

08 a 11 de Outubro de 2018
Instituto Federal Fluminense
Búzios - RJ

PREDIÇÃO DE OCORRÊNCIA DE DESCARGAS ELÉTRICAS NUVEM-SOLO UTILIZANDO PREVISÕES DO MODELO NUMÉRICO BRAMS

Luan Gaspar Pinto de Melo¹ – luanlgpm@gmail.com

Alex de Almeida Fernandes² – alex.fernandes@cptec.inpe.br

Glauston Roberto Teixeira de Lima³ – glauston.lima@cemaden.gov.br

Stephan Stephany⁴ – stephan.stephany@inpe.br

Alan James Peixoto Calheiros⁴ – alan.calheiros@inpe.br

¹ Programa de Pós-graduação em Computação Aplicada (CAP),

Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, SP, Brasil

² Centro de Previsão de Tempo e Estudo Climáticos (CPTEC),

Instituto Nacional de Pesquisas Espaciais (INPE), Cachoeira Paulista, SP, Brasil

³ Centro Nacional de Monitoramento e Alertas de Desastres Naturais (CEMADEN),

São José dos Campos, SP, Brasil

⁴ Laboratório Associado de Computação e Matemática Aplicada (LAC),

Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, SP, Brasil

Resumo. A predição de ocorrência de eventos convectivos severos permite a emissão de alertas meteorológicos, possibilitando a mitigação de diversas catástrofes em potencial. Em muitos casos, os modelos numéricos de previsão do tempo não conseguem simular de forma eficiente esse tipo de eventos. Em contrapartida, dado o grande volume e diversidade de dados meteorológicos, a aplicação de técnicas de mineração de dados torna-se cada vez mais comum em Meteorologia. No caso de atividade convectiva, utilizando-se dados passados, é possível identificar padrões característicos nas previsões de um modelo ao associá-las aos correspondentes campos de densidade de ocorrência de descargas elétricas atmosféricas nuvem-solo por meio de um algoritmo de aprendizado de máquina, no caso, um conjunto de redes neurais. Assim, estes padrões podem ser detectados nas futuras previsões geradas pelo modelo, de forma a fazer a predição de ocorrência de descargas, as quais estão associadas a atividade convectiva. Neste trabalho, a abordagem proposta foi aplicada para o modelo BRAMS, sendo analisada a capacidade de predição de ocorrência de descargas pelas redes neurais propostas para um conjunto de eventos e sua utilidade para predição de eventos convectivos como ferramenta auxiliar na previsão de tempo operacional.

Palavras-chave: Eventos convectivos, Modelos numéricos de previsão de tempo, Descargas elétricas atmosféricas, Redes neurais artificiais, Mineração de dados

1. INTRODUÇÃO

Eventos convectivos atmosféricos podem causar tempestades severas que implicam em ventos muito fortes e grande volume de precipitação, o qual pode causar inundações ou deslizamentos de terra, entre outros efeitos. Os danos socioeconômicos e ambientais podem ser grandes, enfatizando a importância de sua previsão em tempo hábil para emissão de alertas e para eventuais medidas preventivas. Modelos numéricos de previsão de tempo simulam muito bem a dinâmica das massas de ar, mas ainda tem capacidade limitada de prever eventos convectivos, sendo sua previsão condicionada à experiência do meteorologista previsor e à disponibilidade de outros dados e de imagens meteorológicas. Em contrapartida, a atual disponibilidade de supercomputadores com crescente capacidade de processamento permite a execução de modelos numéricos com resoluções espaciais correspondentes a grades geométricas cada vez mais finas, melhorando a capacidade de prever esses eventos, mas ainda insuficientemente.

