

## 5.3 PRECIPITATION FORECASTS OF THE CANADIAN ENSEMBLE PREDICTION SYSTEM.

Syd Peel\* and Laurence J. Wilson  
Meteorological Research Branch  
Meteorological Service of Canada

### 1 INTRODUCTION

The prediction of precipitation amount is arguably the greatest challenge facing the weather forecaster today. Moreover, this weather element often holds the greatest interest to the general public; indeed, the amount of precipitation anticipated is often the only thing people care about when they consult the forecast. It is also of paramount importance to a number of sectors vital to the economy: agriculture, transportation, and utilities to name but a few.

Crude attempts to quantify the uncertainty inherent in precipitation forecasts, conveyed in the form of the probability of precipitation occurrence, date back well before the development of ensemble prediction systems (EPS's). However, with the advent of the EPS into operational use, a much more complete elucidation of the probabilistic nature of a weather forecast is now possible, with the potential of supplying a great deal more information than is contained in a deterministic forecast.

An EPS produces a set of precipitation forecasts, each corresponding to a perturbation of the initial state from the analyzed state of the atmosphere, in an attempt to capture the uncertainty intrinsic to the meteorological observations used to initialize the integration of the NWP model. A single deterministic precipitation forecast is thus replaced with a distribution intended to model forecast uncertainty, the latter arising from the inevitable error in the meteorological observations needed to initialize the model and, in the case of the Canadian EPS, imperfections in the model itself.

In this paper we examine EPS precipitation forecasts from the Canadian ensemble forecast system by validating these precipitation forecasts against surface station reports from across the country. In addition we explore the impact of constructing smoothed pdf's using nonparametric asymmetric kernel density estimators. In the second section of the paper the forecast and observational data used in the study are briefly described. In section 3 we outline the nonparametric models used to smooth the ensemble samples, while in section 4 some preliminary results are presented. The final sec-

tion includes a discussion of the results and outlines some possible avenues for future work.

### 2 DATA

EPS forecasts are generated each day at the Canadian Meteorological Centre (CMC) of the Meteorological Service of Canada (MSC), integrated from an analysis at 00 UTC. The control solution is obtained from the Spectral Finite Element model (Ritchie and Beaudoin, 1994; Buizza et al., 2005), as are 8 of the perturbed members of the ensemble. The other 8 perturbations are produced from the same Global Environmental Multiscale (GEM) model (Côté et al., 1998) which generates the short- and medium-range deterministic forecasts disseminated by the CMC, albeit on a slightly coarser grid. In 2001 the resolution of the spectral members was improved from  $T_L95$  to  $T_L149$  and that of the GEM members from  $1.875^\circ$  to  $1.2^\circ$  (Pellerin et al., 2003; Buizza et al., 2005). In January 2005 the ensemble Kalman filter method was incorporated into the assimilation cycle for the operational EPS. We have therefore elected to restrict the EPS forecast data considered in this analysis to the period between August 1, 2001 and December 31, 2004.

The gridded EPS precipitation forecasts were interpolated to the locations of observation stations from across the country (see Fig 1.). The 6-hourly precipitation amounts in the synoptic reports from these stations were summed to produce the observational record of precipitation amount, for accumulations over 1 to 10 days. Some rudimentary screening was applied to this observational data - station reports of unrealistically large precipitation amounts were culled, and certain stations were excluded from consideration based on egregiously unreasonable time series or suspected systemic instrumentation defects.

---

\*Syd.Peel@ec.gc.ca, 4905 Dufferin St., Toronto, Ontario, Canada M3H 5T4



Figure 1. Locations of stations supplying observational data.

### 3 MODELLING THE PDF

Producing only 17 precipitation forecasts for a given forecast projection, at a given location, the Canadian EPS makes available a forecast sample which is but a third the size of the sample available from the operational ECMWF system which currently consists of 51 members (Buizza et al., 2005). Hence if we were to restrict ourselves to the empirical cumulative distribution function (ecdf) derived directly from the sample of member forecasts then the resolution of the associated probabilistic forecast model would be coarser than 5%. In light of this consideration, as well as the encouraging results reported in Wilks (2002), we decided to explore the feasibility of modelling the pdf associated with the sample of ensemble forecasts.

