Advances in Geosciences

The use of MOGREPS ensemble rainfall forecasts in operational flood forecasting systems across England and Wales

J. Schellekens[1], A. H. Weerts[1], R. J. Moore[2], C. E. Pierce[3], and S. Hildon[4]

[1]Inland Water Systems, Deltares, Delft, The Netherlands
[2]Joint Centre for Hydro-Meteorological Research, Centre for Ecology and Hydrology, Wallingford, UK
[3]Joint Centre for Hydro-Meteorological Research, Met Office, Wallingford, UK
[4]Flood Risk Science, Environment Agency, Leeds, UK

Abstract. Operational flood forecasting systems share a fundamental challenge: forecast uncertainty which needs to be considered when making a flood warning decision. One way of representing this uncertainty is through employing an ensemble approach. This paper presents research funded by the Environment Agency in which ensemble rainfall forecasts are utilised and tested for operational use. The form of ensemble rainfall forecast used is the Met Office short-range product called MOGREPS. It is tested for operational use within the Environment Agency's National Flood Forecasting System (NFFS) for England and Wales. Currently, the NFFS uses deterministic forecasts only. The operational configuration of the NFFS for Thames Region is extended to trial the use of the new ensemble rainfall forecasts in support of probabilistic flood forecasting. Evaluation includes considering issues of model performance, configuration (how to fit the ensemble forecasts within the current configurations), data volumes, run times and options for displaying probabilistic forecasts. Although ensemble rainfall forecasts available from MOGREPS are not extensive enough to fully verify product performance, it is concluded that their use within current Environment Agency regional flood forecasting systems can provide better information to the forecaster than use of the deterministic forecasts alone. Of note are the small number of false alarms of river flow exceedance generated when using MOGREPS as input and that small flow events are also forecasted rather well, notwithstanding the rather coarse resolution of the MOGREPS grid (24 km) compared to the studied catchments. In addition, it is concluded that, with careful configuration in NFFS, MOGREPS can be used in existing systems without a significant increase in system load.

1 Introduction

From the early 1990s, meteorologists have been providing ensemble predictions of rainfall and an increasing number of hydrologists have begun to use these in (semi) operational flood forecasting systems (e.g. Gouweleeuw et al., 2005; Pappenberger et al., 2005). One major source of uncertainty in hydrological forecasting systems is the forecasted rainfall. As such, the use of ensemble rainfall forecasts may lead to increased understanding of the uncertainty associated with hydrological forecasts.

The Met Office is the main provider for the Environment Agency of meteorological forecast products. These forecast products are being used in the Environment Agency's current operational forecasting system, the National Flood Forecasting System (NFFS). The numerical weather prediction (NWP) capability of the Met Office is continuously being enhanced and new ensemble products are now being made operationally available. Currently, the nowcasting system STEPS – Short Term Ensemble Prediction System (Bowler et al., 2006) – produces deterministic rainfall forecasts at a 2 km resolution. In the near future these will be available to the Environment Agency in ensemble form. Also, for longer-term numerical weather prediction (NWP) a new ensemble forecasting system has been developed called MOGREPS – Met Office Global and Regional Ensemble Prediction System (Bowler et al., 2008) – which uses a coarser model resolution of 24 km (now 18 km). These developments offer interesting opportunities for the Environment Agency and open the door to probabilistic flood forecasting. Operational research is required to realise the potential benefits of these developments for flood warning. Many existing ensemble prediction systems across Europe are based on ECMWF ensembles (Cloke et al., 2009; Cloke and Pappenberger, 2009; Thielen-del Poze et al., 2009) and are aimed at medium-range forecasting although MAP D-PHASE (Rotach et al., 2009) is focused on Alpine systems and shorter lead-times.

The MOGREPS product, investigated here for the first time in a hydrological forecasting application context, is aimed at the shorter range of one to three days.

Hydrological models have the capability to provide useful river-flow forecasts supporting flood warnings if the rainfall information they are supplied with is sufficiently accurate. These models have generally been used with raingauge data, radar analyses and extrapolated radar forecasts. More recently longer-term NWP model rainfall forecasts have also been used. All of these have now been incorporated in the NFFS for regional flood forecasting across England and Wales and are used to drive a large number of hydrological models.

The Environment Agency's R&D project "Hydrological Modelling using Convective Scale Rainfall Modelling" (Schellekens et al., 2010) aimed to investigate what hydrological model concepts and associated computational methods allow for making best use of the latest Met Office developments in probabilistic rainfall forecasting. All the hydrological modelling concepts explored were trialled within the current National Flood Forecasting System (NFFS) infrastructure that is based on Delft-FEWS (Werner et al., 2004). Currently, the new Flood Forecasting Centre (FFC), a joint venture between the Environment Agency and Met Office, is trailing MOGREPS in combination with a grid-based hydrological model (Grid-to-Grid; Bell et al., 2009) to deliver flood risk outlooks across England and Wales (Price et al., 2011).

