P6B.5 QUANTIFYING AND PREDICTING THE ACCURACY OF RADAR-BASED QUANTITATIVE PRECIPITATION FORECASTS

Frédéric Fabry[1] and Alan Seed[2]
[1]McGill University, Montreal, Quebec, Canada
[2]Bureau of Meteorology Research Centre, Melbourne, Victoria, Australia

THE SHORT VERSION

On the basis that hydrological users need to know the event-by-event forecast uncertainty (Fig. 1), we computed distributions of radar rainfall forecast uncertainty as a function of forecast lead time, basin size, and forecasted rainfall intensity (e.g., Figs. 5 to 7) using data from the U.S. 3-D National Mosaic of radar data (Fig. 3). Since these uncertainties are also weather dependent, we tried to find good predictors (Fig. 4, Table 1) to help either reduce the forecast uncertainty or better define it (Fig. 2). The value of some predictors was significant though modest (Fig. 9), the predictors being more skillful at characterizing forecast uncertainty than at improving forecast accuracy. The value of predictors also depended on forecast lead time, basin size, and forecasted rainfall accuracy, different predictors performing best in different conditions (Fig. 8). For details, read on!

1. WHY PREDICT FORECAST ACCURACY?

There is always an uncertainty associated with every weather forecast issued, even though it is rarely mentioned. Sometimes, one is fairly confident that the forecast will be accurate; sometimes, the forecast is a lot less certain. What is true for weather forecasts in general is also true for short-term extrapolation forecasts of precipitation in particular: not all forecasts have the same accuracy. This accuracy decreases with the forecast lead time and increases with the scale of the forecast region in a way that depends on the regional climatology, time of day, and precipitation dynamics.

Operators of flash flood warning systems need to take decisions based on the expected level of the river and the uncertainty in that expectation. Because of this, they do not need to know the expected or climatological uncertainty of radar-based rainfall forecasts that has been estimated over a long period of time as much as the uncertainty of the specific forecast that will be used as the basis for some decision, for example an evacuation order. Still, in general, short-term quantitative precipitation forecasts (QPF) made using radar data come only with a mean value but no uncertainty estimates, let alone an uncertainty estimate relevant to the current event and the basin of interest (Fig. 1).

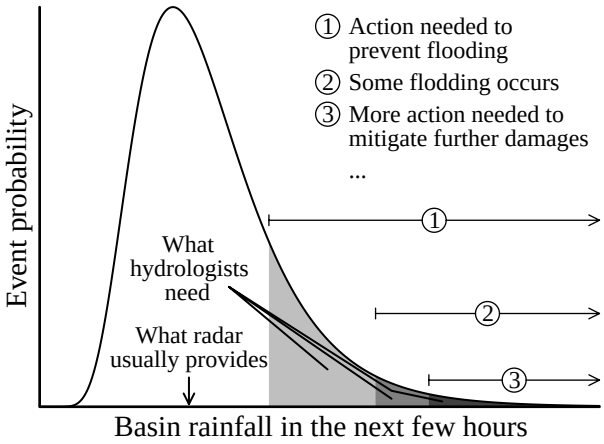

FIG. 1. Schematic illustration of the probability distribution of a rainfall forecast for a specific event, the information generally provided by radar-based QPF systems (the average expected rainfall), and the information usually needed by basin managers (the probability that an event exceeds a specific threshold, shaded areas).

Yet there has been work to address some of these issues. At longer timescales, operational methods for producing probabilistic QPF have been developed using guidance from ensembles of weather prediction models (Krzysztofowicz 1998). At shorter timescales, efforts have been made to characterize the predictability of reflectivity patterns as a function of scale (Turner et al. 2004), event type (Wilson et al. 1998), and geographical location (Germann et al. 2006), itself influenced by which precipitation dynamics processes affect the most each region. Forecasts

Corresponding author address: Frédéric Fabry, Dept. of Atmospheric and Oceanic Sciences, McGill University, Montreal, QC, H3A 2K6 Canada; frederic.fabry@mcgill.ca

of average rain rates over a basin (Berenguer et al. 2006) and short-term accumulations over points and basin locations (Ebert et al. 2004) have been evaluated for a variety of radar QPF algorithms and lead times. The accuracy of a precipitation forecast for a particular location based on a particular technique can change significantly in time (Turner et al. 2004, Berenguer et al. 2006), therefore there remains an unfulfilled need to determine as precisely as possible the probable accuracy of QPF using information available at the time the forecast is issued. A space-time model of forecast error that is conditioned on the current situation remains elusive at this time.

