4.7 PROBABILISTIC QUANTITATIVE PRECIPITATION FORECAST CALIBRATION OVER SOUTH AMERICA: EXPERIMENTS WITH A SHORT RANGE ENSEMBLE.

Juan J. Ruiz[1,2], Celeste Saulo[1,2], Eugenia Kalnay[3]
[1]University of Buenos Aires, Buenos Aires, Argentina [2]Centro de Investigaciones del Mar y de la Atmósfera (CONICET-UBA), Buenos Aires, Argentina [3]University of Maryland, College Park, USA.

Abstract

Different techniques for obtaining probabilistic quantitative precipitation forecasts (PQPFs) over South America are tested during the 2002-2003 warm season. Some of the techniques are based on a parametric representation of the conditional probability of precipitation over a particular threshold given a certain value of forecasted precipitation and the other uses a non-parametric estimation of the probabilities. The results are also compared with the calibration algorithm based on the rank histogram. A number of experiments were performed to compute a reasonable size for the training period.
The PQPFs of a short range ensemble forecast system (SREF) based on the WRF model and the breeding technique were calibrated using the different approaches and the resulting PQPFs scores were compared.
The calibration is performed using two different data sets, one derived from a high resolution rain gauge network and the other one based on passive microwave satellite estimates. This is done in order to evaluate if the use of satellite derived rainfall produces a significant degradation of PQPFs reliability and resolution.

1. INTRODUCTION

Quantitative precipitation forecast (QPF) is one of the most difficult and least accurate products available from numerical weather prediction (NWP) (Ebert 2001). Continuous efforts are devoted to improve forecast quality, ensemble forecasting being an example of one possible strategy to deal with errors arising from uncertainties in the initial and boundary conditions. An interesting characteristic of ensemble systems is that probability forecasts can easily be created, leading to the generation of PQPFs.

Different methodologies for obtaining PQPFs, and corresponding measures to quantify their usefulness, have been developed. Of particular interest is how to obtain a *reliable* PQPF i.e., a system where the forecasted frequency of a particular weather phenomenon is close to the observed probability. The importance of PQPF reliability is directly related to its effect upon the economic value of the forecast, as discussed by Zhu et al. (2002).

Several techniques have been developed to generate reliable PQPFs. For example, Hamill and Colucci (1998), Gallus et al. 2007, Stensrud and Yussouf 2007, Sloughter et al (2007) (hereafter S2007) among many others introduced different techniques for forecast calibration with the aim of improving forecast quality.

The main objective of this work is to evaluate the skill of different calibration algorithms over South America. In a former study (Ruiz et al. 2008), some of these strategies have been tested for a regional ensemble based on the SLAF technique and on a multi model ensemble. Yet, there were other alternatives that we wanted to test –both for ensemble generation and calibration- and also use a denser precipitation network for validation and calibration that became available after that previous work. Accordingly, this work progresses on the previous one, including more methods for PQPF calibration (i.e. the one based on the paper of S2007) applied to an alternate regional ensemble system as will be explained subsequently.

To compare the skill of the different calibration strategies, 2 months of accumulated precipitation obtained from 48-hr ensemble forecasts have been analyzed. The ensemble system is a regional ensemble with perturbations in the initial conditions obtained trought the Breeding of the Growing Modes method (Toth and Kalnay, 1993).

PQPFs obtained via the combination of the above mentioned calibration techniques are analyzed through the computation of the Brier Skill Score (hereafter BSS) and its components (Wilks 1995) which also allows for comparison with results obtained in previous works.

Corresponding author address: Juan J. Ruiz, CIMA/University of Buenos Aires, Ciudad Universitaria, Buenos Aires, Argentina.
E-mail: jruiz@cima.fcen.uba.ar

Even though a large amount of precipitation data from the SALLJEX experiment raingauge network (Vera et al., 2006) and from the Brazilian Water Agency was available for the present study, the lack of enough rain gauge precipitation data is the main constrain for the operational implementation of calibration algorithms. However, in the last decade, precipitation datasets such as CMORPH (Joyce et al. 2004), which combine microwave estimates of precipitation with high temporal resolution IR estimates of cloud motion, have become available. CMORPH has a homogeneous regional coverage, and high spatial as well as temporal resolution (30 min-accumulated precipitation). For this reason, we have explored the potential of using CMORPH data for PQPF calibration, since it could be an interesting alternative for PQPF calibration over this and other regions where the gauge network is too coarse.

2. METODOLOGY

2.1 ENSEMBLE GENERATION

Global and regional short range ensemble systems are used in this work. The global ensemble uses the MRF (Medium Range Forecasts) model with T62L28 resolution (approximately 2.5º horizontal resolution). The breeding of the growing modes technique (hereafter breeding) (Toth and Kalnay 1993) is used to introduce perturbations in the initial conditions with a rescaling period of 6 hours. The ensemble consists of 11 members (5 pairs of perturbed members and a control run) integrated up to 48-hour lead time.