2 Methodological approach

2.1 MOGREPS

In 2005, the Met Office introduced MOGREPS for short-range ensemble prediction, providing a 24-km-resolution regional ensemble for the Atlantic and Europe domain. Ensemble forecasting is based on the principle of adding small perturbations to the "best estimate" of the initial state of the atmosphere. The model is then run forward from the perturbed starting conditions to generate an ensemble of different forecasts. The regional model which is used here (MOGREPS-R) is designed to provide ensemble forecasts for the short-range (days 0–3) for the United Kingdom and Ireland. It provides a 24-member ensemble of 3-hour rainfall totals with a grid resolution of 24 km for a forecast length of 54 h (36 h are used in this research). Boundary conditions for the regional model are provided by a global model (MOGREPS-G) with a 90-km grid and a forecast time of 72 h producing a 24-member ensemble. Both models are run twice daily, the global at 00:00 and 12:00 UTC and the regional 6 h after these times. Due to spin-up issues, and the fact that only two forecasts are available per day, the first hours of the MOGREPS runs are generally not used. The ensembles consist of 1 control run and 23 additional members. The control forecast is run at the same resolution as the other ensemble members but does not contain any perturbations to account for initial condition or model uncertainties – as such it runs from the best analysis of the initial state of the atmosphere. The control run can be compared with the standard deterministic weather forecast that is run at a 12-km resolution.

2.2 Implementation of MOGREPS ensembles in NFFS Thames Region

Thames Region, one of the eight regions within the NFFS, was chosen to test the use of ensemble rainfall forecasts from both MOGREPS and STEPS in an emulation of the operational flood forecasting system. The Region has fairly long lead times to its most important forecasting locations but has, on the other hand, significant fast responding urban areas within its forecasting responsibility. Hydrological modelling within Thames region employs the Thames Catchment Model or TCM (Wilby et al., 1994; CEH, 2005) with a total of 148 TCMs used and that cover most of the region. The forecasting time-step is 15 min and models have quick run-times. Each TCM is used in combination with an ARMA (AutoRegressive Moving Average) error predictor model (CEH, 2005) with the forecast flow formed as the sum of the TCM simulated flow and a prediction of the error that utilises flow observations up to the time the forecast is made. A total of 112 raingauges and 537 river gauges is used in Thames Region to support the forecasting process. TCM plus ARMA combinations are used to forecast flow at key locations or lateral inflows to the main rivers which are modelled separately using hydrodynamic ISIS models. Due to the nested configuration of the TCMs, the larger of the currently available models cover a long lead-time which makes them less beneficial if used with nowcasting ensembles from STEPS and in some cases also for short-range ensembles from MOGREPS.

To evaluate the performance of MOGREPS and STEPS as they would be used in the operational setup in NFFS, a complete copy of the operational system was setup and run on a day-to-day basis using all available operational data as input to the system. After extending the configuration of the system to include the MOGREPS and STEPS ensemble rainfall forecasts, the system could be run in hindcast mode with the available historical forecasts.