2. PREDICTING QPF PERFORMANCE

2.1 *QPF Uncertainty*

Let us consider a QPF issued for a basin B at time T and valid over a period between T_{start} and T_{end}. The expected standard deviation σ_{QPF} of that forecast can be written as

$$\underbrace{\sigma^2_{QPF}\left(B,T,T_{start}\rightarrow T_{end}\right)}_{\text{QPF variance}} = \underbrace{\sigma^2_{QPE}\left(B,T_{start}\rightarrow T_{end}\right)}_{\text{QPE variance}} + \underbrace{\sigma^2_{R(Z)}\left(B,T,T_{start}\rightarrow T_{end}\right)}_{\text{Radar forecast variance}} + \underbrace{2\sigma_{QPE}\sigma_{R(Z)}r_{QPE,QPF}}_{\text{Covariance term}}, \quad (1)$$

where σ_{QPE} is the expected standard deviation of a rainfall estimate using radar over that same basin for the forecast period, $\sigma_{R(Z)}$ is the expected standard deviation of the radar-based forecast of rainfall accumulation over the basin and the forecast period, and $r_{QPE,QPF}$ is the correlation between the errors in quantitative precipitation estimate (QPE) and in the nowcast. Even if one could somehow do a perfect forecast of the time sequence of reflectivity over a basin, there still remains the uncertainty associated with the conversion from reflectivity aloft to rainfall at the surface. This uncertainty will also depend on the type of precipitation, the duration of the rainfall accumulation period $T_{end}-T_{start}$, the basin size, and the range of the basin from the radar. The QPE uncertainty comes from a variety sources: biases for the vertical profile of reflectivity and their correction (Bellon et al. 2005), conversions from the reflectivity Z to the rain rate R (Lee and Zawadzki 2005), attenuation of radar waves (Hitschfeld and Bordan 1954), biases and imperfect corrections for beam blockage or radar calibration, sampling (Fabry et al. 1994), etc. Decades of radar-raingauge comparisons suggest that hourly radar and gauge accumulations differ by of the order of 50%, but it is not really well known how accurate QPE over a basin are because of a lack of proper ground truth since it is not possible to measure mean areal rainfall directly. All that being said, it is likely that the second term in the right hand side of (1), due to the radar QPF, will dominate the QPE error for most basins, forecast durations, and forecast lead times. The focus of this work will be on trying to determine the magnitude of that QPF term.

Several approaches can be used to determine the QPF term of the error equation (1). A simple one is to use radar data to generate forecasts using standard extrapolation approaches over a variety of basins, verify them using radar estimates of rainfall at the verification time, and compute the forecast error statistics over a number of years, e.g. Bellon and Austin (1978). However, it is clear that the QPF error term varies substantially around this climatological value so it is necessary to estimate or predict the error for the current forecast based on the recent past performance and other predictors.

2.2 *Predictors of QPF performance*

What makes a good predictor of QPF and of its accuracy? A good predictor has at least one of two characteristics (Fig. 2): it can be used to improve a QPF thereby reducing the uncertainty (case A of Fig. 2), or it can be used to better quantify that uncertainty (case B). As a result, it must either provide information on future precipitation intensity, or on how predictable or unpredictable a given weather situation is.

Classical extrapolation-based forecasts are a good example of the use of predictors. Using fields of current rainfall intensity and its motion as predictors, forecasts are being made that are clearly superior to climatology-based forecasts. In this work, we do not seek to replace these predictors as they are probably hard to supplant, but we want to try to find additional predictors that, if used in conjunction with traditional extrapolation-based forecasts, can either improve QPF or improve the estimates of its uncertainty.