#### 3.1 Gamma kernel density estimates

Precipitation amount is often treated as a Gamma random variable with probability density function, in its two-parameter form, given by

$$f_{\gamma}(x; \alpha, \beta) = \begin{cases} \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-\frac{x}{\beta}} & \forall x \geq 0 \\ 0 & \forall x < 0 \end{cases}$$

where  $\Gamma$  is the gamma function and  $\alpha, \beta$  are referred to as the shape and scale parameters, respectively. The Gamma distribution evinces pronounced positive skewness; indeed for  $\alpha < 1$  there is a singularity at the origin, while for  $\alpha > 1$  the distribution possesses a non-zero mode which increases with increasing  $\alpha$ .

While precipitation climatologies are typically well rendered by the Gamma distribution, an EPS member's QPF can be considered as a random variable whose pdf describes the probability of precipitation amount conditional upon the state of the control used to initialize the ensemble forecast. As such, they do not necessarily follow a Gamma distribution. We therefore resorted to kernel density estimators to construct nonparametric approximations to the pdf representing the EPS QPF's.

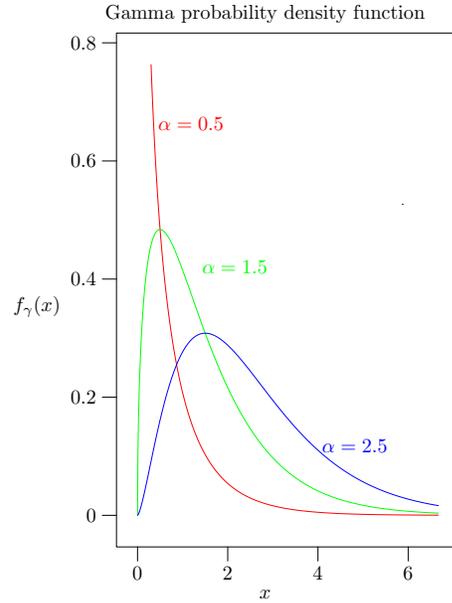


Figure 2. Dependence of  $f_{\gamma}$  on  $\alpha$ .

While the EPS member QPF's need not in general be Gamma random variables, their support is obviously confined to  $[0, \infty)$ , thereby rendering Gaussian kernels unsuited to the construction of a kernel density estimator for the pdf. This is particularly true for precipitation accumulations over smaller time intervals, in which case the mass of the distribution preponderates near the origin (for forecasts of precipitation over sufficiently long intervals the centre of the distribution will be displaced sufficiently far from the origin as to permit adequate representation with Gaussian kernels). Rather than apply a transformation such as Wilks employed for surface winds and proposed for precipitation amount, whereby the distributions of the original weather elements are mapped to Gaussian distributions (Wilks, 2002), we opted for a more direct approach

using gamma kernels. In particular the pdf of the EPS QPF's was modelled by the gamma kernel density estimators prescribed in Chen (2000).

In kernel density estimation a probability density function, or kernel, is associated with each data point in the sample selected from the population whose pdf we seek. Like the more familiar Gaussian kernel, Chen's Gamma kernels possess a single maximum located at the associated data point. Unlike their Gaussian counterparts the Gamma kernels are asymmetric, this asymmetry becoming increasingly pronounced, and the width of the kernels increasingly narrow, as the data point on which they are centred approaches the origin. For data points right at the origin, corresponding to a forecast of nil precipitation, the shape factor of the associated Gamma kernel is less than 1 (see Fig 2.) and the concomitant maximum at the origin is actually a singularity.

The kernel density estimator  $\hat{f}$  is obtained from the individual kernels by taking their average over the data points available from the sample

$$\hat{f}_{\Gamma}(x) \equiv \frac{1}{n} \sum_{i=1}^n f_{\gamma}(x; \frac{x_i}{h} + 1, 1/h) \quad (1)$$

where  $f_{\gamma}$  is the gamma probability density,  $h$  is the bandwidth,  $n$  is the size of the sample used to estimate the probability density, and the  $x_i$  are the sample points, which in our application are just the ensemble member precipitation forecasts. An example of a Gamma kernel density estimator, with its constituent kernels, can be seen in Fig 3.

