Precipitation forecast skill of numerical weather prediction models and radar nowcasts

Charles Lin,[1] Slavko Vasić,[1] Alamelu Kilambi,[1,2] Barry Turner,[1,2] and Isztar Zawadzki[1,2]

Received 10 May 2005; accepted 10 June 2005; published 16 July 2005.

[1] Short term precipitation forecasts based on Lagrangian advection of radar echoes are robust and have more skill than numerical weather prediction models over time scales of several hours. This is because the models do not generally capture well the initial precipitation distribution. We will refer to the advection-based methods as radar nowcasts. Over longer time scales, we expect the models to perform better than nowcast methods as they resolve dynamically the large scale flow. We verify this conceptual picture of the relative accuracy of radar nowcasts and model forecasts using conventional skill scores. We identify the cross-over point in time where model forecasts start to have more skill than nowcast methods. This occurs at about 6 hours after the forecast is initiated. **Citation:** Lin, C., S. Vasić, A. Kilambi, B. Turner, and I. Zawadzki (2005), Precipitation forecast skill of numerical weather prediction models and radar nowcasts, *Geophys. Res. Lett.*, *32*, L14801, doi:10.1029/2005GL023451.

1. Introduction

[2] For very short term prediction of precipitation (0–3 hours), Lagrangian advection of radar echoes performs best while for longer periods, forecasts based on numerical models may be better. This concept is shown schematically in Figure 1, adapted from *Golding* [1998], which depicts qualitatively the loss of forecast skill as a function of forecast lead time. Figure 1 is in turn adapted from an earlier publication by *Austin et al.* [1987]. A similar figure also appears in *Wilson et al.* [1998]. We examine this conceptual picture quantitatively in our study. For the purpose of this paper, we will refer to precipitation forecast by Lagrangian advection of radar echoes as radar nowcasts.

[3] Precipitation predictability also depends on spatial scale. *Germann and Zawadzki* [2002] examined the lifetime of precipitation patterns derived from US continental scale radar images from the storm to synoptic scales. They developed a nowcast methodology that combines variational echo-tracking with semi-Lagrangian advection that allows for large scale rotational motion, which can be important for the synoptic scales. We use their methodology to compare the skill of radar nowcasts with model forecasts of precipitation over the continental US. Four numerical weather prediction models are used: two versions of the Global Environmental Multiscale model (GEM) from the Meteorological Service of Canada [*Côté et al.*, 1998], the ETA model from the US National Weather Service [*Janjić*, 1994] and the WRF model from the National Center for Atmospheric Research (NCAR), [*Michalakes et al.*, 1998]. We focus on the quantitative identification of the cross over point in forecast lead time where the model would perform better than the nowcast, as depicted schematically in Figure 1.

2. Methodology

[4] We compare the skill of radar nowcasts with precipitation forecasts of 1-hour accumulated precipitation from four models: GEM (two versions), ETA and WRF. The two versions of GEM refer to the operational version that is currently run at a resolution of 15 km, and an earlier version (GEM/HIMAP) that is run over a smaller domain at 10 km resolution. Table 1 presents a summary of the analysis methodology. We report on the results of two separate studies done with a slightly different methodology. The first examines the GEM/HIMAP and ETA models in a 2,160 km $\times$ 2,160 km domain over the central and eastern continental US, which is the domain studied by *Germann and Zawadzki* [2002]. A total of 21 days is examined, covering the summer and fall of 2003 (September 12, 13, 18, 27; October 14, 17, 25, 26, 28) and 2004 (May 21, 22, 30, 31; June 11, 24; July 3, 4, 5, 6; August 19, 20). These dates were chosen due to the availability of ETA model output. The first study period is shown as the second and third columns of Table 1. The second period examined has 35 days of precipitation during the period August 13 to September 25, 2004, and covers the continental US east of 102°W longitude; this was done as radar coverage is poor west of the Rocky Mountain Front Range. This is shown as the fourth and fifth columns of Table 1. The radar precipitation data for both studies are taken from US radar composites provided by Weather Decision Technologies, Inc. (WDT), with 10-minute time resolution and 5-km spatial resolution. For the first study period, we use the 12-km resolution of the ETA model for comparison of radar nowcasts and model forecasts. For the second study period, model results are projected onto the radar grid for analysis. Although the methodology is slightly different, the results on the comparison of the skill of the radar nowcasts and model forecasts for both study periods are essentially the same, as we will see later.

