionalmente, o cisalhamento é obtido pelo diferença da velocidade Doppler entre dois pontos, dividido pela distância entre estes dois pontos, sendo possível descrever 3 campos de cisalhamento: vertical, radial e azimutal [16].

Seja $V_{(\theta, \phi_i, r)}$ a velocidade Doppler nas coordenadas esféricas (θ, ϕ_i, r) , então:

$$\text{Cisalhamento azimutal} = \frac{V_{(\theta, \phi_i, r)} - V_{(\theta, \phi_{i+1}, r)}}{dV_{(\theta, \phi_i, r), V_{(\theta, \phi_{i+1}, r)}}}; \quad (5)$$

$$\text{Cisalhamento radial} = \frac{V_{(\theta, \phi, r_i)} - V_{(\theta, \phi, r_{i+1})}}{dV_{(\theta, \phi, r_i), V_{(\theta, \phi, r_{i+1})}}}; \quad (6)$$

$$Cisalhamento \quad vertical = \frac{V_{(\theta_i, \phi, r)} - V_{(\theta_{i+1}, \phi, r)}}{dV_{(\theta_i, \phi, r), V_{(\theta_{i+1}, \phi, r)}}}. \quad (7)$$

As equações 5, 6 e 7 são campos escalares, que representam as variações de velocidade azimutal, radial e verticalmente. Essas variações podem indicar rotação, ou mesmo que existe diferença significativa entre a velocidade da tempestade em baixa, média e alta altitude.

F. Descargas elétricas atmosféricas

As descargas elétricas atmosféricas são resultado do desenvolvimento de intensos centros de cargas na nuvem, que excedeu a capacidade de isolamento elétrico do ar resultando na dissipação dos centros de carga. Os tipos de descargas elétricas atmosféricas são classificados de acordo com local de onde se originam e onde terminam [17]: Intra Cloud (IC) e Cloud to Ground (CG).

Os dados de DEA podem indicar a severidade de uma tempestade por sua posição e quantidade. Conforme apresentado por [18], os raios tendem a se concentrar nas regiões convectivas e durante a maturidade de uma tempestade convectiva para depois decair novamente na fase de dissipação da tempestade.

O crescimento na densidade total de descargas elétricas atmosféricas (Total Lightning- TL), soma das densidades de CG e IC, é um bom indicador de SWE [19], [20]. Este fenômeno é chamado de Lightning Jump (LJ), e ocorre minutos antes da ocorrência de um SWE [19].

No estudo de [21] 10% dos SWEs não foram precedidos por LJ, bem como 37% dos LJ foram alarme falso, assim um modelo de identificação de SWE por estudo de LJ não possui acurácia de 100%, embora continue sendo um bom indetificador [20].

Assim, pode-se afirmar que ocorreu tempestade elétrica para alta densidade de CG, mas que é muito provável a ocorrência de SWE onde é detectado LJ.

III. SUPPORT VECTOR MACHINE (SVM)

Support Vector Machine (SVM) ou máquina de vetores de suporte, é uma técnica de aprendizado supervisionado, desenvolvida por V. Vapnik que busca minimizar o erro no conjunto de treinamento (risco empírico), bem como o erro no conjunto de teste (risco na generalização) [22].

De acordo com sua formulação matemática, a SVM corresponde a obter um hiperplano ótimo tal que melhor segregue as classes, ou seja, encontrar um hiperplano com a máxima distância euclidiana do conjunto de treinamento [23].

Seja o conjunto de treinamento finito tal que a k -ésima instância é da forma (x_k, y_k) , $x_k \in \mathfrak{R}^n$, $y_k \in \{+1, -1\}$, onde x_k é o

padrão de entrada k e y_k é a correspondente resposta desejada. A forma geral da função de decisão linear é dada por

$$D(x) = w^T x + b, \quad (8)$$

onde $w \in \mathfrak{R}^n$ é um vetor de pesos e $b \in \mathfrak{R}$ é o termo independente (bias).

