

Ciência e Natura

ISSN: 0100-8307

cienciaenaturarevista@gmail.com

Universidade Federal de Santa Maria

Brasil

Guzmán Rodríguez, Lissette; Anabor, Vagner; Puhales, Franciano; Dal Piva, Everson
Estimativa da probabilidade de ocorrência de precipitação, a partir de técnicas
estatísticas não paramétricas aplicadas a simulações numéricas de WRF. Um caso de
estudo.

Ciência e Natura, vol. 38, -, 2016, pp. 491-497
Universidade Federal de Santa Maria
Santa Maria, Brasil

Disponível em: <http://www.redalyc.org/articulo.oa?id=467547689080>

Estimativa da probabilidade de ocorrência de precipitação, a partir de técnicas estatísticas não paramétricas aplicadas a simulações numéricas de WRF. Um caso de estudo.

Estimation of rainfall probability, through the use of non parametric statistical techniques, applied to numerical simulations of WRF. A case of study.

Lissette Guzmán Rodríguez¹, Vagner Anabor², Franciano Puhales³, Everson Dal Piva⁴
lissette19901109@gmail.com

^{1,2,3,4} Grupo de Modelagem Atmosférica, Departamento de Física- Universidade Federal de Santa Maria, Brasil.

Resumo

No presente trabalho emprega-se a estimativa da densidade do kernel (KDE), um método não-paramétrico, para estimar a função densidade de probabilidade de uma variável aleatória, na obtenção de um prognóstico probabilístico de precipitação, a partir de uma previsão por conjunto do modelo WRF. Os nove membros da previsão por conjunto foram obtidos variando a parametrização convectiva do modelo, para um evento de precipitação intensa no sul do Brasil. Na avaliação dos resultados, os estimados de probabilidade obtidos para períodos de 3 e 24 horas, e vários limitares de precipitação, foram comparados com os valores de precipitação estimada pelo TRMM, sem mostrar correspondência morfológica entre ambos. Para acumulados em 24 horas, foi possível comparar com os valores pontuais das observações do INMET, encontrando-se melhor coerência entre as observações e as probabilidades previstas. Foram calculadas medidas de desempenho a partir de tabelas de contingência, para diferentes intervalos de probabilidades, sendo que a previsão da chuva intensa teve maior proporção correta em todos os intervalos de probabilidades, e ao prever precipitação com 75% de probabilidade para qualquer acumulado, não ocorreram falsos alarmes. Além do mais, a precipitação de menor intensidade com probabilidade marginal foi sobre-prevista, apresentando maior índice de falsos alarmes.

Palavras-chave: KDE, previsão probabilística, precipitação intensa.

Abstract

In this paper was used the kernel density estimation (KDE), a nonparametric method to estimate the probability density function of a random variable, to obtain a probabilistic precipitation forecast, from an ensemble prediction with the WRF model. The nine members of the prediction were obtained by varying the convective parameterization of the model, for a heavy precipitation event in southern Brazil. Evaluating the results, the estimated probabilities obtained for periods of 3 and 24 hours, and various thresholds of precipitation, were compared with the estimated precipitation of the TRMM, without showing a clear morphological correspondence between them. For accumulated in 24 hours, it was possible to compare the specific values of the observations of INMET, finding better coherence between the observations and the predicted probabilities. Skill scores were calculated from contingency tables, for different ranks of probabilities, and the forecast of heavy rain had higher proportion correct in all ranks of probabilities, and forecasted precipitation with probability of 75%, for any threshold, did not produce false alarms. Furthermore, the precipitation of lower intensity with marginal probability was over-forecasted, showing also higher index of false alarms.

Keywords: KDE, probabilistic forecast, heavy rainfall.

: KDE, probabilistic forecast, heavy rainfall.

