

# A Ferramenta EDDA de Geração de Campos de Densidade de Descargas Atmosféricas para Mineração de Dados Meteorológicos

**Cesar Strauss**

INPE - Coordenação de Ciências Espaciais e Atmosféricas (CEA)  
12227-010, São José dos Campos, SP  
E-mail: cstrauss@cea.inpe.br

**Stephan Stephany**

INPE - Laboratório Associado de Computação e Matemática Aplicada (LAC)  
12227-010, São José dos Campos, SP  
E-mail: stephan@lac.inpe.br

**Mirian Caetano**

INPE - Centro de Previsão de Tempo e Estudos Climáticos (CPTEC)  
12630-000, Cachoeira Paulista, SP  
E-mail: miriam.caetano@cptec.inpe.br

**Resumo:** *A ferramenta EDDA estima a densidade de ocorrência de descargas elétricas atmosféricas a partir do registro de eventos individuais, para uma extensão geográfica e intervalo de tempo selecionados. O método utilizado é o estimador de núcleo gaussiano com janela adaptativa, sendo gerados arquivos em formato ASCII adequados a algoritmos de mineração e em formato de grade binário para a ferramentas de visualização meteorológica GRADS. Parâmetros específicos podem ser ajustados de forma a se poder correlacionar a densidade com outros dados meteorológicos. A ferramenta permite também visualizar animações varrendo os registros de descargas por meio de uma janela deslizante, de forma a acompanhar a evolução temporal de estruturas convectivas. A pronta disponibilidade de dados de descargas permite animações que permitem monitorar os eventos meteorológicos com atividade elétrica em tempo quasi-real.*

**Palavras-chave:** *mineração de dados, previsão meteorológica, eventos convectivos, estimador de densidade*

## 1 Introdução

Este trabalho apresenta a ferramenta EDDA, que estima a densidade de ocorrência de descargas elétricas atmosféricas para uma extensão geográfica e intervalo de tempo selecionados. Os dados brutos de descargas, contendo os registros individuais em formato ASCII são gerados pela Rede Integrada Nacional de Detecção de Descargas Atmosféricas (RINDAT), fornecidos pelo CPTEC/INPE. A ferramenta implementa o estimador de núcleo gaussiano com janela adaptativa, sendo gerados arquivos em formato ASCII adequados a algoritmos de mineração e em formato de grade binário para a ferramentas de visualização meteorológica GRADS. Parâmetros específicos podem ser ajustados de forma a se poder correlacionar a densidade com outros dados, objetivando seu uso na mineração de dados meteorológicos. A ferramenta permite também visualizar animações varrendo os registros de descargas por meio de uma janela deslizante, de forma a acompanhar a evolução temporal de estruturas convectivas. A pronta disponibilidade de dados de descargas permite animações que permitem monitorar os eventos meteorológicos com

atividade elétrica em tempo quasi-real, possibilitando ao meteorologista uma visão instantânea de estruturas convectivas e de sua evolução recente. Assim, a ferramenta proposta tem potencial operacional em meteorologia.

O campo de densidade é calculado por uma técnica de estimação de densidade, a estimação de núcleo (*kernel estimation*), descrita em [8] e [9], no caso adotando-se uma função gaussiana, a qual permite gerar um campo de densidade de ocorrências que é suave e permite delimitar mais claramente a região de atividade convectiva a partir das descargas, as quais são muito esparsas no espaço e no tempo. A aplicação dessa abordagem foi proposta originalmente em [6] e [7], com o intuito de rastrear a atividade convectiva eletricamente ativa por meio das descargas nuvem-solo, tendo sido originalmente implementada na linguagem de programação MatLab. Posteriormente, foi desenvolvido o software CAC [3] [4], escrito na linguagem C++ para compilação com o compilador GNU/gcc e que gera arquivos de saída referente à densidade de descargas (ou outros parâmetros) em formato ASCII. Este software faz também a visualização dos campos gerados por meio do pacote GNU/Gnuplot. Este software foi empregado num estudo de correlação entre atividade convectiva e ocorrência de descargas elétricas atmosféricas nuvem-solo [1] e foi empregado no escopo de um projeto de mineração de dados meteorológicos.