Adicionando uma margem $\delta > 0$, admitindo que $y_k = +1$ se $D(x) \geq \delta$ e $y_k = -1$ se $D(x) \leq -\delta$ e redimensionando w e b adequadamente tem-se a estratégia de decisão dada por:

$$\begin{cases} w^T x + b \geq +1 & \text{para } y_k = +1 \\ w^T x + b \leq -1 & \text{para } y_k = -1, \end{cases}$$

onde, as classes são separadas pelo hiperplano $w^T x + b$, com margem δ [24].

Obter o hiperplano ótimo da SVM corresponde a obter um hiperplano que maximize a distância δ . Ou seja, minimização do erro de treinamento através da solução do problema [23]:

$$\begin{aligned} \min_{w, b, \xi} \quad & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^p \xi_i \\ \text{Sujeito a:} \quad & y_i (w^T x_i + b) \geq 1 + \xi_i, \\ & \xi_i \geq 0, \quad \forall i = 1, \dots, p. \end{aligned} \quad (9)$$

No problema 9 o termo $C \in \mathfrak{R}_+$ é denominado constante de regularização, a qual determina a importância maior ou menor às folgas ξ_k , possibilitando que o modelo seja mais ou menos sensível a pontos com padrão não identificado, ou mesmo mal classificados [23].

Seja a função lagrangeana associada ao primal 9 dada por

$$L(w, b, \xi, \alpha) = \frac{1}{2} w^T w + C \sum_i \xi_i - \sum_i (\alpha_i (y_i (w^T x_i + b) + 1 + \xi_i)),$$

então:

$$\begin{cases} \frac{\partial L(w, b, \xi, \alpha)}{\partial w} = w - \sum_i y_i \alpha_i x_i \\ \frac{\partial L(w, b, \xi, \alpha)}{\partial b} = \sum_i y_i \alpha_i \\ \frac{\partial L(w, b, \xi, \alpha)}{\partial \xi} = - \sum_i \alpha_i + C \end{cases}.$$

De onde segue que

$$\begin{cases} w - \sum_i y_i \alpha_i x_i = 0 \Rightarrow w = \sum_i y_i \alpha_i x_i \\ \sum_i y_i \alpha_i = 0 \Rightarrow \sum_i y_i \alpha_i = 0 \\ - \sum_i \alpha_i + C = 0 \Rightarrow \sum_i \alpha_i = C \end{cases}. \quad (10)$$

Substituindo 10 na função lagrangeana $L(w, b, \xi, \alpha)$ tem-se que

$$L(w, b, \xi, \alpha) = \frac{1}{2} \sum_{i,j} y_i y_j \alpha_i \alpha_j x_i x_j - \sum_i \alpha_i. \quad (11)$$

Substituindo 11 e suas condições em 9 segue a forma dual de 9:

$$\begin{aligned} \max_{\alpha} \quad & \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} y_i y_j \alpha_i \alpha_j (x_i x_j) \\ \text{Sujeito a :} \quad & \sum_i y_i \alpha_i = 0 \\ & 0 \leq \alpha_i \leq C, \quad i, j = 1, \dots, p. \end{aligned} \quad (12)$$

Sejam $\alpha^* = (\alpha_1^*, \alpha_2^*, \dots, \alpha_N^*)$ a solução para 12 e w, d e ξ solução para 9. Da teoria de otimização (condições de Kühn-Tucker), segue que

$$\alpha_i^* (y_i (w^T x + b) - 1 + \xi_i) = 0, \quad \forall \quad i = 1, 2, \dots, N,$$

ou seja, os vetores de suporte são os únicos a contribuir para a determinação do plano separador, pois α^* é diferente de zero apenas se α^* pertence aos planos $D(x) + \delta$ ou $D(x) - \delta$. Desta forma, o parâmetro w apresentado na superfície de decisão 8 pode ser obtido pela primeira equação de 10 e b é dado por [25]:

$$b = \frac{1}{N} \sum_{\alpha_i^* \in \alpha^*} \frac{1}{y_i} - w^T \alpha_i^*.$$