1 Introdução

A previsão quantitativa da precipitação (Quantitative Precipitation Forecast, QPF) é um dos desafios mais importantes e significativos da previsão do tempo. Embora muitos aspectos da previsão numérica do tempo (Numerical Weather Prediction, NWP) têm feito grandes progressos nas últimas décadas, os avanços na QPF tem sido relativamente lentos (Fritsch et al., 1998). A incerteza da previsão devido a erros na inicialização e problemas nas aproximações e parametrizações do modelo, pode ser diminuída a partir dos sistemas de previsão por conjunto de membros (Ensemble Prediction System, EPS), os quais diferem na configuração e/ou condições iniciais. A dispersão dos EPS fornece uma medida objetiva da incerteza da previsão, tanto na média do conjunto como na solução determinística correspondente (Kalnay e Dalcher, 1987), sendo as saídas dos EPS melhor aproveitadas através do cálculo de previsões probabilísticas (Anderson, 1996). A estimativa de densidade de kernel (*kernel density estimation*, KDE), é um método não-paramétrico para estimar a função densidade de probabilidade de uma variável aleatória. Nos últimos anos, várias aplicações desta técnica tem sido registradas na meteorologia (Wilks, 2006; Brooks et al., 2003; Peel and Wilson, 2007), mostrando as vantagens do KDE sobre outros métodos como o emprego de histogramas, ou ajuste a funções de densidade de probabilidade clássicas.

A região Sul do Brasil não possui uma estação chuvosa bem definida (Rao e Hada, 1990) e os eventos de precipitação intensa podem ocorrer ao longo do ano (Texeira e Satyamurty, 2007), estando associados principalmente com frentes frias (Andrade, 2007) e sistemas convectivos de mesoescala (Anabor et al. 2008). Frente esta realidade é importante contar com uma previsão que forneça uma ideia da probabilidade com que poderiam ocorrer esses eventos de chuva com grande impacto na população.

O objetivo deste estudo é, portanto, implementar uma técnica de previsão

probabilística a partir de um EPS, através do emprego de estimativas de densidade (da probabilidade) de kernel. Usou-se a técnica KDE para um caso de precipitação intensa nos estados do sul do Brasil.

2 Materiais e metodologia

2.1 Materiais

O período de estudo selecionado foi para os dias 7 e 8 de Setembro de 2009, nos quais o sul do Brasil foi afetado por fortes chuvas, registrando acumulados significativos (superiores a 50 mm em 24 horas) em pelo menos 10 estações da região (segundo dados do Instituto Nacional de Meteorologia, INMET).

Na simulação dos campos meteorológicos foi usado o Weather Research and Forecasting (WRF) Model. As condições iniciais foram fornecidas pelo NCEP FNL (Final) Operational Global Analysis, com uma resolução de 1 grau, e atualização cada 6 horas das condições de contorno.

Para a avaliação foram usados os dados de precipitação estimada (tri-horários) pelo produto 3B42 V7 do Tropical Rainfall Measuring Mission (TRMM), assim como as observações de precipitação acumulada em 24 horas de 37 estações do INMET localizadas no domínio de estudo, e as imagens do canal infravermelho do satélite GOES 10.

2.2 Estimativa de densidade de kernel (KDE).

A estimativa da densidade do kernel é uma das técnicas não-paramétricas de suavização de dados (*kernel smoothing*) mais amplamente usadas. Neste trabalho foi empregada a variante suavização de kernel multivariado, implementada no pacote 'ks' (Duong, 2007) da plataforma R, que contém funções para estimar a densidade do kernel e para o análise discriminante do kernel.

Para uma amostra aleatória d-variada elaborada a partir de uma densidade f , a estimativa da densidade do kernel é definida por:

$$f'(x, H) = \frac{1}{n} \sum_{i=1}^n K_H(x - X_i)$$

onde $x=(x_1, x_2, \dots, x_n)$, $X_i=(X_{i1}, X_{i2}, \dots, X_{in})$, e $i=(1,2,\dots,n)$. $K(x)$ é o kernel, uma função de densidade de probabilidade simétrica, H é a largura de banda, uma matriz simétrica positiva dxd, e $K_H(x)=|H|^{-1/2}K(H^{-1/2}x)$.

A escolha de K não é um fator determinante para o cálculo da densidade (Duong, 2007), neste trabalho foi empregado o kernel normal: $K(x) = (2\pi)^{-d/2} \exp(-1/2 x^2)$. Em contraste, a escolha de H é crucial para determinar o desempenho de f' .