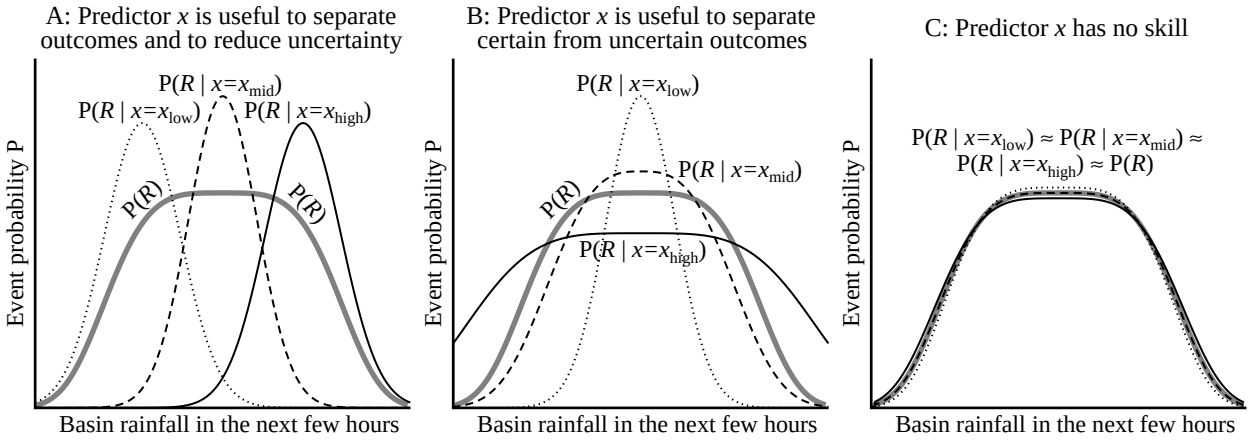

FIG. 2. Curves of the probability P of observing a rainfall accumulation R over a basin in the absence of a predictor (thick gray curves) and conditional to three values x_{low} (dotted line), x_{mid} (dashed line), and x_{high} (solid line) of a potential predictor x. An excellent predictor (case A, left) would reduce the uncertainty on the forecast; also useful would be a predictor (case B, middle) that could be used to better quantify the uncertainty in a forecast; a useless predictor (case C, right) would not provide any additional clue about either the forecast outcome or its uncertainty.

It is questionable whether one can find additional predictors that will significantly and reliably improve QPF accuracy. Several attempts have been made to improve extrapolation forecasts by trending precipitation intensities (Tsonis and Austin 1981) or by smoothing smaller unpredictable scales (Bellon and Zawadzki 1994, Seed 2003), but gains are limited because of the difficulty of the problem. But finding predictors of QPF accuracy might be easier. For example, some types of weather events such as precipitation from large scale systems are bound to be more predictable than disorganized air-mass convective storms. This paper evaluates several possible predictors of the accuracy of advection forecasts and draws conclusions on how the accuracy of the forecasts may be predicted in real-time.

3. APPROACH

3.1 *Data and Baseline QPF Calculations*

The data from two particularly rainy weeks at the end of a summer (19–31 August 2006) from the U.S. National 3-D Mosaic of radar data (Zhang et al. 2004; Zhang et al. 2006) were used in this analysis. We focused on four of the eight tiles covering the center of the country (Fig. 3), and that "limited" our evaluation region to approximately 5000000 km^2. Reflectivity data from the 3.5-km MSL reflectivity CAPPIs from tiles 2, 3, 6, and 7 were remapped onto a Gnomonic projection at 2.5-km resolution. Every half hour, maps of accumulations of precipitation for the past hour were generated using the 5-min data. These were made by tracking the motion of reflectivity patterns using the Bowler et al. (2004) tracking algorithm, converting the reflectivity Z into rainfall rates R using $Z = 300R^{1.5}$, and accumulating rainfall over hourly periods. Six forecasts of hourly rainfall accumulations were generated at 30-minute

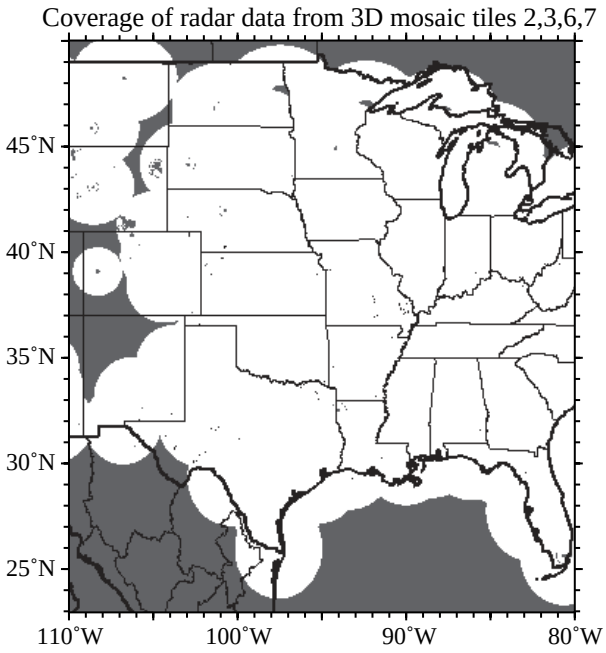

FIG. 3. Coverage of the radar data used in this study. Areas shaded in gray are beyond radar range and do not contain valid data.