This global ensemble is used to provide initial and boundary conditions to a regional ensemble based on the WRF model version 2.0 (Skamarock et al. 2005) which has been run with 40 km horizontal resolution and 31 sigma vertical levels. The convective parameterization selected is Kain-Fritsch (Kain 2004), the boundary layer parameterization is the Yonsei University scheme (Hong and Pan 1996), and the surface processes are modeled using the NOAH surface model (Dudhia 2001).The regional ensemble has the same number of members as the global one, and each member of the regional model is nested in its corresponding global ensemble member and is integrated to obtain 48.hr forecasts.

Both ensembles were initialized twice a day at 00 and 12 UTC: the global ensemble uses NCEP-NCAR Reanalysis (Kalnay et al. 1996) as unperturbed initial condition while the regional ensemble uses the Global Data Assimilation System analysis with a resolution of 1ºx1º. The experiment starts on 15th December 2002 and ends on 15th February 2003.

The results presented in this paper correspond to the regional (WRF based) ensemble forecasts.

2.2 DATA

In this work rain gauge data and passive microwave precipitation estimates were used for forecast calibration. On the other hand, in the case of forecast verification, only rain gauge stations were taken into account.

The rain gauge network used in this work is a combination of the South America Low Level Jet Experiment (SALLJEX) and the Brazilian Water Agency (ANA) rain gauge networks. This network is far denser than the operational precipitation network, particularly over Argentina. For this reason this period is particularly interesting to conduct this kind of experiments. A quality control has been performed over SALLJEX data as described in Penalba et. al (2004).

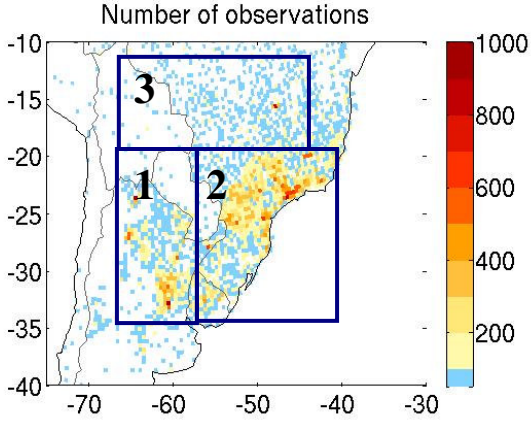

Figure 1: Total number of observations available for each grid box between 15th December 2002 and 15th February 2003. (White means no data available for that grid box) The blue squares shows the location of the regions discussed in the text.

To compare model forecasts with observations, rain gauge data was interpolated to the model grid using box averaging, with boxes centered on each model grid point. Figure 1 shows the number of observations available for the whole period under consideration at each 40km by 40 km box: it can be seen that, in many places, there are less than 100 observations, indicating grid boxes

with only one rain gauge station in this case the box averaging technique is equivalent to a nearest neighbor interpolation from the model grid to the rain gauge station.

For calibration and verification purposes the region has been subdivided into 3 sub regions (Figure 1): Region 1, northern and central Argentina, Region 2, South Eastern Brazil and Region 3 Northern South America. The main reason for this sub-division is to evaluate the impact of forecast calibration over regions with different precipitation regime and to assess the sensitivity of forecast skill to the considered region.

2.3 CALIBRATION METHODS

Several calibration methods are compared in this work. A parametric calibration method similar to that used by S2007, a non parametric method based on the methodology proposed by Gallus et al. (2007) and two versions of the algoritm based on the rank histogram (Hamill and Colucci 1998)

Rank histogram algorithm:

This method was first introduced by Hamill and Colucci (1998) and uses the rank histogram of a variable to compute the probability of occurrence of precipitation above a certain threshold. The implementation adopted here is very similar to the one described in Hamill and Colucci (1998).

Parametric algorithm:

To apply this method, precipitation values are first transformed by raising them to a power. In this work a power of 1/3 is used following the previous works of S2007 and Hamill (2007). The model is very similar to that proposed by S2007: the probability of having precipitation above a certain threshold is calculated considering the probability of the observed precipitation being cero and the probability of having precipitation above a certain threshold given that the precipitation is greater than cero The relationship between the conditional probability of having a cero in the observed precipitation given the forecasted precipitation is modeled through a logistic regression approach as in S2007 (Equation 1). In this expression f_k is the power transformed ensemble mean forecast, $P(y=0| f_k)$ is the conditional probability of the observation being 0 given that the ensemble mean forecast is f_k. The case where f_k is 0 is not considered for the logistic regression (and is not used in the model). This is considered trough the variable δ_k which is 0 if f_k is 0 and one is f_k is greater than 0. The variable *y* represents the observed value and a_o, a_1 and a_2 are constants.