IV. METODOLOGIA

A metodologia aplicada pode ser descrita pelos seguintes passos:

- 1) Coleta dos dados volumétricos e cálculo do produto CAPPI;
- 2) Pré-processamento dos dados (correção de dados faltantes e resolução, normalização, cálculo dos cisalhamentos, retirada de ruído e redução da dimensionalidade da entrada);
- 3) Inclusão das classes ao conjunto de dados;
- 4) Separação do conjunto de teste e treinamento;
- 5) Treinamento da técnica SVM, gerando os modelos M-SVM3D e M-SVM2D ;
- 6) Identificação de regiões de SWE por LJ e densidade de CG (modelo M-ENTLN);
- 7) Comparação dos modelos obtidos entre si e com M-ENTLN;
- 8) Estudo e análise dos resultados obtidos.

A. Especificações do Radar

Neste trabalho, foram utilizados os dados coletados pelo radar operado pelo Sistema Meteorológico do Paraná (SIMEPAR), um radar polarimétrico, com alcance de 240 km em coleta volumétrica.

Os momentos Z e V são coletados por este radar com emissão e recepção na mesma direção. Ou seja, Z_{HH} , Z_{VV} , V_{HH} e V_{VV} .

Para este trabalho optou-se por utilizar Z_{HH} e V_{HH} que são mencionados como Z e V respectivamente.

B. Especificações dos dados de Descargas Elétricas Atmosféricas

A Earth Networks Total Lightning Network (ENTLN) é a mais ampla rede global de detecção de descargas atmosféricas dos tipos IC e CG, com alta densidade de sensores e operando em tempo real [20]. Neste trabalho são utilizados os dados de IC e CG, fornecidos pela ENTLN.

C. Especificações dos Eventos de Tempo Severo

Dentre os eventos relatados pela Defesa Civil, no período de janeiro de 2015 a junho de 2016, serão estudados os SWEs que se enquadram nas especificações de disponibilidade (existência do dado) e localização (até 240 km do radar estudado).

Para estudo de curtíssimo prazo, apenas dados coletados em até 30 minutos antes da ocorrência do SWE. Assim, cada SWE possui entre 3 e 5 arquivos com dados de radar correspondentes a instantes de tempo diferentes, com diferença de 7.5 ou 15 minutos. Cada arquivo possui a coleta volumétrica do dado de radar iniciando no instante de tempo que descreve, totalizando 43 arquivos com os dados de radar.

A fim de padronizar a resolução dos dados e tratar valores faltantes, os mesmos foram interpolados, por uma função de interpolação exponencial, em uma grade cartesiana, com o auxílio do software Pyart [26].

Após a interpolação, excluiu-se os pontos onde Z é inferior a 30dBZ por se tratarem de dados que não são representativos em uma tempestade convectiva [7].

Para levar em consideração o crescimento vertical da tempestade, incluiu-se ainda a cada ponto de grade o parâmetro HMAX, que representa a altitude do ponto de maior refletividade por coluna (mesmo par de latitude e longitude).

O objetivo de “identificar regiões de ocorrência de ETS dentro de 30 minutos”, implica que a região de interesse é a que possui o atributo “ocorreu ETS dentro de 30 minutos”. Assim, podemos estudar o problema de identificação de ETS como um problema de classificação binária, ou seja, indica-se com 1 as regiões de interesse e zero as que não são de interesse. Neste sentido, convém indicar como região que não é de interesse onde “é desconhecida a informação da ocorrência de ETS dentro de 30 minutos”.

Inicialmente, a proporção de dados na classe 1 era de menos de 10%, a fim de aumentar a representatividade destes dados esta proporção foi aumentada, escolhendo menos dados entre aqueles classificados como 0.

D. Treinamento

Dos 43 arquivos disponíveis na forma descrita anteriormente, foram utilizados 33 para o treinamento das técnicas e 10 para verificar a generalização (conjunto de teste).

