



Kernel smoothing dos dados de chuva no Nordeste¹

Nyedja F. M. Barbosa², Borko D. Stosic³, Tatijana Stosic⁴,
Fabrício M. O. Lopes⁵, Geber B. de A. Moura⁶ & José S. P. Melo⁷

¹ Parte da dissertação de mestrado do primeiro autor apresentada ao Programa de Pós-Graduação em Biometria e Estatística Aplicada da UFRPE

² DEINFO/UFRPE. Recife, PE. E-mail: nyedjaf@gmail.com (Autora correspondente)

³ DEINFO/UFRPE. Recife, PE. E-mail: borkostosic@gmail.com

⁴ DEINFO/UFRPE. Recife, PE. E-mail: tastosic@gmail.com

⁵ DEPA/UFRPE. Recife, PE. E-mail: pabricio@depa.ufrpe.br

⁶ DEPA/UFRPE. Recife, PE. E-mail: geber@depa.ufrpe.br

⁷ DEPA/UFRPE. Recife, PE. E-mail: jspmelo@gmail.com

Palavras-chave:

interpolação
suavização
precipitação

RESUMO

O regime de chuvas sobre o Nordeste do Brasil é bastante complexo, sendo considerado sazonal, além de sofrer fortes influências dos fenômenos El Niño, La Niña e outros sistemas meteorológicos, como o dipolo, atuantes sobre as bacias do oceano Atlântico Tropical. Neste trabalho foi aplicada a técnica matemática-computacional de interpolação do Kernel Smoothing nos dados de precipitação pluviométrica sobre o Nordeste, coletados no período de 1904 a 1998, provenientes de 2.283 estações meteorológicas. Os cálculos foram desenvolvidos por meio do software “Kernel”, escrito em linguagem C e Cuda o que possibilitou fazer a interpolação de mais de 26 milhões de medidas de precipitação pluviométrica, permitindo gerar mapas de intensidade de chuva sobre toda a região e calcular estatísticas para a precipitação do Nordeste em escalas mensais e anuais. De acordo com as interpolações realizadas foi possível detectar, dentro o período estudado, os anos mais secos e mais chuvosos, a distribuição espacial das chuvas em cada mês, bem como a característica da precipitação pluviométrica em épocas de El Niño e La Niña.

Key words:

interpolation
smoothing
precipitation

Kernel smoothing of rainfall data from the Northeast of Brazil

ABSTRACT

The rainfall over Northeast Brazil is quite complex and is considered seasonal, besides suffering strong influences of El Niño, La Niña and other weather systems such as dipole, acting on the ocean basins Tropical Atlantic. In this study the mathematical-computational interpolation technique Kernel Smoothing was applied to rainfall data collected over the Northeast in the period from 1904 to 1998, from 2,283 weather stations. The calculations were developed using the software “Kernel”, written in C language and Cuda, enabling interpolation of over 26 million rainfall measurements, allowing to generate maps of rainfall intensity over the whole region, and to calculate statistics for precipitation in the Northeast on monthly and annual scales. The resulting interpolations were used to identify the years with the most and least precipitation along the studied period, the spatial distribution of rainfall in each month, as well as the characteristics of rainfall in times of El Niño and La Niña.

INTRODUÇÃO

A Região Nordeste do Brasil é conhecida por grande diversidade climatológica. Seu regime pluviométrico tem sido estudado por muitos pesquisadores na tentativa de explicar o comportamento deste fenômeno que é tão inconstante e altamente influenciado pelas perturbações de anomalias de temperatura dos oceanos Atlântico e Pacífico.

Fenômenos como o El Niño e La Niña, geralmente afetam o índice de precipitação causando épocas de estiagem e época de grande volume de precipitação pluviométrica (Beserra et al., 2011). Além disso, o Nordeste contém regiões semiáridas com concentração de chuvas em curto período do ano, além de grandes intervalos secos que dificultam a vida da população nordestina.

Os dados de precipitação pluviométrica sobre o Nordeste são muitos, porém sua organização, entendimento e visualização,

fundamentais para modelagem deste fenômeno, ainda representam grande desafio para a ciência.

