

Using KDE to Obtain a Probabilistic QPF from an EPS

Lisette Guzmán Rodríguez, Vagner Anabor*, Franciano Scremin Puhales, Everson Dal Piva

Grupo de Modelagem Atmosférica, Departamento de Física-Universidade Federal de Santa Maria, Santa Maria, RS, Brasil

Abstract In this paper was used kernel density estimation (KDE), a nonparametric method to estimate the probability density function of a random variable, to obtain a probabilistic quantitative precipitation forecast (QPF) from an ensemble prediction of the WRF model, for a case study. The nine members of the ensemble prediction system (EPSm) were obtained by varying only the convective parameterization of the model. The case study corresponded to a heavy precipitation event in Southern Brazil. Evaluating the results, estimated probabilities obtained for a 24 hours period, and various thresholds of precipitation, were compared with the observations of ANA (Agência Nacional de Águas, Brasil) stations, finding better coherence between the observations and the predicted probabilities for the lower amounts of precipitation. Some skill scores were calculated for the KDE product (for different ranks of probability), all the EPSm and for the ensemble mean precipitation product (MPP). Even though KDE forecasts with >75% of probability had better FAR and B than the EPSm and MPP forecasts, according to the values of PC and PS, KDE forecast for this category was not superior to the others. However, KDE forecasts with probability of >25% and >50% had the best values of PC, FAR, B and PSS for almost every threshold, except for the heavy rainfall (>50mm/24h), that was in general the worst forecasted by the KDE and also by the EPSm.

Keywords EPS, KDE, Probabilistic QPF

1. Introduction

Quantitative Precipitation Forecast (QPF) is one of the most significant challenges of weather forecast, due to the large impact of this variable in the development of human activities. Although many aspects of Numerical Weather Prediction (NWP) have made great progress in recent decades, the QPF still sensitive to the uncertainties in the forecasts (mainly produced by initialization errors and inaccuracies of approaches and parameterizations used in the models). Those uncertainties can be reduced or quantified by using EPS, in which, instead of running the model once (to make a deterministic forecast), the model runs several times from different initial conditions and/or physical parameterizations (multi-physics EPS), or different models are used within the EPS (multi-model EPS).

The range of different solutions (or ensemble members, EPSm) provides an objective measure of the forecast uncertainty and allows us to evaluate how confident we should be in a deterministic forecast [1]. According to [2], a EPS is designed to obtain the probability density function (PDF) of the prediction, so that the outputs of the EPS can be better used in obtaining probabilistic forecasts, to assess the possibility of occurrence of certain events.

The classical statistical approach to achieve this goal, relies heavily on parametric models or assumptions on the distribution of the data, which completely fix the distribution of this except for the value of one or more parameters to be estimated. Although, there are many practical situations where a simple exploratory data analysis shows that the assumption of normality is inadequate, as in the case of precipitation.

To overcome these deficiencies, non-parametric estimators have been developed, as is the case of kernel estimators, whose original idea goes back to the work of [3] and [4]. Kernel density estimation (KDE), also known as kernel smoothing, is a simple way to find structures in data without the restrictions imposed by a parametric model. In recent years several applications of this technique has been registered in meteorology [5], [2], [6] showing KDE advantages over other classical methods, such as the use of histograms or adjustments to PDFs.

Southern Brazil does not have a well-defined rainy season and events of heavy rainfall occur well distributed over the year [9], mainly associated with cold fronts [11] and mesoscale convective systems [10]. In front of this reality it is important to have a forecast that allow us to assess the possibility of occurrence of these events. In order to achieve that goal, the objective of this paper was to explore the performance of KDE in obtaining probabilistic precipitation forecasts from an EPS, for a case of study.

* Corresponding author:
vanabor@gmail.com (Vagner Anabor)

Published online at <http://journal.sapub.org/ajee>

Copyright © 2016 Scientific & Academic Publishing. All Rights Reserved

2. The Ensemble Prediction System

In this study was used a multi-physics EPS integrated by 9 member, only differentiated by their convective parametrization schemes. The integrations were made with the Weather Research and Forecasting (WRF, V3.6). Two nested domains of 24 and 12 km of resolution, 27 vertical levels and lambert projection, were created over part of South America (Figure 1). Initial and boundary conditions of 1° resolution were from NCEP FNL.