Por outro lado, o enorme volume de dados e imagens meteorológicas gerados por sensores embarcados em satélites ou no solo, como no caso de estações meteorológicas automáticas, torna difícil, senão impossível sua pronta análise pelo meteorologista no monitoramento em tempo real do estado da atmosfera ou na sua previsão. Isso ocasionou a aplicação de técnicas diversas de mineração de dados para análise e interpretação dessa grande massa de dados, constituindo um tema de pesquisa corrente em Meteorologia. Entre outros objetivos, a mineração de dados permite identificar padrões ocultos nos dados, relativos à ocorrência de uma dada condição ou de um fenômeno relacionado a esses dados.

Neste trabalho, explora-se a correlação entre atividade convectiva e ocorrência de descargas elétricas nuvem-solo (Lang & Rutledge, 2011). Propõe-se identificar nas previsões do modelo padrões característicos da ocorrência de descargas, correspondentes a um conjunto de variáveis atmosféricas com valores em faixas determinadas utilizando um algoritmo de aprendizado de máquina, no caso, um conjunto de redes neurais artificiais. Na fase de treinamento, podem-se encontrar esses padrões a partir de dados de previsões passadas do modelo numérico e dos correspondentes dados de descargas elétricas atmosféricas. Uma vez que as redes neurais foram treinadas, os seus pesos foram ajustados para detectar os padrões desejados nas futuras previsões do modelo, os quais indicarão a ocorrência de descargas e, indiretamente, a ocorrência de evento convectivo. Utilizou-se o modelo numérico BRAMS (Freitas et al., 2017) com resolução espacial de 5 km e campos de densidade de ocorrência de descargas gerados pelo software EDDA (Strauss et al., 2010, Strauss et al., 2013) a partir de dados da Rede Integrada Nacional de Detecção de Descargas (RINDAT, Naccarato & Pinto Jr., 2009). A abordagem proposta constitui um aperfeiçoamento de uma abordagem anterior (Lima & Stephany, 2013a, Lima & Stephany, 2013b) que havia sido desenvolvida para dados do modelo ETA (Mesinger et al., 1988) com resolução espacial de 20 km, e que também utilizava o software EDDA e dados da rede RINDAT, mas com uma única rede neural. Note-se que a previsão é dificultada ao se trabalhar com uma resolução espacial menor.

A abordagem aqui apresentada é ainda inédita em termos de pesquisa em Meteorologia, exceto pela produção bibliográfica anterior do próprio grupo de pesquisa. Foram encontradas dificuldades típicas da mineração de dados, tais como manipulação de um grande volume de dados, dados incompletos, alta capacidade de processamento (demandada aqui no treinamento das redes neurais) e a necessidade de ajustar iterativamente parâmetros relativos ao pré-processamento dos dados em função dos resultados, repetindo inúmeras vezes o ciclo de pré-processamento, aprendizado, teste e análise.

2. DADOS E METODOLOGIA

2.1 Dados meteorológicos

A abordagem proposta foi testada utilizando-se uma área de $1^{\circ} \times 1^{\circ}$ entre os municípios de Bauru e Presidente Prudente, no estado de São Paulo, definida pela faixa de latitudes entre $21,5^{\circ}$ e $22,5^{\circ}$ Sul e pela faixa de longitudes $49,5^{\circ}$ e $50,5^{\circ}$ e correspondendo a aproximadamente 598 pontos de grade do modelo BRAMS na resolução espacial de 5 km. Esse modelo gera saídas a cada 6 horas (horário UTC), correspondentes a análise da 0 h, e às previsões das 6, 12, 18, 24, ... 72 h. Campos de densidade de ocorrência de descargas elétricas atmosféricas nuvem-solo foram gerados para a mesma área e horários utilizando o software EDDA e dados da rede RINDAT. Para cada saída/previsão do modelo, os campos de densidade de descargas foram gerados integrando-se o número de ocorrências de descargas num intervalo de tempo de 1 hora centrado no horário da previsão. As instâncias da base de dados de treinamento foram particionadas aleatoriamente em 5 conjuntos, de forma a se fazer o treinamento do mesmo número de redes neurais, com base num esquema proposto em Gupta et al. (2006). Cada instância é composta pela densidade de ocorrência de descargas nuvem-solo normalizada num ponto de grade do modelo e pelos valores de 55 variáveis do modelo para um determinado horário de previsão/análise. Essas variáveis foram selecionadas com auxílio de meteorologistas como sendo indicativas da presença/ausência de atividade convectiva. A lista com essas 55 variáveis aparece na Tabela 1. Além disso, foi feita uma análise de sensibilidade de cada variável no resultado da rede neural, constatando-se que todas as 55 variáveis têm influência semelhante. Os dados destinados ao treinamento foram sanitizados, ou seja, filtrados para a remoção de campos constantes nulos ou indefinidos, de forma que as datas/horários correspondentes foram excluídos.