Na literatura científica mundial se encontram diversos estudos, como os de Li & Shao (2010) e Garcia-Pintado et al. (2009) na direção de aproveitar os benefícios e diminuir as falhas das fontes de dados com o objetivo geral de estabelecer padrões e regras para o tratamento dos dados extraído o máximo de informações neles contidas, bem como estabelecer os procedimentos e regras para medidas futuras.

Além da técnica matemática/computacional, Parzen-Rosenblatt Window Method (Rosenblatt, 1956), (Parzen, 1962), (Epanechnikov, 1969) ter sido proposta várias décadas atrás, o desafio computacional para sua implementação prática só se tornou possível no caso atual, com recentes avanços revolucionários de hardware para processamento paralelo, em geral GPGPU (General Purpose computing on Graphics

Processing Units) e em particular no caso atual com o uso do GPU cluster "Neumann" instalado no Departamento de Estatística e Informática da UFRPE.

O trabalho atual teve por objetivo utilizar a técnica Kernel Smoothing para construção da estimativa contínua da densidade espaço-temporal da chuva para a região Nordeste, no período de 1904 a 1998. Para tal serão investigadas as características da precipitação pluvial sobre o Nordeste ao longo de todo o período estudado e em épocas específicas quando constam registros sobre incidências dos fenômenos El Niño e La Niña (Trenberth, 1997). Também será analisado o comportamento espacial da chuva sobre o Nordeste em cada mês do ano e em épocas de chuvas sazonais (Rao et al., 1993).

Com o mapeamento é possível ter uma visão geral da distribuição da chuva em escala espacial e temporal estabelecendo-se, assim, base para estudos futuros da dinâmica da chuva.

MATERIAL E MÉTODOS

Os dados pluviométricos utilizados neste trabalho são provenientes de observações coletadas pela Superintendência de Desenvolvimento do Nordeste (SUDENE) e cedidos pela Agência Pernambucana de Águas e Clima, APAC (2014).

O banco de dados é composto por medições de precipitação pluvial extraídas de 2.283 estações meteorológicas convencionais localizadas em todos os estados da Região Nordeste, entre os anos de 1904 a 1998, correspondendo a 26.496.444 de observações.

Um método bastante eficaz para fazer interpolação de dados pontuais utilizando estatística não paramétrica, estudado por (Rosenblatt, 1956), (Parzen, 1962), é o Estimador de Núcleo, (Kernel Estimator, em Inglês). Esta técnica, computacionalmente intensiva, tem sido utilizada recentemente com maior frequência em virtude do crescimento de poder computacional (Duong, 2007).

O método do Kernel é muito utilizado para mapear e estimar a distribuição dos pontos no espaço utilizando a estatística não paramétrica por meio da função de Núcleo. Seu objetivo é interpolar um valor de intensidade para cada célula de uma grade considerando uma função simétrica, centrada na célula, utilizando-se para o cálculo os pontos situados até certa distância do centro da célula (Carneiro & Santos, 2003; Ali, 1998).

Suponha a existência de uma variável aleatória de onde foi obtida uma amostra aleatória X_1, \dots, X_n , cujos valores observados são independentes e identicamente distribuídos. O estimador de kernel, $\hat{f}_h(x)$, para esta amostra é dado por meio da expressão matemática dada pela Eq. 1:

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \quad (1)$$

em que n representa o tamanho da amostra, $K(\cdot)$ representa a função de Kernel escolhida para a interpolação, e h é o raio de abrangência, também chamado parâmetro de suavização.

Esses parâmetros são os únicos necessários para o cálculo das estimativas (Wand & Jones, 1995).

Muitas funções densidades de probabilidade são utilizadas neste método não paramétrico, dentre elas as mais utilizadas são Uniforme, Epanechnikov, Biweight, Guassiana e triangular (Jou et al., 2012).

A estimativa de densidade Kernel possui as mesmas propriedades da função de núcleo escolhida, razão por que é imprescindível a escolha de um Kernel suave e claramente unimodal e simétrico ao redor do zero.

