

IDENTIFICAÇÃO DA BANDA BRILHANTE EM DADOS DE RADAR METEOROLÓGICO UTILIZANDO A TÉCNICA DE AGRUPAMENTO K-MEANS

Camila de Oliveira

Programa de Pós-Graduação em Métodos Numéricos em Engenharia, UFPR.
Centro Politécnico da UFPR, Caixa Postal 19011, CEP:81531-980, Curitiba, Paraná, Brasil.
cmlolv@gmail.com

Sérgio Scheer

Programa de Pós-Graduação em Métodos Numéricos em Engenharia, UFPR.
Centro Politécnico da UFPR, Caixa Postal 19011, CEP:81531-980, Curitiba, Paraná, Brasil.
scheer@ufpr.br

Cesar Beneti

Sistema Meteorológico do Paraná.
Centro Politécnico da UFPR, Caixa Postal 19100, CEP:81531-980, Curitiba, Paraná, Brasil.
beneti@simepar.br

RESUMO

A banda brilhante ocorre quando uma camada de água no estado líquido se forma em uma partícula de gelo em processo de derretimento, formando uma estreita camada horizontal de elevada refletividade, próxima à isoterma de 0°C, podendo assim, superestimar a previsão de precipitação. Para que não ocorram erros na interpretação dos dados, a banda brilhante precisa ser identificada e corrigida. Este trabalho utiliza a técnica de agrupamento *k-means* para identificar a região afetada pela banda brilhante e em seguida, aplica uma correção nessa região. Após a correção os dados são convertidos em taxa de precipitação para que se possa fazer uma comparação com os dados de precipitação observados. Nota-se grande melhora nos dados após a correção, o que indica que o *k-means* é um método eficaz para a identificação da região afetada.

PALAVRAS CHAVE. K-Means, Radar Meteorológico, Banda Brilhante.

Área Principal: Programação Matemática.

ABSTRACT

The bright band occurs when a layer of water in the liquid state is formed in an ice particle melting process, forming a narrow horizontal layer of high reflectivity, near the 0°C isotherm and may thus overestimate the forecast precipitation. To prevent any errors in data interpretation, the bright band needs to be identified and corrected. This work uses the *k-means* clustering technique to identify the region affected by the bright band and then apply a correction in that region. After correction, the data is converted to precipitation rate so that a comparison with the observed precipitation data can be made. A great improvement in the data after correction is noted, which indicates that the *k-means* is an effective method for the identification of the affected region.

KEYWORDS. K-Means, Weather Radar, Bright Band.

Main Area: Mathematical Programming.

1. Introdução

O termo banda brilhante (BB) é utilizado na meteorologia para se referir a uma camada de alta refletividade associada com o derretimento de gelo (ZHANG; LANGSTON; HOWARD, 2008). Ou seja, acima da isoterma de 0°C está presente basicamente gelo, que ao cair e atingir essa isoterma entra em processo de derretimento, formando-se em sua volta uma camada de água no estado líquido. Isto provoca um aumento considerável no valor da refletividade obtida pelo radar meteorológico.

Essa região com alta refletividade pode causar superestimação na previsão de precipitação se uma correção não for aplicada (ZHANG; QI, 2010). O objetivo desse trabalho é identificar as áreas afetadas pela banda brilhante nos dados do radar meteorológico do Sistema Meteorológico do Paraná (SIMEPAR), para que uma correção possa ser feita evitando erros na previsão de precipitação.

Na seção seguinte é apresentado o funcionamento do radar meteorológico para a coleta dos dados de refletividade. A seção 3 apresenta a técnica de agrupamento *k-means* e, sua aplicação para fazer a identificação da região afetada pela banda brilhante é apresentada na seção 4.

2. Radar Meteorológico

O SIMEPAR possui um radar meteorológico Banda-S Doppler, instalado em Teixeira Soares - PR, realizando medições numa área de até 480 km de raio, abrangendo os estados do Paraná e Santa Catarina, parte do Centro-Sul de São Paulo e norte do Rio Grande do Sul. Ele monitora chuvas e ventos em intervalos de tempo de até 5 minutos e possui algoritmos para identificação de tempestades severas, estimativa de chuvas e previsão de deslocamento de tempestades de muito curto prazo (até 2 horas), possibilitando interação do usuário para integração com outros sistemas de monitoramento do tempo (raios e satélites) e emissão de alertas em curto intervalo de tempo (BENETI et al., 2004).