As instâncias de treinamento referem-se a 16 meses, correspondendo ao período de novembro de 2016 até janeiro de 2017 e parte de fevereiro de 2018, enquanto que, para os testes, foram selecionados alguns dias do mês de fevereiro de 2018 que não foram usados no treinamento. Os dados foram selecionados de forma a se utilizar um esquema de hold out "cronológico" simulando o uso operacional, ou seja, para uma determinada data de previsão, utilizam-se instâncias de dias/meses anteriores. Os testes aqui apresentados referem-se à semana de 11 a 17 de fevereiro de 2018, em que houve grande instabilidade na área considerada, havendo dias de alta densidade de descargas em quase todos os pontos da grade e também dias com densidade nula de descargas em todos os pontos da grade. Essa alternância de densidades dificulta muito sua previsão, razão pela qual esses dias foram escolhidos para teste. Neste trabalho, optou-se por utilizar apenas as previsões de 24 horas do BRAMS como estudo de caso para a prova de conceito da abordagem proposta.

2.2 Redes neurais empregadas

As redes neurais empregadas executam regressões, produzindo resultados numéricos e não categóricos, sendo empregadas 5 redes de mesma configuração, correspondentes às partições aleatórias do conjunto de treinamento. Essas redes são do tipo Perceptron multicamadas com retropropagação disponibilizadas no pacote Neuralnet (Fritsch & Guenther, 2016) da linguagem e ambiente computacional R (R Core Team, 2018) para computação estatística e visualização. Esse pacote possibilita configurar a rede neural escolhida por meio da escolha de funções de ativação, do número de camadas ocultas e do número de neurônios de cada uma, entre outros. Foi escolhida a tangente hiperbólica como função de ativação. A arquitetura de rede foi otimizada empiricamente, avaliando-se os resultados com diferentes

Tabela 1 - As 55 variáveis meteorológicas do modelo BRAMS que foram selecionadas (variáveis multinível são definidas pelo nível de pressão em hPa).

Pressão reduzida ao nível do mar	Temperatura 1000 hPa
Temperatura em 2m	Temperatura 925 hPa
Temperatura do ponto de orvalho em 2m	Temperatura 800 hPa
CINE	Temperatura 700 hPa
CAPE	Temperatura 500 hPa
Água precipitável	Temperatura 300 hPa
Componente V do vento 1000 hPa	Temperatura 250 hPa
Componente V do vento 925 hPa	Umidade relativa 1000 hPa
Componente V do vento 800 hPa	Umidade relativa 925 hPa
Componente V do vento 700 hPa	Umidade relativa 800 hPa
Componente V do vento 500 hPa	Umidade relativa 700 hPa
Componente V do vento 300 hPa	Umidade relativa 500 hPa
Componente V do vento 250 hPa	Umidade relativa 300 hPa
Componente U do vento 1000 hPa	Umidade relativa 250 hPa
Componente U do vento 925 hPa	Omega 1000 hPa
Componente U do vento 800 hPa	Omega 925 hPa
Componente U do vento 700 hPa	Omega 800 hPa
Componente U do vento 500 hPa	Omega 700 hPa
Componente U do vento 300 hPa	Omega 500 hPa
Componente U do vento 250 hPa	Omega 300 hPa
Altura geopotencial 1000 hPa	Omega 250 hPa
Altura geopotencial 925 hPa	Umidade específica 1000 hPa
Altura geopotencial 800 hPa	Umidade específica 925 hPa
Altura geopotencial 700 hPa	Umidade específica 800 hPa
Altura geopotencial 500 hPa	Umidade específica 700 hPa
Altura geopotencial 300 hPa	Umidade específica 500 hPa
Altura geopotencial 250 hPa	Umidade específica 300 hPa
	Umidade específica 250 hPa