$$f_k = ensemblemean^{1/3}$$
$$P(y=0 \mid f_k) = \frac{1}{1+e^{-z}} \qquad (1)$$
$$z = a_0 + a_1 f_k + \delta_k a_2$$

On the other hand to compute the probability of having precipitation above a certain threshold it is assumed that given that the observed precipitation is not equal to 0 then the PDF of the observed precipitation is a Gamma and that the parameters of the distributions are a function of the forecasted precipitation. Equation 2, shows the functional form of the relationship between the distribution parameters and the ensemble mean forecast.

$$mean = m_p f_k + m_b$$
$$sigma = s_p w + s_b \qquad (2)$$
$$w = \log(f_k + 1)$$

where mean and sigma are the mean and standard deviation of the fitted gamma distribution. The mean is assumed to be a linear function of the power transformed ensemble mean as in S2007 (where m_p and m_b are computed through linear regression between the ensemble mean and the observed precipitation). The standard deviation is assumed to be a function of the logarithm of the power transformed ensemble mean. This part of the model is different from that developed by S2007. In this case some tests were performed to asses the relationship between sigma and the forecasted precipitation and the proposed relationship gives a better fit. To explore this relationship the forecasted precipitation range was divided into several bins (each bin has approximately the same amount of elements). At each bin the standard deviation of the gamma distribution fitted to the observed precipitation corresponding to that bin was computed. The results are shown in Figure 2. As can be seen in this Figure the proposed relationship produces good results particularly for low precipitation values. For larger values the fit is not so

good and the values of standard deviation are almost independent of the forecasted precipitation. For this range of values the shape of the probability density function also is closer to a normal distribution (not shown). As can be seen there is also a difference between the computed standard deviation and the fitted gamma standard deviation which is usually higher.

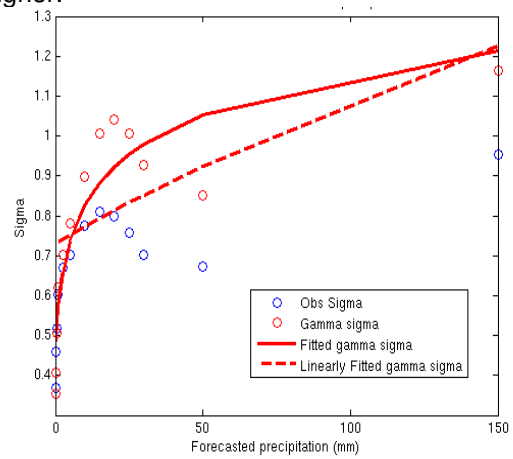

Figure 2: Standard deviation of the fitted gamma as a function of the forecasted precipitation (red circles), standard deviation of the observations as a function of the forecasted precipitation (blue circles), logarithmic fit to the Gamma standard deviation (red line) and linear fit to the Gamma standard deviation (dashed red line).

The above procedure is also used to determine s_p and s_b (from Equation 2) in the calibration process.

When the relationship between the mean forecasted precipitation and the mean of the gamma distribution fitted to the observations is computed using the approach above described, the results obtained are similar to that obtained in the case were the relationship is computed as the regression between the transformed forecasted precipitation and the observed precipitation. This suggests that the present implementation could be a good alternative for the computation of the regression between forecasted precipitation and standard deviation of the fitted gamma distribution.

S2007 applied this model to each ensemble member of an ensemble system based on different models. In our case, as all the members are based on the same model and only the initial conditions are perturbed then little differences are observed if the model is applied individually to each ensemble member or only to the ensemble mean.

Non-parametric algorithm:

This method is based on the work of Gallus et al. (2007). The range of the forecasted precipitation is divided in several bins. At each bin the probability of having precipitation above a certain threshold is computed directly from the observations without fitting a PDF to the data. This probability is assumed to take place at the center of the corresponding bin. Probabilities for other forecasted precipitation values are computed through linear interpolation.

Figure 3, shows an example of the computed relationship between the ensemble mean forecasted precipitation and the probability of precipitation occurrence above 2.5 mm.

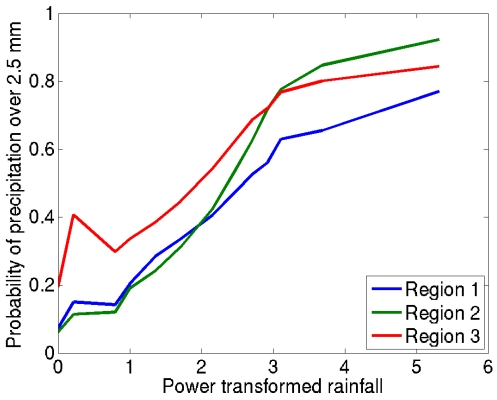

Figure 3: Relationship between the ensemble mean forecasted precipitation and the probability of occurrence of precipitation above 2.5 mm.