[5] The period of comparison is over the duration of each radar nowcast. More specifically, for every 24-hour model forecast of the first study period, five radar nowcasts each of duration 9 hours are initialized at times 3, 6, 9, 12, 15 hour. Thus the first radar nowcast covers the period from hours 3 to 12, and the last nowcast from hours 15 to 24. The skill of the five radar nowcasts is compared with the model forecast

[1] Department of Atmospheric and Oceanic Sciences, McGill University, Montreal, Quebec, Canada.
[2] Marshall Radar Observatory, McGill University, Montreal, Quebec, Canada.

0094-8276/05/2005GL023451$05.00

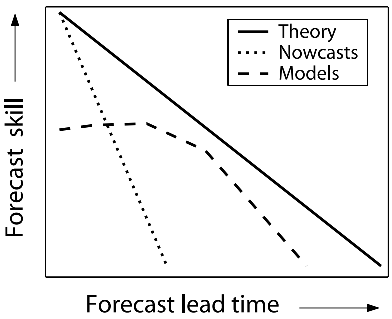

Figure 1. Schematic representation of the loss of forecast skill as a function of forecast lead time. The solid line represents the theoretical limit of predictability. The dashed and dotted lines correspond respectively to numerical weather prediction models and nowcasting methods (from *Golding* [1998], copyright 1998, Royal Meteorological Society).

over the corresponding 9-hour duration of the nowcast. For the second study period, two nowcasts each of duration 12 hours are performed and evaluated. A threshold of 10 dBZ in units of logarithmic radar reflectivity is used, with rain rate (mm/hour) inferred from reflectivity as $(Z/300)^{2/3}$. Six 10-minute rain rates are used to obtain the hourly accumulation.

[6] The skill measures probability of detection (POD), false alarm rate (FAR) and critical success index (CSI) are used for the first study period. They are categorical scores based on a contingency table applied at each analysis grid point over the verification period [*Johnson and Olsen*, 1998]. The radar-retrieved precipitation is taken as "truth", with different thresholds of 0.1, 0.5 and 1.0 mm for hourly accumulation. For a perfect forecast, POD = 1, FAR = 0 and CSI = 1. Following *Germann and Zawadzki* [2002], we also calculate the conditional mean absolute error (CMAE; based on the logarithm of the reflectivity) for the second study period. This score measures the domain average absolute error of the forecast at a particular time. For a perfect forecast, CMAE = 0.

3. Results

[7] For the first study period (summer and fall of 2003–04), we evaluate the skill scores POD, FAR and CSI for five daily 9-hour radar nowcasts initiated at different times, and compare them with the scores obtained from the corresponding 9-hour periods of the 24-hour model forecasts (GEM/HIMAP, ETA). There is a total of 21 days of 24-hour model forecasts available for study. Figure 2 shows the averaged skill scores over these cases based on a 0.1 mm precipitation threshold, together with the ±1 standard deviation curves for the radar nowcast and GEM/HIMAP forecast. The radar nowcast starts with high initial skill (POD > 80%, CSI > 60%, FAR < 25%), and the skill decreases with forecast lead time. This is because the initial rain field is well captured by the nowcast, but the skill decreases relatively rapidly as development and decay are not taken into account in an advection-based nowcast. The models start with lower skill than the nowcasts, as the initial precipitation field is not captured as well. However, the model skill remains almost constant over the 9-hour period (POD ~ 50%, CSI ~ 30%, FAR ~ 65%). This results in a crossing of the skill curves of the radar nowcast and the model forecast for all three scores. The cross over point occurs at about 5–6 hours for POD, and about 7–8 hours for FAR and CSI. Thus the radar nowcast has more skill than the models for forecast lead times shorter than about 6 hours, with the models becoming more accurate after this period. This is in agreement with the conceptual picture of *Golding* [1998] presented in Figure 1. The ±1 standard deviation curves give an idea of the spread of the skill in the cases, and the cross over point inferred from the +1 and −1 standard deviation curves yield similar results as for the average. A difference between the GEM/HIMAP and ETA models is the latter performs better with a higher POD score, resulting in a cross over point which is about 1.5 hours earlier. We attribute this to the generally sharper precipitation features of GEM/HIMAP precipitation, resulting in a higher possibility of error. This difference becomes insignificant for FAR and CSI. All the above results are robust with respect to the precipitation threshold: the conclusions regarding the cross over point remain unchanged when higher thresholds are used (0.5 and 1.0 mm; figures not shown). Further analysis was also performed where the radar and model results were smoothed by filtering out scales smaller than 4 or 8 times (4Δ, 8Δ) the grid resolution (Δ = 12 km), as the small scales have less predictability. The skill scores are uniformly improved upon smoothing for both radar nowcasts and model forecasts, but the cross over point in time between the two remain largely unchanged (figures not shown).