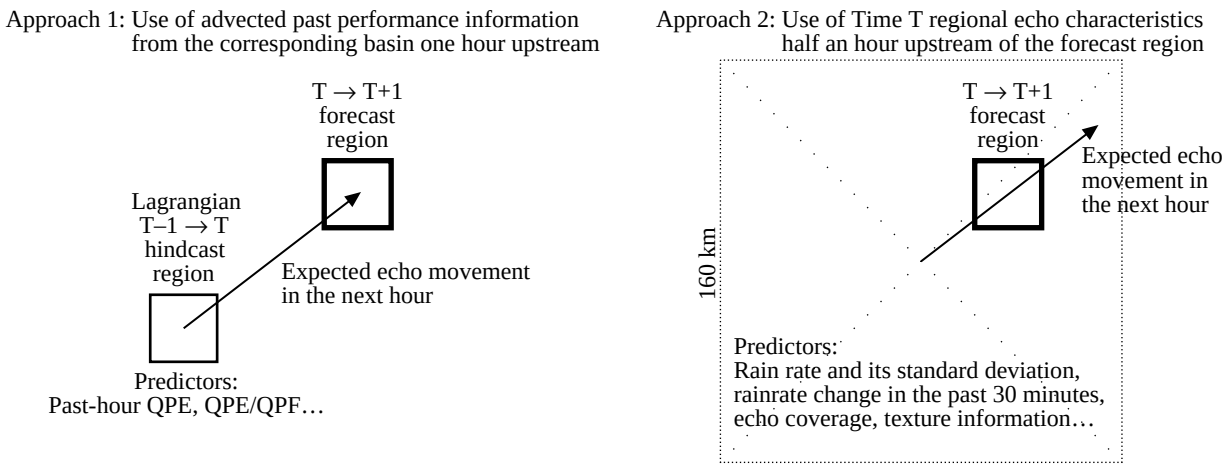

FIG. 4. The two conceptual approaches used to guide the choice of potential predictors as applied to forecasts for the next hour.

intervals, the first one valid from the current time *t* to *t*+1 hr, and the last one valid from *t*+5 hrs to *t*+6 hrs. The various forecasts and the computed accumulations valid for the same time period were compared, and statistics of QPF accuracy were generated as a function of forecast lead time and basin size, with the size of hypothetical square basins varying from 5 km by 5 km to 160 km by 160 km.

3.2 *Choice of Predictor Candidates*

A variety of predictors were tested in this study. They were selected for their potential to provide clues about forecast uncertainty using two distinct conceptual models as basis (Fig. 4). The first conceptual model is based on classical nowcasting and Lagrangian persistence approaches: If the previous forecast was accurate, then there is a good chance that the next forecast should also be accurate, and conversely. Fields of accuracy of past QPF of hourly accumulations are hence computed, advected following the echo motion, and tested as a predictor of future QPF accuracy. In the same vein, the cross-correlation of dBZ patterns between the present time and half an hour before was also tested, on the grounds that the magnitude of the correlation in time of the reflectivity field in the past should provide clues about the reliability of extrapolation forecasts for these echoes in the future.

The second conceptual model used to select predictors that are based on considerations of precipitation dynamics. Namely, rainfall patterns are likely to be predictable if they are forced by a single, slowly-evolving, larger scale process and not by multiple interacting, fast-evolving, small scale triggers. Clues that the former is the case would include a low mean rain rate, slow changes in time of that rain rate, and either large stratiform precipitation patterns or linear patterns arising from a single dominating forcing process such as a front. Conversely, disorganized patterns and small intense cells especially during daytime hours would arise from unsettled conditions that, by the very nature of their complexity, would be difficult to forecast. A variety of predictors possibly capable of detecting one or more of these characteristics were tested, and Table 1 lists them along with the rationale for their choice. All of them were computed on the latest radar image available. In order to get predictors relevant to the mesoscale weather condition at hand, all predictors were calculated over regions of 160-km by 160-km and not on smaller sub-regions. These predictors were then also advected with the precipitation field and tested for each basin and forecast lead time.