A escolha da função de densidade escolhida para a estimação é importante; contudo, a escolha do raio é muito mais pois, a partir do raio se definirá a vizinhança do ponto a ser interpolado e se controlará quão suave será a estimação da densidade.

A escolha do parâmetro h vai depender do experimento que se esteja realizando não sendo bom utilizar um tamanho muito grande para não suavizar nem muito pequeno, para não suavizar a interpolação.

As estimativas de kernel smoothing são encontradas por meio da Eq. 2:

$$\hat{Y}(x_0) = \frac{\sum_{i=1}^n K\left(\frac{x_0 - X_i}{h}\right) y(X_i)}{\sum_{i=1}^n K\left(\frac{x_0 - X_i}{h}\right)} \quad (2)$$

Os cálculos numéricos para realização deste trabalho foram implementados no GPU "Cluster Neumann" utilizando-se o software "Kernel" escrito em linguagens C e CUDA. Tal ferramenta possibilitou realizar cálculos complexos que, realizados apenas pelo CPU, necessitariam de muito tempo ou até nem seriam possíveis de se realizar, já que os dados correspondem a aproximadamente 26 milhões de observações.

O software é preparado para ler o banco de dados, calcular a média diária de precipitação das estações meteorológicas e, através da função de Kernel Smoothing, utilizando a função Gaussiana, estimar a curva de densidade que representa a precipitação pluvial para todo o mapa da região Nordeste, preenchendo os espaços com dados ausentes, e estimar a média diária de chuva sobre a região.

RESULTADOS E DISCUSSÃO

Serão apresentados, a seguir, os resultados de interpolações feitas para dados de chuva no Nordeste do Brasil aplicando a técnica de kernel Smoothing nos dados oriundos da antiga SUDENE, desde 1904 até 1998.

Primeiramente, fez-se a interpolação de todas as observações que constam no banco de dados. Por meio do software utilizado foi possível localizar a posição da cada Estação Meteorológica que contribuiu para a pesquisa, conforme a Figura 1. Os dados das estações que não pertencem à região Nordeste, na Figura 1^a, foram utilizados nos cálculos do Kernel Smoothing para gerar resultados na Figura 1B, porém o efeito da presença desses dados é limitada só às regiões vizinhas desses pontos, controlado pelo

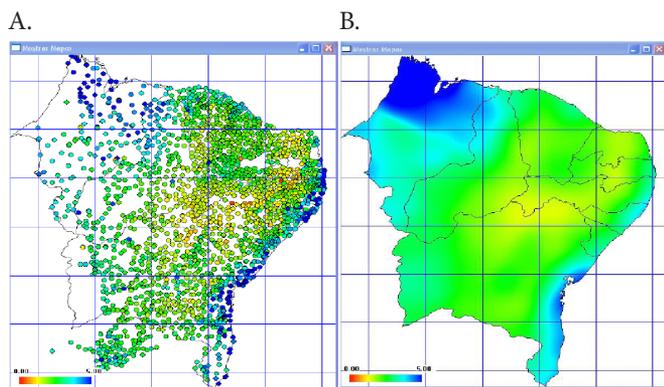


Figura 1. Localização das estações meteorológicas no Nordeste de Brasil entre 1904 e 1998 (A); Interpolação dos dados de Precipitação Pluviométrica média diária na Região Nordeste a partir das informações das Estações Meteorológicas (B)

parâmetro de suavização h . A escala de cores apresentada varia de vermelho a azul, correspondendo à média diária de chuva sobre cada região, em que o vermelho representa 0 mm e azul 5 mm de chuva, em média.

Percebe-se, através Figura 1, que há uma grande área sobre o estado do Maranhão em que a quantidade de precipitação se destaca das demais áreas, seguido pela costa leste da Bahia; e, como previsto, verifica-se o baixo nível de precipitação sobre a região semiárida do Nordeste.