### 3.2 Bandwidth determination

While the kernel density estimator (KDE) is a non-parametric model for the pdf, one parameter does remain to be determined, namely the bandwidth  $h$ . The magnitude of the bandwidth controls the amount of smoothing effected by the KDE on the raw data comprising the sample of ensemble member forecasts. The larger the bandwidth, the larger the relative width of the Gamma kernels, the greater the influence of each associated data point on the entire KDE, and the greater the smoothing.

For Gaussian KDE's a simple closed-form expression for the bandwidth which minimizes the mean integrated square error (MISE) of the density estimator,

in the asymptotic limit of arbitrarily large sample sizes, is given by

$$h_{opt} = \left(\frac{4}{3}\right)^{\frac{1}{5}} \sigma n^{-\frac{1}{5}} \quad (2)$$

where  $\sigma$  is the standard deviation of the population and  $n$  is the size of the sample.

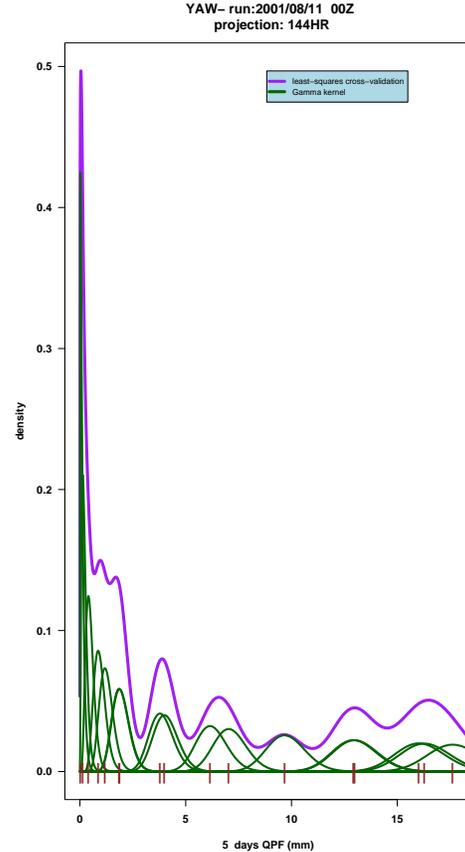


Figure 3. Construction of the gamma kernel density estimator

The above bandwidth generally works well in the construction of KDE's to distributions which are unimodal and roughly symmetric. However, the distributions of the EPS forecasts cannot a priori be assumed to be unimodal. Furthermore, their distributions can possess considerable positive skewness, particularly for precipitation accumulated over shorter intervals. For distributions evincing such deviations from normality the *normal scale bandwidths* such as are obtained from (2) tend to be too large and thereby oversmooth the data

(Wand and Jones, 1995). We thus explored the cross-validation methods prescribed in Silverman (1986) for the determination of a more appropriate bandwidth.

The first step in cross-validation methods is the construction of the functions  $\hat{f}_{-i}$  defined by

$$\hat{f}_{-i}(x) \equiv \frac{1}{16} \sum_{j \neq i} f_{\gamma}(x; \frac{x_j}{h} + 1, \frac{1}{h}) \quad (3)$$

which are simply the KDE's obtained from the original sample of EPS forecasts except that the data point  $x_i$  has been withheld.  $x_i$  can then be employed as an independent data point to test the density estimate  $\hat{f}_{-i}$ , yielding the likelihood  $\hat{f}_{-i}(x_i)$ . The least-squares cross-validation (lscv) bandwidth is obtained by minimizing the expression  $M_0(h)$  defined by

$$M_0(h) \equiv \int \hat{f}^2 - \frac{2}{n} \sum_i \hat{f}_{-i}(X_i) \quad (4)$$

with respect to  $h$ , which is equivalent to determining that bandwidth which minimizes the integrated square error  $\int (\hat{f} - f)^2$  in  $\hat{f}$  (Silverman, 1986).