números de camadas ocultas e diferentes números de neurônios por camada, obtendo-se o melhor resultado com 3 camadas ocultas que têm, respectivamente, 14, 10 e 9 neurônios, sendo que todos os neurônios recebem os dados de todos os outros da camada anterior. Nessas avaliações, os dados foram divididos para treinamento e teste utilizando-se um esquema de validação cruzada simples. As 55 entradas correspondentes às variáveis do modelo alimentam a camada de entrada e são direcionadas a todos os neurônios da primeira camada oculta e a última camada fornece uma saída numérica única, correspondente ao valor numérico da densidade de descargas normalizada que foi estimada pela rede. Adicionalmente, no treinamento de cada uma dessas 5 redes, utilizam-se 3 sementes aleatórias diferentes para geração aleatória dos pesos iniciais, resultando então em 15 redes neurais.

Após a fase de treinamento, as redes já tem os pesos ajustados e podem ser utilizadas na fase de generalização, onde vão fornecer como resultado a densidade de descargas para cada pixel da previsão considerada. Idealmente, num cenário operacional, a cada novo dia, novos dados tornam-se disponíveis, devendo ser inseridos na base de dados e realizado um novo treinamento. Neste trabalho, optou-se por considerar duas classes de densidade de descargas: alta (classe F) e moderada/baixa/nula (classes NF), de forma a tentar identificar apenas atividade convectiva forte, que tipicamente apresenta alta densidade de descargas. As saídas numéricas das 15 execuções das redes são categorizadas para uma dessas duas classes e são então combinadas num esquema de votação (polling) gerando uma classificação conjunta única (F ou NF). Na categorização, utilizou-se o mesmo limiar utilizado na categorização dos dados das densidades de descargas: maiores que 0.01 como F e o complemento como NF. Para a votação, adotaram-se dois critérios de classificação para a classe F considerando a concordância dos votos: abaixo de 10% ou acima de 90%. Deste modo, tenta-se prever eventos fortes específicos, detectados por poucas redes, ou então aqueles mais comuns, que foram detectados pela maioria das redes. Além disso, como na maior parte do tempo não ocorrem descargas elétricas, há uma preponderância de valores nulos ou muito baixos de densidade, que podem levar a uma tendência de as redes neurais na fase de generalização a classificar corretamente apenas as instâncias NF. Essa característica foi corrigida por meio de um balanceamento de 1:2 entre as classes F e NF, distinguidas pelo limiar de densidade normalizada de 0,01. Dessa forma, 1/3 do total de instâncias são da classe F, sendo esse balanceamento assegurado nas partições aleatórias.

3. RESULTADOS DOS TESTES DE PREDIÇÃO

Apresentam-se a seguir os resultados da predição para a semana de 11 a 17 de fevereiro de 2018, ressaltando que o objetivo é fazer a predição dos pontos de grade com densidade de descargas alta, classificados como fortes (F). A Figura 1 permite comparar o número de pontos de grade (instâncias) classificados/preditos pelas redes como fortes com o número real/observado. Observa-se que, apesar da subestimação ou da superestimação, as redes foram capazes de prever a tendência de aumento ou diminuição e, considerando o caso de teste específico, essa tendência constituiria uma informação útil ao meteorologista previsor.