A média diária de chuva sobre o Nordeste para o período de 1904 a 1998, obtida por meio de interpolação, foi de 2,83 mm com desvio padrão de 0,97 mm. Também foi analisada a quantidade de precipitação acumulada em cada ano, exceto nos anos de 1904 a 1909, por apresentarem pouquíssimas informações.

A Figura 2 apresenta o comportamento da precipitação anual acumulada para o Nordeste, desde 1910 até 1998.

A média de precipitação acumulada anual para o período estudado é de 962,9 mm, com desvio padrão de 222,5 mm. Observa-se que alguns anos apresentam precipitação acumulada muito inferior, e outros, muito superior à média.

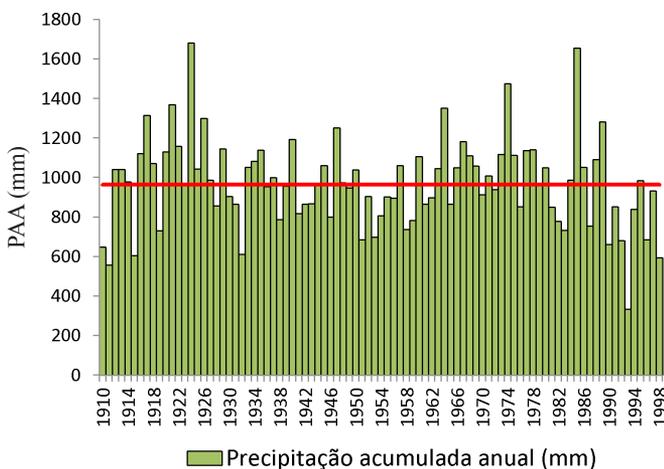


Figura 2. Precipitação acumulada anual (PAA) para os períodos de 1910 a 1998 no nordeste

Dos anos observados é possível destacar os anos de 1911 e 1993 como os anos mais secos e 1924, 1974 e 1985 como os anos que apresentaram maior volume de chuva.

A grande oscilação entre a quantidade de chuva sobre a região Nordeste se deve à influência de fenômenos meteorológicos a climáticos atuantes sobre a região.

Por meio do software também foi possível estimar a precipitação média diária em meses isolados descartando-se, do período informado pelo usuário na caixa de entrada, os meses desnecessários.

A estimativa da precipitação média diária em cada mês e durante o período de 1904 a 1998 sobre o Nordeste, pode ser observada na Figura 3.

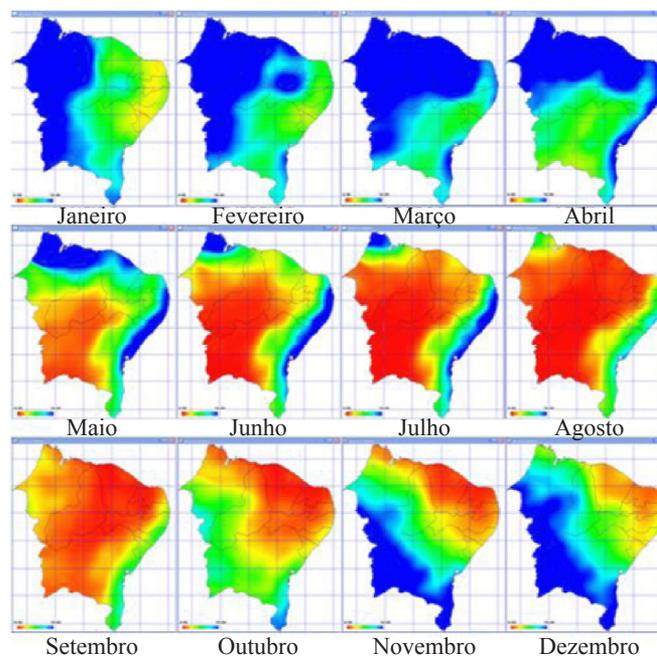


Figura 3. Interpolação dos dados de precipitação pluvial mensal na Região Nordeste de Brasil, no período de 1904 a 1998

Por meio da Figura 3 observa-se que a distribuição espacial das chuvas não é uniforme sobre o Nordeste, apresentando meses nos quais a chuva se concentra em áreas específicas.