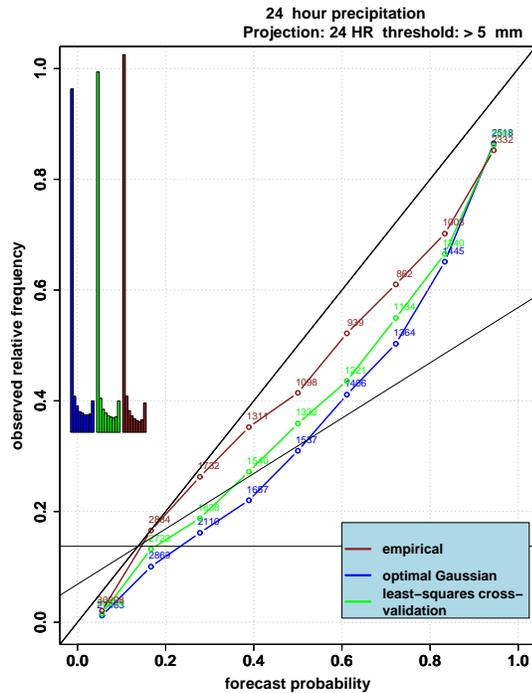


Figure 4. Reliability table for 1-day precipitation/24-hour projection

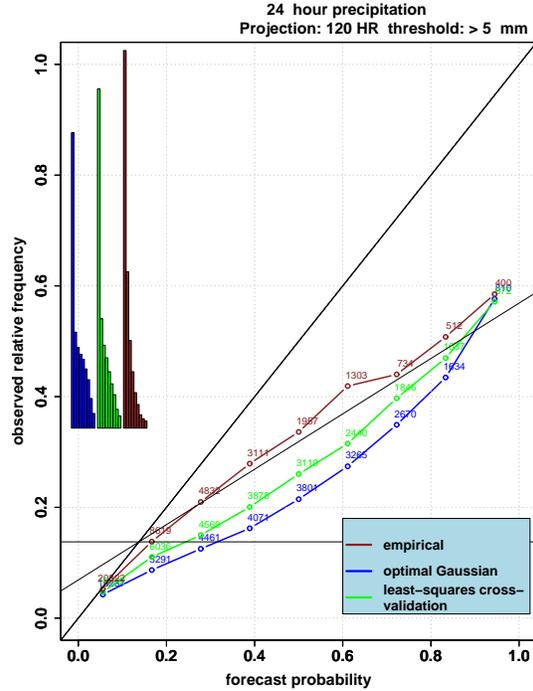


Figure 5. Reliability table for 1-day precipitation/5-day projection

## 4 PRELIMINARY RESULTS

In order to validate the precipitation forecasts from the Canadian EPS we computed reliability diagrams, Brier Skill scores, and ROC areas (Stanski et al., 1989) to measure the quality of the EPS forecasts against surface station reports and gauge the impacts of fitting the EPS member forecasts to a Gamma KDE model. Thus in Fig. 4 can be seen the reliability diagram for the raw EPS probability forecasts of 24-hour precipitation at a forecast projection of 24 hours, along with the forecasts obtained from the Gamma KDE models using the optimal Gaussian and least-squares cross-validation bandwidths. Since the Canadian EPS produces samples of 17 members, resulting in a discrete range of 18 possible probability forecasts, 9 bins were used to construct the reliability tables in order to ensure a fair comparison between the empirical and fitted forecasts while attempting to keep the noise in the tables to a minimum (Candille and Talagrand, 2005). The horizontal line corresponding to an observed relative frequency of

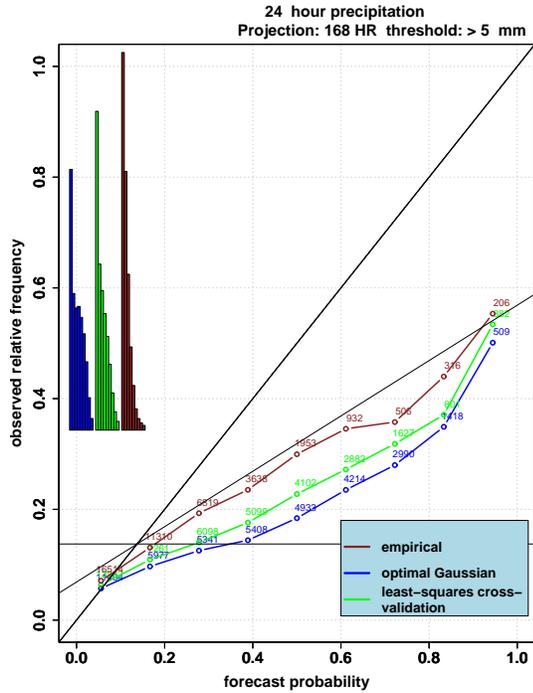


Figure 6. Reliability table for 1-day precipitation/7-day projection

roughly 15% represents the sample climatology, while the line bisecting the angle between this line and the line having unit slope delineates the skill vs. no-skill boundary.