The main difference between this algorithm and the parametric one is that in this case the shape of the PDF is not assumed to be a gamma. Although the fit of the gamma distribution to the transformed precipitation observations is very good, it has been observed that for high forecasted precipitation amounts the shape of the distribution is closer to a normal distribution which suggests that better results can be obtained with a more general approach. More over, when these algorithms are used in a multi model ensemble, both algorithms can be easily applied independently to each ensemble member to take into account the bias of each model. However, in the case of the parametric algorithms, it is easier to compute weights associated to each ensemble member in order to calculate a Bayesian average of the probability associated with each ensemble member.

In the case of the algorithms based on the rank histograms, it should be remarked that they are not adequate for a multi model ensemble system, since they can't take into account the individual bias of each ensemble member.

3. RESULTS

Estimation of the calibration training period:

To estimate the length -in days- of the optimum training period, two different approaches were taken into account. First the dispersion of the distribution of the estimated parameters was evaluated as a function of the length of the training period. To explore this relationship, one hundred groups for each length –N- days were randomly selected, and the parameters were estimated for each group. Then the parameter mean and standard deviation was computed for each value of N. Figure 4 shows the results for the estimation of the parameters m_p and m_b from Equation 2. The estimation of the parameters became more stable for training periods larger than 20 days. The uncertainty in this estimation is, however, not the same for the three regions: for example at region 1 the uncertainty in the parameter estimation is larger, what can be due to a stronger temporal variability of the precipitation over this region or perhaps due to the size of this region which is smaller than the two others.

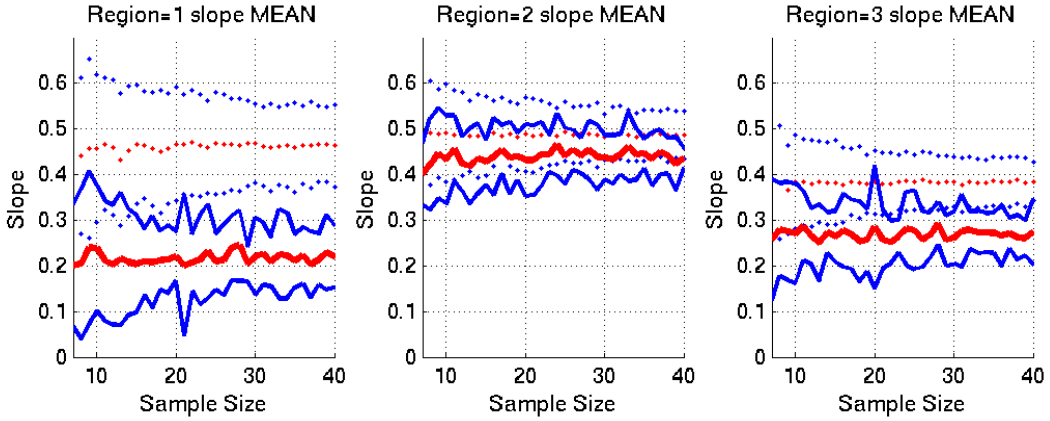

Figure 4: Mean slope (m_p) as a function of training period length (days) (red line): mean plus and minus one standard deviation (blue line) estimated from the rain gauge data (solid lines) and CMORPH estimates (doted lines) at region 1 (left panel), 2 (center panel) and 3 (right panel).

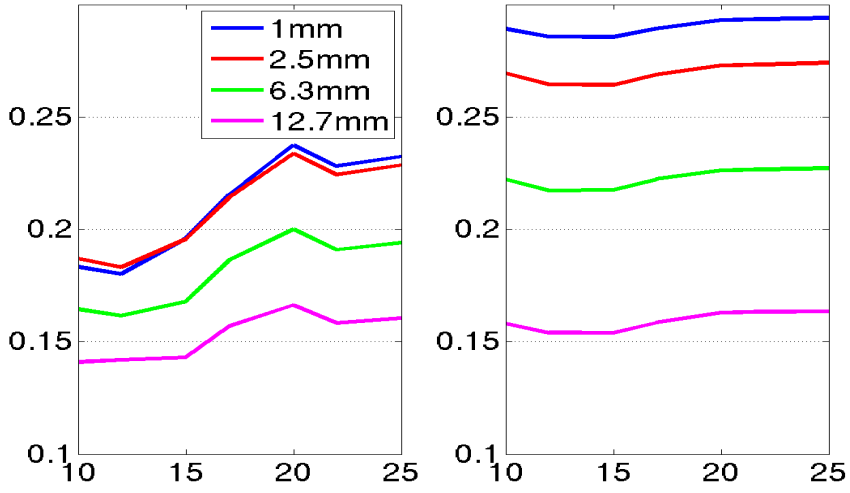

Figure 5: Brier Skill Score as a function of the length of the training period over regions one (left) and two (right) for different thresholds: 1 mm (blue line), 2.5 mm (red line), 6.3 mm (green line) and 12.7 mm (pink line).