A antena do radar rotaciona sobre o eixo vertical (varredura azimutal), com diferentes elevações da antena. Uma sequência completa de várias varreduras azimutais com diferentes elevações é chamada de volume de varredura (CALVETTI; BENETI; FILHO, 2003). Os dados que são enviados pelo radar estão em formato numérico e depois são transformados em informações tais como: data, hora, localização, volume de chuva e a altura (em graus) da elevação da antena e uma matriz de dados, onde estão armazenados os valores obtidos pelo radar (Neto, 2008).

Esses valores possuem coordenadas esféricas com três graus de liberdade (r , θ , ϕ), onde r é o *range* (distância do alvo até radar), θ é o azimute (ângulo em relação ao norte) e ϕ é a elevação da antena.

Uma das variáveis medidas pelo radar é a refletividade (z), que é o fator entre as ondas eletromagnéticas emitidas e as que voltam para o radar, depois de passarem por uma nuvem e serem irradiadas em todas as direções pelas gotas de chuva. Devido ao grande domínio que os valores de refletividade possuem, é conveniente utilizar sua representação em uma escala logarítmica:

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

onde Z é o valor de refletividade em escala logarítmica, medida em dBZ, e z é o valor de refletividade em escala linear (mm^6/m^3).

Para uma melhor percepção humana, os dados de refletividade são convertidos em imagens (produtos), as quais podem ser visualizadas de diferentes modos. Um desses produtos é o PPI (*Plan Position Indicator*) que fornece a projeção num plano horizontal, obtido através de uma varredura em azimute em coordenadas polares, para um ângulo de elevação constante. Deve-se observar que, conforme a distância em relação ao radar aumenta, a altitude do feixe com relação ao solo também aumenta. A Figura 1 apresenta a varredura do radar em um determinado ângulo de elevação com plano fixo e ilustra também como a altitude do feixe aumenta com a distância.

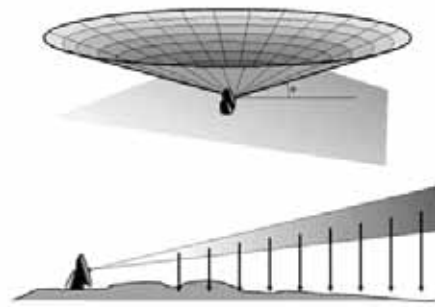


Figura 1: Produto PPI. Fonte: MeteoPT, 2013.

Os dados em coordenadas esféricas são transformados para coordenadas cartesianas e então, são plotados com o centro da imagem indicando a localização do radar. A Figura 2 mostra o PPI da variável refletividade obtidos da varredura realizada pelo radar no dia 02/12/2009 às 23:00, num ângulo de elevação da antena de 0,5° e alcance de 200 km. As cores quentes representam valores de refletividade mais elevados, enquanto que cores frias representam valores mais baixos.

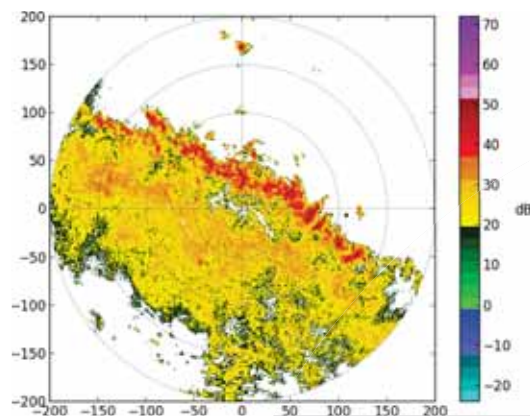


Figura 2: PPI da variável refletividade do dia 02/12/2009 às 23:00, num ângulo de elevação de 0,5°. Fonte: A autora, 2015.

A partir dos dados de refletividade (em coordenadas esféricas) de um volume de varredura obtidos pelo radar, serão aplicados os métodos para a detecção e correção da banda brilhante, que serão descritos na seção 4.