3.3 *Predictor Evaluation*

Individual predictors may have value both by improving the QPF and by helping to characterize its accuracy. In both cases, a predictor is useful if the distribution of QPF conditioned on one value of a predictor is different from the distribution of that QPF conditioned on another value of that predictor. As seen in Fig. 2, if conditional QPFs have large overlaps for different values of a predictor, then that predictor does not provide much information. Based on this, we can compute

the value V of a predictor x by testing the extent with which the curves of probability of rainfall amounts conditional on different values of the predictor do not overlap, e.g.,

$$V = 1 - \int P(x') \int P(x'') \int \min\left[P(R \mid x = x'), P(R \mid x = x'')\right] dR\, dx''\, dx' \quad (2)$$

where $P(x)$ is the probability distribution of the predictor x itself and $P(R \mid x = x')$ is the probability of observing a rainfall amount R given a value x' of a predictor x. Equation (2) computes for each possible pair of values of a predictor x the extent with which the probability distributions of rainfall overlap. In practice, $P(R \mid x = x')$ was calculated over intervals of R and x large enough for the probability distributions to converge (hundreds of data points per interval). Also note that the usefulness of a predictor may be a function of basin size, and will depend on the forecast lead time.

Possible predictors	Rationale
Previous QPF and its accuracy	Past performance may be a good indicator of future performance
Mean rain rate (overall and of rainy areas only) and their standard deviations	Light or heavy rain environment? Stratiform or convective?
Fractional coverage of precipitation (>15 dBZ) and of convective echoes (>35 dBZ)	Less predictable isolated patches or wide coverage? Stratiform or convective?
Lagrangian correlation between *dBZ*(*t*) and *dBZ*(*t* – ½ hr); 30-min mean rain rate trend	Is the weather evolving rapidly or not?
Mean autocorrelation of dBZ fields after a 20 km displacement	Large-scale or small-scale patterns dominate?
Energy of power spectrum of rain at 5 km scale; slope of power spectrum at scales larger than 40 km	Stratiform or convective?
Echo motion velocity and direction	A clue of the strength of large scale forcing?
Time of day	Stabilizing or destabilizing conditions?

TABLE. 1. Possible predictors of QPF performance tested.

4. BASELINE QPF ACCURACY

Three hundred sixty probability distributions of forecast accuracy were computed as a function of forecast time, forecasted rainfall intensity range, and basin size. For obvious reasons, only a few will be shown here. Figure 5 explores how forecast accuracy varies as a function of forecast time for a 10-km by 10-km basin and a moderate rainfall intensity range.

First, please note that all the distributions were plotted in terms of the ratio between the verification and the forecast (QPE/QPF), and not the other way around. This was done because, at forecast time, the only data that is available is the forecast. The information of interest to a potential real-time user is not "given a verification, what was the forecast", but rather "given a forecast, what can the likely outcomes be". Probabilities of QPE/QPF less than 1, on the left side of the curves, correspond to overforecasts or false alarms, while probabilities of QPE/QPF greater than 1 correspond to missed severe events.

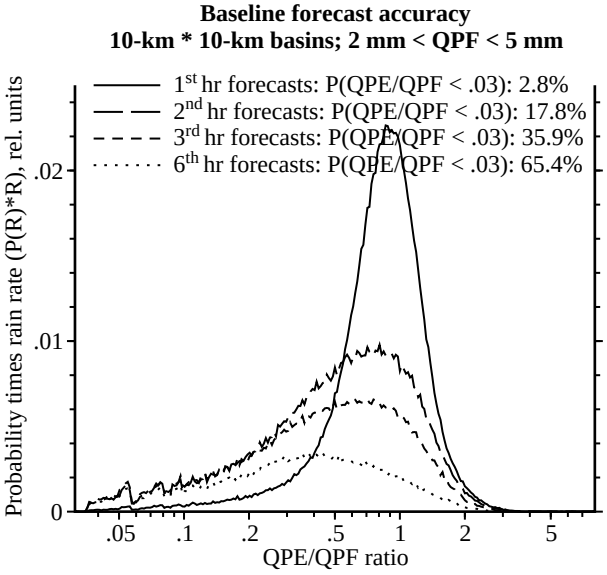

FIG. 5. Probability distributions of QPE/QPF ratio for 10-km by 10-km basins and basin hourly forecasts between 2 mm and 5 mm for the first hour (solid line), second hour (long dashes), third hour (short dashes) and sixth hour (dotted line). The y-axis was chosen so that a unit area anywhere in the plot corresponds to a fixed probability of occurrence. In addition, the probability that a forecast verified with almost no rainfall (verification smaller than 3% of the forecast) is indicated next to the legend.

