P2.38

THE PERFORMANCE OF WEATHER FORECASTS FOR VARIOUS FORECAST PROVIDERS

Patrick McCarthy
Prairie and Arctic Storm Prediction Centre
Meteorological Service of Canada

1. INTRODUCTION

Weather forecasts are not only provided by national weather services, but by many private weather forecast companies, as well. Much of this information is freely available to the general public. There are a wide number of potential forecast sources available to the public. Do the forecasts of national weather services still hold up well to the other providers?

This presentation examines the precipitation and temperature forecast performance of a number of major forecast providers for the city of Winnipeg, Manitoba, Canada. Their performance is weighed against the expectations of the public as indicated in public surveys.

The commercial providers chosen were picked from a larger group. The final group of providers assessed represented the better performers and encompassed a broad range of NWP input.

The temperature forecasts were assessed for the day-time high. The primary assessment used was Mean Absolute Error, since the end user normally compares the predicted values versus actual values to ascertain error.

The precipitation forecasts from each provider were categorized and these categories were assessed. The categorical forecasts were assessed against the observed precipitation.

Finally, this paper examines one approach that could significantly improve the utility of medium to long range temperature forecasts.

2. METHODOLOGY

Choosing the forecast providers

Over a dozen commercial providers were used at various times throughout the project. Many were dropped for one of three reasons:

1) the forecasts were routinely of poor quality
2) the forecasts were routinely unavailable
3) the forecasts were essentially identical to another provider

In the end, only 4 providers in addition to Environment Canada (EC) were used. Since the aim of this paper is not to critique any of the 4 unofficial forecast providers, their names will not be used in the document or references. The providers did represent companies from a number of countries who incorporated information from a number of major national weather service organizations into the forecast process.

The forecast approach for each provider typically contained some level of automation and human intervention. EC's forecasts typically had significant human intervention for days 1-2 while the forecasts beyond day-2 were routinely automated using model output statistics (MOS) and performance optimizing algorithms.

Some providers were fully automated relying solely on the information provided by the national weather service of their choice. Other providers further modified this information automatically with their own algorithms, etc., or at times with selective intervention by a meteorologist. Some providers could select the national weather service dataset "of the day", or even use one

Corresponding author address: Patrick McCarthy, Prairie and Arctic Storm prediction Centre, 123 Main Street – Suite 150, Winnipeg, Manitoba, Canada, R3C4W2; email: patrick.mccarthy@ec.gc.ca

dataset for part of their forecasts and another dataset for another part.

Initially, the project was simply an assessment of Environment Canada's official forecasts in comparison to readily available forecasts from commercial providers. Early in the project it was decided that this was a good opportunity to test the old forecasters' trick of the "poor man's ensemble" (PME). Normally, this ensemble approach is used for the averaging of independent operational numerical weather prediction (NWP) models. For this project, the PME, referred to in the project as "Pat's Ensemble", is the mean of the model-based forecasts, not the direct output from the models themselves. Climatology was used as a control.

Temperature verification

The project used a user-based approach. The forecasts had to be meaningful to the public. Public surveys (e.g. Decima Research Inc., 2002) indicated that the majority of the public used the morning forecasts as their primary decision-making information. Therefore, all temperature forecasts had to arrive by 0900 local time to be used in the project.

Environment Canada produces forecasts at 0500 local time. Most provider forecasts arrived in the following 4 hours. If no new forecasts were available at 0900 local time, the existing forecasts were used. Forecasts were recorded for all day-time highs provided. Some providers only had forecasts to day-5 while some went to day-15. For verification purposes, the day-time high was assumed to the highest temperature recorded between 0500 – 2400 local time. This approach was used for two reasons:

1) The same public surveys showed that the public's decision-making is made primarily in the morning and that they expect the "day-time" high will occur during the day and not overnight.
2) To ensure that the forecasts for day-1 and the following days were equally assessed, the 0500 local time forecast for day-1 requires all subsequent forecasts have the same period of time. This point was generally moot since almost all day-time highs occurred between 0500-2400 local time over the course of the experiment. However, when exceptions occurred, this rule was applied.

The original project assessed the daytime high temperature forecasts for five Canadian Prairie cities (Winnipeg, Regina, Saskatoon, Calgary, and Edmonton). The official forecasts from Environment Canada were used plus the forecasts from a small number of commercial providers whose information was readily available from the Internet.

After approximately 6 months, it was clear the performance of temperatures forecasts was essentially the same for all 5 cities. To reduce workload, four of the cities were dropped from the project, leaving only the forecasts for the city of Winnipeg being verified.

The measure of performance was also user-based. Typically, the public will assess a forecasts performance as the difference between the forecast and what was observed. Temperature bias tends to be secondary. The Mean Absolute Error (MAE) of the temperature forecasts most closely matches the public's assessment and this approach was used in this project. Other statistics can be derived from the dataset, but they will not be provided here.

Precipitation verification

Canadian public surveys (e.g. Decima Reseach Inc., 2002) indicate the day-time highs and precipitation forecasts are the most important forecast elements. Over a year into the project, precipitation forecasts were added to the assessment.

Assessing precipitation forecasts turned out to be rather challenging. While each provider readily provides a daytime high, the each have a different approach to communicating precipitation. Some only use icons, some use descriptions, some use a mix of icons and descriptions, while others may use probability of precipitation (POP).