3. Algoritmo K-Means

O *k-means* é uma técnica de Análise de Agrupamentos, que permite classificar um conjunto de dados em um determinado número k de grupos, baseado na dissimilaridade entre os dados. Uma vantagem principal deste algoritmo é a sua simplicidade computacional, o que o torna um bom candidato para uma variedade de aplicações (THEODORIDIS; KOUTROUMBAS, 2009).

A técnica inicia com a escolha do número de grupos e o posicionamento do centro de cada grupo no espaço de atributos. Esse centro é chamado de centroide, e é o ponto médio do grupo. Após a definição da localização destes centroides, cada dado é associado ao centroide mais próximo através do cálculo do quadrado da distância euclidiana.

Quando todos os dados já tiverem sido associados a algum centroide, estes são recalculados (SOUZA, 2008). Este procedimento é repetido inúmeras vezes até que não haja mais diferença (ou que ela seja mínima) nos centros dos grupos.

Assim, dado um conjunto de dados $X = (x_1, x_2, \dots, x_N)$, o método segue os seguintes passos (ARTHUR; VASSILVITSKII, 2006):

1. Escolher arbitrariamente k centros iniciais $C = \{C_1, C_2, \dots, C_k\}$;
2. Atribuir cada amostra x_i ($i = 1, \dots, N$) para o centro mais próximo;
3. Determinar os novos centroides de cada grupo;
4. Repetir 2 e 3 até não ocorrer mudanças nos centroides de duas iterações consecutivas.

A parte mais importante desta técnica é a sua inicialização, isto é, a localização dos centroides iniciais, pois diferentes centroides irão gerar diferentes resultados (SOUZA, 2008). Isto pode ser observado na Figura 3, onde é feito o agrupamento dos pontos A, B, C, D, E, F e G utilizando o algoritmo *k-means* com duas inicializações diferentes dos centroides.

Na Figura 3(a) foram escolhidos como centroides iniciais os pontos A, B e C e os grupos resultantes foram $\{A\}$, $\{B, C\}$ e $\{D, E, F, G\}$. Já na Figura 3(b) os centroides iniciais foram A, D e F e obteve-se os grupos $\{A, B, C\}$, $\{D, E\}$ e $\{F, G\}$. O valor encontrado pela função objetivo é bem maior para o agrupamento obtido na Figura 3(a) do que para o agrupamento obtido na Figura 3(b).

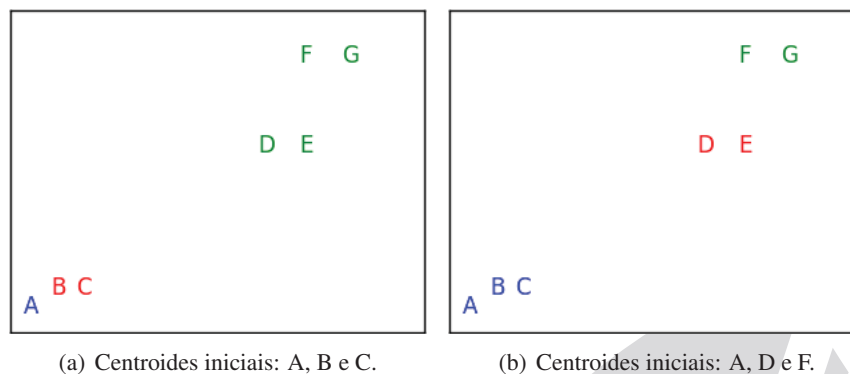


Figura 3: Exemplo do algoritmo *k-means* com diferentes inicializações dos centroides. Fonte: Adaptado de Jain, 1999.

Sendo assim, Arthur e Vassilvitskii (2006) propuseram uma maneira de escolher centroides iniciais para o algoritmo *k-means*, dando origem ao algoritmo *k-means++*. Este algoritmo escolhe os centroides iniciais baseado nos seguintes passos:

- 1a. Escolher aleatoriamente um centro inicial C_1 dentro do conjunto de dados;
- 1b. Escolher um novo centro C_i , tal que $C_i = x' \in X$ com probabilidade $\frac{D(x')^2}{\sum_{x \in X} D(x)^2}$;
- 1c. Repetir o passo 1b até se obter k centros;
- 2-4. Prosseguir com o algoritmo *k-means* padrão.