Hindcasts using MOGREPS were performed for Thames Region for the period July 2008 to February 2009 for which MOGREPS ensemble rainfall forecasts were available twice a day. Hydrometric data between January 2008 and the start of the hindcast were used to hotstart the TCMs used for river flow forecasting. After warming up the models, runs where made ending each day at 09:00 and for which time a set of initial conditions were saved. Repeating this for the whole period provided a set of initial conditions to be used for starting the forecasts. Subsequently, the forecasts were run twice a day at 09:00 and 21:00 coinciding with current operational
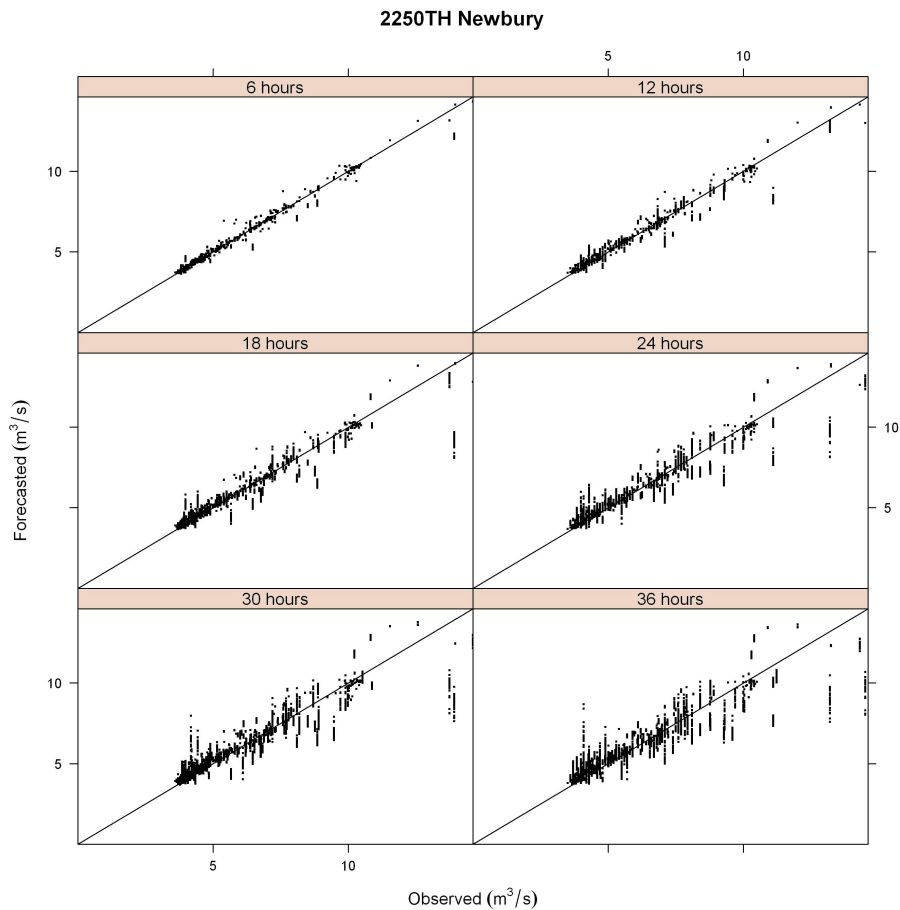
2250TH Newbury

Fig. 1. Scatter plot of forecast and observed flow for different lead-times for Newbury.

practice in Thames Region. Results from the operational system were exported and employed in post-processing analyses that used the R programming language (Ihaka and Gentleman, 1996) outside the operational environment.

3 Results and discussion

The eight-month period used in hindcasting proved to be too short to determine reliable statistics for assessing performance over flood events. Ideally, verification would measure the performance of the system in forecasting the crossing of important flood warning thresholds. However, to establish meaningful verification statistics, a sufficiently large number of observed "events" is needed, where an event is understood to be the crossing of a certain flow (or level) threshold. Typically, thresholds that are meaningful within the context of operational flood forecasting are relatively high. As the verification period considered here is relatively short, crossing of these thresholds may not have occurred at all or only so infrequently that the number of events is not large enough to give a meaningful performance statistic. Other work on

MOGREPS (Bowler et al., 2007) has look at verification of the rainfall forecast itself. Therefore, here the results of the hindcast assessment will be presented in three ways: (i) scatter plots of forecast versus observed flow for a given lead time, (ii) Rank Probability Skill Score (RPSS), of limited value because of sample size, and (iii) individual forecast hydrograph plots to get an impression of the behaviour of the short-term ensemble prediction system for the small catchments involved.

For the analysis of the results, eight locations were chosen (that have warning thresholds set) to provide a representative cross section of the TCMs used in Thames Region. The name, station identifier and catchment area of these locations are given in Table 1 together with their warning thresholds.

Figures 1 and 2 show typical scatter plots of the flow forecasts using MOGREPS rainfalls as input versus the corresponding observations for different forecast lead times. These scatter plots show how well the forecasted flows correspond to those observed. As one would expect, the spread of the forecasts narrows and the bias falls as the lead-time of the forecast decreases. For some locations, this is more the case than for others. This has mainly to do with the response
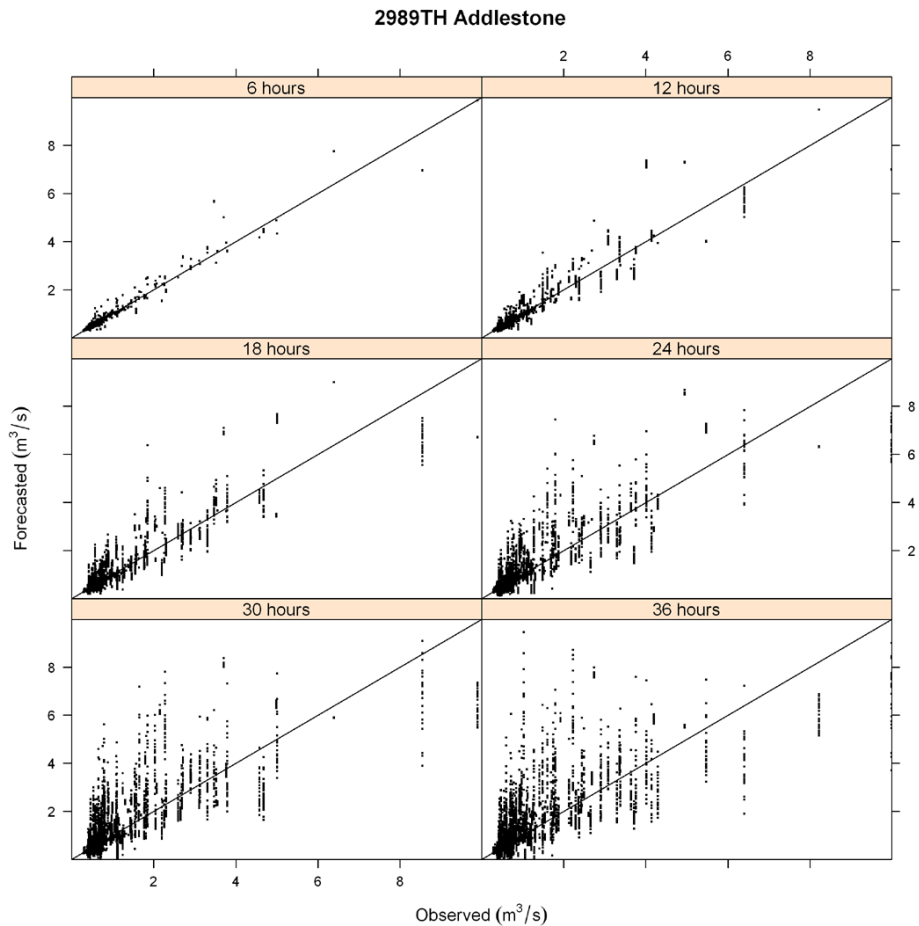
2989TH Addlestone

