

CORRIGENDUM

T. P. LEGG AND K. R. MYLNE

Met Office, Exeter, United Kingdom

(Manuscript received and in final form 5 October 2006)

ABSTRACT

In an earlier paper the authors described the use of ensemble information for the generation of early warnings of defined severe-weather events within the United Kingdom. A comprehensive verification of the system was also included in this study. However, an error was later found within the verification code for relative operating characteristic and reliability, which affects most of the results (though the Brier skill scores were not affected). The purpose of the present corrigendum is to provide amended verification results. Briefly, what was found before was that skill appeared to exhibit a maximum for these severe-weather events at 4 days ahead, but, although the results for day 4 remain good, the authors underestimated the skill at other days and so the 4-day skill maximum is no longer clear; instead, skill is useful at days 1–4, and tails off only slowly at days 5–6.

1. Introduction

In our earlier paper (Legg and Mylne 2004, hereafter LM04) we gave a detailed description of how we have used ensemble information from the European Centre for Medium-Range Weather Forecasts (ECMWF) 51-member Ensemble Prediction System (EPS) to generate probabilistic forecasts of severe weather events, in support of the Met Office's National Severe Weather Warning Service. These early warnings [first-guess early warnings (FGEWs)] are offered in probabilistic form and are issued up to 5 days ahead, as described more fully in LM04. The severe events we are concerned with here are severe gales, heavy snowfall, and heavy rainfall; 12 defined geographical areas are used to cover the United Kingdom, and probabilities are also estimated for an event happening anywhere in the United Kingdom.

Our earlier paper claimed an apparent maximum in skill [measured by both relative operating characteristic (ROC) and reliability, as well as by using cost–loss analysis] at 4 days ahead for FGEWs. This does not appear intuitive, and we attempted to explain it in terms of possible spinup effects and the ability of the ensemble perturbations to divert the model evolution

toward or away from severe weather development in the short range (1–3 days), supposing that restricted spread at days 1–3 may have produced overconfident forecasts (i.e., too many event probabilities close to 0 or 1).

Unfortunately, some time after LM04 was published we uncovered an error in our verification system, which resulted in a misalignment of forecast times with verifying times in the FGEW verification database for forecast lead times other than 4 days. This affected the verification code for ROC and reliability, which affected most of the results (though the Brier skill scores were not affected). The revised results are thus unchanged at 4 days, and for the forecaster-issued warnings with which we compare, but the skill of FGEWs for lead times other than 4 days has improved. Skill was previously underestimated because of the error. Thus we see more of a plateau in FGEW skill across days 1–4, and the tailing off beyond 4 days now takes place at a slower rate. This would seem intuitively more plausible, and indicates that our previous explanation of the apparent skill maximum must be withdrawn.

2. Generation of first-guess early warnings

The method by which early warnings are generated, and then used by Met Office forecasters, was described fully in LM04, and need not be repeated in full here. Basically, the process consists of counting how many

Corresponding author address: Timothy Legg, Met Office, FitzRoy Rd., Exeter, EX1 3PB, United Kingdom.
E-mail: tim.legg@metoffice.gov.uk