The 24 hour forecast shows quite good reliability, with an over-forecasting bias at increasing forecast probabilities but skill over the whole domain except for the bin corresponding to the lowest forecast probability, where the EPS also evinces a slight over-forecasting bias. By 120 hours the bias in the lowest probability forecasts has disappeared while the reliability in the rest of the bins has degraded, but the forecasts are still demonstrating skill for all but the lowest probability bins. In Fig. 6 it is clear that by 168 hours reliability has deteriorated to the point that the forecasts possess little skill.

Whereas for density estimators constructed from symmetric kernels the asymptotic bias is proportional to the curvature of the true density (Wand and Jones, 1995), the bias in the Gamma kde's is slightly more complicated, given by (Chen, 2000)

$$E\hat{f}_\Gamma(x) - f(x) = h \left\{ f'(x) + \frac{1}{2}x f''(x) \right\} + o(h) \quad (5)$$

For precipitation accumulation over one day the behaviour of the distribution near the origin can reasonably be expected to resemble the Gamma density with shape less than 1, resulting in a negative bias near the origin. Hence the population of the lowest probability bin of the KDE models is smaller than that of the unsmoothed EPS forecast, as is evident in the histograms displayed in the reliability table of Fig. 4. With increasing forecast probability the slope of the true density diminishes while the moment of its curvature becomes increasingly positive, and the populations of the bins for the KDE forecasts become larger than those of the corresponding bins of the raw EPS forecast.

Hence the effect of the smoothing introduced by the KDE models is to push the distribution towards higher probabilities. This results in somewhat sharper forecasts from the KDE models at the upper end of the spectrum with no significant degradation in reliability, at the cost of lower reliability for forecasts of intermediate probabilities. The least-squares cross-validation bandwidth is generally around half the normal scale bandwidth, thus its performance is intermediate between that of the empirical EPS forecast and the KDE model fitted using the normal scale bandwidth.