No passo 1b, $D(x)$ é a distância dos dados para o centroide mais próximo do grupo ao qual está incluído. Este passo atribui uma probabilidade de cada amostra ser um bom centroide inicial.

Segundo Arthur e Vassilvitskii (2006) o novo método produz um algoritmo com melhor desempenho e de forma consistente encontra um agrupamento melhor que o *k-means* padrão.

4. Aplicação e Resultados

Para a identificação da região afetada pela banda brilhante foi utilizada a técnica de agrupamento *k-means*, aplicado utilizando a linguagem de programação *Python*.

O *Python* possui uma vasta biblioteca de módulos e pacotes que é distribuída com o interpretador. E existem também disponíveis, muitas bibliotecas escritas por terceiros, que implementam diversos recursos além da biblioteca padrão (BORGES, 2010). Uma delas é o *Scikit-learn*, que oferece uma ferramenta simples e eficiente para a mineração e análise de dados. Neste trabalho foi utilizado o módulo `sklearn.cluster.KMeans()` desta biblioteca, que agrupa os dados utilizando

o algoritmo *k-means*. Mais informações da biblioteca podem ser encontradas em Pedregosa et al. (2011).

Este módulo tem como principais parâmetros de entrada a matriz de dados, a quantidade de grupos desejados e a forma de inicialização dos centroides. Esta inicialização pode ser feita de forma aleatória, através do algoritmo *k-means++* ou pode ser passado um vetor com as posições desejadas dos centroides.

Para este trabalho foi escolhida a inicialização utilizando o algoritmo *k-means++*, pois como citado na seção 3, ele oferece uma convergência mais rápida do algoritmo *k-means*, assim como um melhor agrupamento dos dados.

4.1. Aplicação do Algoritmo K-Means

Os dados de refletividade obtidos pelo radar meteorológico podem ser classificados em dois tipos de precipitação: convectiva e estratiforme. Os sistemas convectivos estão associados a fortes campos verticais de vento e altas taxas pluviométricas, enquanto os sistemas estratiformes possuem extensas áreas cobertas, baixas velocidades de vento e taxas de precipitações menores, facilitando a formação da banda brilhante.

Devido a essas grandes diferenças, escolheu-se fazer a classificação dos dados em três grupos distintos, um para a região convectiva, um para a região estratiforme não afetada pela banda brilhante e outro para a região afetada.

Os dados utilizados para esta classificação são provenientes da variável refletividade captada pelo radar meteorológico, em um raio de alcance de até 200 km. Através desses valores de refletividade, em coordenadas esféricas, pode-se obter algumas variáveis úteis para a identificação das três regiões. Essas variáveis são:

- Z : valor da variável refletividade para o PPI de elevação $0,5^\circ$ (em dBZ). Quanto maior a refletividade, maior a quantidade de gotas de chuva, ou maior o diâmetro das gotas, e, portanto, mais intensa é a precipitação.
- Z_{\max} : é o valor de máxima refletividade vertical. Ou seja, para cada posição (r, θ) dos dados do radar, encontra-se o maior valor de refletividade de todas as elevações do volume de varredura.
- $h_{Z_{\max}}$: indica a altitude correspondente ao valor Z_{\max} . Valores de Z_{\max} próximos à altitude da isoterma de zero grau estão relacionadas à existência de banda brilhante, já quando Z_{\max} é encontrado em altitudes mais elevadas, significa que a chuva é convectiva.
- h_{30} : maior altitude onde se encontra o valor de refletividade de 30 dBZ. Alturas elevadas mostram que a tempestade em grandes altitudes é mais intensa, o que está associado à existência de chuva convectiva.
- h_{45} : análogo ao anterior, porém com valor de refletividade de 45 dBZ.
- VIL: o termo VIL, do inglês *Vertically Integrated Liquid-Water*, representa o conteúdo de água líquida integrada verticalmente. Ele é definido por:

$$VIL = 3.44 \times 10^{-6} \sum \left(\frac{z_i + z_{i+1}}{2} \right)^{4/7} \Delta h, \quad (2)$$

onde z_i e z_{i+1} são os valores de refletividade (em mm^6/m^3) no limite inferior e superior, respectivamente, de uma camada de amostra e Δh é a diferença de altura (em metros) entre os limites dessa camada, a variável é expressa em Kg/m^2 . Valores maiores estão relacionados a maior quantidade de precipitação, ou seja, precipitação convectiva.