The Canadian public attributes POP and descriptive precipitation terms to specific precipitation events (Environics, 1999). All forecast providers were contacted to understand what was meant by their icons, descriptions, etc. Sometimes, the commercial providers treated this information as “trade secrets”. Still, there was adequate information provided to help interpret their precipitation forecasts. With this information, a period of assessment was employed to gain familiarity with approaches the providers used to communicate their precipitation forecasts. From that assessment, a consistent process was defined. The precipitation approach breaks the forecasts down into 6 categories:

0 <30 % POP (no precipitation mentioned)
1 30 to <50% POP (e.g. *chance of showers/flurries*)
2 50 to <80% POP (e.g. *scattered showers, chance of rain*)
3 ≥ 80% POP (e.g. *showers, snow, periods of rain*)
4 ≥ 60% of warning criteria amounts *(*e.g*. “heavy” rain, specific amounts mentioned.*)
5 warning level accumulations (e.g. *“heavy” snow, specific amounts mentioned*)

The categorization of the observed precipitation followed the following guidelines:

0 no precipitation
1 brief precipitation with minimal amounts or observations reporting precipitation nearby
2 light precipitation reported for at least 2 hours
3 measurable (at least 1 mm/1 cm) precipitation and reported for 3 or more hours
4 ≥ 60% of warning criteria measured
5 warning level accumulations measured

The assessment was similar to the temperature approach. The period of 0500 to 2400 local time was used for all forecast periods. Icons and precipitation descriptions were changed to a POP and the highest POP category for the day was used as the category forecast.

For precipitation, Winnipeg typically experiences measurable precipitation for 30%-40% of the days annually. If one forecast no precipitation all year, you would be correct 56%-70% of the time. For dry years, your performance would be even better. During the 500+ days of this project, the 500+, the weather was actually wetter than normal, with some form of precipitation 52% of the time. Climatology would suggest a slight chance of precipitation (30%-40%) every day. Both climatology and “no precipitation” (category 0) were used for comparison.

Since the public uses these categories in their decision-making, the difference between the predicted category and the observed category was the measure of error. A linear relationship was used for this error. Again, other traditional measures could be utilized, but none are presented here.

Daily forecasts of maximum temperature and POP were collected for over 500 days of data were collected with each day having a 5-day forecast for each provider. Each forecast for days 1-5 were verified for both temperature and precipitation.

3. RESULTS

Temperature

Figures 1 and 2 summarize the temperature performance of EC’s forecasts, 4 internet forecast providers, PME, and climatology. The project assessed the performance only out to day-5 because all providers produced forecasts for this period. Some providers extended their forecast for varying lengths beyond day-5 and their performance was also captured and presented here. However, only the results for day-1 to day-5 should be considered to have complete and equal datasets.

Temperature forecasts are in Celsius and recent public surveys (e.g. Decima, 2002) note the the public can tolerate errors up to 4 degrees Celsius. The majority of the public considered forecast errors within 2 degrees Celsius were very good while errors greater than 4C were unacceptable.

For the day-1 (today) and day-2 (tomorrow) forecasts, the best performer was the human-produced official forecasts by EC. The commercial providers had a larger MAE of almost 1 degree Celsius. This should be expected as the forecasters' role is to bring to bear additional resources to enhance the automated output. This advantage disappears by day-3 and the EC performance is comparable to the others. By day-5 all providers have crossed into the "unacceptable MAEs of 4 degrees C.

Winnipeg is in the heart of the North American continent and it faces the typical extremes of a temperate continent air mass. Daytime highs can be -30C in the winter and over +35C in the summer. Wide ranges in daily maximum temperature mean the climatology tends to be a very poor predictor. The average MAE error for climatology in the non-summer months ranged from approximately 5C in autumn to approximately 7C in winter. The winter forecasts also suffered with unacceptable performance being reached in the day-3 to day-4 period, while spring and autumn climatology performance exceeded this threshold near day-5 and day-6, respectively.

During the summer, climatology proved to be a strong performer during this project with a MAE only near 3C. Climatology, therefore, would provide acceptable forecasts for the public in most situations, and it outperformed all providers beyond day-4. Other than the summer period, most providers outperformed climatology out to day-10.

It is interesting to note that the performance of all providers seems to level off around day-8, as if hitting some sort of "predictability wall".

Like EC, the PME was a very strong performer, outperforming everyone but EC for day-1 and day-2. Beyond day-2, the PME dominated, consistently outperforming the rest over the entire 10-day period. More importantly, the PME had "acceptable" forecasts out day 6-7 for most months, the longest of the group.

The PME also had more consistent predictions from forecast to forecast. The PME forecasts tended to trend to the correct solution over time, whereas the provider forecasts were often prone to more pronounced changes in their day-to-day forecasts, particularly beyond day-3.

Precipitation

Figure 3 and Table 1 show the distribution of each category forecast for each provider. For the majority of the forecasts, the lowest probability for precipitation was forecast. This matches climatology since, as indicated earlier, the majority of the days over the course of the year in Winnipeg have no precipitation.

Provider #2 was a notable exception. Its dominant forecast was the 30% to <50% category. If fact, this provider was the least prone to forecast a dry day.

Provider #5 rarely forecast category 2 and this was an artifact of their dissemination system. They used the "dominate weather of the day" approach and a lower chance of precipitation was rarely deemed dominant. The result was the high frequency of category 0 days and the highest frequency of category "4" (≥ 80% POP). However, their approach also meant that they were most prone forecast for the highest POPs (category 3) and were the closest to meeting the observed frequency in this category.

Provider #3 was most prone to forecast precipitation with a probability ≥ 50%.

Environment Canada's forecasts were the only one to have decreasing frequency towards higher categories. However, their category 3 forecasts were well under what was observed. Overall, EC's forecasts had a dry bias.

The PME offset these varying biases to also produce a decreasing frequency towards higher POPs. However, the result was an almost linear reduction rather than a significant low POP frequency as reality would dictate. Therefore, the PME tends to over-forecast POPs in the middle categories while under-forecasting in the extreme categories.

Cat.	Obs	EC	2	3	4	5	PME
0	52.2	67.5	30.0	56.5	64.1	74.9	39.6
1	14.6	11.7	56.1	14.9	17.5	0.3	32.0
2	9.1	10.9	9.8	20.6	7.4	8.2	18.1
3	20.9	9.1	3.7	7.9	10.8	14.9	9.1
4	1.9	0.6	0.3	0.0	0.3	1.3	1.1
5	1.0	0.1	0.2	0.0	0.0	0.5	0.2

Table 1. Percentage of precipitation forecasts for each category per provider versus what was observed (Obs.).