The impact of the training period length upon the probabilistic forecast skill was also analyzed. The BSS was computed for forecasts calibrated using different training period lengths (both calibration and verification were performed using rain gauge data only), keeping the verification period the same for all the experiments. Figure 5 shows the relationship between the Brier skill score and the size of the verification sample at region 1 and 2. As expected, the brier skill score gets larger as the size of the training period increases, however some noise in the estimation of the BSS as a function of the verification period length has been observed. Some experiments were also performed randomly choosing the verification period. In this case the result was even noisier and the variability associated with the increase in the length of the training period was hard to detect, given the variability associated with different model skill over different periods.

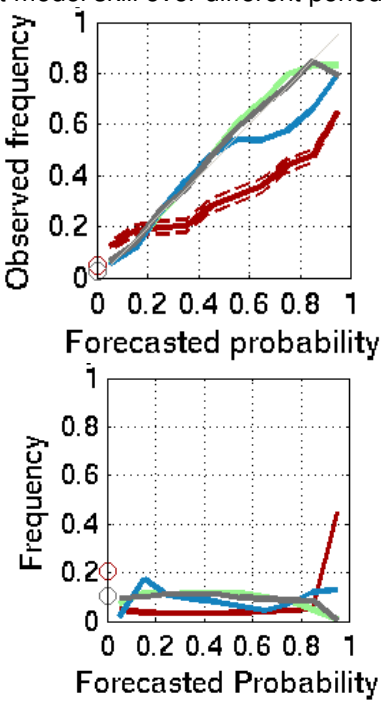

Figure 6: Upper panel: Reliability diagram for the 2.5 mm threshold over region 2. Uncalibrated forecast (red solid line) and its confidence intervals (red dashed line), rank histogram calibration (blue solid line), parametric calibration (green solid line) and non-parametric calibration (grey solid line). The lower panel shows the frequency of occurrence of each probability range in the forecast.

Based on the results discussed so far, the subsequent experiments have a 20 days training period length, which seem to provide a reasonable stability of the estimated parameters while little improvement is observed in the Brier skill score for values larger than those. Also, as the length of the experiment is relatively short (around 2 months) a training period larger than 20 days would significantly reduce the length of the period available for verification.

Comparison between different calibration strategies

As stated before, three different types of calibration strategies have been tested in this work. The performance of these algorithms has been tested over the three selected regions.

Figure 6 shows the reliability diagram for the 2.5 threshold over region 2. As can be seen, all the strategies reduce the amount of overestimation of the probability of having rainfall above this threshold. In the case of the algorithm based on the rank histogram, there is still an overestimation of the probability, particularly for high values. This over estimation is mainly due to the fact that the cases where the observed precipitation is 0 usually fall in the lower ranks of the histogram, producing higher forecasted probabilities when the selected threshold falls in the firsts ranks. In the lower panel of Figure 6, it can be seen that all the calibration strategies significantly reduce the frequency of forecasted probabilityes near 1, however the algorithm based on the rank histogram shows a higher frequency of this forecasts that the other two.

On the other hand, the parametric and non-parametric algorithms described before exhibit a similar behavior as can be seen in both panels of Figure 6.

Figure 7 shows the BSS and its reliability and resolution components for the different calibration strategies over the three regions. The best results in terms of the BSS are obtained over region 2 where the rain gauge network used for verification and calibration is denser, the worst results are observed over region 3 where the precipitation regime is more tropical. The parametric and non-parametric calibration approaches give similar results over the three regions. The non-parametric approach seems to be slightly better than the parametric one over regions 1 and 2 but the differences are non significant. Moreover, at 48-hour forecast length the result is the opposite (not shown). Mainly, the improvement of the BSS through calibration algorithms is explained by an

increase in the reliability of the forecasts. The methodology based on the rank histogram shows less reliability than the other two approaches.

Use of CMORPH estimates for forecast calibration:

One of the main problems for the implementation of this type of forecast calibration over South America is the lack of a dense rain gauge network. Most of the precipitation data used in this work is not part of the operative precipitation network, meaning that the data quality control performed over this data is not regularly done and so the data is not available in real time.