O conjunto de entrada consiste de vetores que contém os valores normalizados de Z, ZDR, RHOHV, KDP, HMAX, cisalhamento azimutal em módulo, cisalhamento radial em módulo e cisalhamento vertical em módulo para cada ponto de grade para cada arquivo e nesta ordem.

Para treinamento da SVM foi utilizado o software livre Scikitlearn, do Python [28]. Esse pacote possui vasta gama de técnicas de AM implementadas e de fácil utilização.

V. RESULTADOS

Como descrito anteriormente, o treinamento da técnica ocorreu com a apresentação dos dados de 33 arquivos do radar utilizado, obtendo os melhores resultados a proporção de aproximadamente 25% dos dados positivos. As figuras 3 e 2 apresentam exemplos de saída dos modelos obtidos por SVM como dado volumétrico (M-SVM3D), pelo dado de CAPPI (M-SVM2D) e pela análise de densidade de CG ou lightning jump (M-ENTLN).

A Fig. 2 apresenta os modelos aplicados no dado de 28/12/2015 às 21:15, momento em que ocorria um SWE do tipo vendaval na região de Francisco Beltrão (interna ao círculo). Neste exemplo tem-se que tanto M-SVM3D quanto M-SVM2D identificam a ocorrência de SWE na região pré-identificada. No entanto, M-SVM2D identifica apenas uma região enquanto M-SVM3D identifica 3 regiões e nenhum dos dois modelos identifica todas as regiões identificadas por M-ENTLN (Fig. 2 (c)). Neste exemplo, M-SVM3D possui melhor concordância com M-ENTLN para regiões não conhecidas pelos modelos, embora M-SVM2D também identifique a região pré-identificada.

Por outro lado, a Fig. 3, representa os modelos M-SVM3D (Fig. 3 (a)), M-SVM2D (Fig. 3 (b)) e M-ENTLN (Fig. 3 (c)) aplicados ao dado de 13:30 de 23/11/2015. Neste exemplo o modelo M-SVM2D identifica menos regiões que M-SVM3D. No entanto, as regiões identificadas por M-SVM3D que não são identificadas por M-SVM2D não necessariamente são regiões identificadas por M-ENTLN. Além disso, a região de identificação do SWE (região circulada na Fig. 3 (b)) não é identificada por M-SVM2D. Desta forma, a Fig. 3, apresenta um exemplo de dado em que M-SVM2D exclui identificações em que M-SVM3D não possuía concordância com M-ENTLN, o que seria um ponto positivo ao modelo, porém a exclusão de regiões também ocorre tanto nos dados pré-classificados quanto aqueles que são identificados por M-ENTLN.

Tanto a Fig. 2 quanto a Fig. 3 sugerem que M-SVM2D identifique menos regiões que M-SVM3D, quer sejam estas regiões classificadas ou não. Como apresentado por [29], o modelo M-SVM3D, em [29] apresentado apenas como M-SVM, possui boa concordância com M-ENTLN (M-DEA em [29]) mas possui variabilidade maior em relação às altitudes das entradas com saída positiva. Assim, ao tomar apenas uma altitude (CAPPI em 3km), e não um subconjunto de altitudes como sugerido por [29] é possível que regiões identificadas com altitude diferente de 3 km deixem de serem identificadas mesmo havendo novo treinamento.

De fato, em relação às regiões previamente identificadas, a identificação por M-SVM2D é inferior a aplicada M-SVM3D, conforme apresentado pela Tabela I.

Tabela I: Proporções de acerto para os modelos M-SVM3D e M-SVM2D

Modelo	Conjunto de Treinamento	Conjunto de Teste	Conjunto Geral
M-SVM3D	90%	100%	93,02%
M-SVM2D	72,72 %	80%	74,41 %

Para o treinamento dos modelos M-MLP3D e M-MLP2D, as regiões em que é conhecida a ocorrência de SWE (exemplificado pelo círculo nas figuras 3 e 2) são referente a 10 eventos específicos. No entanto, pode-se ver pelas figuras 3 e 2 que existem outras regiões identificadas pelos modelos, mas cuja informação de ocorrência de SWE é desconhecida.