Fig. 2. Scatter plot of forecast and observed flow for different lead-times for Addlestone.

Table 1. Warning thresholds (m^3 s^{-1}) for locations (name, station identifier and area) used in the evaluation.

Warning Level	Colindeep Lane 2989 TH 29.0 km^2	Marsh Farm 3680 TH 81.0 km^2	Redbridge 5480 TH 303.3 km^2	Addlestone 2989 TH 90.1 km^2	Binfield 2250 TH 50.2 km^2	Kinnersley Manor 3240 TH 142.0 km^2	Newbury 2250 TH 548.1 km^2
Standby	7.6	2	12.5	3.9	16	18	12.4
Bankfull	12.4	30.2	18	6	23	29	16
Floodplain	14.9	33	24.5	6.9	31.5	33	18.5
Property	17.68	39	35.1	10	45	61.5	22.4

time of the specific catchment and if a rainfall event was predicted by MOGREPS. For example, plots for Newbury (Fig. 1) hardly show any spread because no large rainfall event was predicted (and observed) during the period July 2008 to February 2009. Plots for smaller catchments, such as Addlestone (Fig. 2) and Kinnersley Manor (Fig. 3), already exhibit larger spreads at 6 h lead-time due to catchment size/response time.

In order to verify the whole range of possible outcomes, the Ranked Probability Score (RPS) can be used (Wilks, 2006). For verifying flow rates, a number of categories are defined whose flow ranges cover all possible outcomes. The squared difference between the cumulative forecast probability and the corresponding cumulative observation for each category is averaged across all categories to obtain the RPS. The RPS is sensitive to category distance: for example, if a forecast falls into a more distant category than the
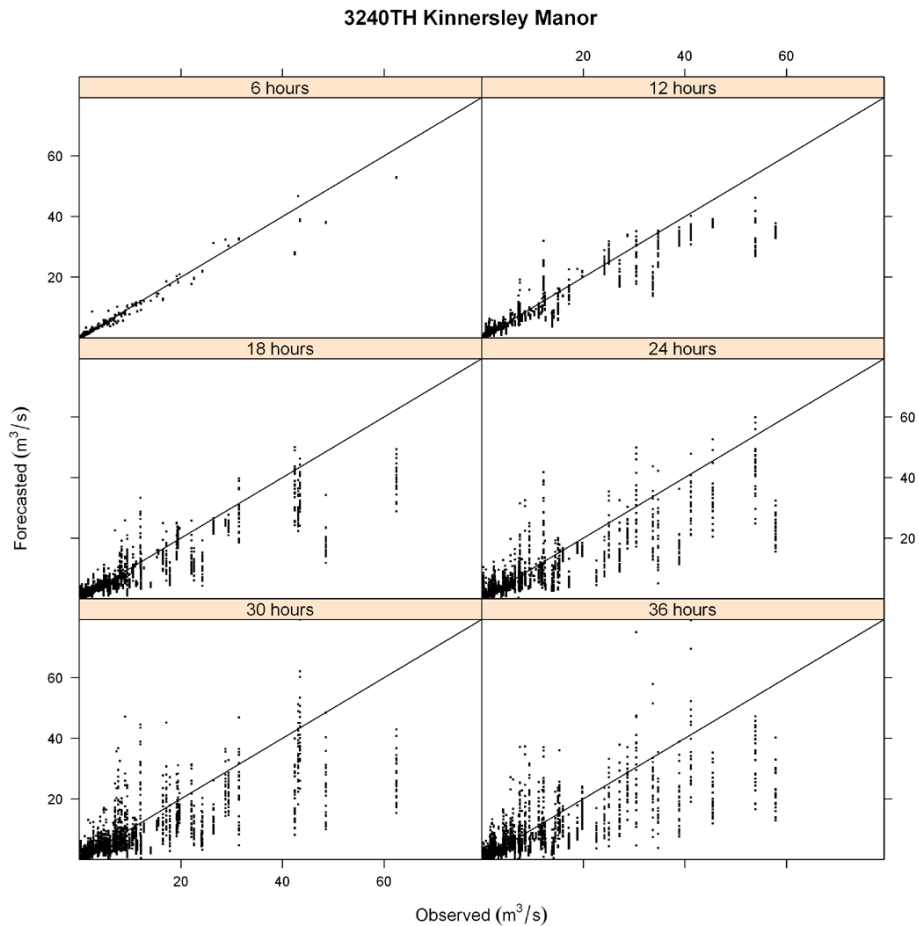
3240TH Kinnersley Manor