Table 1. Summary of Analysis Methodology[a]

Model	GEM/HIMAP	ETA	GEM (Operational)	WRF
Agency	MSC, EC	NWS, US	MSC, EC	NCAR, US
Reference	*Coté et al.* [1998]	*Janjić* [1994]	*Coté et al.* [1998]	*Michalakes et al.* [1998]
Domain	Continental US $2{,}160 \times 2{,}160$ km^2	Continental US $2{,}160 \times 2{,}160$ km^2	Continental US east of 102°W	Continental US east of 102°W
Resolution	10 km	12 km	15 km	28 km
Period	Summer, fall of 2003–04, 21 days	Summer, fall of 2003–04, 21days	Summer of 2004, 35 days	Summer of 2004, 35 days
Skill measures	POD, FAR, CSI	POD, FAR, CSI	POD, FAR, CSI, CMAE	POD, FAR, CSI, CMAE

[a]The models used are GEM (two versions), ETA and WRF. The analysis domains, model resolution of forecasts, period of analysis and skill measures used are shown. The first study period uses GEM/HIMAP and ETA models (second and third columns), while the second study period uses GEM (operational) and WRF models (fourth and fifth columns). See text for further details. (MSC: Meteorological Service of Canada; EC: Environment Canada; NWS: National Weather Service; NCAR: National Center for Atmospheric Research).

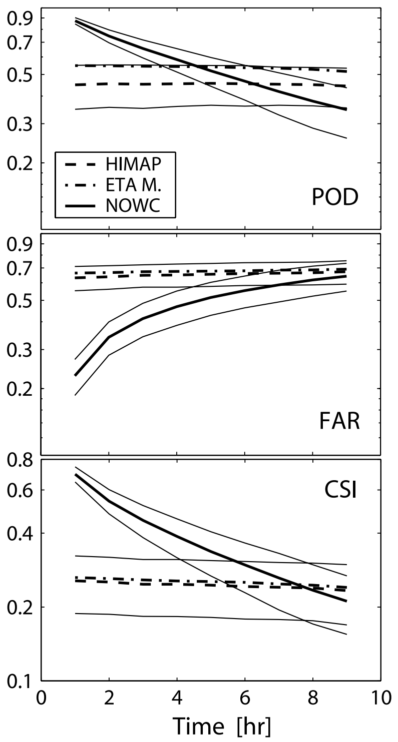

Figure 2. Skill scores (POD, FAR, CSI) for 9-hour radar nowcasts (solid) and model precipitation forecasts (dashed for GEM/HIMAP, dash-dot for ETA), averaged for all cases over the 21 days of the first analysis period. The threshold for hourly precipitation is 0.1 mm. The thin solid lines are the ±1 standard deviation for the radar nowcasts and GEM/HIMAP model, taken over all cases.