Overall, the individual providers were unlikely to forecast a near warning or warning event with all falling well below the observed frequency (Figure 7). Much of this has to do with the reduced amount of detail within the forecasts beyond Day-2, employed by most providers. Another factor is that warning level information may be added in the forecasts not issued in the morning. For Environment Canada, the warning information is often added outside the normal issue times for the public forecasts once the likelihood of the event has been assessed. Still provider “5”, with its “dominant weather” approach was most likely the one to predict this category. Surprisingly, the PME was the second most likely, suggesting that on many of these days the majority were forecasting high category POPs with at least one provider forecasting significant amounts.

It terms of performance, Figure 4 shows the categorical error for day-1 forecasts. Included in these charts is the performance of climatology and the performance of simply forecasting no precipitation (category “0”). Generally, there is a high level of accuracy simply because most days are dry. The figure shows that most providers over-forecast precipitation. The PME is the best provider overall.

In the medium range, Figure 5 shows the performance for Day-3 forecasts. The performance is weaker such that most providers do little better in performance that simply picking a dry day. Since climatology is usually Category 1 and provider 2 tends not to forecast Category 0, both tend to over-forecast for dry periods. In spite of that, the majority still tend to under-forecast precipitation. PME remains competitive though it no longer dominates as in the day-1 forecasts.

In the longer term, Figure 6 shows the day-5 forecast performance. The results are similar to the day-3 performance, with the majority over-forecasting precipitation and with the PME competitive with the group. However, overall skill is comparable to the controls (climatology and simply forecasting dry) other than on days with high POPs.

4. DISCUSSION

This project demonstrated that taking the average of temperature forecasts from various providers could produce better performing forecasts. This should not be unexpected. Over 30 years earlier, Sanders (1973) noted that “consensus” forecasts (mean of forecaster forecasts) showed on average more skill than any individual forecaster. Forecaster consensus also showed some improvement to temperature and precipitation forecasts out to day-4, though it was noted that this approach could hinder day-one quality (Bosart, 1975). Indeed, in this project the human forecasts were typically best performers for day-1. However, this advantage was relatively small.

With the growing capabilities of numerical weather prediction around that time, Leith (1974) proposed that a “Monte Carlo” (ensemble) mean was a practical and objective approach for NWP. It could produce better skill over the long term by reducing small scale variability while preserving easy to predict large scale features. Thompson (1977) mathematically demonstrated that combining independent numerical predictions shows better skill on average. He argued that this approach is very similar to the human “consensus” approach, but the NWP system would be objective.

Gyakum (1986) continued to find that consensus temperature and precipitation forecasts were “virtually unbeatable” by any individual forecaster.

Operational ensemble forecasts emerged in the 1990’s. More concerted efforts were made to explore the value of the “poor man’s ensemble” (multi-model), especially for the potential of a cost-effective (time/money) objective standard to measure the quality of EPS systems. In this project, the PME required few resources and was simply the mathematical average of provider output.

Atger (1999) examined a small member PME and demonstrated that this approach had “impressively high skill”. Skill over standard EPS systems was demonstrated beyond 5 days. This was consistent with this project for temperature as the small number of members was sufficient to produce a consistently better product on average. In fact, when additional members were used, there was no significant improvement in overall performance; a mathematical reality when averaging a large number of similar values. This advantage was more limited with precipitation.

Ziehmann (2000) explored the performance of a 4 member PME against the ECMWF EPS. She found that while an individual EPS member could have the most accurate solution, the PME routinely provided the better and most consistent solutions. Similar results were found by Fritsch et al (2000), and Chessa and Lalaurette (2001). Often in this project, individual providers were the best performing. Still the averaging process ensured that the PME was always better than at least one provider.

Ebert (2001a, 2001b) found that a 7-member PME out-performed the ECMWF’s EPS for precipitation for days 1-2 and speculated that this skill would extend beyond day-2. The results in this project demonstrated that for some elements, this was likely, as the PME’s temperature forecast showed skill beyond 5 days. However, the PME for precipitation lost its advantage by day-3.

Weighting of ensemble members improves the ensemble mean (van den Dool et al, 1994, Woodcock and Engel, 2005), though the weighting approach can be challenging (Chien et al, 2004). Eckel and Mass (2005) demonstrated the high skilled members add the most value to the ensemble while poorer performers need to perform well often to be of value. In this project, occasional attempts were made to remove an apparent weak member to improve the ensemble. That approach was often unsuccessful since sometimes the majority of members turned out to be in error, particularly for longer range forecasts.

But what about extreme weather? It is often suggested (e.g. Young, 2002) that ensemble means are less sharp and thus are not sensitive to extremes. Mathematically with is true, except, of course, when all members are forecasting an extreme event. Still, this project found that for precipitation, the mean was more likely to forecast a significant precipitation more often than the majority of providers (Figure 7). Figure 8 shows the categorical error for category 4 (near warning) and category 5 (warning level) precipitation events. In general, EC and provider 5 forecast these categories the most. This was likely due to EC’s extensive human intervention is days 1-2 and provider 5’s penchant to forecast extremes. The remaining provider forecasts, including the PME, were weaker in this area and performed very similarly.

Mylne and Robertson (2002) and Arribas et al (2005), however, did find that a PME had its greatest for probabilistic skill for extreme events at T+24. For this project, Figure 9 shows the performance of each provider for categories 4 and 5 for each forecast period. The PME was the second best performer for day-1 but it did not exhibit any particularly advantage beyond that.