O software CAC mostrou-se útil como ambiente de experimentação, entretanto, um novo software foi desenvolvido visando a operacionalização da estimação da densidade de descargas em tempo quasi-real. Assim, foi desenvolvida a ferramenta EDDA - Estimação de Densidade de Descargas Atmosféricas. Esse é também o nome de um compêndio de contos e poesias da literatura pré-nórdica, que inclui Thor, o deus do trovão.

## 2 Metodologia

### 2.1 Estimação de densidade

O estimador de núcleo gaussiano bidimensional clássico pode ser escrito como [9]:

$$\hat{f}(x) = \frac{1}{nh^2} \sum_{i=1}^n K\left(\frac{d(x, x_i)}{h}\right) \quad (1)$$

onde:

$\hat{f}(x)$  é a estimativa de densidade do ponto de grade  $x$ ,

$n$  é o número de amostras,

$d(x, x_i)$  é a distância euclidiana do ponto de grade  $x$  à amostra  $x_i$ ,

$K(r) = \frac{1}{2\pi} \exp(-\frac{1}{2}r^2)$  é a função de núcleo gaussiano bidimensional,

$h$  é a janela de influência centrada no ponto de grade  $x$

A distância euclidiana  $d(x, x_i)$  pode ser estimada a partir das diferenças de latitude e longitude dos pontos, incorporando uma correção geométrica devido à curvatura da Terra:

$$d(A, B) \approx (\text{lat}_A - \text{lat}_B)^2 + (\text{lon}_A - \text{lon}_B)^2 \cos\left(\frac{\text{lat}_A + \text{lat}_B}{2}\right) \quad (2)$$

O parâmetro  $h$  pode ser escolhido de forma manual, caso a caso, de maneira a encontrar um resultado visual satisfatório. Entretanto, um esquema automático é desejável para poder incorporar o método a um fluxo contínuo de dados. Uma idéia é otimizar o valor de  $h$  para otimizar o valor médio do erro quadrático integrado (MISE):

$$\text{MISE}(\hat{f}) = E \int \{\hat{f}(x) - f(x)\}^2 dx \quad (3)$$

Assumindo uma distribuição gaussiana, o valor ótimo que minimiza o MISE será [9]

$$h_{\text{opt}} = n^{-1/6} \sigma \quad (4)$$

Uma possível escolha para  $\sigma$  é a média da variância marginal [9]:

$$\sigma^2 = \frac{1}{2}(s_{xx}^2 + s_{yy}^2) \quad (5)$$

onde:

$$s_{xx} = \sigma_x^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \left( \frac{1}{n} \sum_{i=1}^n x_i \right)^2$$

$$s_{yy} = \sigma_y^2 = \frac{1}{n} \sum_{i=1}^n y_i^2 - \left( \frac{1}{n} \sum_{i=1}^n y_i \right)^2$$

Um esquema análogo [8], calcula a largura de janela  $h$  e o desvio padrão  $\sigma$  separadamente para o conjunto de coordenadas  $x$  e  $y$ :

$$\hat{f}(x) = \frac{1}{nh_x h_y} \sum_{i=1}^n K_1 \left( \frac{x - x_i}{h_x} \right) K_1 \left( \frac{y - y_i}{h_y} \right) \quad (6)$$

onde:

$$K_1(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right)$$

$$h_x = n^{-1/6} \sigma_x$$

$$h_y = n^{-1/6} \sigma_y$$

Os esquemas acima descritos calculam a largura de janela  $h$  a partir dos dados e o mantém fixo para todos os pontos da grade. Em princípio, melhores resultados podem ser obtidos permitindo variar a janela  $h$ , de acordo com a densidade local.