A Tabela II apresenta a porcentagem de regiões em que M-SVM3D e M-SVM2D diferem na identificação. Nesta tabela, nota-se que M-SVM2D identifica menos regiões que M-SVM3D, em geral, 33,37% regiões a menos.

Tabela II: Comparação entre M-SVM3D e M-SVM2D

Conjunto	M-SVM3D identifica e M-SVM2D não identifica
Treinamento	26,36 %
Test	46,77 %
Geral	33,37 %

Todas regiões que não correspondem aos eventos estudados não possuem informação prévia sobre a ocorrência de SWE, no entanto, esta informação pode ser obtida de forma indireta ao se comparar as regiões identificadas por M-ENTLN.

A fim de validar a identificação de M-MLP3D e M-MLP2D em regiões em que é desconhecida a informação da ocorrência de SWE e também verificar se as regiões desconsideradas por M-MLP2D são relevantes, foi obtido o modelo M-ENTLN (baseado na densidade de raios do tipo CG e LJ). Desta forma, nas regiões em que a ocorrência de SWE é desconhecida, os modelos M-SVM3D e M-SVM2D serão comparados a M-ENTLN, modelo que obteve 100 % de identificação correta para os eventos conhecidos.

A Tabela III apresenta a porcentagem de regiões identificadas cujo modelo M-ENTLN confirma o SWE e a porcentagem de regiões em que os modelos identificam SWE mas M-ENTLN não confirma. Nesta tabela R1, R2, R3, R4, R4 e R5 representam:

- **R1:** Regiões identificadas por M-SVM3D e por M-ENTLN com raio de até 240 km do radar;
- **R2:** Regiões identificadas por M-SVM2D e por M-ENTLN com raio de até 240 km do radar;
- **R3:** Regiões identificadas por M-ENTLN e não identificadas por M-SVM3D com raio de até 240 km do radar;

- **R4:** Regiões identificadas por M-ENTLN e não identificadas por M-SVM2D com raio de até 240 km do radar;

Tabela III: Comparação entre identificação de ETS por M-ENTLN, M-MLP3D e M-MLP-2D

Conjunto Treinamento	R1	R2	R3	R4
Test	75,00%	73,64%	23,94%	45,27%
Geral	55,42%	53,23%	23,33%	40,00%
	67,84%	66,28%	23,76%	43,84%

Assim, pela Tabela III, é possível afirmar que o modelo M-MLP2D possui correlação, com as regiões identificadas por M-ENTLN até um raio de 240 km do radar, inferior a M-SVM3D. Pois, deixa de identificar, em geral, 43,84 % regiões que M-ENTLN identifica, enquanto que M-SVM3D deixa de identificar 23,76 % no mesmo raio. Como apresentado na Tabela II, M-SVM2D identifica menos regiões que M-SVM3D. Com base nas tabelas III e I, pode-se afirmar que as regiões não identificadas por M-SVM2D são relevantes e que M-SVM3D possui classificação superior.

VI. CONCLUSÕES

O modelo M-SVM3D, conforme apresentado por [29] se mostrou uma boa ferramenta de apoio a decisão, com concordância com M-ENTLN acima de 89% até 120 km do radar. O trabalho de [29], sugere que, para a identificação de SWE por SVM, basta o estudo do subconjunto de pontos até 4 km. No entanto, neste estudo utilizou-se o produto CAPPI a 3 km de altitude, podendo ser este um grande contribuinte ao baixo desempenho de M-SVM2D.

A identificação feita por M-SVM2D, mostrou resultado inferior tanto de classificação, quanto de generalização, podendo ser necessário uma nova abordagem de treinamento, pois a tabela I mostra que o treinamento da técnica foi insuficiente.

Desta forma, M-SVM2D não se mostrou uma boa ferramenta de apoio a decisão, não sendo possível, até o presente momento, substituir a identificação realizada por M-SVM3D.