Essas variáveis formam as colunas da matriz de dados (X) que será um dos parâmetros de entrada para a função `sklearn.cluster.KMeans()`. Aplica-se então, a técnica *k-means* para o volume de dados. Como já citado, os outros parâmetros de entrada para a função são o modo de inicialização dos grupos (*k-means++*) e a quantidade de agrupamentos desejado, neste caso, $k = 3$.

4.2. Correção da Banda Brilhante

Após a identificação da região de banda brilhante é necessário fazer a correção desses dados, com o intuito de realizar comparações com dados observados de precipitação. Assim, é possível determinar se o *k-means* conseguiu identificar corretamente a área afetada.

Para a correção utiliza-se a técnica descrita em Zhang e Qi (2010), a qual parametriza o perfil vertical de refletividade (PVR) obtido com os dados afetados pela banda brilhante.

Nesta técnica, é obtido um PVR para cada elevação. Estes PVRs são calculados a partir da média azimutal para cada *range*. Faz-se o cálculo apenas a partir dos dados da área de BB ao invés de todos os dados do radar, pois, muitas vezes a banda brilhante não está uniformemente distribuída em todas as regiões de cobertura do radar. Assim, se um único PVR médio for calculado a partir de todo o volume, a intensidade do pico da banda brilhante no PVR pode ser reduzida devido à contribuição de dados não afetados. Se este PVR for aplicado para correções em todos os lugares, então haveria subcorreções em áreas de BB e supercorreções em áreas não afetadas. Então, isolar a área de BB do resto da precipitação torna a correção mais eficaz.

Após o cálculo do PVR, ele é então parametrizado para a obtenção de cinco parâmetros, descritos a seguir e ilustrados na Figura 4:

- BB_{top} : Maior altitude encontrada no PVR;
- BB_{pico} : Altitude correspondente ao maior valor de refletividade encontrado no PVR;
- BB_{base} : Altitude correspondente ao menor valor de refletividade do PVR, abaixo do BB_{pico} . Se o menor valor de refletividade for menor que o limite de 29 dBZ, então BB_{base} receberá o valor da altitude correspondente a esse limite. Esse limite é usado para evitar correções excessivas (ZHANG; QI, 2010);
- α : É obtido pelo método de mínimos quadrados através dos pontos do PVR entre BB_{topo} e BB_{pico} ;
- β : É obtido também pelo método de mínimos quadrados, porém entre os pontos do PVR entre BB_{pico} e BB_{base} .

Após o cálculo desses parâmetros, pode-se então calcular um fator de correção para os dados da região da banda brilhante. Esse fator de correção, representado por FC , é calculado da seguinte maneira:

$$FC = \begin{cases} \alpha[h(r) - BB_{pico}] + \beta[BB_{pico} - BB_{base}], & \text{se } h(r) > BB_{pico} \\ \beta[h(r) - BB_{base}], & \text{se } h(r) \leq BB_{pico} \end{cases} \quad (3)$$

Assim, o valor de refletividade corrigido (dBZ_c) é obtido pela equação:

$$dBZ_c = dBZ_o - FC, \quad (4)$$

onde dBZ_o é o valor de refletividade original do radar na região da banda brilhante.