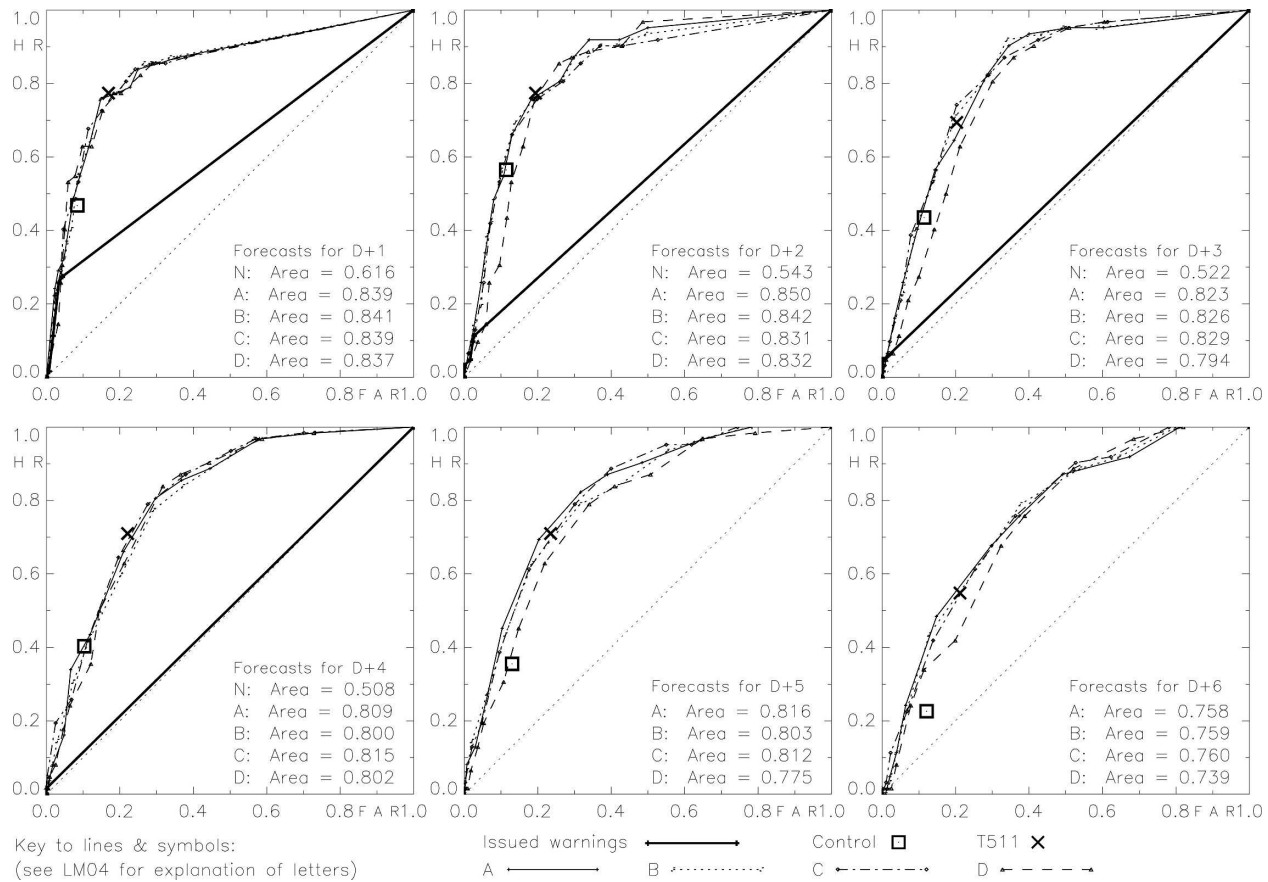


FIG. 1. ROC curves for probabilities of heavy-rainfall events occurring anywhere in the United Kingdom. Forecasts for 1–6 days ahead. Data period: 1 Oct 2001–12 Feb 2003. Letter codes denote different versions of the FGEW system.

(out of 51) ensemble members exceed a given threshold, for at least 1 grid point within each of the 12 U.K. areas, at each 6-hourly time frame. These event thresholds have been calibrated to allow for the limited ability

of the model to resolve extreme events by balancing the mean forecast probability with the sample frequency of severe events. Depending on the event probabilities produced, one of a series of FGEW alert levels is then

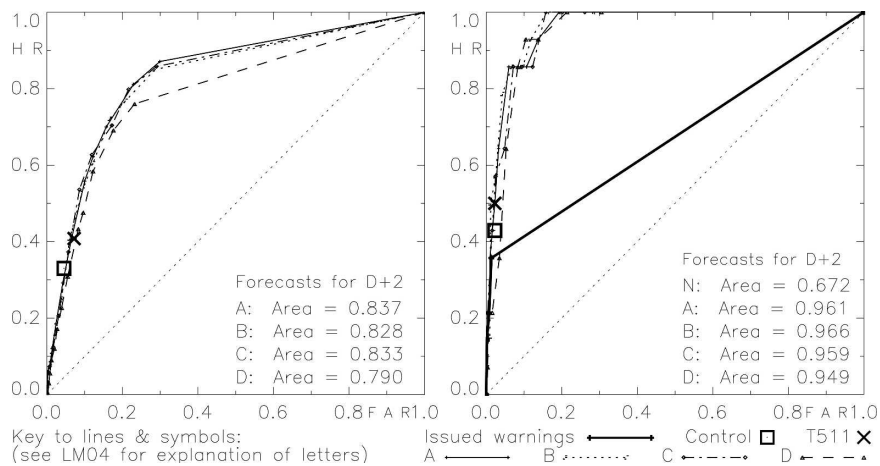


FIG. 2. ROC curves for (left) probabilities of heavy-rainfall events occurring in individual areas of the United Kingdom for FGEW 2 days ahead, and (right) probabilities of severe-gale events occurring anywhere in the United Kingdom for FGEW 2 days ahead and issued 1 day ahead. Data period: 1 Oct 2001–12 Feb 2003.

offered to the forecaster, who considers this in conjunction with all other available information and can then actually issue an early warning if appropriate.