A matriz de confusão é formada pelo número de instâncias de cada classe preditos acertadamente ou erroneamente, e permite calcular métricas tais como a acurácia, a probabilidade de detecção (*probability of detection*, POD) e a taxa de alarme falso (*false-alarm ratio*, FAR) Considerando-se que o objetivo é detectar instâncias da classe F, as instâncias corretamente classificados como da classe F constituem os verdadeiros positivos (VP) e aquelas incorretamente classificados assim, os falsos positivos (FP). Por outro lado, instâncias corretamente classificadas como da classe NF são os verdadeiros negativos e as incorretamente classificados assim, os falsos negativos (FN). A acurácia é dada pelo quociente $(VP+VN)/(total)$, o POD por $VP/(VP+FN)$ e o FAR por $FP/(VP+FP)$. A Figura 2, a seguir, ilustra a acurácia e a probabilidade de detecção para essa mesma semana. Os valores da acurácia tendem a ser altos devido à facilidade de a rede prever instâncias da classe NF, que ocorrem em maior número, enquanto que os valores de POD são baixos, mas úteis para identificar a tendência de ocorrerem instâncias F. As Tabelas 2 e 3 mostram as matrizes de confusão para a predição/classificação referentes a dois dias da semana considerada com características opostas: o dia 15/02/2018, que só apresentou instâncias da classe NF e dia 16/02/2018, com muitas instâncias da classe F. As mesmas tabelas mostram os valores correspondentes das métricas de classificação, acurácia, POD e FAR. Pode-se observar a alta

acurácia da matriz de confusão para a primeira data, com apenas 9 instâncias fora da diagonal da classe NF que foram mal classificadas como F (falsos negativos). Na segunda data, ocorreram 409 instâncias da classe F, mas apenas 143 foram corretamente classificadas (verdadeiros positivos), enquanto que as restantes 266, incorretamente como NF (falsos negativos), enquanto que 169 das 189 instâncias da classe NF foram corretamente classificadas (verdadeiros negativos), restando 20 mal classificadas como F (falsos positivos).

