# COMPARING THE CHARACTERISTICS OF STEPS NOWCASTS TO SHORT-RANGE NWP

Marion Mittermaier and Clive Pierce
Weather Science, Met Office, Exeter, EX1 3PB, United Kingdom

ABSTRACT

The Fractions Skill Score (FSS) and related metrics such as the skilful spatial scale (L), distribution of physical thresholds (when using frequency thresholds), and skill as a function of scale are used to compare the characteristics of hourly totals from the STEPS nowcast system for to two km-scale NWP models, the 4 km UK4 and 1.5 km UKV. All are compared to the 1 km radar composite over the UK. In particular the evolution of metrics with lead time will be examined, considering the "cross over" lead time where nowcasts become less skilful than NWP.

## 1. INTRODUCTION TO STEPS

STEPS (Bowler *et al.*, 2006) exploits a multiplicative cascade model representation of precipitation fields to generate a high resolution (2 km) ensemble of precipitation nowcasts with a range of 7 hours. These ensembles are a blend of an extrapolated rain analysis derived from weather radar and satellite observations, precipitation forecasts from the 4 km UK4 Met Office Unified model, and noise with space-time statistical properties inferred from the rain analysis.

An extrapolation nowcast is produced using an extrapolation velocity field diagnosed from a time sequential pair of rain analyses using optical flow techniques (Bowler and Pierce, 2002).

## 2. FRACTIONS SKILL SCORE

The Fractions Skill Score (FSS) was introduced by Roberts and Lean (2008) and is a routinely computed spatial verification metric at the Met Office. It enables the comparison of forecasts of different resolutions against a common spatial truth (radar rainfall analyses) in such a way that high-resolution forecasts are not penalised for representativeness errors that arise from the so-called “double penalty” problem and precise matching of forecasts in space and time. The method therefore takes the inherent uncertainty of small scales into account. The FSS enables “close” forecasts to receive some credit, instead of none, and has been used to assess the benefit of high-resolution forecasts (Mittermaier *et al.*, 2011).

## 3. RESULTS

This study focuses on a 6-month period from 1 October 2011 to 31 March 2012. Both direct (comparing a t+1h nowcast with t+1h forecast etc up to t+6h) and lagged comparisons (comparing the t+1h nowcast to the previous model run, i.e. t+7h forecast etc) are made. This is important because output from the latest model run is typically only available 2-3h after initialisation time. This implies that from a user perspective the comparison should be between the t+1h and t+7h, and t+2h with t+8h because this is what users will have available to them.

The direct and lagged comparisons for the largest 5% of hourly totals are shown in Fig. 1. At t+1h the superiority of the STEPS nowcast is clear. At t+2h the skill of the t+2h NWP forecasts is the same as the nowcasts but the t+8h forecasts are still inferior, showing there is value in the nowcasts for the user. The NWP skill deficit has been largely eliminated by t+3h. However, given that t+3h STEPS is worse than t+9h UK4 at this lead time, implies sub-optimal calibration of STEPS. STEPS should always be at least as good as the UK4 run from t-6h at all scales and lead times.

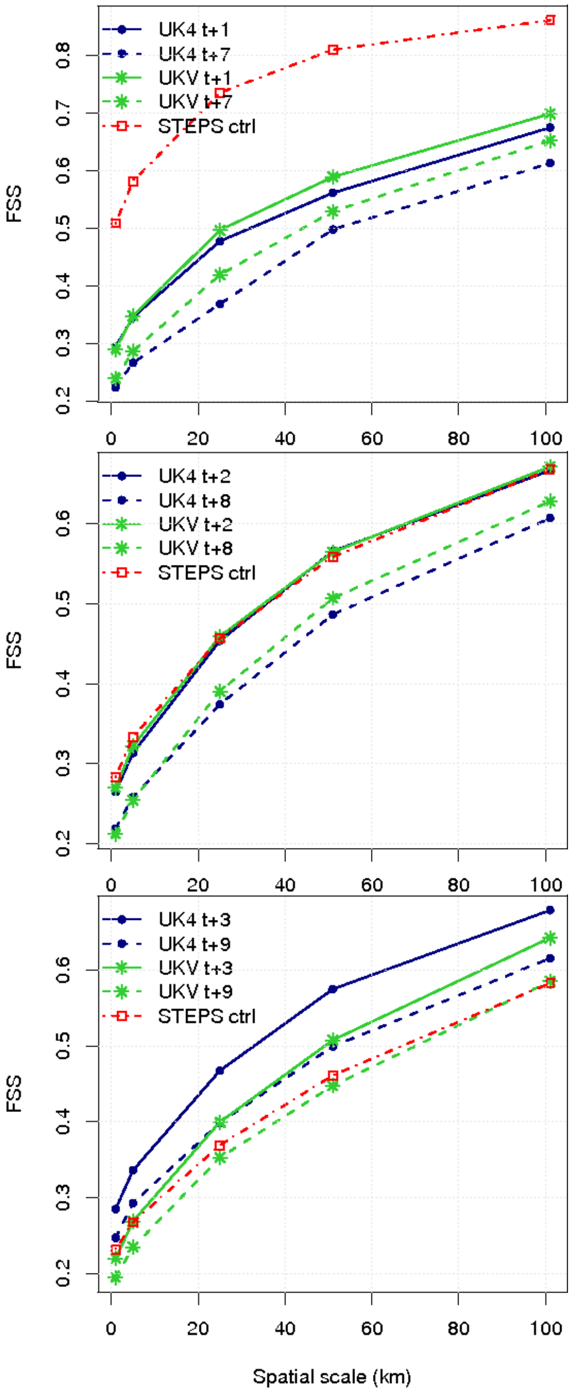


Fig. 1 FSS as a function of neighbourhood size for 1 October 2011 to 31 March 2012.

The verification period spanned the winter months only when the performance of NWP models precipitation forecasts and nowcasts tends to be better.

## 4. CONCLUSIONS

Arguably, the comparison of STEPS to the radar composite gives STEPS an advantage over the UK4 because the radar composite is not an independent observation.

STEPS nowcasts are twice exposed to radar data quality, first by being based on the radar analysis, and when verified against the radar analysis.

More generally fluctuating radar data quality is very detrimental for verification use. An unstable radar baseline leads to subtle observation biases being introduced (which have nothing to do with the forecast), and subsequently misleading results. Arguably an absolutely stable long-term baseline of radar quality is not achievable. Hence the use of frequency thresholds for verification using radar QPE is encouraged, because this removes the bias from the assessment. Note that this means the bias must then be assessed separately.

## 5. REFERENCES

Bowler, N.E. and C.E. Pierce, 2002: Development of a short range QPF algorithm based upon optical flow techniques. Met Office Forecasting Research Technical Report 390.

Bowler, N.E., C.E. Pierce and A.W. Seed, 2006: STEPS: A probabilistic precipitation forecasting scheme which merges an extrapolation nowcast with downscaled NWP. *Q. J. R. Meteorol. Soc.*, **132**, 2127–2155.

Mittermaier, M.P., N. Roberts and S.A. Thompson, 2011: A long-term assessment of precipitation forecast skill using the Fractions Skill Score. *Meteorol. Apps*, DOI=10.1002/met.296.

Roberts, N. and H. Lean, 2008: Scale-Selective Verification of Rainfall Accumulations from High-Resolution Forecasts of Convective Events**,** *Mon. Wea. Rev.,* **136**, 78-97**.**