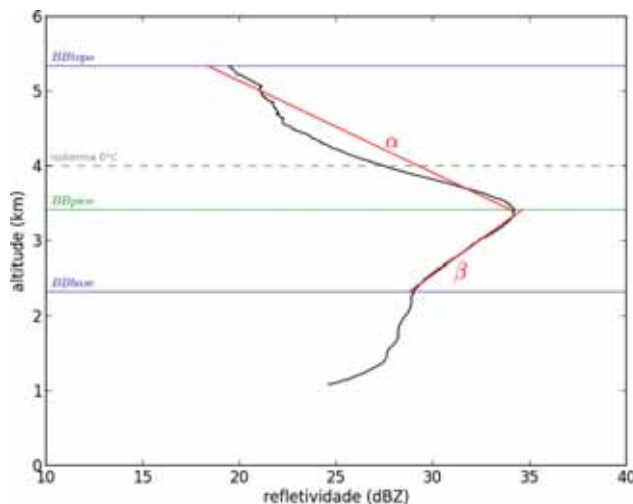


Figura 4: Parâmetros obtidos a partir de um PVR de dados afetados pela banda brilhante. A isoterma de 0°C foi obtida através de radiossondagem. Fonte: A autora, 2015.

4.3. Resultados

Foram selecionados 68 eventos de chuva no Paraná com possível contaminação por banda brilhante. Para cada um desses casos, foram aplicados os métodos de identificação e correção a todos os volumes de dados coletados no intervalo de uma hora de ocorrência do evento. A correção é aplicada para todas as elevações dos volumes de dados, porém, para comparação com dados pluviométricos utilizou-se apenas a primeira elevação (0,5°), por ser a mais próxima da superfície terrestre, e por consequência, mais próxima dos pluviômetros.

A Figura 5 apresenta os resultados obtidos para o evento ocorrido em 16/07/2010 às 03:00. Os PPIs mostrados são da elevação de 0,5° para o primeiro volume de varredura da hora. Em 5(a) é mostrado o PPI sem correção, em 5(b) a classificação dos hidrometeoros pelo algoritmo *k-means* e em 5(c) tem-se o PPI dos dados de reflectividade após a correção.

Devido a diferença de intervalo de tempo no qual são coletados os dados do radar (10 minutos) e dos pluviômetros (15 minutos), optou-se por utilizar os valores de precipitação no intervalo de uma hora. Para os dados de precipitação obtidos pelos pluviômetros, os valores são acumulados. Já para os dados do radar, decidiu-se por utilizar o valor de reflectividade máxima dos PPIs de 0,5° em cada hora medida.

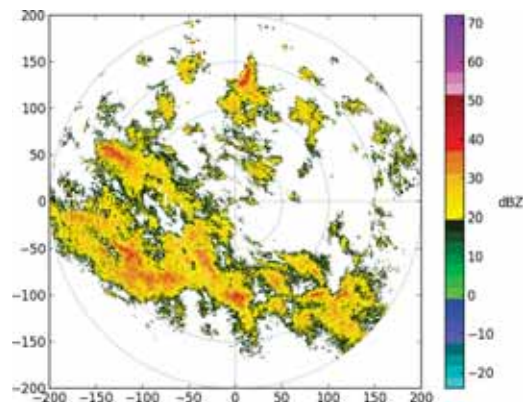
Para a avaliação dos resultados os dados de reflectividade horários foram convertidos para taxa de precipitação, utilizando-se a relação *Z-R* dada pela seguinte relação empírica exponencial:

$$z = aR^b, \quad (5)$$

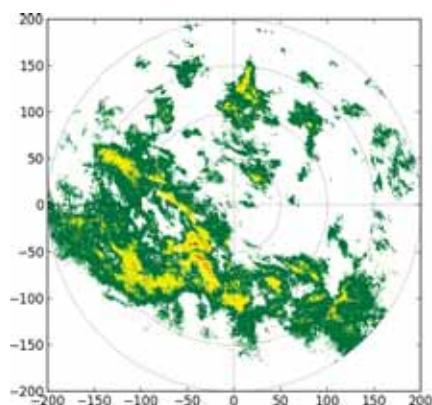
onde *R* é a taxa de precipitação em mm/h, *z* é o fator de reflectividade do radar em mm⁶/m³, e *a* e *b* são constantes empíricas. Os valores destas constantes variam de acordo com a região do globo e tipos de precipitação (DAMIAN, 2011). Existem várias relações usadas na literatura, com diferentes valores para *a* e *b*. Porém, Santos (2014) desenvolveu recentemente no SIMEPAR, pesquisas relacionadas à relação *Z-R* e concluiu que para os dados do radar meteorológico de Teixeira Soares, a melhor relação para o cálculo da taxa de precipitação é com as constantes empíricas *a* = 52 e *b* = 2,7.