Woodcock and Engel (2005) found that a well designed weighting system (bias correction and optimal weighting algorithms) could overcome some of this insensitivity to extremes. They also noted that the ensemble mean is not to be used in isolation by forecasters. Large ensembles are more likely to have members capturing extreme events than the much smaller PME. However, these are likely of lower probability. They are important for forecasters to keep in mind and as useful information for certain decision-makers.

Beyond the mathematical and statistical nuances of a PME, ultimately the quality of the output comes down to the value perceived by the end user. Environment Canada forecasts are the most widely received forecasts by the Canadian public. Other forecasts by private providers, including some of the providers in this project, are also widely used. The different predictions do cause confusion occasionally. For many people, all forecast products, regardless of the provider, are assumed to be official Environment Canada forecasts. Depending on who they listen to, the forecasts can appear very different. Environment Canada receives complaints from the public about forecast quality and often the source of the problem relates to provider confusion.

Another source of criticism is significant changes to forecasts, particularly in the longer term. The user makes plans or expectations based upon one forecast only to have them dashed by a new forecast. Sometimes the next forecast flips back to the original prediction, giving the impression that the forecasters do not seem to understand what is going on.

This “flip-flopping” of forecasts is a common characteristic of automated medium to long range forecasts. This is because the numerical weather models produce forecasts true to the data assimilated and by the physics encoded within the system. If the new data produces different results, the model will dutifully output a new and different forecast. Statistical modification of the model output can also introduce new changes. A human forecaster often considers the impact of flip-flopping, and will try to temper the changes. However, with automated forecasts common by commercial providers and beyond day-2 by Environment Canada, flipping of forecasts is common.

The PME is much less prone to flipping since it will offset the various opposing forecasts. Therefore in the longer term, the PME is less extreme and it will trend to the correct solution over time, as the various members of the PME form a consensus. The result is a less volatile product that forms a “reasonable” solution rather quickly, given the public’s willingness to allow some measure of error. Amoung staff at the PASPC, this project’s PME continues to be produced internally and it is the preferred forecast beyond day-2 for temperature.

5. SUMMARY

This project examined the performance of a number of providers against the performance of the official forecast for one major city in Canada. In general, the official forecast was the top performer, though its advantage was confined to the first two days when the weather service’s meteorologists could add value. Since the majority of the public uses forecast information for decisions in this period, the value of human intervention is clear. The best performer for temperature overall was the PME approach. It produced good forecasts, as deemed by the Canadian public well beyond the individual providers

For the most part, all providers provide quality precipitation forecasts in the short term. Because of the averaging process, the PME tended to over forecast precipitation for Winnipeg, which climatology had a majority of days with no precipitation. Still, the PME performed well for extreme precipitation events for the 1st 2 days.

The PME also has value but maintaining a level of credibility with the public. Which the forecasts may not be the most accurate, they tend to be the “least wrong”, consistently. The PME significantly reduces the magnitude of flip-flopping and, instead, trends to the correct solution over time.

The value of using different models in an ensemble is also evident. Each brings its own bias to the ensemble and the PME helps smooth those out. The PME is a good starting point for the public and it is a good starting point for forecasters. Their analysis and diagnosis should help them understand the likelihood of one solution over another, and whether they have enough confidence to make an appropriate adjustment from the PME. However, it is important that forecast realize that the PME is purely a mathematical process and contains no meteorology.

Finally, the PME is a simple and cost-effective way to improve existing forecasts beyond day-2. Trying to improve forecasts by continually creating bigger and faster ensembles may not be cost-effective. Refining a smaller set of multi-modal ensemble members to perform better, utilizing the PME to improve overall performance, and giving forecasters the training and tools to excel in the short-term forecasts would likely be a far more cost-effective approach.

6. REFERENCES

Arribas, A., K.B. Robertson, and K.R. Mylne, 2005: Test of a Poor Man's Ensemble Prediction System for Short-Range Probability Forecasting. *Mon. Wea. Rev.*, **133**, 1825–1839.

Atger, F., 1999: The Skill of Ensemble Prediction Systems. *Mon. Wea. Rev.*, **127**, 1941–1953.

Bosart, L.F., 1975: SUNYA Experimental Results in Forecasting Daily Temperature and Precipitation. *Mon. Wea. Rev.*, **103**, 1013–1020.

Chessa. P.A., F. Lalaurette, 2001: Verification of the ECMWF Ensemble Prediction System forecasts: A study of large-scale patterns. *Wea. Forecasting*, 1**6**, 611–619.

Chien, F.C., and B.J.D. Jou, 2004: MM5 Ensemble Mean Precipitation Forecasts in the Taiwan Area for Three Early Summer Convective (Mei-Yu) Seasons. *Wea. Forecasting*, **19**, 735–750.

Decima Research, 2002: National Survey on Meteorological Products and Services – 2002. Final Report. Prepared for Environment Canada.

Ebert, Elizabeth E., 2001. Ability of a poor man's ensemble to predict the probability and distribution of precipitation. *Mon. Wea. Rev*., Amer. Met. Soc., Vol. 129, 10, 2461-2480.

___________, 2002. Corrigendum (Ability of a poor man's ensemble to predict the probability and distribution of precipitation). *Mon. Wea. Rev*., Amer. Met. Soc., Vol. 129, 10, 1661-1663.

Environics Research Group, 1999. *Evaluation of Precipitation*. Prepared for Environment Canada by Environics Research Group Limited, Toronto, Canada.

Fritsch, J.M., J. Hilliker, J. Ross, and R.L. Vislocky, 2000: Model Consensus. *Wea. Forecasting*, **15**, 571–582.

Gyakum, J.R., 1986: Experiments in Temperature and Precipitation Forecasting for Illinois. *Wea. Forecasting*, **1**, 77–88.