2.2 Metodologia

Os membros da previsão por conjunto, foram nove simulações do WRF, inicializadas nas 00UTC do 06/09 até 18UTC do 10/09 abrangendo o período de estudo. Os membros tiveram a mesma configuração do domínio e física variando apenas o esquema da parametrização de cúmulos (Tabela 1). Algumas das configurações restantes foram: esquema de microfísica de Lin et al. (1983), esquemas RRTM e Dudhia para radiação de onda longa e curta respectivamente, e Yonsei University na parametrização de camada limite planetária.

Tabela 1: Parametrizações.

Cu_PHYSICS	Esquema
1	Kain-Fritsch
2	Betts-Miller-Janic
3	Grell-Freitas
5	Grell-3
6	Tiedtke
14	New SAS
84	New SAS (HWRF)
93	Grell-devenyi
99	Old Kain-Fritsch

Foram criados dois domínios: de 24 e 12 km de resolução, respectivamente, centrados perto da cidade de Uruguaiana, RS. O segundo domínio com coordenadas entre os 22-35°LS e os 45-61°LW, de 119x137 pontos de grade, foi aninhado ao primeiro. As saídas das simulações foram escritas cada 1h.

Os dados de precipitação foram separados em

classes com limiar ascendente, considerando chuva maior de 3 mm, de 8 mm e de 16 mm, para intervalo de 3 horas (de 8mm/3h em adiante é equivalente ao critério de precipitação intensa de mais de 50mm em 24h, de Texeira e Satyamurty, 2007). Dessa forma foi testado o mesmo esquema para o cálculo de densidade de probabilidade para diferentes intensidades da chuva.

Na análise de 24 horas usaram-se os limiares de 10, 25 e 50 mm para a precipitação acumulada nesse intervalo.

A estimativa de densidade de kernel foi feita apenas para o domínio interior, especificamente com as funções contidas no pacote 'ks' do R. A função 'kde', realiza a estimativa de densidade de kernel bidimensional usando um H (largura de banda), calculado pela função 'hpi'. No cálculo do H foram testadas combinações que incluíram várias transformações prévias dos dados (*sphering, scaling*) e seletores do H ótimo (AMSE, SAMSE) (Duong, 2007).

As distintas variantes não mostraram diferenças significativas na distribuição da densidade de kernel obtida, mas sim no tempo de cômputo, pelo que foi escolhida para o presente trabalho, a combinação *sphere + SAMSE*, a mais econômica computacionalmente.

Para o caso de 24 horas foram criadas tabelas de contingência e calculadas várias medidas de desempenho (*skill scores*) da previsão: o PC (*proportion correct*), o TS (*threat score*), o FAR (*false alarm ratio*), o H (*hit rate*), o F (*false alarm rate*) e o B (*bias*) (Wilks, 2011).

3 Resultados e discussão

3.1 Precipitação em 3 horas

A densidade de probabilidade obtida pelo KDE para os diferentes limiares de precipitação em 3 horas (3mm, 8mm, 16mm, para o caso de 8 de setembro de 2009, das 06UTC até as 09UTC) são mostrados na Figura 1 (a, b e c). As cores amarelo, laranja e vermelho indicam probabilidades de ocorrência do evento de 25%, 50% e 75%, respectivamente, e a cor verde se corresponde com o valor estimado de precipitação do TRMM em mm, também para cada limiar.