As forecast time increases, forecast accuracy rapidly diminishes (Fig. 5). For 10-km by 10-km basins and moderate rainfall, one hour forecasts have a 40% uncertainty that increases to 100% by hour 2. A less obvious result is the gradual shift of the distributions towards lower values of QPE/QPF. As forecast uncertainty increases, the likelihood that a mistaken moderate or intense rainfall forecast verifies as an even rarer more intense rainfall accumulation is smaller than the likelihood that it will verify as a more common weaker rainfall. As a result, uncertain forecasts of moderate to heavy rainfall tend to overpredict the severity of the rainfall. In this case, by the sixth hour, almost two thirds of the forecasts of moderate rainfall will actually verify as negligible rain or no rain at all (verification smaller than 3% of the forecast), and only 3% of the forecasts underpredict rainfall intensity. This pattern of shifting distributions towards low QPE/QPF ratios as forecast uncertainty increases will reappear in many of the other results.

But not in the dependence of forecast accuracy with rainfall intensity (Fig. 6): Here, the poorer forecasts were for weaker rain accumulations, and they were mostly under-predictions. Distributions of QPE/QPF ratios were narrowest for the heavy rainfall forecasts, but almost all of those were overpredictions. The explanation for the biases is the same as in the previous paragraph: missed forecasts of rare events will verify as more common events, and this brings QPE/QPF ratios down for heavy rain predictions and up for very light rain ones. Very light rain forecasts were the most uncertain percentage-wise because they occurred on the edges of precipitation systems, and small errors on the forecasted motion of these systems resulted in large forecast accumulation errors.

As basin size increases, forecast accuracy improves. But for the smallest basins, forecast accuracy is identical irrespective of basin size, and QPE/QPF distributions for the 5-km by 5-km and 10-km by 10-km are so identical that the two curves can hardly be distinguished in Fig. 7. This is due to the fact that hourly accumulations and QPF errors have correlation distances exceeding the size of the basins. At the large basin end, accuracy improves rapidly as independent QPF errors are combined.

We are currently trying to fit curves with two or three parameters through the many probability

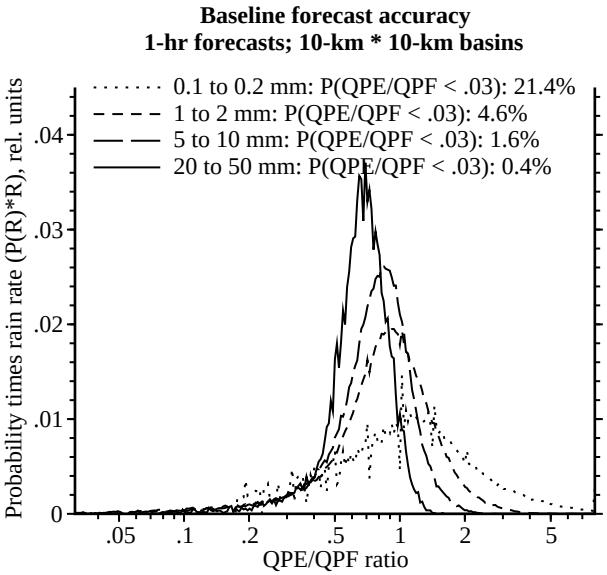

FIG. 6. Probability distributions of QPE/QPF ratio for 1-hr forecasts over 10-km by 10-km basins for very light rain (0.1 to 0.2 mm forecast for the basin, dotted line), light rain (1 to 2 mm forecasts, short dashes), moderate rain (5 to 10 mm, long dashes), and heavy rain (20 to 50 mm, solid line).

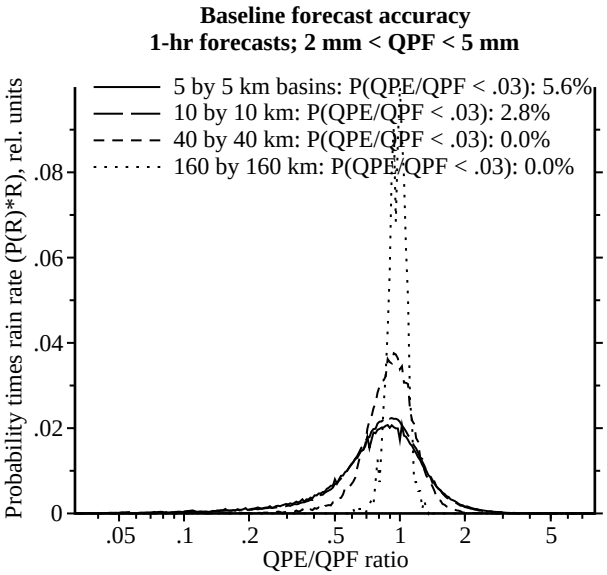

FIG. 7. Probability distributions of QPE/QPF ratio for 1-hr forecasts and basin hourly forecasts between 2 mm and 5 mm for basin sizes of 5 km by 5 km (solid line), 10 km by 10 km (long dashes), 40 km by 40 km (short dashes), and 160 km by 160 km (dotted line).

distribution curves we have computed. Stay tuned for the results of that exercise (look for a likely journal paper by Fabry and Seed 2008 or 2009).