Hansen, B. U., 2007: Internet Weather Forecast Accuracy. *Omninerd*. Available online at: http://www.omninerd.com/2007/02/08/articles/69 (accessed on Feb. 12, 2007)

Leith, C., 1974: Theoretical Skill of Monte Carlo Forecasts. *Mon. Wea. Rev.*, **102**, 409–418.

Mylne, K.R., K.B. Roberston, 2002: Assessment of a Multi-centre "poor man's" ensemble prediction system for shart-range use. Symposium on observations, data assimilation, and probabilistic prediction, Preprints,*16th Conf. on Probability and Statistics in the Atmospheric Sciences*, J1.5.

Rousseau D. and P. Chapelet, 1985: A test of the Monte-Carlo method using the WMO/CAS intercomparison Project Data. Report of the second session of the CAS working group on short-and-medium-range weather prediction research, Belgrade, 26-30 August 1985. PSMP Report Series n° 18, WMO/TD n° 91, 53-58.

Sanders, F., 1973: Skill In Forecasting Daily Temperature and Precipitation: Some Experimental Results. *Bull. Amer. Meteor. Soc.*, **54**, 1171–1178.

Van Den Dool, H.M., and Z. Toth, 1991: Why Do Forecasts for "Near Normal" Often Fail? *Wea. Forecasting*, **6**, 76–85.

Van Den Dool, H., and L. Rukhovets, 1994: On the Weights for an Ensemble-Averaged 6–10-Day Forecast. *Wea. Forecasting*, **9**, 457–465.

Woodcock, F., and C. Engel, 2005: Operational Consensus Forecasts. *Wea. Forecasting*, **20**, 101–111.

Young, G., 2002: Combining forecasts for superior prediction. Preprints, *16th Conf. on Probability and Statistics in the Atmospheric Sciences*, 107-111.

Ziehmann, C. (2000): Comparison of a single model EPS with a multi-model ensemble consisting of a few operational models *Tellus* **52A**, 280-299*.*

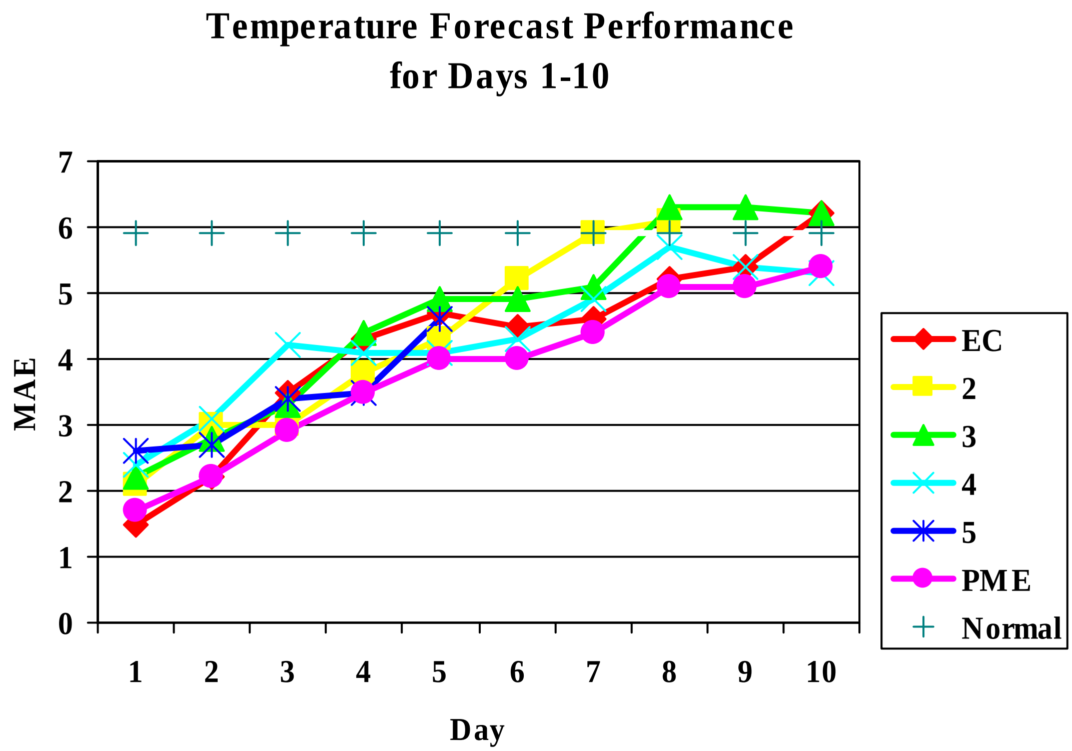

Figure 1. Temperature forecast performance for Environment Canada, 4 commercial providers, the PME, and climatology.