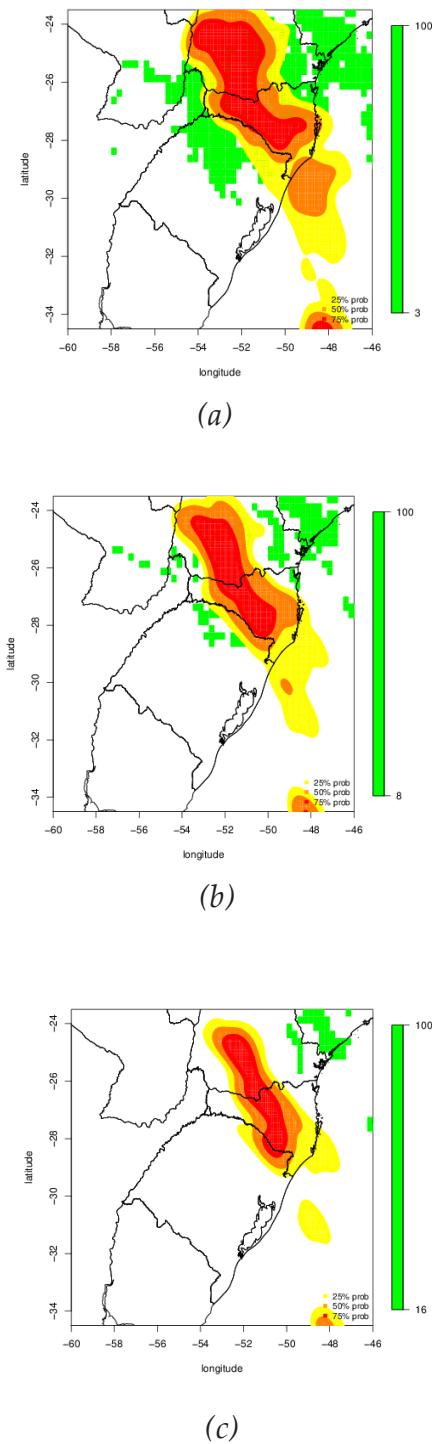


Figura 1: Probabilidade estimada de kernel e precipitação estimada pelo TRMM para mais de 3 (a), 8 (b) e 16 (c) mm/3h.

Pode se observar que a localização das probabilidades obtidas com o KDE não correspondem exatamente com a dos valores de precipitação estimados pelo TRMM. Essa situação piora na medida que aumenta o limiar

da chuva. Esses resultados são coerentes com os de Fensterseifer (2013), para uma bacia no Rio Grande do Sul. Este estudo mostra que as médias diárias estimadas pelo 3B42 V7, superestimam os volumeis médios observados, porém, a medida que aumenta o limiar da precipitação, diminui a detecção do evento de 68% (>5mm) para 58 (>50mm).

A Figura 2 mostra a média da precipitação obtida a partir dos nove membros do ensemble. As áreas de maiores valores promédios mostram coincidência com as áreas de maior probabilidade prevista da Figura 1, como é esperado.

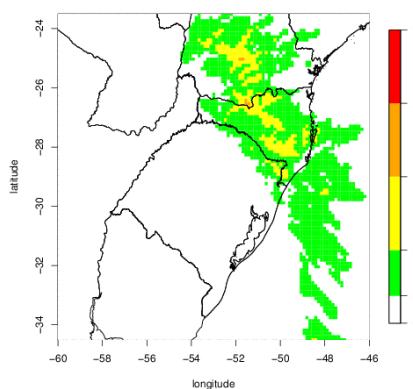


Figura 2: Média da precipitação dos nove membros do ensemble.

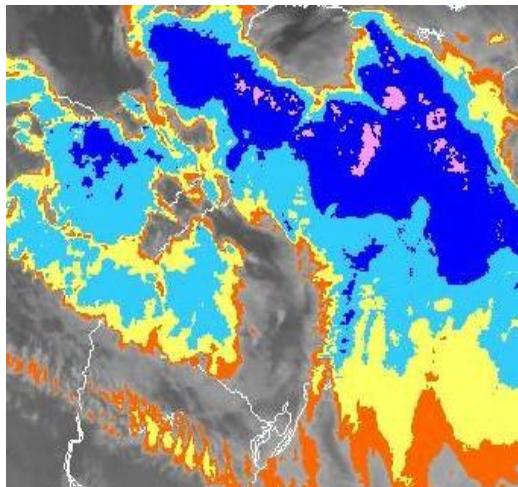
Ao analisar as imagens de satélite do infravermelho realçado, correspondentes aos horários inicial e final do período considerado (Figura 3, (a) e (b), respectivamente), é possível ver a correspondência de áreas de maiores probabilidades estimada mediante o KDE, com os núcleos de temperatura brilho mais baixas (zonas de convecção profunda) em cor roxa, que persistiram sobre os estados de Santa Catarina e Paraná. Isto avalia positivamente as simulações feitas e o resultado obtido a partir da técnica empregada.