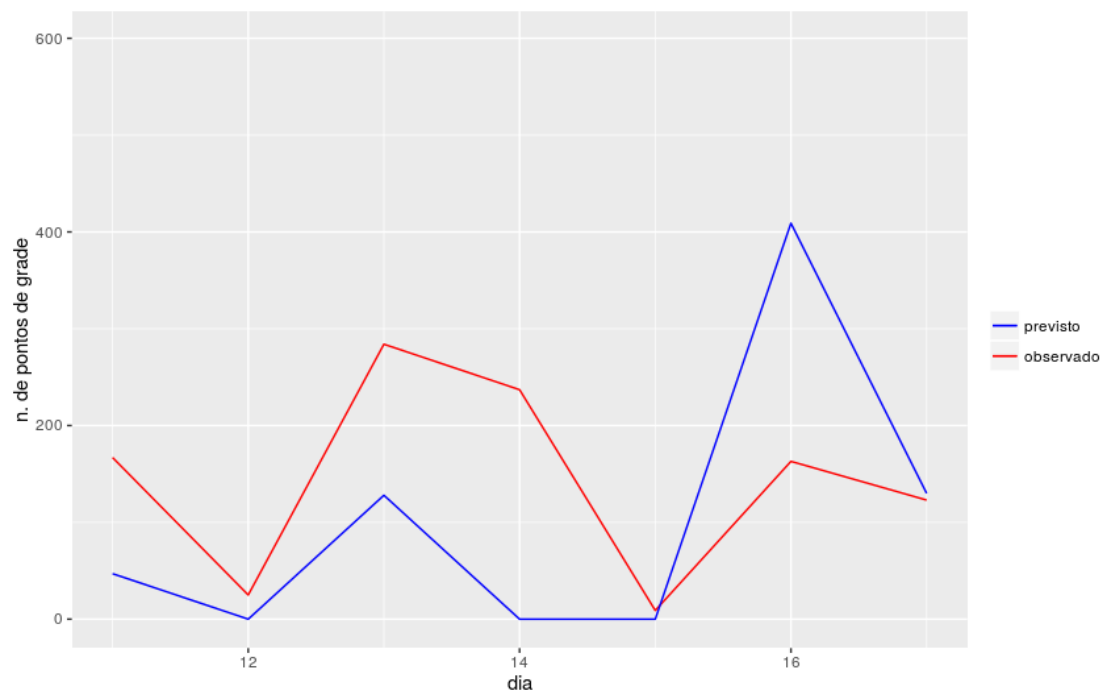


Figura 1 - Comparação do número de instâncias previstas corretamente como sendo da classe F e o número de instâncias observadas como sendo da classe F para a semana 11-17 de fevereiro de 2018 para a área considerada.

Pode-se ressaltar que a semana escolhida para testes apresenta grande variabilidade de atividade convectiva, alternando predominância de instâncias da classe F e da classe NF, constituindo casos de difícil predição. Além disso, o *hold out* “cronológico” representa um teste muito mais difícil para o classificador, no caso as redes neurais, do que a validação cruzada, uma vez que a classificação tem que ser feita para eventos completamente novos, que não apareceram no treinamento. Outra observação relevante é que no treinamento, a acurácia das redes sempre foi superior a 99% (0.9). Assim, reproduziram-se condições da previsão de tempo operacional, e os resultados apresentados demonstram o potencial da abordagem proposta como ferramenta auxiliar para o meteorologista predictor. O *spin-up* do modelo BRAMS corresponde às 12 horas iniciais de previsão após as quais o modelo gera previsões mais precisas, mas que ainda assim não reproduzem as condições reais da atmosfera, enquanto que os dados de descargas são reais, embora com pequenos erros de localização. Esse descompasso entre dados do modelo e dados reais dificulta a classificação.

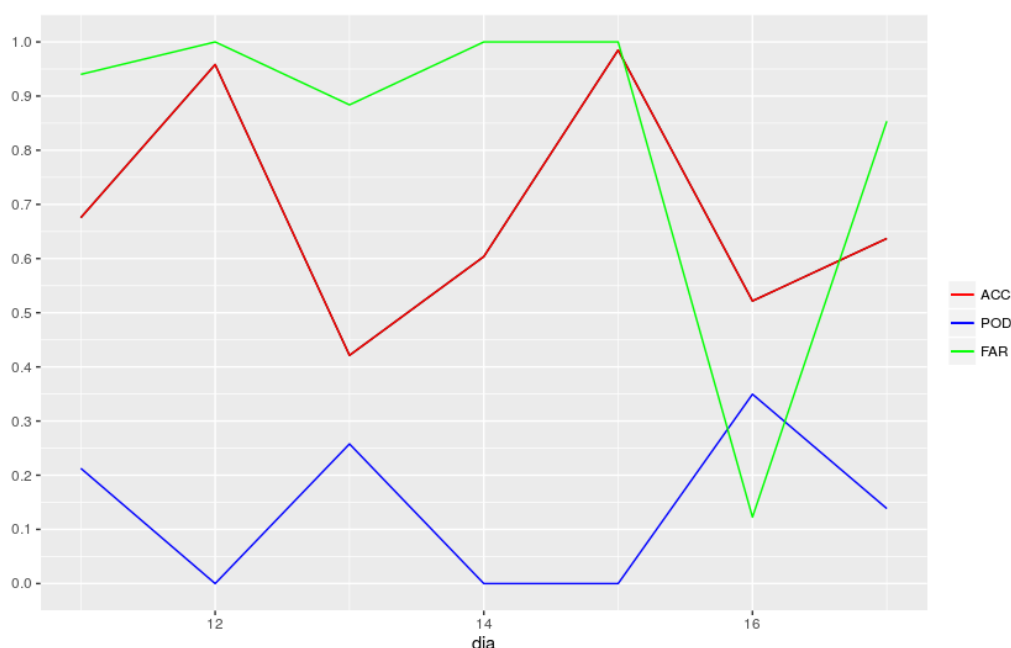


Figura 2 - Evolução da acurácia (ACC), da probabilidade de detecção (POD) e da taxa de alarme falso (FAR) referentes às instâncias da classe F para a semana 11-17 de fevereiro de 2018 na área considerada.