Consequently, the use of precipitation estimates from satellites could provide denser and more homogeneous data distribution for parameter estimation. The caveat of using these products is that they have errors. In the past decades, estimations based on passive microwave sensors has been developed. These algorithms are based in the relationship between the microwave emitted radiance at different wave lengths and the content of liquid water and ice in clouds. In particular the CMORPH algorithm (Joyce et al. 2004) uses passive microwave radiances for precipitation rate estimations and geostationary satellites for tracking precipitating systems. The skill of CMORPH precipitation estimates has been assessed over North America by Joyce et al. (2004) and over South America by Ruiz (2009). In the later, also a statistically based calibration is proposed to reduce CMORPH systematic errors as a function of the estimated precipitation amounts.

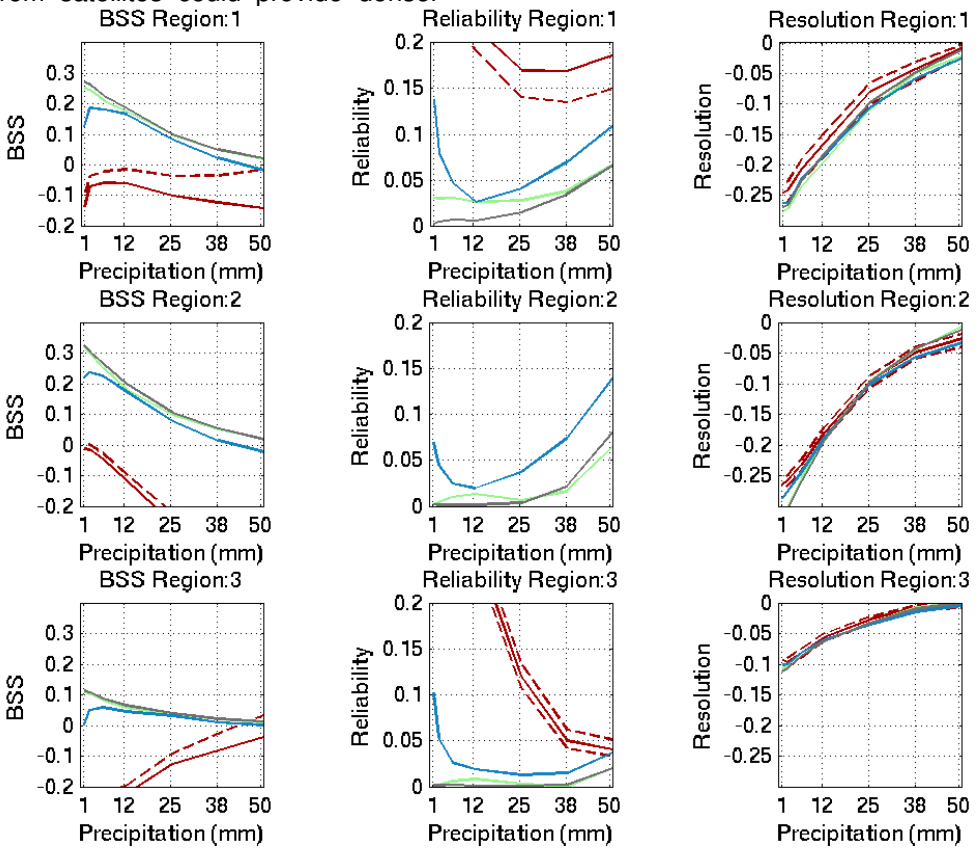

Figure 7: BSS as a function of precipitation threshold (mm) first column. Second column: as in the first column but for the reliability component of the BSS. Third column: as in the first column but for the resolution component of the BSS. The first row is for region 1, the second for region 2 and the third for region 3.

The impact of using CMORPH in the calibration process was estimated in Ruiz et al. 2009, using only CMORPH estimates interpolated to the rain gauge locations. This procedure does not take advantage of the larger number of precipitation estimates available in CMORPH data for locations where no rain gauges are present. In order to obtain a better measure of

increased data availability through CMORPH data, we now used the full CMORPH estimates for forecast calibration.