[8] Similar results are obtained for the second study period (summer and fall of 2004) with two different models (GEM operational, WRF) over the continental US east of 102°W (Figure 3). Details of the WRF can be found in the work of *Carpenter et al.* [2004]. Two 12-hour radar nowcasts are performed for each 24-hour model forecast, with a total of 35 days of model forecasts available for analysis. MAPLE refers to the ''McGill Algorithm for Precipitation Nowcasting by Lagrangian Extrapolation'', the radar nowcast algorithm based on the work described by *Germann and Zawadzki* [2002] and *Turner et al.* [2004] that is used in this study. A fourth skill measure (CMAE) is included in this analysis. The results for the first study period were obtained either with no smoothing or smoothing at 4 or 8 times the grid resolution. For the second study period, we implement a ''near optimal forecast filtering (MAPLE-NOFF)'' for the radar nowcasts, described by *Turner et al.* [2004]. In MAPLE-NOFF, the unpredictable small scales as defined by *Germann and Zawadzki* [2002] are filtered out progressively as the nowcast proceeds. In addition, a thresholding is applied to optimize the CSI score [see *Turner et al.*, 2004]. The optimized nowcasts from MAPLE-NOFF have significantly higher skill than MAPLE (Figure 3), especially for CMAE, where the score is superior to the model forecast for up to 12 hours lead time. However, the overall results regarding the radar nowcasts and model forecasts are similar to our first study period: the models have less skill than the radar nowcast for forecast lead times less than about 6 hours and higher skill subsequently. GEM again has a lower score compared to WRF using the POD score, but this difference becomes marginal for

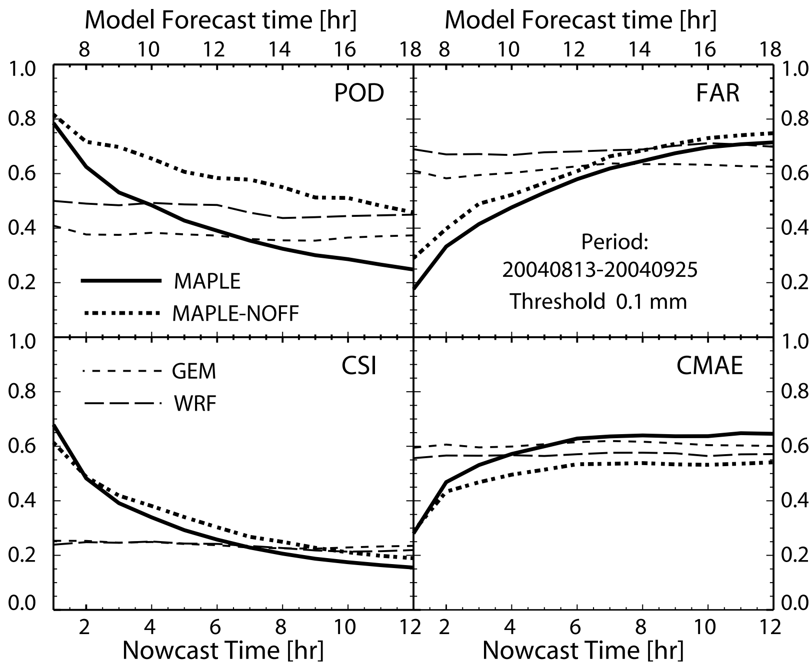

Figure 3. As in Figure 2 but for the second analysis period using two different models (long dashed for WRF and short dashed for GEM operational). The solid (MAPLE) and dotted (MAPLE-NOFF) curves correspond to two versions of 12-hour radar nowcasts; see text for description. The 12-hour forecast period is shown as the abscissa at the bottom, starting at t = 0. The threshold for hourly precipitation is 0.1 mm.

the other scores. The model forecast skill is also approximately constant over the 12-hour evaluation period.