O seguinte esquema [9] atribui um valor de janela  $h_i$  para cada evento:

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h_i^2} K \left( \frac{d(x, x_i)}{h_i} \right) \quad (7)$$

onde:

$$h_i = h \lambda_i$$

$$\lambda_i = (\tilde{f}(x_i)/g)^{-\alpha}$$

onde  $h$  é a janela ótima global,  $h_i$  é a largura local,  $\lambda_i$  é o fator de largura de banda local,  $\tilde{f}(x_i)$  é uma estimativa prévia da densidade local usando janela fixa e  $\alpha$  é um parâmetro entre 0 e 1 que define o grau de influência da densidade local sobre a janela local. Os  $\lambda_i$  são normalizados por  $g$ , que é a média geométrica de  $\tilde{f}(x_i)$

Outra possibilidade [3] [4] [6] [7] é atribuir uma janela  $h(x)$  para cada ponto de grade:

$$\hat{f}(x) = \frac{1}{nh(x)^2} \sum_{i=1}^n K \left( \frac{d(x, x_i)}{h(x)} \right) \quad (8)$$

onde:

$$h(x) = n^{-1/6} \sigma(x)$$

$$\sigma^2(x) = \frac{1}{n} \sum_{i=1}^n d(x, x_i)^2 - \left( \frac{1}{n} \sum_{i=1}^n d(x, x_i) \right)^2$$

## 2.2 Avaliação dos esquemas de janela fixa e adaptativa

Realizou-se uma avaliação dos esquemas de janela fixa e adaptativa para reestimação de densidade expostos na seção anterior. Com essa finalidade, geraram-se eventos aleatórios a partir de uma distribuição arbitrária, supostamente desconhecida, que será estimada pela metodologia presente para os diferentes esquemas de janela adaptativa, os quais foram avaliados pelo valor médio do erro quadrático MISE (Equação 3).

Para gerar essa distribuição, empregamos um algoritmo de geração de terreno fractal, empregado em computação gráfica [5]. Neste método, uma grade bidimensional é preenchida com valores de altitude que lembram uma cadeia de montanhas. No nosso caso, reinterpretemos o eixo z como a densidade de probabilidade local.

No algoritmo, chamado “subdivisão diamante-quadrado”, subdivide-se uma grade quadrada, calculando-se o valor do centro de cada quadrado. O resultado é uma grade de “diamantes”. Novamente, calcula-se o valor do centro de cada diamante. O resultado é uma grade quadrada. Repete-se a operação até o nível de refinamento desejado. Para calcular o valor do centro de cada diamante ou quadrado, toma-se a média dos vértices e adiciona-se um deslocamento aleatório. A amplitude desse deslocamento decresce exponencialmente à medida que a malha é refinada. Esse fator de redução controla a rugosidade (ou dimensão fractal) da superfície final.

Para gerar números aleatórios que obedecem a uma dada distribuição unidimensional  $P(x)$ , Devroye [2] oferece o seguinte método: gera-se um valor aleatório uniforme  $U$  entre 0 e 1, e calcula-se  $F^{-1}(U)$ , onde  $F(x)$  é a função de distribuição acumulada:

$$F(x) = P(X \leq x) = \int_{-\infty}^{\infty} P(x)dx \quad (9)$$

Utilizando esse método nas linhas e colunas da distribuição de teste, conseguimos gerar os eventos aleatórios (Figura 1).

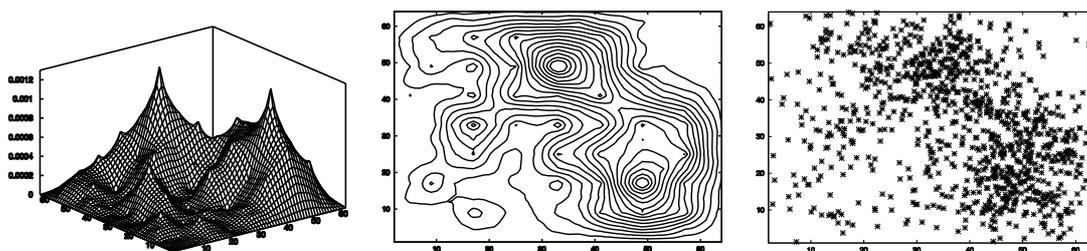


Figura 1: Geração de um distribuição aleatória teste por método fractal

Os melhores resultados foram obtidos com janelas adaptativas (Tabela 1, Figura 2). Os critérios de janela fixa tiveram um pior desempenho, e se mostraram semelhantes entre si.