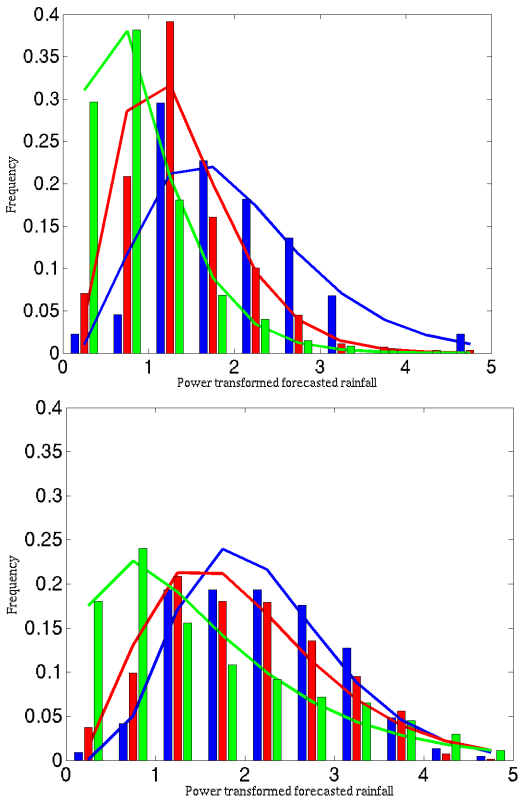

Figure 8: Observed distribution (bars) and fitted distribution (solid lines) for observed rainfall when the mean forecast is (a) between 0 and 0.01 mm and (b) between 5 and 10 mm over region 1. Rain gauge data in blue, CMORPH raw estimates in green and CMORPH calibrated estimates in red.

Figure 4 shows the parameter estimation from CMORPH data as a function of the length of the training period. Results are quite similar over region 2, but differ over region 1 and 3. This difference is due to differences between the PDF of CMORPH and rain gauges. The difference is larger for cases where the forecasted precipitation is small as can be seen comparing Figures 8a and 8b. CMORPH data, both calibrated and uncalibrated, show higher frequencies at small precipitation amounts (around 1 mm), although the calibrated estimates distribution is closer to the observed one, as could be expected. When the comparison between the two distributions is performed taking into account only the grid points where both rain gauge data and CMORPH estimates are available the distributions are much closer (not shown). This suggests that the differences observed in Figure 8 between the calibrated CMORPH and the rain gauge data is mainly due to the spatial variability of the rainfall within region 1 which is poorly represented by the rain gauge network. Over region 2, where the rain gauge network is denser, the differences between the PDF for the observed data and the CMORPH estimates is smaller as suggested by Figure 4 and also from the comparison of the PDFs (not shown).

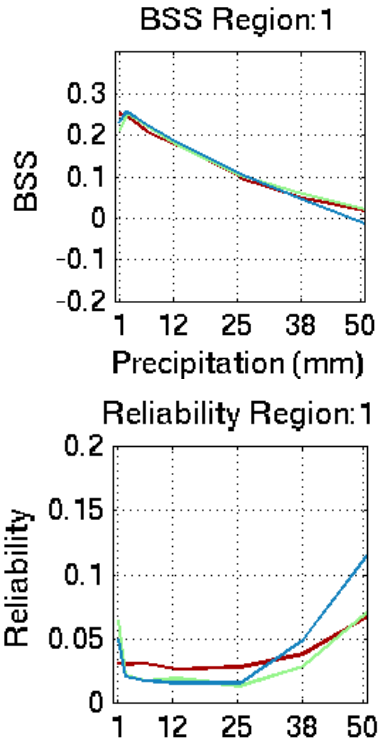

Figure 9: BSS (upper panel) and its resolution component (lower panel) over region 1 for the 24 hour forecast using rain gauge data (dark red), raw CMORPH (blue) and calibrated CMORPH (green) as training data.

The impact of CMORPH data upon probabilistic forecast skill is assessed using the BSS. Figure 9 shows the BSS computed over region 1 using the parametric algorithm for probability calibration with the rain gauge data, the raw CMORPH estimates and the calibrated CMORPH estimates. The verification is performed using the rain gauge data in all the cases. As can be seen, the performance of both calibrated and raw CMORPH are quite close to the case where only rain gauges are used. Similar results where obtained over the other 2 regions and also in the case where the non-parametric algorithm was used (not shown).

In some cases, raw CMORPH estimates produce a degradation of the forecasts for thresholds over 25 mm (not shown), what might be due to the systematic errors in the CMORPH estimates over these regions which tend to over estimate high precipitation events. These results suggest that, even though there are differences between the PDF of CMORPH data and the rain gauge data, this estimate can be used for forecast calibration without loosing skill.

CONCLUSIONS

Three methods for probabilistic forecast calibration have been evaluated over South America. The best results have been obtained with methods that attempt to describe the conditional PDF of the observed rainfall given the amount of forecasted precipitation (both parametric and non-parametric) by the ensemble mean. In this case as the entire ensemble members are based on the same model, the use of the ensemble mean gives the same results as the mean of the probabilities forecasted for each ensemble members.

No difference has been observed between the parametric and non-parametric algorithms suggesting that the fit to a gamma distribution is quite good to compute the calibrated probabilities.

The use of CMORPH estimates produce little impact upon forecast skill. This is particularly important over this region because the operational rain gauge network is far coarser than the one used in this work for forecast calibration and verification. The calibration of CMORPH estimates produces a distribution that is closer to the observed one. Still, the impact upon forecast skill is small.