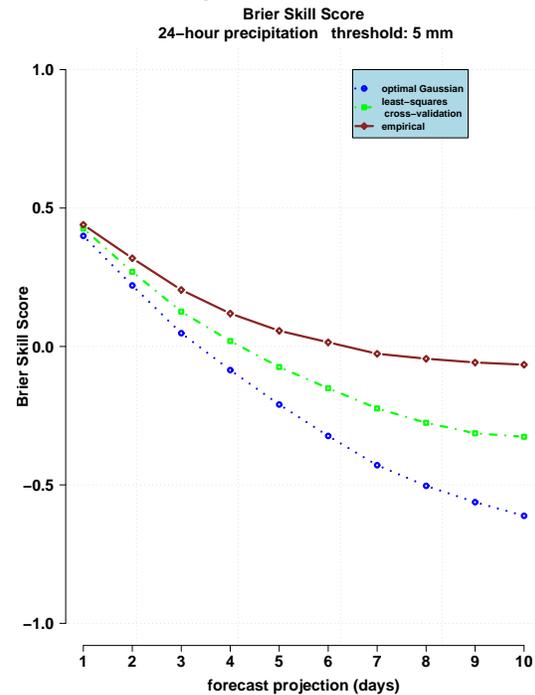


Figure 7. Brier Skill Score as a function of forecast projection

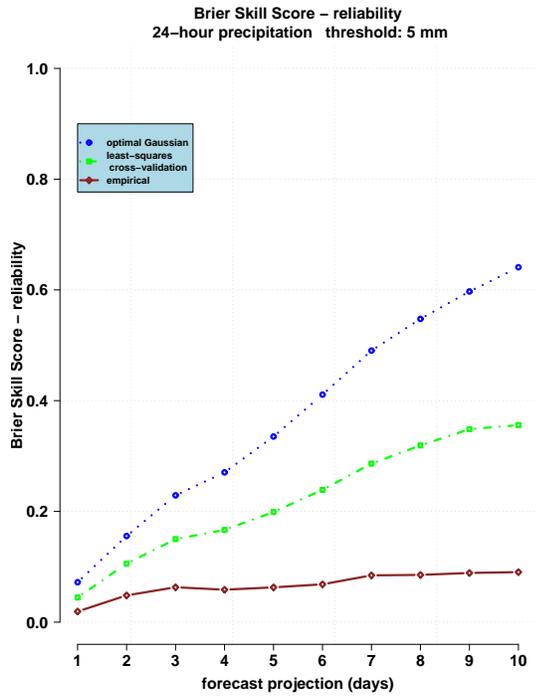


Figure 8. Reliability as a function of forecast projection

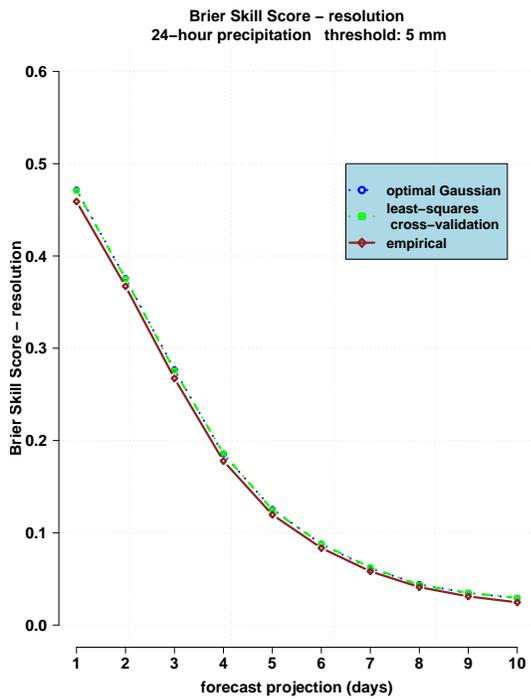


Figure 9. Resolution as a function of forecast projection

The behaviour of the Brier Skill Score with increasing forecast projection can be observed in Fig 7 where it is seen that the EPS forecasts have little skill at a forecast projection of a week or more. The skill of the modelled forecasts is inferior to the unsmoothed EPS forecasts at all projections, becoming increasingly inferior with increasing forecast projection.

The Brier score  $BS$  can be resolved into contributions from the reliability, resolution, and uncertainty of the forecast system (Murphy, 1973; Stanski et al., 1989) :

$$\begin{aligned}
 BS &\equiv \frac{1}{N} \sum_{i=1}^N (f_i - o_i)^2 \\
 &= \frac{1}{N} \sum_{k=1}^b \sum_{i=1}^{N_b} \left\{ \underbrace{(f_k - \bar{o}_k)^2}_{\text{reliability}} - \underbrace{(\bar{o}_k - \bar{o})^2}_{\text{resolution}} + \underbrace{(\bar{o} - o_{ki})^2}_{\text{uncertainty}} \right\}
 \end{aligned}$$