Após a conversão, calculou-se o Viés e o Raiz do Erro Médio Quadrático (REMQ) dos dados antes e após a correção, para os 68 casos.

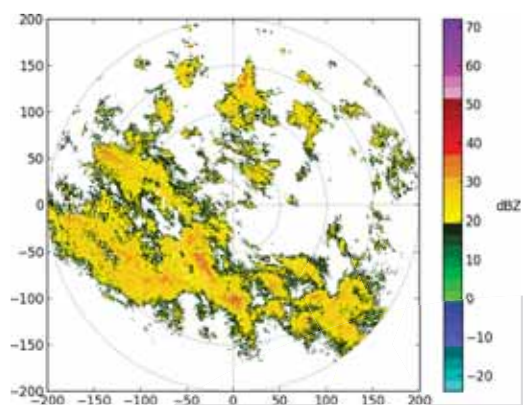
A Figura 6 apresenta a taxa de precipitação obtida com o cálculo da relação *Z-R* para o evento do dia 16/07/2010 às 03:00. Observa-se que os PPIs estão acumulados por hora. Nesta Figura, tem-se também os valores de precipitação obtidos pelos pluviômetros para cada estação,



(a) PPI sem correção.



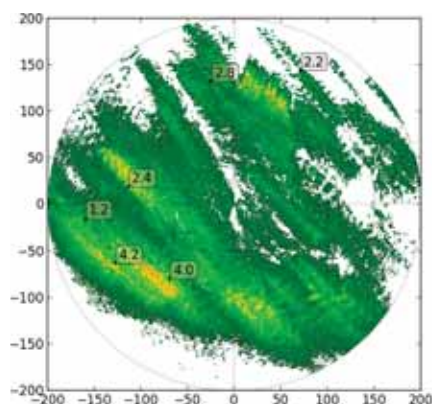
(b) Identificação dos hidrometeoros.



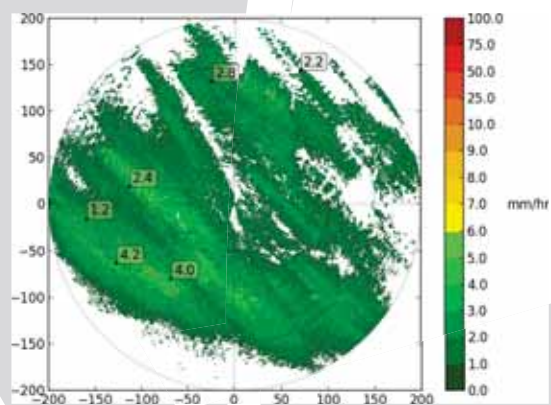
(c) Valores corrigidos.

Figura 5: PPI da elevação de 0,5° do dia 16/07/2010 às 03:00. Em (a) tem-se o PPI dos dados originais, em (b) a identificação da banda brilhante pelo *k-means* e em (c) o PPI após a correção. Fonte: A autora, 2015.

nos pontos onde se tem pares de pontos válidos (ambos com valores maiores que zero). É apresentada também, uma tabela com os valores da taxa de precipitação dos dados originais, dos dados corrigidos pelo algoritmo *k-means* e da precipitação medida pelos pluviômetros para os pares de pontos apresentados nas imagens e outra tabela com o cálculo do Viés e REMQ.



(a) Sem correção.



(b) *k-means*.

Figura 6: Taxa de precipitação (16/07/2010 às 03:00). Fonte: A autora, 2015.

Tabela 1: Valores da taxa de precipitação (em mm/hr) para 16/07/2010 às 3:00.

Dados originais	<i>k-means</i>	Pluviômetros
8,0	4,5	4,2
5,4	3,1	2,4
2,3	2,3	2,2
3,4	3,3	1,2
7,0	4,2	4,0
2,7	2,7	2,8

Tabela 2: Valores de Viés e REMQ para 16/07/2010 às 3:00.