ACKNOWLEDGMENTS

The authors are thankful to Istvan Szunyogh for providing the scripts to run the global ensemble, to Erick Kostelich for his help with the MRF model runs and to Jae Schemm for providing the initial conditions in the required file format.
This study has been supported by the following projects: ANPCyT PICT 2004 25269, UBACyT X204, CONICET PIP 5417 and GC06-085 from NOAA/OGP/CPPA.

REFERENCES

Dudhia J., 1989: Numerical study of convection observed during the winter monsoon experiment using a mesoscale two-dimensional model, *J. Atmos. Sci.*, **46**, 3077-3107.

Ebert, E. E., 2001: Ability of a Poor Man's Ensemble to Predict the Probability and Distribution of Precipitation. *Mon. Wea. Rev.*, **129**, 2461-2480.

Gallus, W. A., M. E. Baldwin and K. L. Elmore, 2007: Evaluation of Probabilistic Precipitation Forecasts Determined from ETA and AVN forecasted amounts. *Wea. Forecasting*, **22**, 207-215.

Hamill, T. 2007: Comments on "Calibrated Surface Temperature Forecasts from the Canadian Ensemble Prediction System Using Bayesian Model Averaging", *Mon. Wea. Rev.*, **135**, 4226-4230.

Hamill, T. and S. J. Colucci, 1998: Evaluation of Eta-RSME Ensemble Probabilistic Precipitation Forecasts. *Mon. Wea. Rev*, **126**, 711-724.

Hong S. and H. Pan, 1996: Nonlocal Boundary Layer Vertical Diffusion in a Medium-Range Forecast Model. *Mon. Wea. Rev.*, **10**, 2322-2339.

Joyce, R. J., J. E. Janowiak, P. A. Arkin, and P. Xie, 2004: CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. *J. Hydromet.*, **5**, 487-503.

Kain J. S., 2004: The Kain-Fritsch Convective Parameterization: An Update. *J. Appl. Meteor*. **43**, 170-181.

Kalnay E., M. Kanamitsu, R. Kstler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen, Y. Zhu, A. Leetmaa, R. Reynolds, M. Chelliah, W. Ebisuzaki, W. Higgins, J. Janowiak, K. C. Mo, C. Roplewski, J. Wang, Roy Jenne, and Dennis Joseph, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, **77**, 437-471.

McLean Sloughter J., J., A. Raftery, T. Gneiting, and C. Fraley, 2007: Probabilistic Quantitative Precipitation Forecasting Using Bayesian Model Averaging. *Mon. Wea. Rev.*, **135**, 3209-3220.

Penalba, Olga, C. Vera, B. Cerne, M. Rusticucci, P. Salio, L. Ferreira, B. Liebmann, D. Allured and A. Diaz, 2004: Daily Rainfall Data over Argentina and Uruguay during SALLJEX. *Clivar Exchanges*, **29**, 29-31.

Ruiz, J., C. Saulo and E. Kalnay, 2009: Comparison of methods to generate probabilistic quantitative precipitation forecasts over South America, *Wea. Forecasting*, **24**, 319-336.

Ruiz, J, 2009: CMORPH precipitation estimates calibration and verification over South America. (in Spanish), *Revista Brasileira de Meteorologia*, in press.

Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, W. Wang, and J. G. Powers, 2005: A description of the Advanced Research WRF Version 2. NCAR Tech Notes-468+STR

Stensrud, D. and N. Yussouf, 2007: Reliable Probabilistic Quantitative Precipitation Forecasts from a Short-Range Ensemble Forecasting System. *Wea. Forecasting*, **22**, 3-17.

Toth, Z. and E. Kalnay, 1993: Ensemble forecasting at NMC: The generation of perturbations. *Bull. Amer. Meteor. Soc.*, **74**, 2317-2330.

Vera, C., J. Baez, M. Douglas, C. B. Emmanuel, J. Marengo, J. Meitin, M. Nicolini, J. Nogues-Paegle, J. Paegle, O. Penalba, P. Salio, C. Saulo, M. A. Silva Dias, P. Silva Dias, and E. Zipser, 2006: The South American Low-Level Jet Experiment. *Bull. Amer. Meteor. Soc.,* **87**, 63-77.

Wilks, D. S., 1995: Statistical Methods in the Atmospheric Sciences: An Introduction. International Geophysics Series, Vol. 59, Academic Press, 467 pp.

Zhu, Y., Z. Toth, R. Wobus, D. Richardson, and K. Mylne, 2002: The Economic Value of Ensemble-Based Weather Forecast. *Bull. Amer. Meteor. Soc.*, **83**, 73-83.