3. Verification of early warnings

Again, the verification methods we have used were described more fully in LM04. Here we concentrate on the revised verification results following the correction of the error in our software. We use standard probabilistic verification scores, including the ROC (Stanski et al. 1989; Wilks 2006), reliability and Brier skill score (Wilks 1995, 2006), and the relative economic value (Richardson 2000). Most of these are event based, using two-by-two contingency tables of whether or not an event occurred and/or was forecast at each of a range of probabilities.

Early warnings are verified against the issue of “flash warnings,” which are issued within a very few hours of the event when confidence is very high.

a. ROC

Figure 1 shows ROC curves for probabilities of heavy-rainfall events occurring anywhere in the United Kingdom at 1–6 days ahead, and is a corrected version of Fig. 2 in LM04. The increase in ROC areas at days 1–3 and 5–6 for FGEW forecasts is evident. This increase is also seen for the deterministic forecasts (ECMWF EPS control member run at T255, and ECMWF high-resolution operational deterministic forecast run at T511, as it was during the period of study); these results are included for comparison.

Figure 2 shows $D + 2$ results equivalent to those shown in LM04’s Fig. 4 (left-hand panel) and Fig. 5 (left-hand panel). These are for heavy-rainfall events occurring in individual areas, and for severe-gale events occurring anywhere in the United Kingdom. In both cases the corrected results show much higher ROC areas at $D + 2$, slightly higher than the equivalent at $D + 4$ shown in LM04.

There is still little to choose between the different versions of the FGEW system, which were tried experimentally (these are described more fully within LM04); we still believe that the version used operationally, which includes the five “multianalysis” members run using the same ECMWF model at T255 resolution, provides the best probabilistic forecasts possible by this method.

b. Reliability

The reliability diagram (Wilks 1995, 2006) plots how well forecast probability relates to observed frequency of events. The reliability diagram for $D + 2$ forecasts of

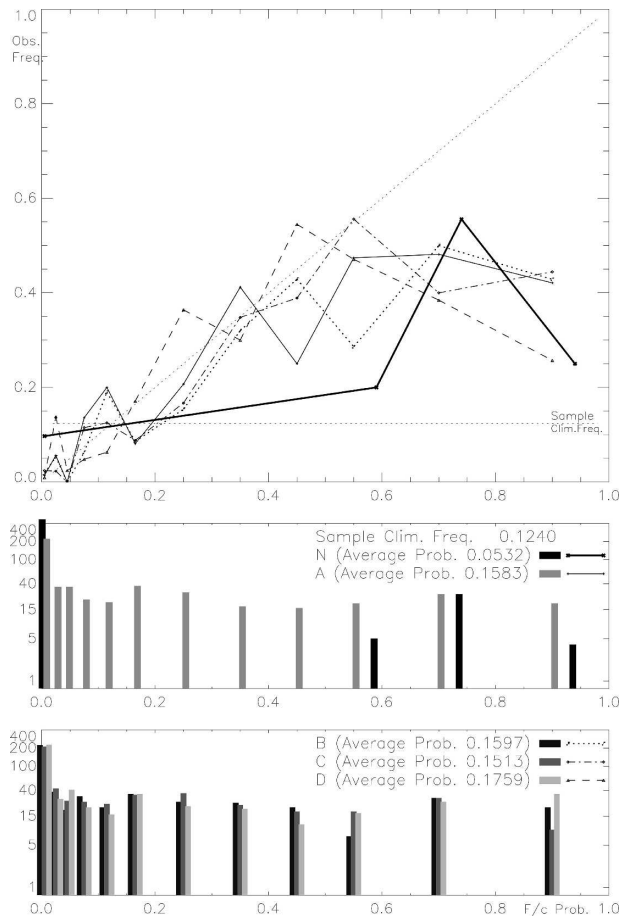


FIG. 3. (top) Reliability, and sharpness for versions (middle) N and A and (bottom) B, C, and D, for probabilities of heavy-rainfall events anywhere in the United Kingdom for FGEW 2 days ahead and issued 1 day ahead. Data period: 1 Oct 2001–12 Feb 2003.