	Viés	REMQ
Dados originais	2,0	2,5
<i>k-means</i>	0,5	0,9

Além dos erros para cada caso, foram calculados os erros para o número total de pontos obtidos dos casos selecionados, mostrados na Tabela 3.

Tabela 3: Viés e REMQ para o total de dados obtidos dos 68 eventos.

	Viés	REMQ
Dados originais	2,0	3,2
<i>k-means</i>	1,0	2,4

5. Conclusões

A estimativa da precipitação de chuva com a melhor acurácia possível é fundamental para uma adequada e realística representação da distribuição espacial da precipitação. Uma maneira de estimar a precipitação é a utilização de dados de refletividade obtidos por radares meteorológicos. Assim, há a necessidade de eliminar ou corrigir fatores nestes dados que podem prejudicar esta estimativa de precipitação.

Um destes fatores é a banda brilhante, que pode gerar uma superestimativa no cálculo da taxa de precipitação. O foco deste trabalho foi justamente identificar esta região para que uma correção pudesse ser aplicada.

Para isto, utilizou-se a técnica de agrupamento *k-means* para identificar nos dados de radar meteorológico regiões afetadas pela banda brilhante e a seguir, utilizou-se um método para a correção dos valores desta região.

A técnica *k-means*, utilizando a inicialização dos centroides com o algoritmo *k-means++*, obteve melhores resultados de Viés e REMQ quando comparados com os dados sem correção. Sendo assim, pode-se dizer que a técnica *k-means* mostrou-se eficaz para identificação da banda brilhante.

Como trabalho futuro, pretende-se comparar os resultados obtidos pelo *k-means* com outros métodos da literatura.

Referências

- Arthur, D. e Vassilvitskii S.** K-means++: The Advantages of Careful Seeding, *Technical Report*, Stanford, 2006.
- Beneti, C. A. A., Jusevicius, M., Calvetti, L. e Filho, A. J. P.** (2004), Características Elétricas e Hidrometeorológicas de uma Tempestade Severa no Paraná, *XIII Congresso Brasileiro de Meteorologia*, Fortaleza, CE.
- Borges, L. E.** *Python para Desenvolvedores*, 2010.
- Calvetti, L., Beneti, C. A. A. e Filho, A. J. P.** (2003), Integração do Radar Meteorológico Doppler do SIMEPAR e uma Rede de Pluviômetros para a Estimativa da Precipitação, *Anais XI SBSR*, pp. 1137, Belo Horizonte, MG.
- Damian, E. A.** *Duas metodologias aplicadas à classificação de precipitação convectiva e estratiforme com radar meteorológico: SVM e k-means*, Dissertação de Mestrado, UFPR, 2011.
- Jain, A. K., Murty, M. N. e Flynn, P. J.** (1999) Data Clustering: A Review, *ACM Computing Surveys*, 31, 264-323.
- METEOPT.** Fórum de meteorologia, Dez 2013, (<http://www.meteopt.com/>).
- Neto, M. A. S.** *Mineração visual de dados: extração do conhecimento a partir das técnicas de visualização da informação e mineração de dados*, Dissertação de Mestrado, UFPR, 2008.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. e Duchesnay, E.** (2011), Scikit-learn: Machine Learning in Python, *Journal of Machine Learning Research*, 12, 2825-2830.
- Santos, T. N.** *Redes Neurais Artificiais e Relação ZR Aplicadas à Estimativa de Chuva*, Dissertação de Mestrado em Métodos Numéricos, UFPR, 2014.
- Souza, T. A.** *Agrupamento de séries temporais de vento para avaliação da disponibilidade de geração de usinas eólicas*, Dissertação de Mestrado em Engenharia Elétrica, UFRJ, 2008.
- Theodoridis, S. e Koutroumbas, K.** *Pattern Recognition*, Academic Press, 2009.
- Zhang, J., Langston, C. e Howard, K.** (2008), Brightband Identification Based on Vertical Profiles of Reflectivity from the WSR-88D, *Journal of Atmospheric and Oceanic Technology*, 25, 1859-1872.
- Zhang, J. e Qi, Y.** (2010), A Real-Time Algorithm for the Correction of Brightband Effects in Radar-Derived QPE, *Journal of Hydrometeorology*, 11, 1157-1171.