Tabela 2 - Matriz de confusão e valores das métricas de classificação para o dia 15/02/2018, que apresentou 598 instâncias da classe NF e nenhuma da classe F.

15/02/2018	Observado		Métricas		
Previsto		F	NF	ACC	0.98
	F	0	9	POD	0.0
	NF	0	589	FAR	1.0

Tabela 3 - Matriz de confusão e valores das métricas de classificação para o dia 16/02/2018, que apresentou 409 instâncias da classe F e 189 da classe NF.

16/02/2018	Observado		Métricas		
Previsto		F	NF	ACC	0.52
	F	143	20	POD	0.35
	NF	266	169	FAR	0.12

4. CONCLUSÕES

A abordagem aqui apresentada constitui uma evolução de uma pesquisa iniciada anos atrás, particularmente aquela apresentada em Lima & Stephany (2013b), que também utilizava redes neurais. Os resultados aqui apresentados, utilizando dados do modelo BRAMS e da rede RINDAT, mostram que essa abordagem tem potencial de ser utilizada como ferramenta auxiliar na previsão de tempo operacional, e pretende-se avaliá-la num futuro próximo em duas instituições governamentais, o Centro de Previsão de Tempo e Estudo Climáticos (CPTEC) do Instituto Nacional de Pesquisas Espaciais (INPE) e o Centro Nacional de Monitoramento e Alertas de Desastres Naturais (CEMADEN), ambas do Ministério da Ciência, Tecnologia, Inovações e Comunicações. É importante ressaltar que essa abordagem foi aplicada ao modelo BRAMS, mas pode ser facilmente portada para qualquer outro modelo meteorológico regional.

No tocante aos resultados apresentados, foi feita a previsão para dias de uma semana de fevereiro de 2018 que apresentou grande variabilidade meteorológica, com alternância de dias com muita ou nenhuma atividade convectiva. O esquema adotado de *hold out* “cronológico”, no qual a classificação/previsão tem que ser feita para eventos completamente novos, isto é, não contidos nos dados de treinamento, reproduz um caso real de previsão. Esses dois aspectos tornam a classificação muito difícil, mas foi possível obter resultados aceitáveis.

Considerando-se os erros inerentes da previsão do modelo numérico BRAMS, pode-se dizer que a capacidade preditiva do conjunto de redes neurais demonstrou ser suficiente para servir de auxílio ao meteorologista previsor, considerando o conjunto de eventos que serviram de teste. No caso, foi feita a previsão da ocorrência de descargas elétricas nuvem-solo da classe forte, as quais podem ser associadas a atividade convectiva severa. Essa rede neural foi obtida pelo treinamento com partições aleatórias das instâncias compostas por um conjunto de variáveis de previsões passadas do modelo BRAMS e pela densidade de ocorrência de descargas elétricas atmosféricas correspondentes ao ponto de grade do modelo e ao horário de previsão. A capacidade preditiva das redes neurais é expressa pelo seu desempenho de classificação ao identificar nas previsões futuras do modelo ocorrência de densidade de descargas forte ou não-forte nos pontos de grade.

Pretende-se continuar os testes com dados relativos a períodos mais extensos, uma vez que quanto o aumento do número de instâncias de treinamento tende a melhorar a acurácia da classificação. Outros aspectos podem ser explorados, tais como sazonalidade ou posição geográfica, ou mesmo uma análise de sensibilidade quanto ao conjunto de variáveis do modelo que foram selecionadas, num enfoque de seleção de atributos. Além disso pode-se explorar melhor essa capacidade preditiva com mais classes de densidade de descargas, por exemplo, com três classes, correspondentes a forte, moderada e fraca/nula. A existência dessa classe intermediária moderada faria com que o número de falsos positivos e falsos negativos relativos à classe forte tenda a cair, resultando num aumento de instâncias classificadas como moderadas, o que representaria um erro qualitativamente menor, a ser comprovada mediante extensivos testes.