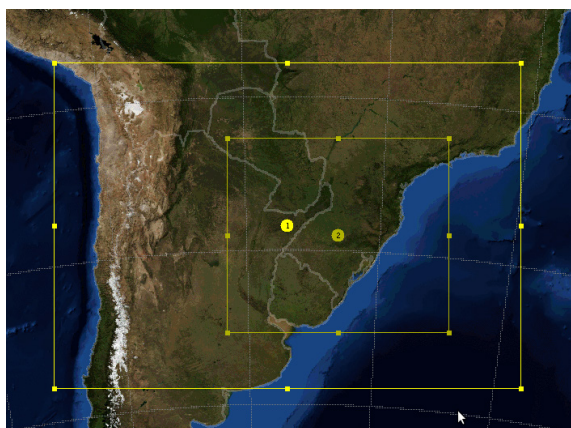


Figure 1. Two nested domains used for simulation: domain 1 (16-38 S, 37-76 W) and domain 2 (22-35 S, 44-61 W)

The common physic parametrizations for all member included Lin microphysics, the rapid radiative transfer model (RRTM) and Dudhia scheme, for long and short wave radiations, and Yonsei University for the planetary boundary layer. The 9 convective parametrization schemes that differentiate the EPSm are resumed in Table 1.

Table 1. Convective Parametrization Schemes Used in the EPS

Option	Scheme
1	Kain–Fritsch (Kain, 2004)
2	Betts–Miller–Janjic (Janjic, 1994)
3	Grell–Freitas Ensemble (Grell; Freitas, 2014)
5	Grell 3D (Grell, 1993) (Grell; Devenyi, 2002)
6	Tiedtke (Zhang; Wang; Hamilton, 2011)
14	New Simplified Arakawa–Schubert (Han; Pan, 2011)
84	New Simplified Arakawa–Schubert HWRF (Han; Pan, 2011)
93	Grell–Devenyi (GD) Ensemble (Grell; Devenyi, 2002)
99	Old Kain–Fritsch (Kain; Fritsch, 1990)

The case of study (22-09-2013-09Z to 23-09-2013-09Z), corresponded to a heavy precipitation event in southern Brazil. The simulations extended for 48 h (initialized at 12UTC of the day before the event, 21-09-2013), with a time step of 60 s and outputs every hour.

3. Kernel Density Estimate

According to [7] for a bivariate random sample X_1, X_2, \dots, X_n drawn from a density f , the kernel density estimate

is defined by:

$$\hat{f}(x; H) = \frac{1}{n} \sum_{i=1}^n K_H(x - X_i) \quad (1)$$

where $x=(x_1, x_2)^T$ and $X_i=(X_{i1}, X_{i2})^T, i=1, 2, \dots, n$. Here $K(x)$ is the kernel which is a symmetric probability density function, H is the bandwidth matrix (or smoothing matrix) which is symmetric and positive-definite, and $K_H(x)=|H|^{-1/2}K(H^{-1/2}x)$. The choice of K is not crucial. In contrast, the choice of H is what determines the performance of \hat{f} . Here was used KDE in two dimensions with a normal kernel $K(x)=(2\pi)^{-1}\exp(-(1/2)x^Tx)$, a sphere pre-transformation of the data and minimizing SAMSE error for the selection of optimal H , recommended by [8].

4. Methodology

Kernel density estimation (KDE) was used to obtain the probability density function for different thresholds of precipitation, from the simulations made with an EPS, in order to generate probabilistic QPF (it was only considered the domain 1).

The precipitation amounts obtained from the 9 WRF simulations or EPSm, were considered as a unique data set. It means that to every one of the 136x118 grid points corresponded 9 forecasted values. KDE was estimated for a 24 hours period (from 21 to 45h of simulation), considering the criteria of exceeding thresholds of precipitation (in 24 hours) of 1mm, 10mm, 25mm, and 50mm.