Não se contou com os dados de precipitação observada para períodos tri-horários para estabelecer a comparação.

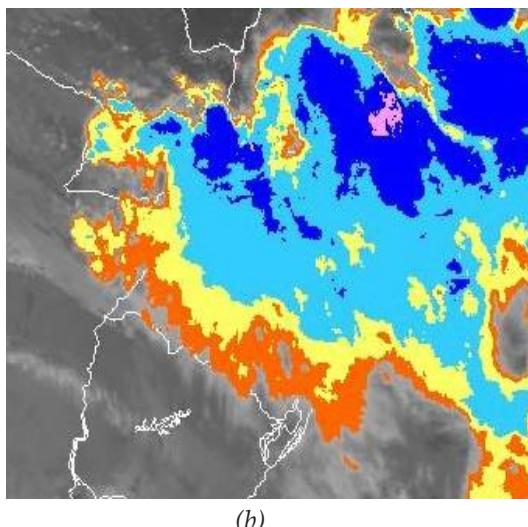
3.2 Precipitação em 24horas

Foram estimadas as probabilidades para a precipitação acumulada em 24h (do dia 07 às 12UTC até o dia 08 às 12UTC) e novamente

comparadas com os dados do TRMM e com os valores observados nas estações do INMET localizadas no domínio de estudo.



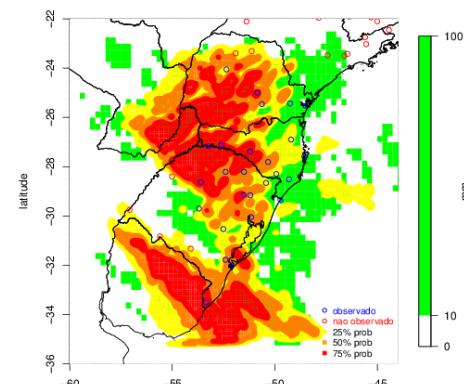
(a)



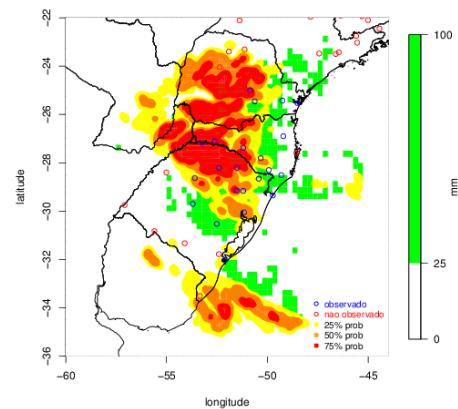
(b)

Figura 3: Imagens do GOES 10 do canal infravermelho realçado correspondentes as 06UTC (a) e 09UTC (b) do dia 8 de setembro de 2009.

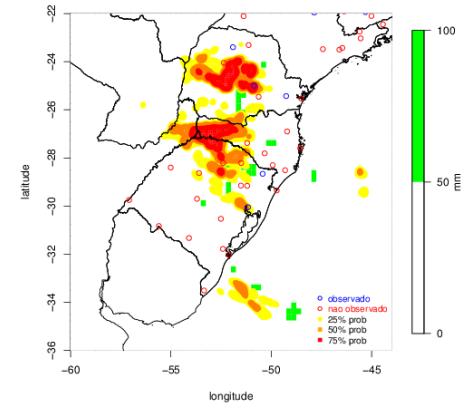
Na Figura 4 as cores representam a mesma informação que na Figura 1 , e adicionalmente aparecem círculos marcando as estações, de cor azul as que reportaram precipitações em correspondência com o limiar estabelecido e em vermelho aquelas que não atingiram esses acumulados. Aqui deve assinalar-se que não sempre existe correspondência entre os valores observados nas estações e os estimados pelo TRMM, o que faz não muito confiável a avaliação a partir desses últimos dados.