De acordo com Meneghetti & Oliveira (2009) o Nordeste possui três regimes de chuva distintos: na parte norte, chove mais entre março e maio; no sul e sudeste, a chuva ocorre de dezembro a fevereiro e no leste, chove entre maio e julho. Este comportamento pode ser confirmado pelos dados apresentados na Tabela 1 e na Figura 4.

Dentre os períodos analisados é possível destacar os dois primeiros períodos, onde a precipitação média diária foi

Tabela 1. Precipitação pluvial sazonal correspondente ao período de 1904 a 1998, no Nordeste do Brasil

Mês	Média	Desvio padrão
Dezembro, janeiro, fevereiro	4,3	1,9
Março, abril, maio	4,4	2,3
Junho, julho, agosto	1,1	1,3
Setembro, outubro, novembro	1,6	1,0

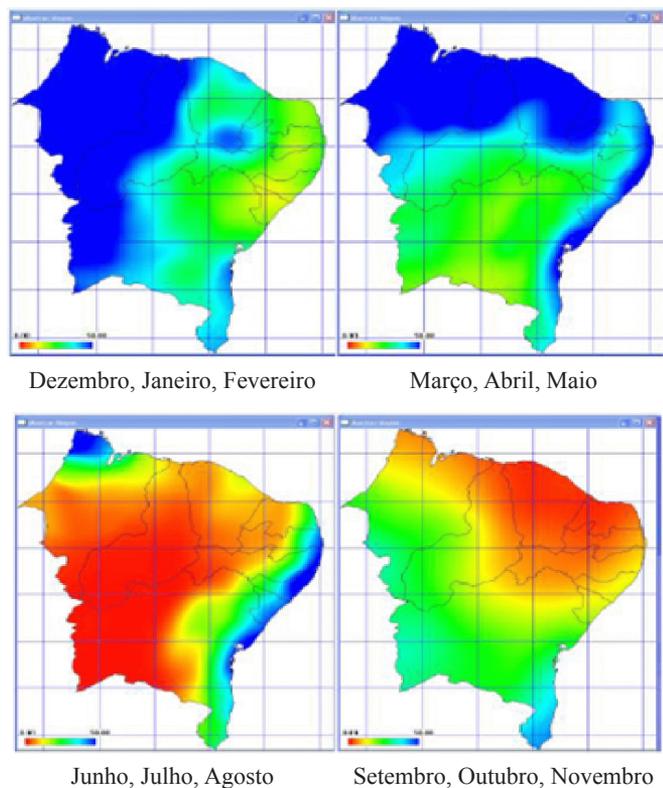


Figura 4. Interpolação dos dados de precipitação pluviométrica média diária na Região Nordeste de Brasil, no período de 1904 a 1998, em cada estação chuvosa local

superior ao restante dos períodos verificados. Para visualizar a distribuição espacial da precipitação pluvial durante os períodos estudados, observe a Figura 4.

Pelo que pôde ser observado na Figura 4, em cada época analisada ao longo dos anos estudados, há uma concentração maior de chuva em determinadas regiões, ao longo dos meses, tornando o período chuvoso diferente em cada localidade.

A precipitação pluvial sobre a região Nordeste também sofre influência de alguns fenômenos climáticos.

Para analisar a distribuição da chuva sobre épocas de El Niño e La Niña, foram feitas interpolações dos dados de chuva registrados nos períodos indicados por Trenberth (1997), como sendo períodos de incidência de tais fenômenos, classificados por intensidade de acordo com as normas adotadas pelo INPE-CPTEC (2014).

A Tabela 2 apresenta a precipitação média diária, em milímetros, dos períodos de El Niño, lembrando que os períodos citados correspondem a eventos fracos, moderados e fortes.

Verifica-se, de acordo com a Tabela 2, que entre os 16 eventos analisados, em 10 eventos deles a precipitação foi inferior à média climatológica e em 3 eventos a precipitação foi superior à média. Esta variabilidade pode ser influenciada pelo Gradiente de Temperatura da Superfície do Mar meridional ou inter-hemisférico (GRADM), (Beserra et al, 2011).