5. PREDICTORS AND THEIR VALUE

The probability distributions in Section 4 were computed without using predictors. To test the usefulness of potential predictors, one needs to recompute them conditional to different values of all the potential predictors, and compute the value of predictors using (2). This was completed, and Fig. 8 summarizes our findings.

Unsurprisingly, no predictor greatly improved forecast quality. But several of them helped to separate somewhat forecast outcomes and more certain from less certain forecasts. Some like the verification for the basin one hour upstream (1-hr old QPE) proved to be better for the first hour, while others such as echo coverage or the 30-min change in area rain rate proved more useful for longer forecasts. We are still processing the data for larger basins and will soon determine the extent with which value depends on basin size and which predictor is stronger under which circumstances. The value of all predictors decreases with time beyond three hours, except for time of day because of its consistency and its large scale effect. Echo coverage (Fig. 9) and mean rain rate at the mesoscale seem to provide the most information, especially on forecast quality. In all cases, predictors help a bit to separate more certain forecasts from less certain ones.

What if one combines more than one predictor? The computation of the value using (2) is sensitive to noise in the probability distribution; if one uses more than one predictor, one must split the forecasts in many more categories, each with much fewer samples. Hence, the possibility of combining multiple predictors was not explored here. Different predictors have different strengths and their skill peaks at different forecast times, so they must have some complementary information. But additional gains will probably be smaller than those obtained using the first predictor, except perhaps if a potentially good predictor has strong geographical dependence such as the effect of time of day on precipitation, in which case important gains may be possible.

That's all for now…

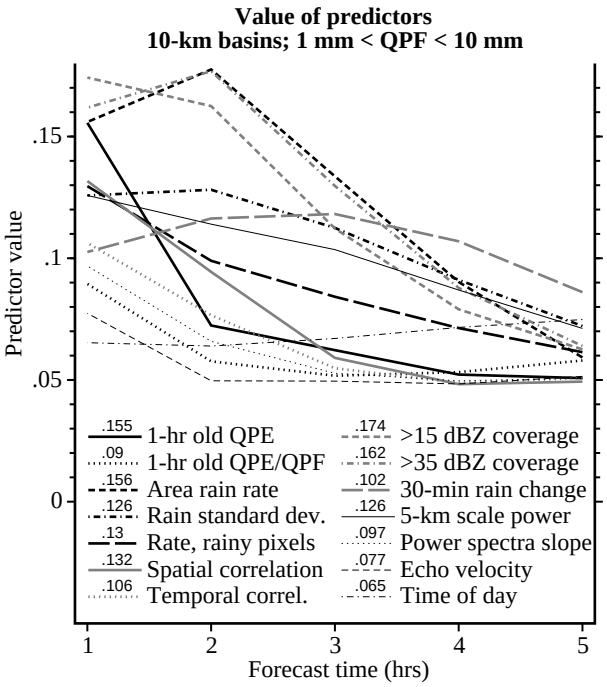

FIG. 8. Value of 14 potential predictors as computed from (2) as a function of forecast time for moderate rainfall for 10-km by 10-km basins. To help locate the many curves, the value computed for the 1-hr forecast is indicated on top of the legend for each curve.

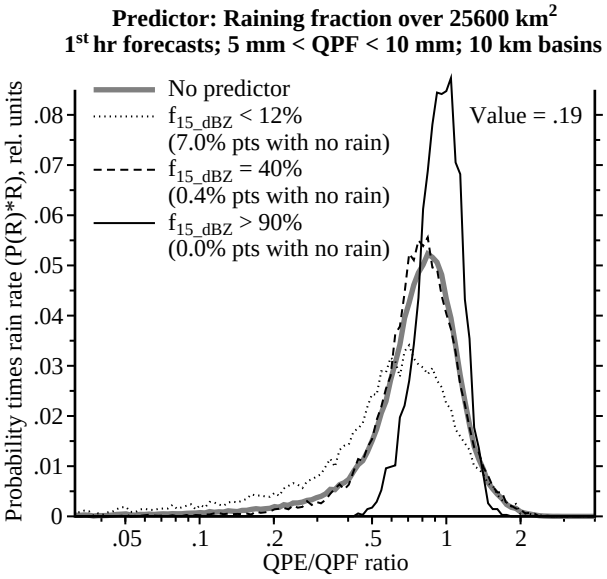

FIG. 9. Probability distributions of QPE/QPF ratio for first hour forecasts for 10-km by 10-km basins and forecasts between 5 mm and 10 without and with the use of the raining fraction as a predictor.