Esquema adotado	MISE x $10^{-5}$
Janela fixa (Equação 5)	2,17
Janela fixa (Equação 6)	2,13
Janela Adaptativa (Equação 7)	1,71
Janela Adaptativa (Equação 8)	1,87

Tabela 1: Erro médio quadrático integrado (MISE) dos vários esquemas de janela fixa e adaptativa

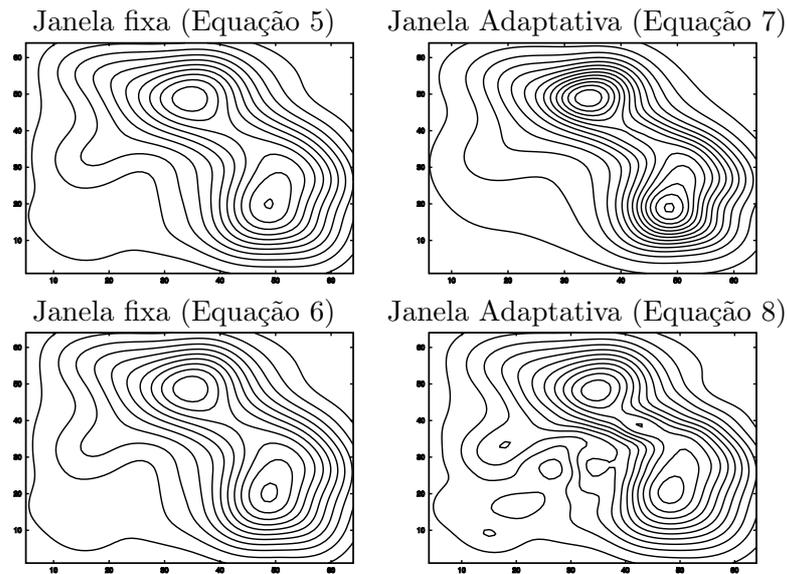


Figura 2: Campos de densidade gerados pelos vários esquemas de janela para a distribuição aleatória teste considerada

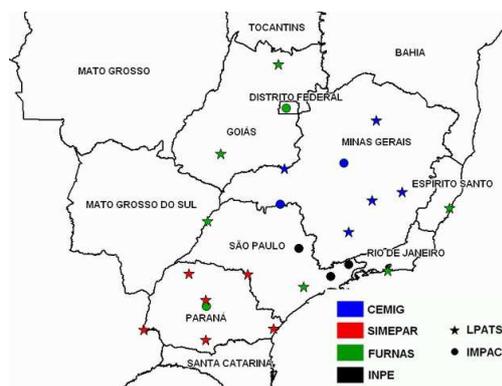


Figura 3: Localização dos sensores da rede RINDAT  
 Fonte: <http://www.rindat.com.br>

### 3 Uso da ferramenta EDDA

#### 3.1 Geração de campos de densidade estáticos

A ferramenta EDDA gera campos de densidade estáticos, sendo que um arquivo de parâmetros de configuração permite especificar o intervalo de tempo desejado, a faixa de latitudes e longitudes, e o espaçamento da grade, além da polaridade (positiva, negativa, todas), bem como o tipo das descargas (nuvem-solo ou nuvem-nuvem) desejados para geração do campo de densidade de ocorrências. Os dados de descargas, cedidos pelo CPTEC/INPE, são fornecidos pela rede RINDAT, composta por uma rede de sensores de descargas distribuída no território nacional (Figura 3). Esses dados utilizam o formato UALF (*Universal ASCII Lightning Format*), que consiste em arquivos texto formatados em colunas de atributos tais como latitude, longitude e instante de ocorrência, polaridade e tipo, sendo cada linha correspondente a uma descarga particular. Além de calcular a densidade de ocorrências, a ferramenta possibilita calcular a densidade de carga elétrica, uma vez que a carga  $q$  de cada descarga esta pode ser calculada a partir dos dados em formato UALF, conforme se segue:

$$q = \int I dt \approx \frac{1}{2} I_{\text{pico}} (t_{\text{subida}} + t_{\text{descida}}) \quad (10)$$

onde  $I_{\text{pico}}$  é a corrente de pico, e  $t_{\text{subida}}$  e  $t_{\text{descida}}$  são respectivamente o tempo de subida e descida da corrente, aproximadamente lineares.