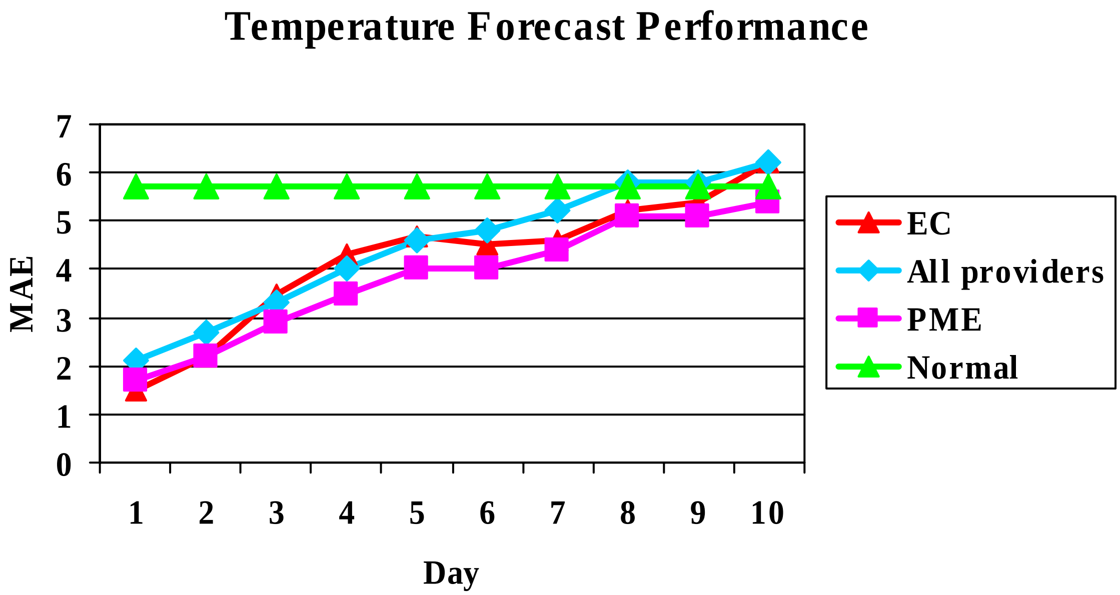

Figure 2. Temperature forecast performance for EC, the average of the 4 commercial providers, PME, and climatology.

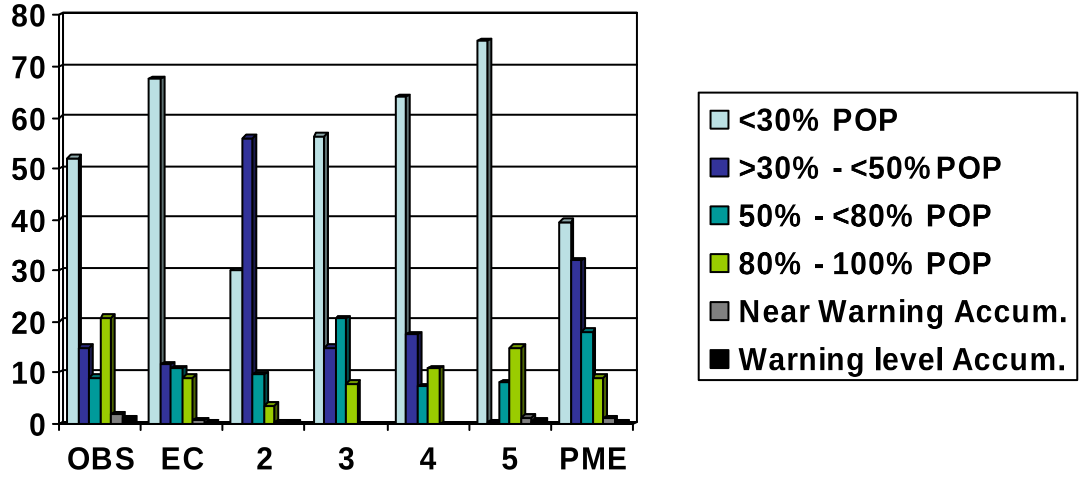

Figure 3. Categorical precipitation forecasts for all providers plus the observed frequency of each category.

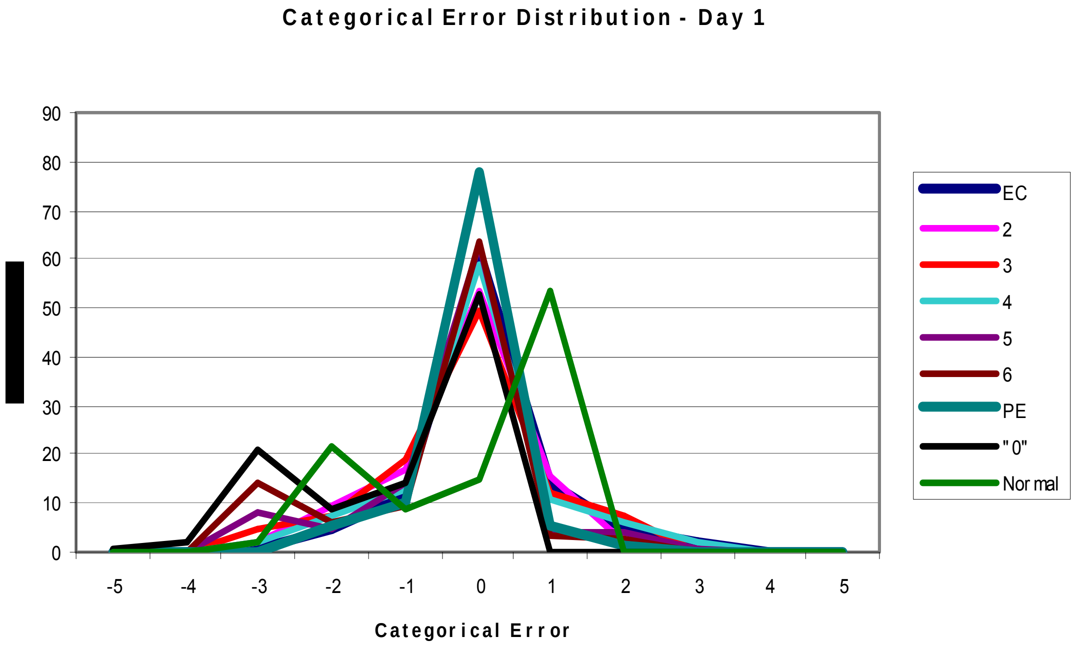

Figure 4. Categorical errors for precipitation for each provider, climatology and forecasting category 0 for day-1.