REFERENCES

Bellon, A., and G.L. Austin, 1978: The evaluation of two years of real-time operation of a short-term precipitation forecasting procedure (SHARP). *J. Applied Meteor.*, **17**, 1778–1787.

Bellon A., and I. Zawadzki, 1994: Forecasting of hourly accumulations of precipitation by optimal extrapolation of radar maps. *J. Hydrol.*, **157**, 211–233.

Bellon, A., G.W. Lee, and I. Zawadzki, 2005: Error statistics of VPR corrections in stratiform precipitation. *J. Applied Meteor.*, **44**, 998–1015.

Bowler, N.E., C.E. Pierce, and A. Seed, 2004. Development of a precipitation nowcasting algorithm based on optical flow techniques. *J. Hydrol.*, **288**, 74–91.

Ebert, E.E., L.J. Wilson, B.G. Brown, P. Nurmi, H.E. Brooks, J. Bally, and M. Jaeneke, 2004: Verification of nowcasts from the WWRP Sydney 2000 forecast demonstration project. *Wea. Forecasting*, **19**, 73–96.

Fabry, F., A. Bellon, M.R. Duncan, and G.L. Austin, 1994: High resolution rainfall measurement by radar for very small basins: The sampling problem reexamined. *J. Hydrology*, **161**, 415-428.

Germann, U., I. Zawadzki, and B. Turner, 2006: Predictability of precipitation from continental radar images. Part IV: Limits to prediction. *J. Atmos. Sci.*, **63**, 2092–2108.

Hitschfeld, W., and J. Bordan, 1954: Errors inherent in the radar measurement of rainfall at attenuating wavelengths. *J. Atmos. Sci.*, **11**, 58–67.

Krzysztofowicz, R., 1998: Probabilistic hydrometeorological forecasts: Toward a new era in operational forecasting. *Bull. Amer. Meteorol. Soc.*, **79**, 243–251.

Lee, G.W., and I. Zawadzki, 2005: Variability of drop size distributions: Time-scale dependence of the variability and its effects on rain estimation. *J. Applied Meteor.*, **44**, 241–255.

Pierce, C.E., E. Ebert, A.W. Seed, M. Sleigh, C.G. Collier, N.I. Fox, N. Donaldson, J.W. Wilson, R. Roberts, and C.K. Mueller, 2004: The nowcasting of precipitation during Sydney 2000: an appraisal of the QPF algorithms. *Wea. Forecasting*, **19**, 7–21.

Seed, A.W., 2003. A dynamic and spatial scaling approach to advection forecasting. *J. Applied Meteor.*, **42**, 381–388.

Tsonis, A.A., and G.L. Austin, 1981: An evaluation of extrapolation techniques for the short-term prediction of rain amounts. *Atmos.-Ocean*, 19, 54–65.

Turner, B.J., I. Zawadzki, and U. Germann, 2004: Predictability of precipitation from continental radar images. Part III: Operational nowcasting implementation (MAPLE). *J. Applied Meteor.*, **43**, 231–248.

Wilson, J.W., N.A. Crook, C.K. Mueller, J. Sun, and M. Dixon, 1998: Nowcasting thunderstorms: a status report. *Bull. Amer. Meteorol. Soc.*, **79**, 2079–2099.

Zhang, J., K. Howard, W. Xia, C. Langston, S. Wang, and Y. Qin, 2004: Three-dimensional high-resolution national radar mosaic. Preprints, 11th Conference on Aviation, Range, and Aerospace Meteorology. Amer. Meteor. Soc., 4–8 October 2004, Hyannis, MA. CD-ROM, paper 3.5.

Zhang, J., K. Howard, and S. Wang, 2006: Single radar Cartesian grid and adaptive radar mosaic system. Preprints, 12th Conference on Aviation, Range, and Aerospace Meteorology. Amer. Meteor. Soc., 29 January – 2 February 2006, Atlanta, GA. CD-ROM, paper 1.8