Atualmente, a ferramenta EDDA fornece a opção de uma janela fixa definida pelo usuário, ou uma janela adaptativa como descrita na Equação 8.

### 3.2 Geração de animações

A ferramenta EDDA permite a geração de animações a partir dos campos de densidade gerados, por meio de um esquema de janela deslizante, sendo possível também configurar o intervalo de tempo abrangido pela animação, a velocidade em quadros/segundo, e a correlação entre o tempo da animação e o tempo real.

A título de ilustração, aparecem 4 quadros de uma animação das descargas ocorridas no sudeste brasileiro no dia 16/03/2007, referentes a períodos consecutivos de 30 minutos, com velocidade de 15 quadros por segundo. (Figura 4). Obviamente, é possível selecionar uma área menor, de forma a se fazer um *zoom* da área de interesse. O recurso da animação parece ser adequado para acompanhar a evolução de centros de atividade convectiva eletricamente ativos, permitindo monitorar em tempo quasi-real o deslocamento e/ou crescimento das tempestades associadas.

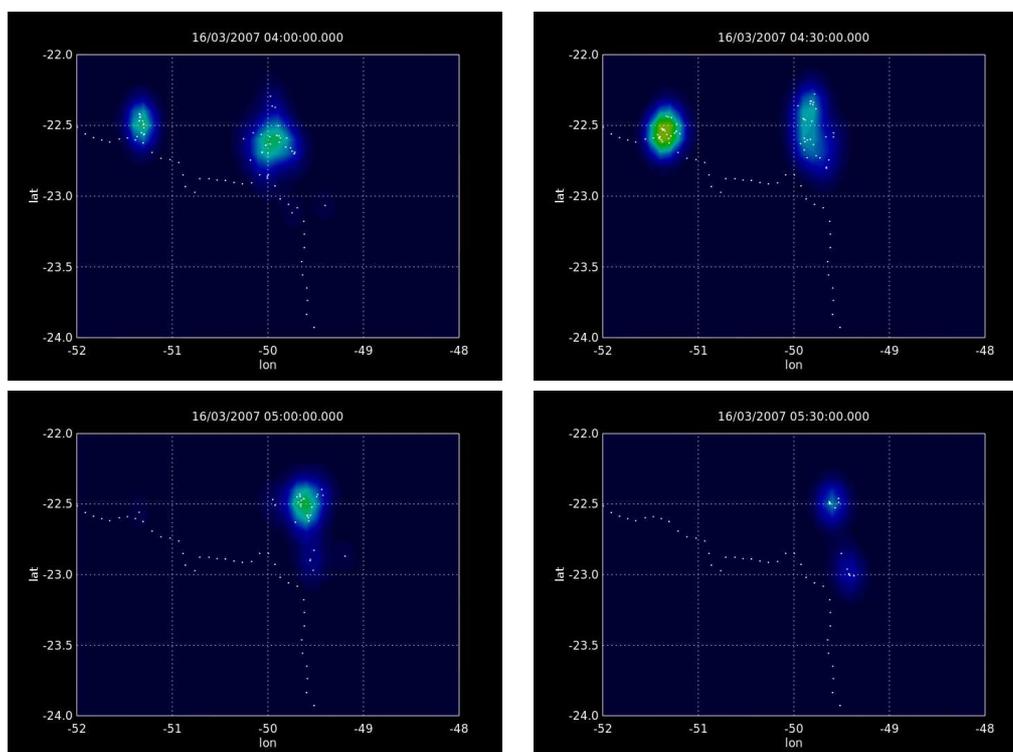


Figura 4: Alguns quadros selecionados da animação das descargas no sudeste brasileiro no dia 16/03/2007

## 4 Comentários finais

A ferramenta EDDA mostrou-se adequada para extrair informação sobre a densidade de descargas atmosféricas a partir dos dados brutos, facilitando a mineração desses dados. Ao mesmo tempo, a animação da densidade de descargas pode ser uma forma útil para a visualização desse fenômeno e ajudar na sua interpretação. O uso de núcleo gaussiano e a escolha de parâmetros adequados permitiu gerar uma animação suave, sem mudanças bruscas, numa velocidade adequada para acompanhar a evolução das estruturas.