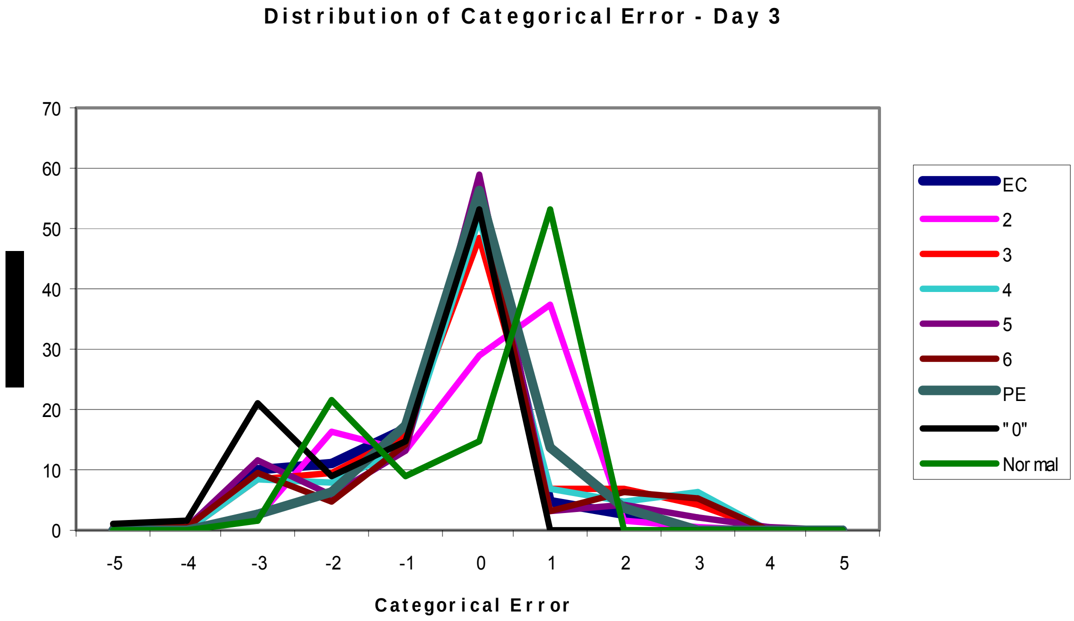

Figure 5. Categorical errors for precipitation for each provider, climatology and forecasting category 0 for day-3.

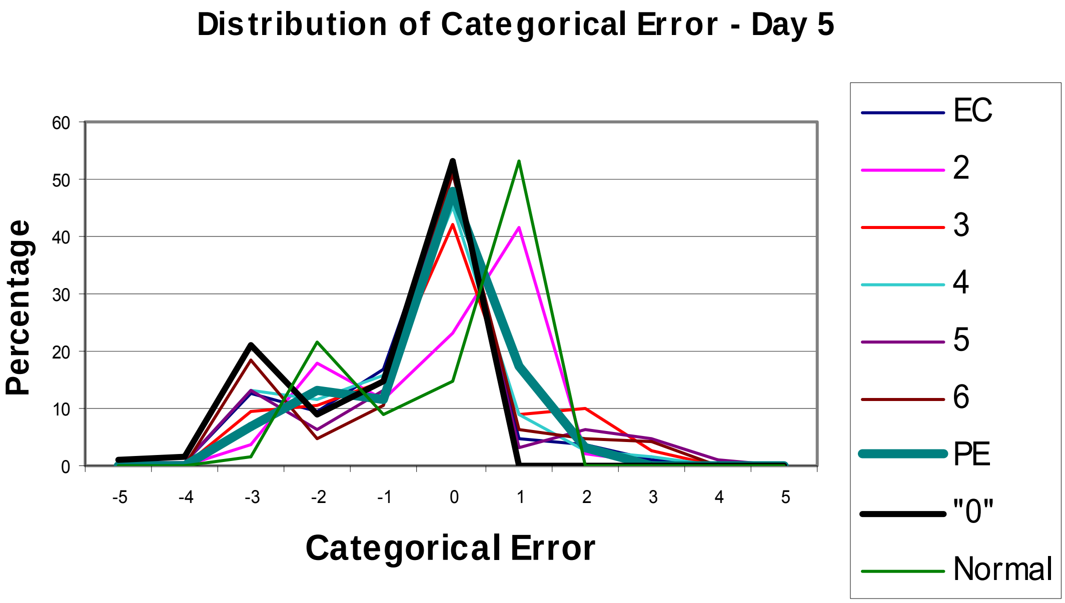

Figure 6. Categorical errors for precipitation for each provider, climatology and forecasting category 0 for day-5.

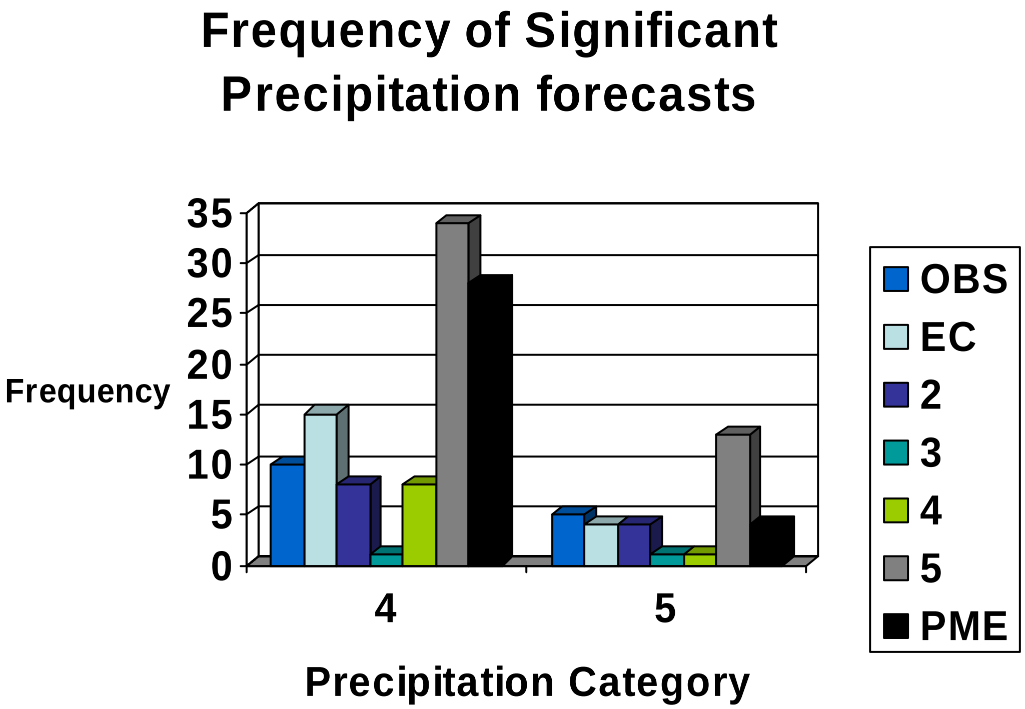

Figure 7. Frequency of forecasts for categories 4 and 5 by each provider, the PME and what was observed.