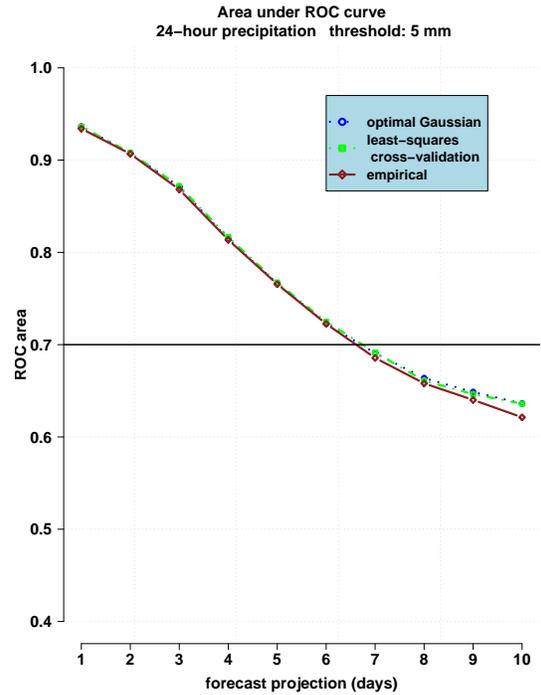


Figure 10. ROC areas for 5mm threshold, as a function of projection

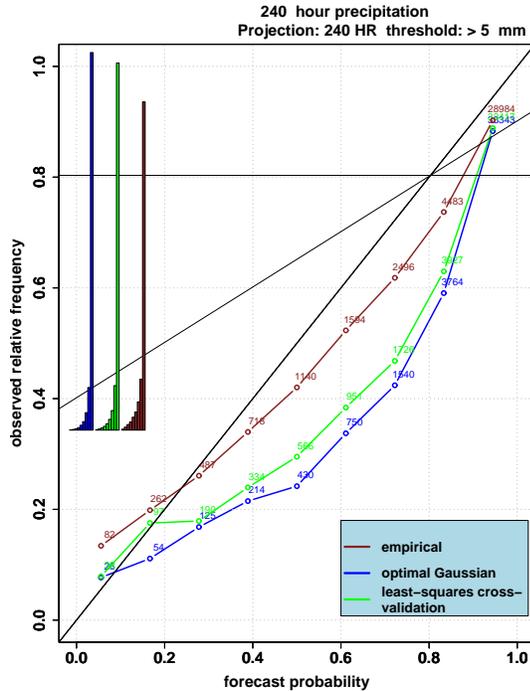


Figure 11. Reliability table for 10-day precipitation/10-day projection

where  $N$  is the total number of pairs of probabilistic forecasts  $f_i$  and verifying binary observations  $o_i$ ,  $b$  is the number of bins,  $f_k$  is the forecast representative of bin  $k$ , given by the forecast at the centre of the bin,  $\bar{o}_k$  is the mean of the observations corresponding to the forecasts falling within bin  $k$ ,  $\bar{o}$  is the mean of all the observations in the sample, and  $o_{ki}$  is the  $i$ th of  $N_b$  observations in bin  $k$ .

Normalizing the reliability and resolution components by the uncertainty, which is just the sample variance, we see upon comparing Figs. 8 and 9 that the superiority in the Brier skill score of the raw EPS forecasts over the smoothed models derives from its superior reliability. The resolution of all three models is virtually indistinguishable and diminishes rapidly with increasing forecast projection. The difference in reliability between the three models increases dramatically with increasing forecast projection, the reliability of the raw EPS forecasts evincing only a very weak dependence upon forecast projection, which is typical of current operational EPS's (Candille and Talagrand, 2005).

The area under the receiver operating characteristic (ROC) curve (Harvey et al., 1992; Stanski et al., 1989; Mason, 1982) is plotted as a function of forecast

projection in Fig. 10. This area measures the ability of the forecast systems to discriminate between the occurrence or non-occurrence of precipitation exceeding the given threshold, in this case 5 mm. As can be seen there is little difference between the ROC areas of the three sets of forecasts. The area diminishes with increasing forecast projection, dropping below the no-skill mark of 0.7 for forecast lead times of 7 days.

In Fig. 11 can be seen the reliability table for forecasts of precipitation accumulated over a period of 10 days, again using 5 mm for a threshold. The EPS evinces good reliability, but skill only in the extreme probability bins. The behaviour of the smoothed models relative to the raw EPS forecasts is as before. The KDE models do appear to be more reliable at the low end of the spectrum, but the population of these bins is quite small so considerable caution must be exercised when drawing conclusions.

## 5 Discussion

Our preliminary results indicate that the precipitation forecasts from the Canadian EPS are quite reliable, evincing only very gradual degradation with increasing forecast projection. The smoothing of the raw EPS forecasts by fitting asymmetric kernel density estimators to them introduces bias into the resulting forecasts (Wand and Jones, 1995; Silverman, 1986), which is reflected in reliabilities which are inferior to the original EPS. The improvement in the resolution of the KDE forecasts with respect to the raw EPS, on the other hand is barely perceptible, and the skill of the fitted KDE forecasts is thus generally inferior to the original EPS output, the smoothed forecast models lagging further behind the raw EPS with increasing forecast projection. However, the reliability of the kernel density estimators is competitive with that of the raw ensemble forecasts at the extremities of the distributions, while at the same time producing somewhat sharper forecasts in the upper end of the spectrum.