A Figura 5 apresenta a distribuição espacial da chuva sobre cada período de El Niño analisado na Tabela 2, segundo ordem cronológica.

Tabela 2. Precipitação pluvial média diária durante os anos de ocorrência de El Niño sobre o Nordeste

Índice	Período	Intensidade	Média	DP#
1	Ago/1951 a Fev/1952	Fraca	1,6	0,7
2	Mar/1953 a Nov/1953	Fraca	1,7	1,7
3	Abr/1957 a Jun/1958	Forte	2,8	0,9
4	Jun/1963 a Fev/1964	Fraca	3,2	1,0
5	Mai/1965 a Jun/1966	Moderada	2,4	0,8
6	Set/1968 a Mar/1970	Moderada	3,1	0,9
7	Abr/1972 a Mar/1973	Forte	2,7	0,9
8	Ago/1976 a Mar/1977	Fraca	2,9	1,0
9	Jul/1977 a Jan/1978	Fraca	2,2	0,9
10	Out/1979 a Abr/1980	Fraca	4,7	1,6
11	Abr/1982 a Jul/1983	Forte	1,9	0,1
12	Ago/1986 a Fev/1988	Moderada	2,1	0,8
13	Mar/1991 a Jul/1992	Forte	2,2	1,1
14	Fev/1993 a Set/1993	Forte	0,6	0,4
15	Jun/1994 a Mar/1995	Moderada	2,0	1,0
16	Abr/1997 a Abr/1998	Forte	2,4	1,4

#DP - Desvio Padrão

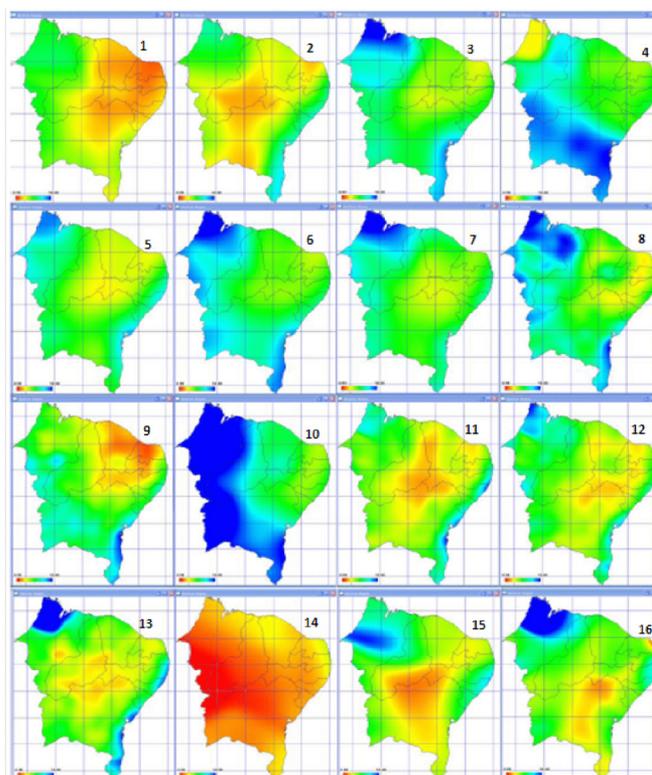


Figura 5. Períodos de El Niño sobre o Nordeste em relação à precipitação pluvial média diária

Constata-se, na Figura 5, que a parte semiárida tende a apresentar precipitação pluvial abaixo da média. Em 1993 toda a região foi atingida pela seca enquanto durante o evento 79/80 o índice de precipitação foi acima da média com grande volume de chuva nas partes norte e centro-sul da região.

Outro fenômeno que também influencia o regime de chuvas sobre o Nordeste é o La Niña, haja vista que a quantidade de precipitação coletada em sua época de atuação foi investigada.