A técnica de estimação de densidade mostrou-se adequada para tratar a esparsidade temporal e espacial das descargas atmosféricas. A correlação entre essa densidade e a atividade convectiva eletricamente ativa vem sendo estudada e a ferramenta EDDA presta-se a esse fim, permitindo acompanhar a evolução do campo de densidade em tempo quasi-real por meio de animações a partir dos dados brutos de descargas adquiridos da rede RINDAT. Assim, constitui-se numa ferramenta auxiliar para o meteorologista. Ademais, os dados pelo EDDA, referentes aos campos de densidade de descargas, já estão sendo usados na mineração de dados meteorológicos no escopo de projetos afins. Neste contexto, o objetivo é monitorar um conjunto selecionado de variáveis do modelo meteorológico Eta para detetar padrões encontrados na mineração associados a possíveis ocorrências de eventos convectivos severos. Estes padrões foram encontrados empregando-se os campos de densidade como atributo de decisão referente à presença de atividade convectiva severa.

**Agradecimentos:** Os autores agradecem o suporte recebido do CNPq por meio do projeto do Edital Universal denominado “Mineração de Dados Associados a Sistemas Convectivos” (“Cb-Mining”, processo 479510/2006-7), bem como o suporte recebido da FINEP por meio do projeto “ADAPT - Tempestades: desenvolvimento de um sistema dinamicamente adaptativo para produção de alertas para a região Sul/Sudeste”, mais especificamente sua Meta 2 - “Mineração de dados para identificação de condições favoráveis à gênese e evolução de tempestades”.

## Referências

- [1] M. Caetano, G.C.J. Escobar, S. Stephany, V.E. Menconi, N.J. Ferreira, M.O. Domingues, O. Mendes Junior, Visualização de campo de densidade de ocorrências de descargas elétricas atmosféricas como ferramenta auxiliar no nowcasting, em: “Proceedings of CLIMET XIII and CONGREGMET X”, v. 1. pp. 1-9. Buenos Aires, 2009.
- [2] L. Devroye, “Non-uniform Random Variate Generation”, Springer-Verlag, 1986.
- [3] V.E. Menconi, M. Caetano, S. Stephany, N.J. Ferreira, M.O. Domingues, O. Mendes Junior, A Tool for the Monitoring of Electrically Active Convective Activity. *Em submissão à Weather and Forecasting*. (2010)
- [4] V.E. Menconi, O. Mendes Junior, M.O. Domingues, M. Caetano, S. Stephany, “Software para Visualização de Centros de Atividade Convectiva (CAC)”. Relatório técnico INPE, 2010.
- [5] G.S.P. Miller, The definition and rendering of terrain maps, em “SIGGRAPH '86: Proceedings of the 13th annual conference on Computer graphics and interactive techniques”, pp 39-48, New York, NY, USA. ACM. 1986.
- [6] J. Politi, “Implementação de um Ambiente para Mineração de Dados Aplicada ao Estudo de Núcleos Convectivos”, Dissertação de Mestrado em Computação Aplicada, INPE, 2005, INPE-14165-TDI/1082.
- [7] J. Politi, S. Stephany, M.O. Domingues, O. Mendes Junior, Mineração de dados meteorológicos associados à atividade convectiva empregando dados de descargas elétricas atmosféricas. *Revista Brasileira de Meteorologia*, 21, n.2, (2006) 232-244.
- [8] D.W. Scott, “Multivariate Density Estimation - Theory, Practice and Visualization”. John Wiley & sons, Inc, 1992.
- [9] B.W. Silverman, “Density Estimation for Statistics and Data Analysis”, Chapman and Hall, 1986.