Acknowledgements

O autor Stephan Stephany agradece ao CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) pela bolsa 307460/2015-0.

REFERÊNCIAS

- Freitas, S.R. and others, (2017), The Brazilian developments on the Regional Atmospheric Modeling System (BRAMS 5.2): An integrated environmental model tuned for tropical areas. *Geoscientific Model Development* 10, 189–222.
- Fritsch, S., Guenther, F., (2016), Neuralnet: Training of Neural Networks. R package version 1.33. <https://CRAN.R-project.org/package=neuralnet>
- Gupta, K. M., Aha, D. W., Moore, P., (2006), "Rough set feature selection algorithms for textual case-based classification", 8th European Conference on Case-Based Reasoning, Olundeniz, Turquia.
- Lima, G.R.T., Stephany, S., (2013), A new classification approach for detecting severe weather patterns. *Computers & Geosciences*, 57, 158–165.
- Lang, T.J., Rutledge, S.A., (2011), A framework for the statistical analysis of large radar and lightning datasets: results from STEPS 2000. *Monthly Weather Review* 139(8), 2536–2551.
- Lima, G.R.T., Stephany, S., (2013), Training a neural network to detect patterns associated with severe weather. *Learning and Nonlinear Models*, 11, 123–152.
- Mesinger, F., et alii, (1988): The step-mountain coordinate: model description and performance for cases of alpine lee cyclogenesis and for a case of an Appalachian redevelopment. *Monthly Weather Review*, 116 (7), 1493–1518.
- Naccarato, K., Pinto Jr, O., (2009) Improvements in the detection efficiency model for the Brazilian lightning detection network (BrasilDAT). *Atmospheric Research* 91(2), 546–563.
- Naccarato, K., Pinto Jr, O., (2009) Improvements in the detection efficiency model for the Brazilian lightning detection network (BrasilDAT). *Atmospheric Research* 91(2), 546–563.
- R Core Team, (2018), R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- Strauss, C., Stephany, S., Caetano, M. (2010), "A ferramenta EDDA de geração de campos de densidade de descargas atmosféricas para mineração de dados meteorológicos", XXXIII Congresso Nacional de Matemática Aplicada e Computacional, Águas de Lindóia, SP, vol. 3, 269–275.
- Strauss, C., Rosa, M.B., Stephany, (2013), Spatio-temporal clustering and density estimation of lightning data for the tracking of convective events. *Atmospheric Research* 134, 87–99.

PREDICTION OF OCCURRENCE OF GROUND-TO-CLOUD ELECTRICAL DISCHARGES USING FORECASTS OF THE NUMERICAL MODEL BRAMS

Abstract. *The prediction of occurrence of severe convective events allows to issue meteorological alerts, allowing to alleviate potential catastrophes. In many cases, weather forecast numerical models are not able to efficiently simulate such events. On the other side, considering the large volume and diversity of meteorological data, the use of data mining techniques has become widespread in Meteorology. In the case of convective activity, by means of past data, it is possible to identify characteristic patterns in model forecasts and to associate them to the corresponding fields of density of occurrence of ground-to-cloud atmospheric electrical discharges using a machine learning algorithm, given here by a set of neural networks. As a consequence, such patterns can be detected in the future model forecasts, in order to predict the occurrence of electrical discharges, which are associated to convective activity. In this work, the proposed approach was employed for the BRAMS model, and the neural networks ability for predicting the occurrence of discharges was analyzed for a set of events, as well as its usefulness for predicting convective events as an ancillary tool in the operational weather forecasting.*

Keywords: *Convective events, Weather prediction numerical models, Atmospheric electrical discharges, Artificial neural networks, Data mining*