Fig. 3. Scatter plot of forecast and observed flow for different lead-times for Kinnersley Manor.

Table 2. Mean RPS for the period July 2008 to February 2009 using the thresholds listed in Table 1.

Lead Time in hours	Colindeep Lane 2989 TH	Marsh Farm 3680 TH	Redbridge 5480 TH	Addlestone 2989 TH	Binfield 2250 TH	Kinnersley Manor 3240 TH	Newbury 2250 TH
6	0.000	0.007	0.004	0.002	0.000	0.004	0.001
12	0.000	0.005	0.006	0.003	0.000	0.004	0.001
18	0.000	0.009	0.006	0.005	0.000	0.005	0.001
24	0.000	0.008	0.006	0.006	0.000	0.008	0.002
30	0.000	0.010	0.005	0.007	0.000	0.008	0.002
36	0.000	0.009	0.007	0.007	0.000	0.008	0.003

observation, it will be penalised more. Zero is the perfect score for RPS. Table 2 shows the mean RPS values for the period July 2008 to February 2009 using the warning thresholds of Table 1. As mentioned previously, these statistics are not very revealing because of the limited period used in the analysis not encompassing many threshold-exceedance events. The ranked probability skill score (RPSS) using the naïve persistence forecast as a reference has been calculated as well and is shown in Table 3. This skill score measures the

improvement of the multi-category probabilistic forecast relative to a reference forecast (usually the long-term or sample climatology, but here the persistence forecast). It is similar to the 2-category Brier Skill Score, in that it takes climatological frequency into account. Because the score denominator approaches 0 for a perfect forecast, this score can be unstable when applied to small datasets. The rarer the event, the larger the number of samples needed to stabilise the score. The unstable behaviour (resulting in ∞ and undefined) is

Table 3. Mean RPSS for the period July 2008 to February 2009 using the thresholds of Table 1 and the naïve forecast (persistence) as reference forecast.