(a)



(b)



(c)

Figura 4: Probabilidade estimada de kernel para mais de 10 (a), 25 (b) e 50(c) mm/24h, precipitação estimada pelo TRMM e observações do INMET.

Com a finalidade de avaliar a metodologia aplicada, foram calculadas diferentes medidas de desempenho, a partir da criação de tabelas de contingência, os quais são mostrados na tabela 2.

O PC, que indica a proporção com que os eventos (ocorrência ou não) são previstos corretamente. O PC mostra melhores resultados (mais próximos a 1) na previsão de precipitação intensa ($>50\text{mm}/24\text{h}$) em todos os intervalos de probabilidades, e não apresenta diferenças significativas entre os limiares inferiores.

O TS, ou razão entre a ocorrência prevista e o número de vezes q foi prevista e/ou observada, apresenta os mesmos resultados desfavoráveis (0.27) para todos os limiares, na hora de prever com mais do 75% de probabilidade, enquanto que nas outras categorias, os resultados mostram-se melhores na medida em que diminui o limite da precipitação analisado, atingindo o melhor valor (0.66) para o 25% de chuvas superior a $10\text{mm}/24\text{h}$.

Tabela 2: Medidas de desempenho da previsão empregando KDE.

Precipitação	>50mm/24h			>25mm/24h			>10mm/24h		
probabilidade	>75%	>50%	>25%	>75%	>50%	>25%	>75%	>50%	>25%
PC	0.78	0.78	0.84	0.57	0.65	0.70	0.49	0.70	0.70
TS	0.27	0.38	0.54	0.27	0.43	0.56	0.27	0.58	0.66
B	0.27	0.64	0.82	0.27	0.50	0.77	0.27	0.58	1.04
FAR	0.00	0.29	0.22	0.00	0.09	0.18	0.00	0.00	0.22
H	0.27	0.45	0.64	0.27	0.45	0.64	0.27	0.58	0.81
F	0.00	0.08	0.08	0.00	0.07	0.20	0.00	0.00	0.55

O B, ou taxa entre as ocorrências previstas e observadas, mostra como quase todas as categorias são sub-previstas, melhorando a proporção na medida que diminui o valor da probabilidade e o limiar de chuva, sendo o 25% de probabilidade de 10mm de chuva, a única que é prevista mais vezes das que em realidade é observada.

Ao calcular o FAR, pode-se ver a fração das previsões de ocorrência do evento que não chegaram a ser observadas. O melhor valor possível é 0, ou seja, em todos os casos em que foi previsto um evento com mais do 75% de probabilidade, o evento foi observado, e no caso de 10 mm/24h com 50% de probabilidade também não foram reportadas falsos alarmes.

O H ou taxa de acerto, é o número de pronósticos corretos sobre o número de vezes q evento ocorreu. Para acumulados de 50 e 25 mm/24h, os resultados foram similares, mostrando maiores acertos para probabilidades mais baixas. Aqui, nota-se os melhores valores para os menores acumulados de chuvas e as menores probabilidades.

Finalmente o F, ou taxa entre as falsos alarmes e as não ocorrências do evento, também conhecida como probabilidade de não detecção, apresentou valores muito bons (0 ou próximos a ele) exceto nas previsões de chuvas de 10 e 25 mm/24h, com 25% de probabilidade.

4 Conclusões

No trabalho foi utilizada a estimativa de densidade de kernel na obtenção de prognósticos probabilísticos para diferentes limiares de precipitação, a partir de simulações ensembles do WRF.

Os resultados da análise deste caso de estudo ratificam ao cálculo de densidade de kernel como uma técnica efetiva na estimativa de densidades de probabilidade. Embora as probabilidades de ocorrência obtidas para os acumulados de precipitação em 3 horas não correspondam espacialmente com os estimados pelo TRMM, eles sim estão em concordância com os valores médios obtidos nas simulações e com o observado nas imagens de satélite.