Como sequência deste trabalho, será gerado um novo modelo pelo treinamento da SVM, aplicando dados entre 0 e 4 km de altitude, a fim de diminuir o volume de dados e, conseqüentemente, o tempo de processamento, sem comprometer a acurácia do modelo.

REFERÊNCIAS

[1] Defesa Civil, *Reconhecimentos realizados em 2015*, Disponível em: <http://www.mi.gov.br/reconhecimentos-em-2015>. Acessado em 17 de fevereiro de 2016.

[2] Defesa Civil, *Reconhecimentos realizados em 2016*, Disponível em: <http://www.mi.gov.br/reconhecimentos-em-2016>. Acessado em 30 de agosto de 2016.

[3] F. Fabry, *Radar Meteorology Principles and Practice*, 1. ed [S.I.]: Cambridge University Press 2015.

[4] E. A. Damian, *Duas metodologias aplicadas à classificação de precipitação convectiva e estratiforme com radar meteorológico: SVM e K-means*, Dissertação (mestrado) - Programa de Pós-Graduação em Métodos Numéricos em Engenharia, Universidade Federal do Paraná. Paraná, 2011.

[5] M. A. Neto, *Mineração de dados: extração da conhecimento a partir das técnicas de visualização da informação e mineração de dados*, Dissertação (mestrado) - Programa de Pós-Graduação em Métodos Numéricos em Engenharia, Universidade Federal do Paraná. Paraná, 2008.

[6] J. Anochi, *Previsão climática de precipitação por redes neurais autoconfiguradas*, Tese - Pós-Graduação em Computação Aplicada, Instituto Nacional de Pesquisas Espaciais. São José dos Campos, 2015.

[7] R. A. Maddox, *Mesoscale convective complexes*. Bull. Am. Meteorol. Soc., v. 61, p.1374-1387, 1980.

[8] C. Doswell, *Severe Convective Storms* American Meteorological Society, Boston, 2001.

[9] R. E. Rinehart, *Radar for Meteorologists* 4 ed. Columbia: Rinehart Publications, 2004.

[10] E. Novo, *Sensoriamento Remoto: princípios e aplicações*. Edgard Blücher, 1992.

[11] C. G. Collier *Applications of Weather Radar Systems: a Guide to Uses of Radar Data in Meteorology and Hydrology*. 2th. ed. [S.I.]: John Wiley & Sons, 1996.

[12] R. Silveira, *Recognition of clutter in weather radars using polarization diversity information and artificial neural networks*. Tese, University of Essex, 1998.

[13] M. Kumjian, Principles and Applications of Dual-Polarization Weather Radar. Part I: Description of the Polarimetric Radar Variables J. Operational Meteorology, pp. 226-242, <http://mwafiles.nwas.org/jom/articles/2013/2013-JOM19/2013-JOM19.pdf>.

[14] V. Bringi, V. Chandrasekar, *Polarimetric Doppler Weather Radar: Principles and Applications*. Cambridge University Press, 2001.

[15] E. Ruzanski, V. Chandrasekar, *Nowcasting Rainfall Fields Derived from Specific Differential Phase*. Journal of applied Meteorology and Climatology, vol 51, pp. 1950-1959, 2012.

[16] J. Newman, V. Lakshmanan, P. Heinselman, M. Richman, T. Smith, T. Range-Correcting Azimuthal Shear in Doppler Radar Data. Weather and Forecasting, vol. 28, pp. 194-211, 2013.

[17] K. C. Lima, *Descargas Eléctricas Atmosféricas em Sistemas Convectivos de Mesoescala no Sul da América do Sul*. Dissertação (mestrado) - Programa de Pós-Graduação em Meteorologia, Universidade Federal do Pelotas. Rio Grande do Sul, 2005.

[18] M. J. Murphy, R. L. Holle *Warnings of cloud-to-ground lightning hazard based on total lightning and radar information*. Second Conf. on Meteorological Applications of Lightning Data, Amer. Meteorol. Soc., Atlanta, GA, paper 2.2 (CD-ROM).