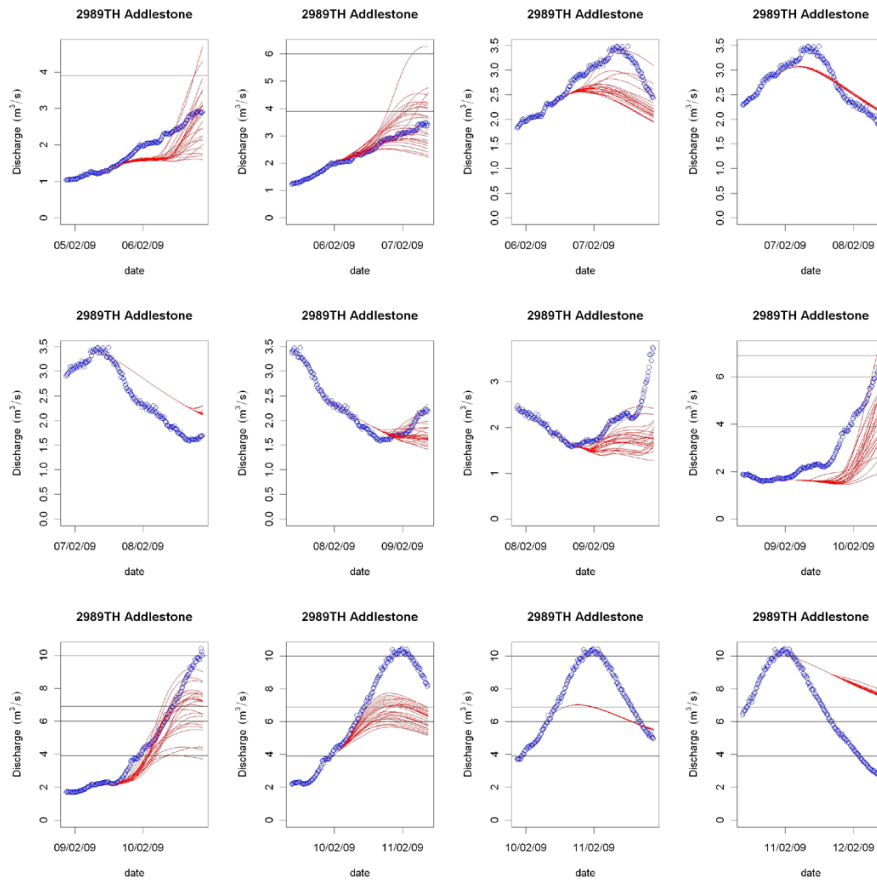| Lead Time in hours	Colindeep Lane 2989 TH	Marsh Farm 3680 TH	Redbridge 5480 TH	Addlestone 2989 TH	Binfield 2250 TH	Kinnersley Manor 3240 TH	Newbury 2250 TH
6	0.8524	0.210473	−0.16667	0.333333	−∞	0.588235	0
12	−∞	0.605379	0.25	0.31304	–	0.770448	0.5
18	0.519032	0.450023	0.451571	0.337467	–	0.782187	0.333333
24	−∞	0.639668	0.539234	0.408015	−∞	0.715205	0.24826
30	0.325112	0.545441	0.666822	0.373642	−∞	0.725603	0.196872
36	−∞	0.642509	0.583501	0.488848	−∞	0.723337	−0.074

Fig. 4. Flow forecast ensembles for 12 consecutive forecasts, 12 h apart, in February 2009 at Addlestone. Observed flows indicated by circles. Horizontal lines represent the thresholds from Table 1.

clearly visible in Table 3. Possibly the only result which has some credibility is that for Kinnersley Manor where there are multiple events over the whole range of observed values. Here, the RPPSSs indicate that there is skill when using MOGREPS ensembles as forecast rainfall input to the TCM compared to use of the naïve persistence forecast. Another insight that may be deduced from Table 3 is that MOGREPS did not produce many false alarms.

For Colindeep Lane, Marsh Farm, Binfield and Newbury no warning threshold, only the Standby threshold, was crossed (both by flood forecasts and observations) during the 8 month hindcast period.

The lack of enough warning thresholds being crossed limits the usefulness of the RPPSS somewhat in this case. Therefore, the Continuous Ranked Probability Score (CRPS) was also calculated to get a better indication of the performance of MOGREPS (Grimit et al., 2006). Table 4 shows the

Table 4. CRPS for the period July 2008 to February 2009 using the naïve forecast (persistence) as reference forecast.