The evaluation of the results obtained with KDE was made against the observation of 608 ANA stations. There were only considered the grid points coincident with ANA stations. Four ranks of probabilities (>1%, >25%, >50%, >75%) were independently analyzed. Some skill scores of the forecast, such as proportion correct (PC), false alarm ratio (FAR), bias (B) and Pierce Skill Score (PSS), were calculated through the used of contingency tables [6]. The performance of every EPSm, as well as the ensemble mean precipitation product (MPP) were also evaluated following the same methodology, and compared with the KDE product.

5. Results and Discussion

The probabilities of precipitation obtained with KDE for ranks of 1, 10, 25 and 50mm/24h, are shown in Figure 2. Probability contours of 25%, 50% and 75% appears in yellow, orange and red. The blue points represent the ANA stations where the threshold of precipitation was observed, and the grey ones where it was not observed.

In general, for all the thresholds of precipitation analyzed there was a morphological correspondance between the areas with probabilities >25% and the stations where that amount of precipitation was observed, which shows the fair

performance of both the simulations and the KDE results. This correspondence is better for lower thresholds. For higher values of precipitation, some of the stations with observations, stay out of the zones with the higher probabilities. It could be explained by the deficiency of NWP in quantification and localization of heavy rainfall.

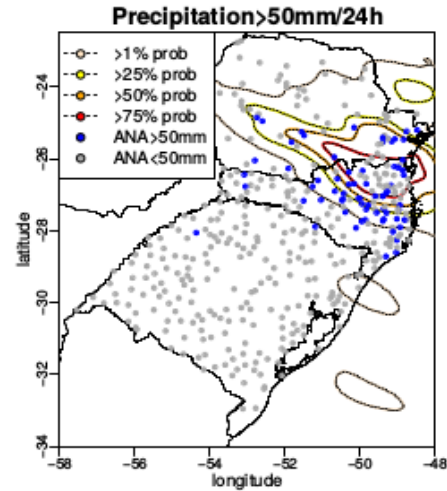
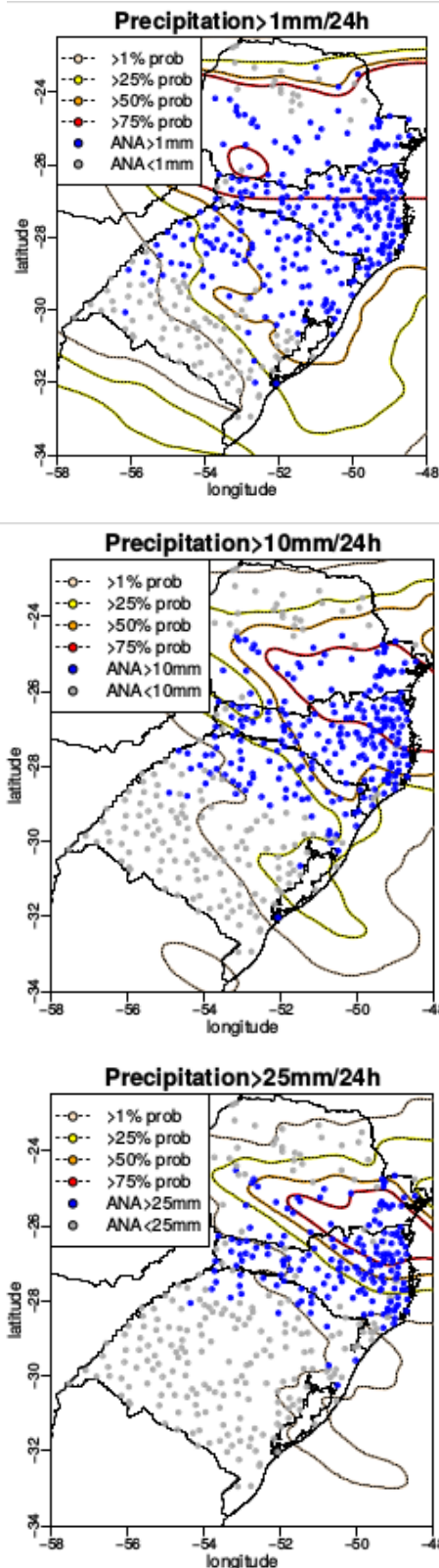