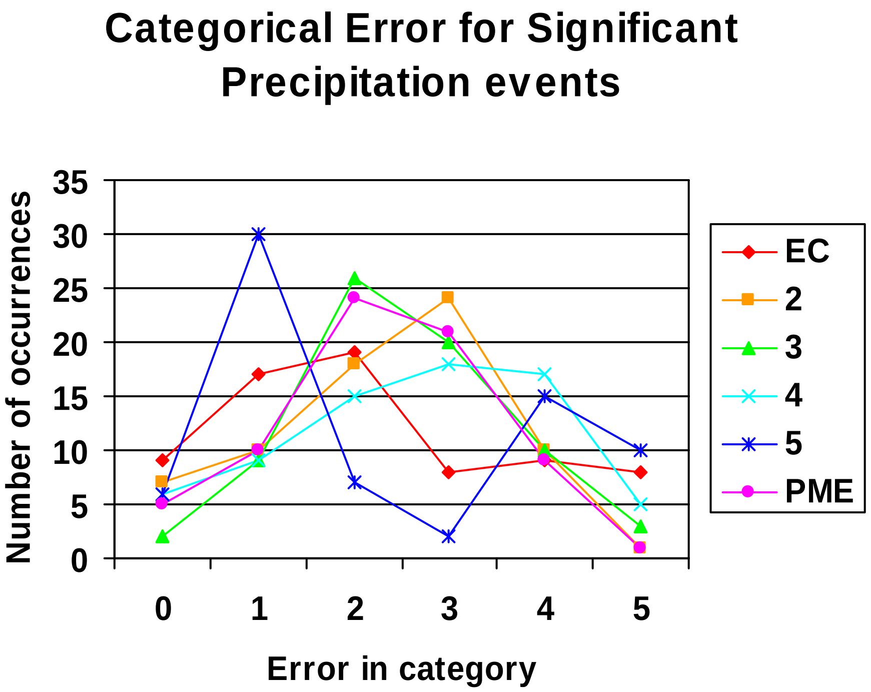

Figure 4. Categorical error for each provider when a category 4 or 5 event is observed.

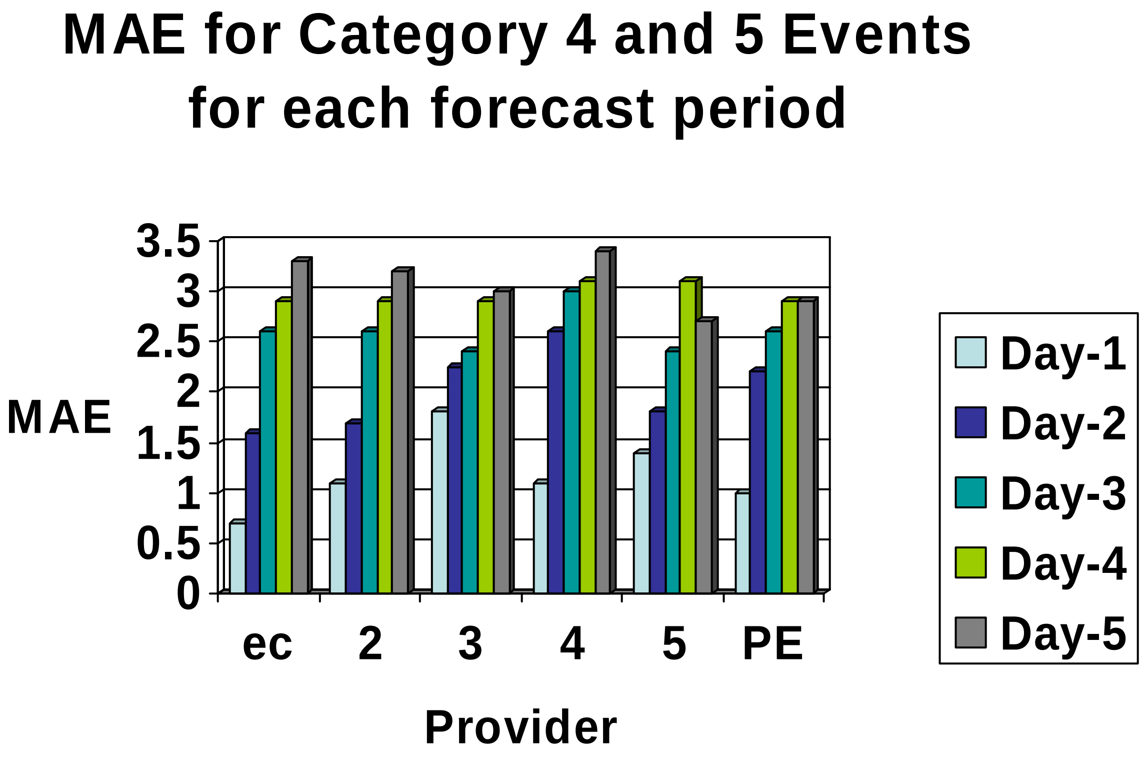

Figure 9. Combined mean absolute error (per category) for category 4 and 5 forecasts for each provider for days 1-2.