[19] Schultz, C., Petersen, W., Carey, L., 2011. Lightning and Severe Weather: A Comparison between Total and Cloud-to-Ground Lightning Trends. *American Meteorological Society*, vol. 26, pp.744-755.

[20] C. Liu, S. Heckman, *Total Lightning Data and Real-Time Severe Storm Prediction*. WMO Technical Conference on Meteorological and Environmental Instruments and Methods of Observation, 2012.

[21] P. Gatlin, S. Goodman, *A Total Lightning Trending Algorithm to Identify Severe Thunderstorms*. Journal of atmospheric and oceanic technology, vol. 27, pp. 744-755, 2010.

[22] V. VAPNIK, A. CHERVONENKIS, *Theory of Pattern Recognition*. Akademie Verlag, 1974.

[23] S. Haykin, *Neural Networks - A Comprehensive Foundation*. Macmillian College Publishing, inc., 1994.

[24] R. Semolini, *Support Vector Machines, Inferência Transdutiva e o Problema de Classificação*. Dissertação (mestrado), Pós-graduação da Faculdade de Engenharia Elétrica UNICAMP, 2002.

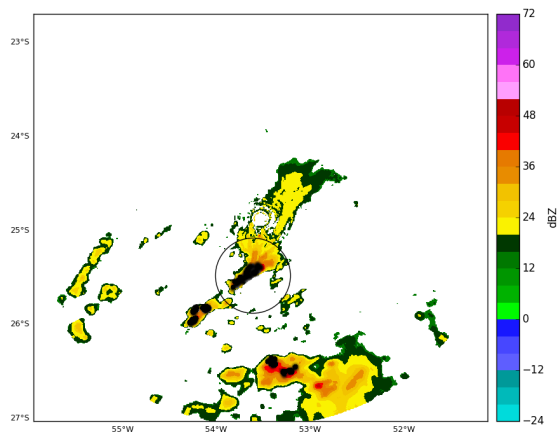
[25] A. Smola, P. Bartlett, B. Schölkopf, D. Schuurmans, *Advances in large margin classifiers* MIT Press, 2000.

[26] J. Helmus, S. Collis, *Dual-polarization phase shift processing with the Python ARM Radar Toolkit* JORS 2016, doi: 10.5334/jors.119

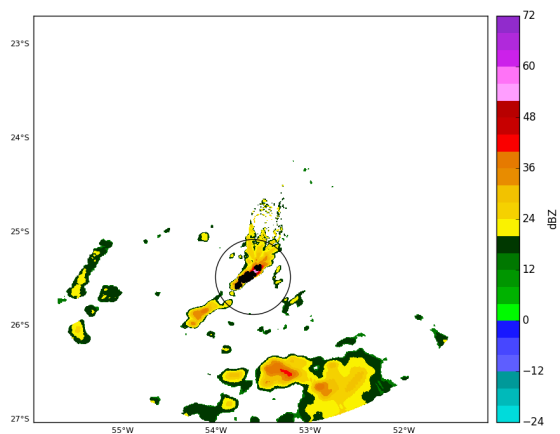
[27] M. Liepert, *Topological fields chunking for German with SVM's: optimizing SVM parameters with GA's* International conference on recent advances, 2003.

[28] A. J. Smola, *A Tutorial on Support Vector Regression*. Bernhard Schölkopf -Statistics and Computing archive Volume 14 Issue 3, August 2004, p. 199-222.

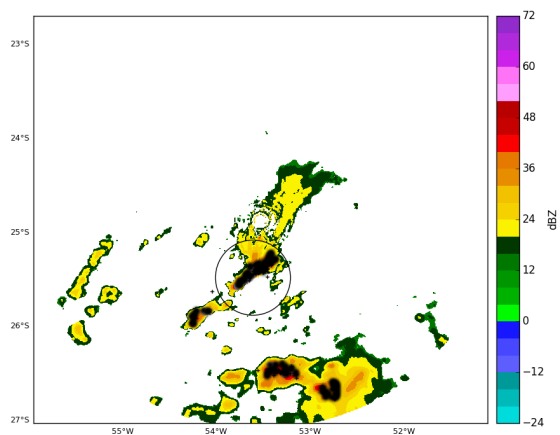
[29] T. Silva, *Identificação de evento de tempo severo utilizando técnicas de aprendizagem de máquina em dados de radar polarimétrico*, Dissertação (mestrado) - Programa de Pós-Graduação em Métodos Numéricos em Engenharia, Universidade Federal do Paraná. Paraná, 2017.