Figure 2. Probability of 24hours-precipitation obtained with KDE from 22-09-2013-09Z to 23-09-2013-09Z

5.1. Comparing KDE with EPSm and MPP

The KDE forecasts with of >25% and >50% of probability had better PC than all the EPSm and the MPP of all the thresholds, except for 50mm with >50% of probabilities. With a high probability (>75%), the PC of 50mm of precipitation is the only showing improvement over the lower probabilities. Between the EPSm, m14 (New Simplified Arakawa–Schubert by Han and Pan, 2011) had the best values. The extreme values of precipitation (1mm and 50 mm) obtain the best values of PC (Figure 3).

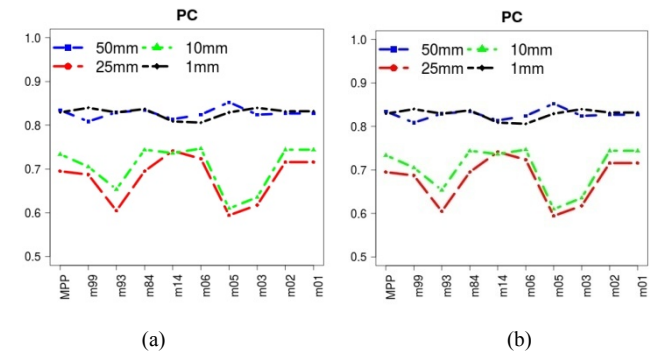


Figure 3. Proportion correct of KDE probabilities (a) and MPP/EPSm (b)

From >25% of probability in ahead, the forecasts obtained with KDE showed lower FAR than the EPSm forecasts and the MPP; and when the probabilities in chances, the values of FAR drops. The only exception is giving by the precipitation >50mm (Figure 4).

With low probabilities the KDE over-forecast every thresholds of precipitation (in greater quantity the higher-ones), but with >50% of probability, all the thresholds are practically unbiased. On the other hand, the EPSm and the MPP product over-forecasted events of >1mm of precipitation and under-forecasted the other thresholds (Figure 5).

PSS (equivalent to do H-F) showed that KDE forecasts with >25% and >50% of probability, were better than

EPSm forecasts. The EPSm m14 was one of the best members in forecasting heavy rainfall (Figure 6).

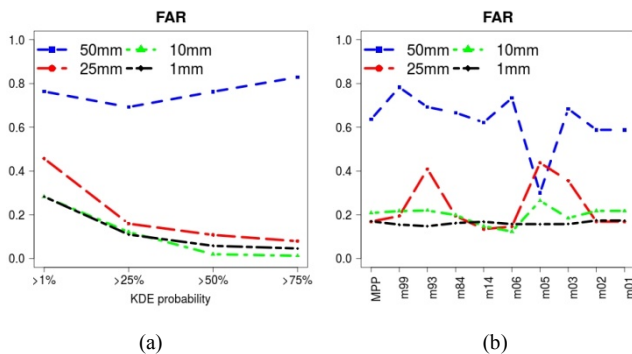


Figure 4. False alarm ratio of KDE probabilities (a) and MPP/EPSm (b)

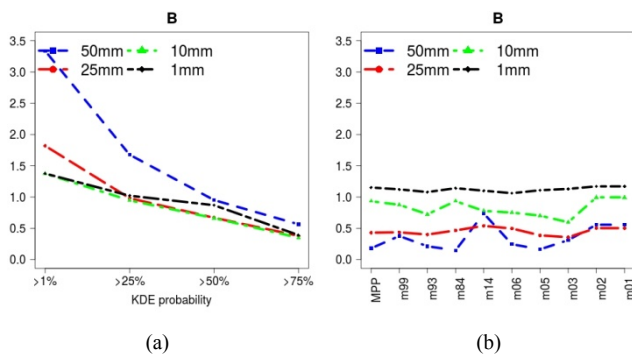


Figure 5. Bias of KDE probabilities (a) and MPP/EPSm (b)