Na análise da probabilidade de ocorrência dos diferentes limiares de precipitação em 24 horas, ao se comparar com o observados nas estações do INMET mostrou como a proporção com que os eventos são previstos corretamente é maior nos casos de precipitação de mais de 50 mm, para os três intervalos de probabilidades considerados.

Encontrou-se que na medida que diminui o limiar da chuva, se incrementa o índice de ameaça, ou seja, os eventos são mais previstos em comparação com o total de vezes que são observados e previstos. Esses resultados mostram-se coincidentes com os obtidos pelo B,

os quais indicam que na medida que diminui tanto o limiar da precipitação como a probabilidade de ocorrência do evento, existe um acréscimo na taxa entre o número total de eventos que são previstos e os observados.

Além do mais, a proporção de eventos previstos corretamente é maior para probabilidades baixas ($>25\%$), porém nesta categoria também foi maior a proporção de falsos alarmes, que em contraste foi zero nos casos de probabilidade elevada ($>75\%$).

É notável que em nenhum dos casos em que foi prevista a chuva com 75% de probabilidade, a previsão esteve errada. Esses resultados reforçam a utilidade da metodologia aplicada na obtenção de previsões de eventos de precipitação extrema com alto grau de certeza.

Agradecimentos

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes) e ao Programa de Pós-Graduação em Meteorologia da Universidade Federal de Santa Maria.

Referências

- Anabor, V., D. J. Stensrud, and O. L. L. De Moraes. Serial upstream-propagating mesoscale convective system events over southeastern South America. *Mon. Wea. Rev.*, 2008, 136, 3087–3105.
- Andrade, K. M. Climatology and behavior of frontal systems in South America (in Portuguese). M.S. thesis, Pos-Graduação em Meteorologia, Instituto Nacional de Pesquisas Espaciais, 2007, 185 pp. [Available from Instituto Nacional de Pesquisas Espaciais, Av. Astronautas 1758, São José dos Campos, São Paulo, Brazil.]
- Brooks, H. E., C. A. Doswell, and M. P. Kay. Climatological estimates of local daily tornado probability for the United States. *Wea. Forecasting*, 2003, 18, 626–640.
- Duong, T. ks: Kernel density estimation and kernel discriminant analysis for multivariate data in R. *J. Stat. Software*, 2007, 21 (7), pp. 1–16.
- Fensterseifer C. A. Qualidade das estimativas de precipitações derivadas de satélites na bacia de Alto Jacuí-RS. M.S. thesis, Pos-Graduação em Engenharia Civil e Ambiental, Universidade Federal de Santa Maria., 2013, 26 pp.
- Fritsch JM, Houze Jr RA, Adler R, Bluestein H, Bosart L, Brown J, Carr F, Davis C, Johnson RH, Junker N, Kuo Y-H, Rutledge S, Smith J, Toth Z, Wilson JW, Zipser E, Zrnic D. Quantitative precipitation forecasting: Report of the eighth prospectus development team , US Weather Research Program. *Bull. Am. Meteorol. Soc.*, 1998, 9: 285 – 299.
- Peel, S., and L. J. Wilson. Modeling the distribution of precipitation forecasts from the Canadian ensemble prediction system using kernel density estimation. *Wea. Forecasting*, 2008, 23, 575–595.
- Ochoa, A., Pineda, L., Crespo, P., Willems, P. Evaluation of TRMM 3B42 precipitation estimates and WRF retrospective precipitation simulation over the Pacific-Andean region of Ecuador and Peru. *Hydrol. Earth Syst. Sci.*, 2014, 18, pp. 3179–3193 .
- Teixeira, M. S., and P. Satyamurthy. Dynamical and synoptic characteristics of heavy rainfall episodes in southern Brazil. *Mon. Wea. Rev.* , 2007, 135, 598–617.
- Velasco, I., and J. M. Fritsch. Mesoscale convective complexes in the Americas. *J. Geophys. Res.*, 1987, 92, 9591–9613.
- Wilks, D. Statistical Methods in the Atmospheric Sciences, 3rd ed0. International geophysics serie, 2011, v.100, 308–311.