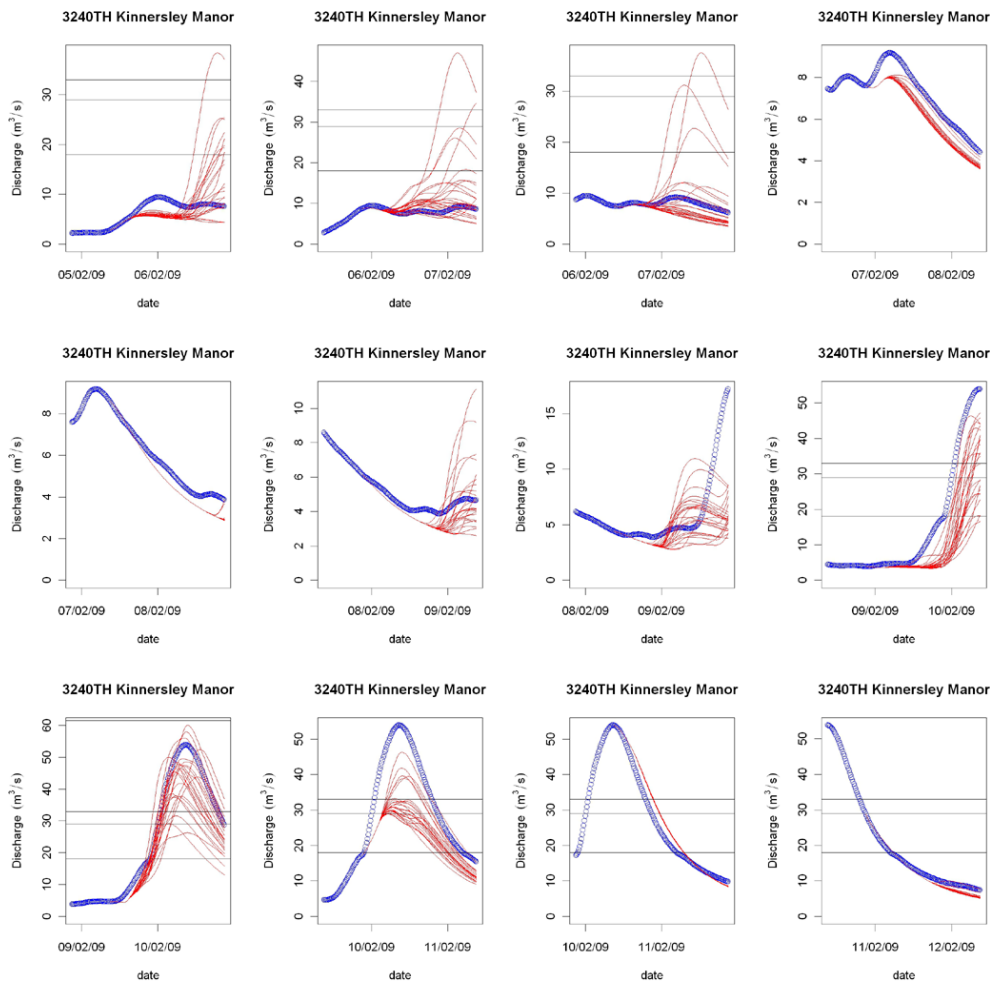| Lead Time in hours	Colindeep Lane 2989 TH	Marsh Farm 3680 TH	Redbridge 5480 TH	Addlestone 2989 TH	Binfield 2250 TH	Kinnersley Manor 3240 TH	Newbury 2250 TH
6	0.0867	0.2019	−0.0292	0.0419	0.1621	0.2552	0.0193
12	0.2567	0.4299	−0.0242	0.0944	0.4178	0.5717	0.1447
18	0.2570	0.4520	0.1413	0.1335	0.3890	0.5836	0.1983
24	0.2955	0.4954	0.2389	0.1510	0.4276	0.6111	0.1933
30	0.3311	0.4839	0.3086	0.1605	0.4612	0.6050	0.2206
36	0.2927	0.4914	0.3085	0.1419	0.4362	0.6325	0.1963

Fig. 5. Flow forecast ensembles for 12 consecutive forecasts, 12 h apart, in February 2009 at Kinnersley Manor. Observed flows indicated by circles. Horizontal lines represent the thresholds from Table 1.

CRPSS (Continuous Rank Probability Skill Score). This is defined as the CRPS of the MOGREPS forecast divided by the CRPS of the naïve forecast. A value of 1 indicates perfect skill and, while zero would indicate that the forecast is equal in skill to the reference forecast, anything below 0 indicates that the skill is less than the reference forecast. Table 4 shows that the skill of the MOGREPS forecast increases with lead time. This is to be expected as the naïve forecast does not include any forecasted precipitation. For Redbridge the naïve forecast performs better up to 12 h but for other, faster responding, stations MOGREPS shows skill even for the shorter lead times (6 to 16 h).

Figures 4 and 5 show the forecast and observed hydrographs for Addlestone and Kinnersley Manor for February 2009. For both locations the event of 9 to 11 February is well predicted, although use of the MOGREPS ensembles tends to lead to underestimation of the flood event to some degree. This behaviour can also be observed in some of the scatter plots (Fig. 1).

4 Conclusions

An ensemble product MOGREPS developed by the Met Office has been tested for operational use by the Environment Agency. The regional model used is designed to provide ensemble forecasts (24 members) for the short-range (days 0–3) for the United Kingdom and Ireland.

A real verification study of the ensemble flood forecast made using MOGREPS is not possible given the period of the hindcast July 2008 to February 2009. But from this hindcast study several observations can be made. The MOGREPS ensembles of forecast rainfall give good ensemble flood forecasts when used with the flood forecasting models used across the Thames Region. The frequency of false alarms is low over this period. Also smaller events below the warning thresholds are often well predicted, despite the rather coarse resolution of the ensemble rainfall forecasts available at the time. Whilst MOGREPS at 24 km resolution does not purport to represent localised convective (or orographic) rainfall, this operational evaluation over the lowland Thames Region of England provides evidence of its utility for probabilistic flood forecasting and warning.

Edited by: F. Pappenberger
Reviewed by: two anonymous referees

References

Bell, V. A., Kay, A. L., Jones, R. G., Moore, R. J., and Reynard, N. S.: Use of soil data in a grid-based hydrological model to estimate spatial variation in changing flood risk across the UK, J. Hydrol., 377(3–4), 335–350, 2009.