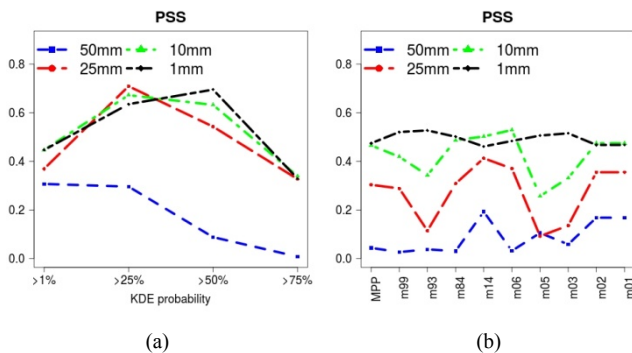


Figure 6. Pierce Skill Score of KDE probabilities (a) and MPP/EPSm (b)

6. Conclusions

A probabilistic QPF was obtained through the use of KDE applied to numerical simulation from an EPS integrated by 9 members with different convective parametrization schemes. In the study of a case of heavy rainfall in southern Brazil, thresholds of precipitation of 1, 10, 25 and 50 mm in 24 hours, were analyzed. Comparing the KDE product, the ensemble members and the MPP with ANA observations,

several advantages of the KDE were observed. KDE forecast with >25% and >50% of probabilities had better proportion correct than all the EPS members and than the MPP. Although when precipitation of >50mm/24h obtained the worst PC, with >25% of probability, KDE forecast is better than the EPSm forecast. KDE with >50% of probability, forecasted all the threshold almost unbiased, showing improvements over the MPP and the EPSm forecasts.

ACKNOWLEDGEMENTS

CAPES; PPGMet-UFSM

REFERENCES

- [1] E. Kalnay "Forecasting forecast skill". *Mon. Wea. Rev.*, v. 115, p. 349–356, 1987.
- [2] S. Peel and L. Wilson, "Modeling the distribution of precipitation forecasts from the Canadian ensemble prediction system using kernel density estimation". *Wea. Forecasting*, v. 23, p. 575–595, 2008.
- [3] M. Rosenblatt, "Remarks on some nonparametric estimates of a density function". University of Chicago, p. 6, 1956
- [4] E. Parzen, "On estimation of a probability density function and mode". Stanford University, p. 12, 1962.
- [5] H. E. Brooks, C. A. Doswell and M. P. Kay, "Climatological estimates of local daily tornado probability for the United States". *Weather Forecasting*, v. 18, n. 8, p. 626–640, 2003.
- [6] D. Wilk, "A sample line graph using colors which contrast well both on. Oxford: Academic Press, 2011. 676 p.
- [7] M. P. Wand and M. Jones, "Kernel Smoothing". New York: Chapman & Hall, 1995. 224 p.
- [8] T. Duong, "ks: Kernel density estimation and kernel discriminant analysis for multivariate data in r". *Journal of Statistical Software*, v. 21, n. 7, p. 1–16, 2007.
- [9] M. S. Teixeira and P. Satyamury, "Dynamical and synoptic characteristics of heavy rainfall episodes in Southern Brazil". *Mon. Wea. Rev.*, v. 135, p. 598–617, 2007
- [10] V. Anabor, D. J. Stensrud and O. L. L. D. Moraes, "Serial upstream-propagating mesoscale convective system events over southeastern South America". *Mon. Wea. Rev.*, v. 68, n. 136, p. 3087–3105, 2008.
- [11] K. M. Andrade, "Climatology and behavior of frontal systems in South America". 2007. 185 f. Dissertação (Mestrado em Meteorologia) — Instituto Nacional de Pesquisas Espaciais, São Paulo, 2007.