4. Conclusion

[9] We have examined the skill of precipitation forecasts from radar nowcasts and numerical weather prediction models. Four models were examined (GEM/HIMAP, GEM operational, ETA, WRF) over two study periods: 21 days over the summer and fall of 2003–04, and 35 days over the summer of 2004. The domain of analysis is the central and eastern US (2,160 km $\times$ 2,160 km) for the former and the continental US east of 102°W for the latter. Hourly accumulated precipitation amounts were examined. For the first period, five 9-hour radar nowcasts were performed for each 24-hour model forecast, while two 12-hour nowcasts were done for the second period. Skill scores for POD, FAR, CSI and CMAE were calculated and compared. Although the details of the methodology of the two analysis periods are slightly different, the results are similar due to the large sample of precipitation cases examined. According to all four measures, radar nowcasts start with high initial skill that decreases with forecast lead time. The models have lower skill at the beginning of the forecast, but the skill remains approximately constant throughout the forecast period. At a lead time of about 6 hours, the skill of the radar nowcast has decreased to approximately the same level as the models. The reason is the radar nowcasts capture well the initial precipitation distribution, thus resulting in high initial skill. The skill decreases with forecast lead time as development/decay processes are not resolved in an advection-based nowcast algorithm. On the other hand, the models start with initial precipitation conditions that are not as good as the radar's, but the models' skill remains approximately constant throughout the forecast period as the models resolve the large scale processes. The implementation of smoothing to filter small unpredictable scales improve the overall scores, but the cross over point in forecast lead time where the models become more skilful than the radar is largely unchanged. Our results verify quantitatively the conceptual picture of the relative skills of radar nowcasts and model forecasts proposed by *Austin et al.* [1987], *Golding* [1998] and *Wilson et al.* [1998]. An ultimate goal is the blending of radar nowcasts and model forecasts to yield an optimum precipitation forecast.

[10] **Acknowledgments.** The authors gratefully acknowledge the help of Dr. Eric Rogers of the US National Centers for Environmental Prediction (NCEP), who made available the 12-km ETA model hourly accumulated precipitation. The precipitation data from the GEM model were obtained from the Canadian Meteorological Centre (CMC) archives. The US composite radar data were provided by Weather Decision Technologies, Inc. (WDT). Dr. Jim Wilson of NCAR provided useful comments on an initial version of this manuscript. Financial support for this study is from the CRTI-02-0041RD project on "Real Time Determination of Influence of CBRN Release" of Health Canada, and the project on "Improving Quantitative Precipitation Forecasts of Extreme Weather" funded by CFCAS (Canadian Foundation for Climate and Atmospheric Sciences).

References

Austin, G. L., A. Bellon, P. Dionne, and M. Roch (1987), On the interaction between radar and satellite image nowcasting systems and mesoscale numerical models, in *Mesoscale Analysis and Forecasting*, edited by B. Battrick and E. Rolfe, *Eur. Space Agency Spec. Publ.*, *ESA-SP 282*, 225–228.

Carpenter, R. L., Jr., G. M. Bassett, K. A. Brewster, D. Weber, Y. Wang, J. A. Brotzge, K. W. Thomas, F. Kong, and D. Jahn (2004), A globally relocatable numerical weather prediction system based on WRF and ADAS, paper presented at 20th Conference on Weather Analysis and Forecasting and 16th Conf. on Numerical Weather Prediction, Am. Meteorol. Soc., Seattle, Wash.

Côté, J., S. Gravel, A. Méthot, M. Roch, and A. Staniforth (1998), The operational CMC-MRB Global Environmental Multiscale (GEM) model. Part I: Design considerations and formulation, *Mon. Weather Rev.*, *126*, 1373–1395.

Germann, U., and I. Zawadzki (2002), Scale dependence of the predictability of precipitation from continental radar images. Part I: Description of the methodology, *Mon. Weather Rev.*, *130*, 2859–2873.

Golding, B. W. (1998), Nimrod: A system for generating automated very short range forecasts, *Meteorol. Appl.*, *5*, 1–16.

Janjić, Z. I. (1994), The step-mountain Eta coordinate model: Further developments of the convection, viscous sublayer, and turbulence closure schemes, *Mon. Weather Rev.*, *122*, 927–945.

Johnson, L. E., and B. G. Olsen (1998), Assessment of quantitative precipitation forecasts, *Weather Forecasting*, *13*, 75–83.

Michalakes, J., J. Dudhia, D. Gill, J. Klemp, and W. Skamarock (1998), *Design of a Next-Generation Regional Weather Research and Forecast Model: Towards Teracomputing*, pp. 117–124, World Sci., Hackensack, N. J.

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

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

A. Kilambi, C. Lin, B. Turner, S. Vasić, and I. Zawadzki, Department of Atmospheric and Oceanic Sciences, McGill University, 805 Sherbrooke Street West, Montreal, QC, Canada H3A 2K6. (slavko.vasic@mcgill.ca)