(a) M-SVM3D

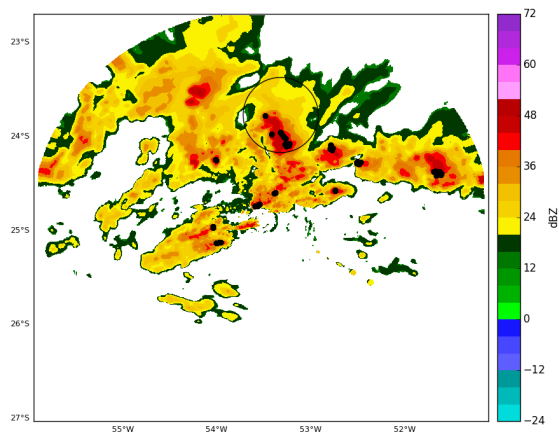


(b) M-SVM2D

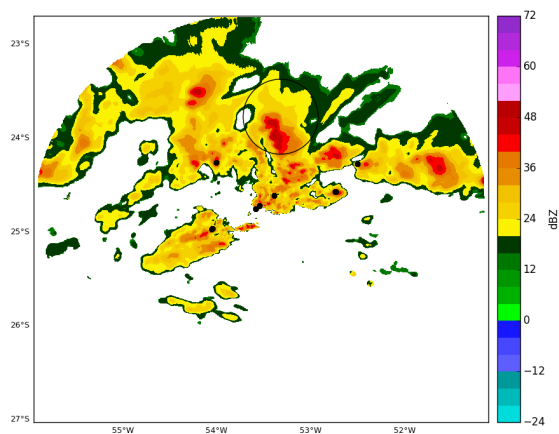


(c) M-ENTLN

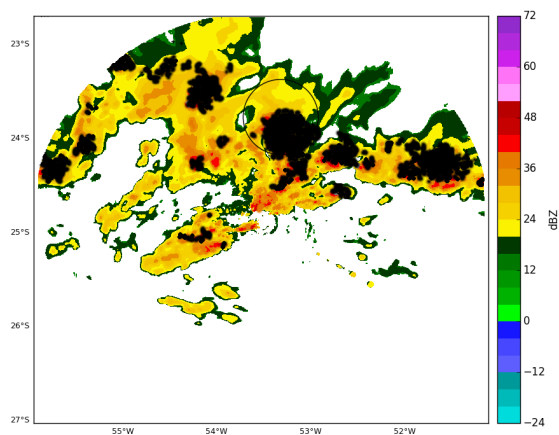
Figura 2: Saída obtida com a utilização dos modelos M-SVM3D (a), M-SVM2D (b) e M-ENTLN (c) no dado de 28/12/2015 às 21:15. As imagens representam os pontos identificados pelos respectivos modelos (em preto) representados sobre a imagem do atributo refletividade, CAPPI para M-SVM2D e PPI para os demais modelos. A região circulada mostra onde ocorreu SWE dentro de 30 minutos.



(a) M-SVM3D



(b) M-SVM2D



(c) M-ENTLN

Figura 3: Saída obtida com a aplicação dos modelos M-SVM3D (a), M-SVM2D (b) e M-ENTLN (c) no dado de 23/11/2015, às 13:30. As imagens representam os pontos identificados pelos respectivos modelos (em preto) representados sobre a imagem do atributo refletividade, CAPPI para M-SVM2D e PPI para os demais modelos. A região circulada mostra onde ocorreu SWE dentro de 30 minutos.