A Tabela 3 apresenta a média diária de precipitação sobre períodos de atuação do La Niña na região Nordeste, de acordo com (Trenberth, 1997) classificados por intensidade segundo as normas adotadas pelo INPE-CPTEC (2014).

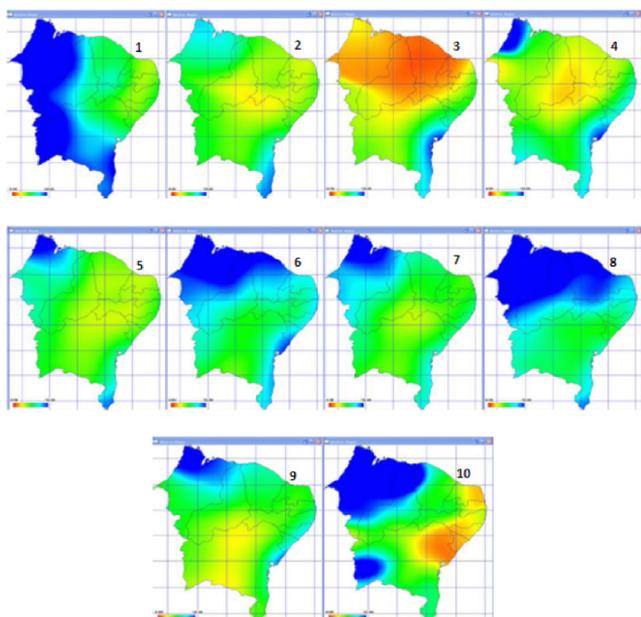
Tabela 3. Precipitação pluvial média diária durante os anos de ocorrência de La Niña sobre o Nordeste

Índice	Período	Intensidade	Média	DP [#]
1	Mar/1950 a Fev/1951	Forte	2,5	0,9
2	Jun/1954 a Mar/1956	Forte	2,3	0,7
3	Mai/1956 a Nov/1956	Forte	1,4	1,0
4	Mai/1964 a Jan/1965	Moderado	2,2	1,0
5	Jul/1970 a Jan/1972	Moderado	2,5	0,8
6	Jun/1973 a Jun/1974	Forte	3,9	1,3
7	Set/1974 a Abr/1976	Forte	3,0	1,0
8	Set/1984 a Jun/1985	Forte	4,4	1,7
9	Mai/1988 a Jun/1989	Forte	2,6	1,0
10	Set/1995 a Mar/1996	Forte	3,4	1,3

[#]DP - Desvio Padrão

Na Tabela 3 se observam dois períodos distintos: com precipitação abaixo da média climatológica, antes de 1973, e com precipitação acima da média, depois de 1973, que podem estar relacionados à mudança na fase da Oscilação Decadal do pacífico e à mudança na fase da Oscilação do Atlântico Norte, ocorridas nesta época (Lucena et al., 2011).

A partir da Figura 6 é possível visualizar a interpolação dos dados de precipitação para a região Nordeste durante os períodos de La Niña, seguindo ordem cronológica na Tabela 3.



Obs.: Para ordem cronológica 1, 2, ..., 10, consultar Tabela 3

Figura 6. Precipitação pluvial média diária nos anos de ocorrência de La Niña sobre o Nordeste

Na Figura 6 observa-se que o setor norte do Nordeste tende a apresentar um índice de chuva em torno ou acima da média climatológica, com exceção do evento em 1956, quando a região foi atingida pela seca.

CONCLUSÕES

1. À técnica Kernel Smoothing possibilitou fazer interpolação de 26 milhões de medidas de precipitação pluvial, extraídas de 2.283 estações meteorológicas no Nordeste, entre 1904 a 1998.

2. A média diária de chuva obtida foi de 2,83 mm com desvio padrão de 0,97 mm. Os anos de 1911 e 1993 foram os mais secos e 1924, 1974 e 1985 os mais chuvosos.

3. Em termos de período do ano, os meses de janeiro, fevereiro, março e abril foram os mais chuvosos para o Nordeste, já os meses de agosto, setembro e outubro foram os mais secos.