Moreover, while the normal scale bandwidth (referred to as the "optimal Gaussian") was expected to be too large, the bandwidth obtained from the least-squares cross-validation algorithm also seems to be significantly too large, resulting in oversmoothing of the resulting KDE. Further work is therefore needed to determine a more optimal bandwidth, either by recourse

to another of the algorithms proposed in Wand and Jones (1995) and Silverman (1986) or, perhaps simply searching for that bandwidth which minimizes the Brier skill score. The bandwidth is crucial to the quality of the kernel density estimators - a more propitious choice of bandwidth could dramatically improve the quality of these forecasts.

The effects of stratification by season and region is currently underway and results of this facet of the investigation should be available soon.

## Acknowledgements

Part of the observational record was extracted from a database constructed by Gerard Croteau at the Canadian Meteorological Centre of the Meteorological Service of Canada.

Brier scores and reliability tables were generated with R code adapted from/inspired by the verification package contributed by the NCAR Research Application Program.

The graphics were generated using R (Ihaka and Gentleman, 1996), MetaPost and MAX (developed and maintained at the Canadian Meteorological Centre).

ROC areas were computed using code supplied by Gérard Pellerin at the Canadian Meteorological Centre of the Meteorological Service of Canada, adapted from software due to D.D. Dorfman and E. Alf, based on code listed in Swets and Pickett (1982).

This document was prepared in L<sup>A</sup>T<sub>E</sub>X.

## References

- Buizza, R., P.L.Houtekamer, Z. Toth, G. Pellerin, M. Wei, and Y. Zhu : 2005, A Comparison of the ECMWF, MSC, and NCEP Global Ensemble Prediction Systems. *Mon. Wea. Rev.*, **133**, 1076–1097.
- Candille, G. and O. Talagrand : 2005, Evaluation of probabilistic prediction systems for a scalar variable. *Quart. J. Roy. Meteor. Soc.*, **131**, 2131–2150.
- Chen, S. X. : 2000, Probability Density Function Estimation Using Gamma Kernels. *Ann. Inst. Stat. Math.*, **52**, 471–480.
- Côté, J. S., S. Gravel, A. Méthot, A.Patoine, M.Roch, and A.Staniforth : 1998, The operational CMC/MRB Global Environmental Multiscale (GEM) model. Part I : Design considerations and formulation. *Mon. Wea. Rev.*, **126**, 1373–1395.
- Harvey, L. O., Jr., K. R. Hammond, C. M. Lusk, and E. F. Mross : 1992, The Application of Signal Detection Theory to Weather Forecasting Behavior. *Mon. Wea. Rev.*, **120**, 863–883.
- Ihaka, R. and R. Gentleman : 1996, R : A Language for Data Analysis and Graphics. *Journal of Computational and Graphical Statistics*, **5**, 299–314.
- Mason, I. : 1982, A model for assessment of weather forecasts. *Aust. Met. Mag.*, **30**, 291–303.
- Murphy, A. H. : 1973, A new vector partition of the probability score. *J. Appl. Meteor.*, **12**, 595–600.
- Pellerin, G., L.Lefavre, P. Houtekamer, and C.Girard : 2003, Increasing the horizontal resolution of ensemble forecasts at CMC. *Nonlinear Processes Geophys.*, **10**, 463–468.
- Ritchie, H. and C. Beaudoin : 1994, Approximations and sensitivity experiments with a baroclinic semi-Lagrangian spectral model. *Mon. Wea. Rev.*, **122**, 2391–2399.
- Silverman, B. W. : 1986, *Density Estimation for Statistics and Data Analysis*. Chapman and Hall/CRC.
- Stanski, H. R., L. J. Wilson, and W. R. Burrows : 1989, Survey of common verification methods in meteorology. Technical report, Atmospheric Environment Service - Forecast Research Division.
- Swets, J. A. and R. M. Pickett : 1982, *Evaluation of Diagnostic Systems : Methods from Signal Detection Theory*. Academic Press.
- Wand, M. P. and M. C. Jones : 1995, *Kernel Smoothing*. Chapman and Hall/CRC.
- Wilks, D. S. : 2002, Smoothing forecast ensembles with fitted probability distributions. *Quart. J. Roy. Meteor. Soc.*, **128**, 2821–2836.