Bowler, N. E., Pierce, C. E., and Seed, A. W.: STEPS: A probabilistic precipitation forecasting scheme which merges an extrapolation nowcast with downscaled NWP, Q. J. R. Meteorol. Soc, 132(620), 2127–2155, 2006.

Bowler, N. E., Dando, M., Bear, S., and Mylne, K.: The MOGREPS short-range ensemble prediction system: Verification report. Trial performance of MOGREPS January 2006–March 2007, Forecasting Research Techinical Report, Met Office, Exeter, 2007.

Bowler, N. E., Arribas, A., Mylne, K. R., Robertson, K. B., and Beare, S. E.: The MOGREPS short-range ensemble prediction system, Q. J. R. Meteorol. Soc, 134(632), 703–722, 2008.

CEH: PSM Penman Store Model Incorporating the TCM and the IEM. Version 2.2, September 2005, Centre for Ecology and Hydrology, Wallingford, UK, 2005.

Cloke, H., Thielen, J., Pappenberger, F., Nobert, S., Balint, G., Edlund, C., Koistinen, A., De Saint-Aubin, C., Sprokkereef, E., Viel, C., Salamon, P., and Buizza, R.: Progress in flood forecasting, ECMWF Newsletter No. 121, Autumn 2009, 20-24, 2009.

Cloke, H. L. and Pappenberger, F.: Ensemble flood forecasting: a review, J. Hydrol, 375(3–4), 613–626, 2009.

Gouweleeuw, B. T., Thielen, J., Franchello, G., De Roo, A. P. J., and Buizza, R.: Flood forecasting using medium-range probabilistic weather prediction, Hydrol. Earth Syst. Sci., 9, 365–380, doi:10.5194/hess-9-365-2005, 2005.

Grimit, E. P., Gneiting, T., Berrocal, V. J., and Johnson, N. A.: The continuous ranked probability score for circular variables and its application to mesoscale forecast ensemble verification, Q. J. R. Meteorol. Soc., 132(621C), 2925–2942, 2006.

Ihaka, R. and Gentleman, R.: A language for data analysis and graphics, J. Comput. Graph. Stat., 5(3), 299–314, 1996.

Pappenberger, F., Beven, K. J., Hunter, N. M., Bates, P. D., Gouweleeuw, B. T., Thielen, J., and de Roo, A. P. J.: Cascading model uncertainty from medium range weather forecasts (10 days) through a rainfall-runoff model to flood inundation predictions within the European Flood Forecasting System (EFFS), Hydrol. Earth Syst. Sci., 9, 381–393, doi:10.5194/hess-9-381-2005, 2005.

Price, D., Hudson, K., Boyce, G., Schellekens, J., Moore, R. J., Clark, P., Harrison, T., and Conolly, E.: Operational application of a distributed grid-based hydrological model for flood forecasting in England and Wales, in review, Water Management, 2011.

Rotach, M. W., Ambrosetti, P., Ament, F., Appenzeller, C., Arpagaus, M., Bauer, H., Behrendt, A., Bouttier, F., Buzzi, A., Corazza, M., Davolio, S., Denhard, M., Dorninger, M., Fontannaz, L., Frick, J., Fundel, F., Germann, U., Hegg, T. G. C, Hering, A., Keil, C., Liniger, M. A., Marsigli, C., McTaggart-Cowan, R., Montani, A., Mylne, K., Ranzi, R., Richard, E., Rossa, A., Santos-Muñoz, D., Schär, C., Seity, Y., Staudinger, M., Stoll, M., Volkert, H., Walser, A., Wang, Y., Werhahn, J., Wulf-meyer, V, and Zappa, M.: MAP D-PHASE: Real-time demonstration of weather forecast quality in the Alpine region, Bull. Amer. Meteorol. Soc, 90(9), 1321–1336, 2009.

Schellekens, J., Moore, R., Cole, S., Reggiani, P., Weerts, A., Minett, A., Bell, V., and Robson, A. J.: Hydrological modelling using convective scale rainfall modelling, Research Contractors: Deltares and CEH Wallingford. Environment Agency, Bristol, UK, 2010.

Thielen, J., Bartholmes, J., Ramos, M.-H., and de Roo, A.: The European Flood Alert System - Part 1: Concept and development, Hydrol. Earth Syst. Sci., 13, 125–140, doi:10.5194/hess-13-125-2009, 2009.

Werner, M., van Dijk, M., and Schellekens, J.: DELFT-FEWS: an open shell flood forecasting system, in Proceedings of the 6 TH International Conference on Hydroinformatics, 1205–1212, 2004.

Wilby, R., Greenfield, B., and Glenny, C.: A coupled synoptic-hydrological model for climate change impact assessment, J. Hydrol, 153(1–4), 265–290, 1994.

Wilks, D. S.: Statistical methods in the atmospheric sciences: an introduction, 2nd edn., Academic Press, Amsterdam, 2006.