heavy rain presented in Fig. 7 of LM04 showed near-horizontal lines indicating no probabilistic discrimination of events. Figure 3 here is a corrected version of this diagram and shows reasonable reliability, particularly for probabilities up to around 0.5, and that forecasts do have useful resolution when it comes to determining whether an event is more or less likely. Other reliability diagrams in LM04 were for $D + 4$ and are not affected by the software error. Also we see that the FGEW forecasts are more reliable than the issued forecasts at $D + 1$ (boldface solid line, included to illustrate the benefit available from FGEWs). This marked improvement in reliability is seen for all days, and is also evident to some degree for severe-gale and heavy-snowfall events (not shown).

c. Cost–loss analysis

Finally, Fig. 14 of LM04 presented relative economic value plotted against user cost–loss ratio (Richardson

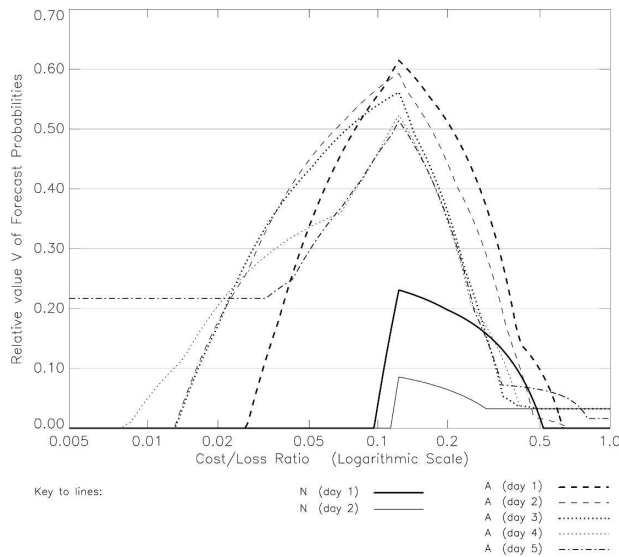


FIG. 4. Cost-loss diagrams for heavy-rainfall events occurring anywhere in the United Kingdom. Issued forecasts 1–2 days ahead, compared with operational FGEW probabilities for 1–5 days ahead. Data period: 1 Oct 2001–12 Feb 2003.

2000) for FGEW forecasts of heavy rain at 1–5 days ahead, showing greatest value for $D + 4$ forecasts. Here, Fig. 4 shows a corrected version of this diagram, with a slow decrease in relative economic value with increasing lead time. Relative value exceeds 0.5 up to day 5, far exceeding the value of the issued forecasts 1 and 2 days ahead.

4. Summary and conclusions

Because of an error in the verification software, some of the results presented by Legg and Mylne (2004) from

the FGEW system have had to be altered. The effect of this error was to underestimate the skill of the system at days 1–3 and 5–6. Hence, the earlier striking result that the skill of the system peaked at 4 days ahead was incorrect, and instead we find that forecasts are skilful throughout days 1–4, with the skill tailing off only slowly with time. There is no change to our conclusion regarding which of the different versions of FGEW that have been tried actually performs best. Our previous suggestions that the effects seen were due to spinup effects or were characteristic of the short-range forecast system for severe weather must now be withdrawn. Finally, we would like to apologize to our friends and colleagues at the ECMWF who invested considerable time and effort in trying to understand the erroneous results that we previously presented.

REFERENCES

Legg, T. P., and K. R. Mylne, 2004: Early warnings of severe weather from ensemble forecast information. *Wea. Forecasting*, **19**, 891–906.

Richardson, D. S., 2000: Skill and relative economic value of the ECMWF Ensemble Prediction System. *Quart. J. Roy. Meteor. Soc.*, **126**, 649–667.

Stanski, H. R., L. J. Wilson, and W. R. Burrows, 1989: A survey of common verification methods in meteorology. WMO WWW Tech. Rep. 8, WMO Tech. Doc. 358, 114 pp.

Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences—An Introduction*. International Geophysics Series, Vol. 59, Academic Press, 467 pp.

—, 2006: *Statistical Methods in the Atmospheric Sciences*. 2d ed. International Geophysics Series, Vol. 91, Academic Press, 627 pp.