4. Em relação ao comportamento sazonal, a precipitação média diária foi superior no período de dezembro a maio, sendo de dezembro a fevereiro na parte norte e centro-sul, e de março a maio na parte leste e sudeste da região.

5. Nos anos de El Niño, a parte semiárida do Nordeste apresenta precipitação pluvial abaixo da média climatológica, enquanto nos anos de La Niña o setor norte do Nordeste apresenta precipitação em torno, ou acima, da média. Toda a região Nordeste foi atingida pela forte seca em 1993 durante o evento do El Niño.

LITERATURA CITADA

- Ali, A. Nonparametric spatial rainfall characterization using adaptive Kernel estimator. *Journal of Geographic Information and Decision Analysis*, v.2, p.34-43, 1998.
- APAC - Agência Pernambucana de Águas e Climas. 2014. <<http://www.apac.pe.gov.br>>, 1 Abr. 2014.
- Beserra, L. D.; Servain, J.; Gomes Filho, M. F. Rainfall response in Northeast Brazil from ocean climate variability during the second half of the twentieth century. *Journal of Climate*, v.24, p.6174-6184, 2011.
- Carneiro, E. O.; Santos, R. L. Análise espacial aplicada na determinação de áreas de risco para algumas doenças endêmicas (calazar, dengue, diarreia, DST - Doenças sexualmente transmissíveis e tuberculose), no bairro de Campo Limpo, Feira de Santana (BA). *Sitientibus*, v.28, p.51-75, 2003.
- Duong, T. Ks: Kernel density estimation and kernel discriminant analysis for multivariate data in R. *Journal of Statistical Software*, v.21, p.1-16, 2007.
- Epanechnikov, V. A. Non-parametric estimation of a multivariate probability density. *Theory of Probability and its Applications*. v.14, p.153-158, 1969.
- Garcia-Pintado, J.; Barberá, G. G.; Erena, M.; Castillo, V. M. Rainfall estimation by rain gauge-radar combination: A concurrent multiplicative-additive approach. *Walter Resources Research*, v.45, p.1-15, 2009.
- INPE-CPTEC - Instituto Nacional de Pesquisas Espaciais - Centro de Previsão de Tempo e Estudos Climáticos. 2014. <<http://enos.cptec.br>>, 19 Mar. 2014.
- Jou, P. H.; Akhoond-Ali, A. M.; Nazemosadat, M. J. Nonparametric kernel estimation of annual precipitation over Iran. *Theoretical and Applied Climatology*, v.112, p.193-200, 2012.
- Li, M.; Shao, Q. An improved statistical approach to merge satellite rainfall estimates and raingauge data. *Journal of Hydrology*, v.385, p.51-64, 2010.

- Lucena, D. B.; Gomes Filho, M. F.; Servain, J. Avaliação do impacto de eventos climáticos extremos nos oceanos pacífico e atlântico sobre a estação chuvosa no nordeste do Brasil. *Revista Brasileira de Meteorologia*, v.26, p.297-312, 2011.
- Meneghetti, G. T.; Oliveira, I. M D. Variabilidade sazonal e interanual da precipitação no nordeste brasileiro. In. *Simpósio Brasileiro de Sensoriamento Remoto*, 14, 2009. Natal. Anais... Natal: INPE, 2009. p.1685-1689.
- Parzen, E. On estimation of a probability density function and mode. *Annals of Mathematical and Statistics*, v.33, p.1065-1076, 1962.
- Rao, V. B.; Lima, M. C. de; Franchito, S. H. Seasonal and interannual variations of rainfall over eastern northeast brazil. *Journal of Climate*, v.6, p.1754-1763, 1993.
- Rosenblatt, M. Remarks on some nonparametric estimates of a density function. *Annals of Mathematical and Statistics*, v.27, p.832-837, 1956.
- Trenberth, K. E. The definition of El Niño. *Bulletin of the American Meteorological Society*, v.78, p.2771-2777, 1997.
- Wand, M. P.; Jones, M.; C. Kernel smothing. New York: Chapimam & Hall. v.60, 1995. 212p.