METEOROLOGICAL APPLICATIONS
Meteorol. Appl. (2009)
Published online in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/met.132

Tracking the uncertainty in flood alerts driven by grand ensemble weather predictions

Y. He[a],* F. Wetterhall,[a] H. L. Cloke,[a] F. Pappenberger,[b] M. Wilson[c], J. Freer[d] and G. McGregor[e]

[a] *Department of Geography, King's College London, Strand, London, WC2R 2LS, UK*
[b] *European Centre for Medium-Range Weather Forecasts (ECMWF), Shinfield Park, Reading, RG2 9AX, UK*
[c] *Department of Food Production, University of the West Indies, St Augustine, Trinidad and Tobago*
[d] *School of Geographical Sciences, University of Bristol, University Road, Bristol, BS8 1SS, UK*
[e] *School of Geography, Geology and Environmental Science, University of Auckland, Private Bag 92019, Auckland, New Zealand*

ABSTRACT: The incorporation of numerical weather predictions (NWP) into a flood warning system can increase forecast lead times from a few hours to a few days. A single NWP forecast from a single forecast centre, however, is insufficient as it involves considerable non-predictable uncertainties and can lead to a high number of false or missed warnings. Weather forecasts using multiple NWPs from various weather centres implemented on catchment hydrology can provide significantly improved early flood warning. The availability of global ensemble weather prediction systems through the 'THORPEX Interactive Grand Global Ensemble' (TIGGE) offers a new opportunity for the development of state-of-the-art early flood forecasting systems. This paper presents a case study using the TIGGE database for flood warning on a meso-scale catchment (4062 km^2) located in the Midlands region of England. For the first time, a research attempt is made to set up a coupled atmospheric-hydrologic-hydraulic cascade system driven by the TIGGE ensemble forecasts. A probabilistic discharge and flood inundation forecast is provided as the end product to study the potential benefits of using the TIGGE database. The study shows that precipitation input uncertainties dominate and propagate through the cascade chain. The current NWPs fall short of representing the spatial precipitation variability on such a comparatively small catchment, which indicates need to improve NWPs resolution and/or disaggregating techniques to narrow down the spatial gap between meteorology and hydrology. The spread of discharge forecasts varies from centre to centre, but it is generally large and implies a significant level of uncertainties. Nevertheless, the results show the TIGGE database is a promising tool to forecast flood inundation, comparable with that driven by raingauge observation. Copyright © 2009 Royal Meteorological Society

KEY WORDS flood forecast; Upper Severn; LISFLOOD-RR; LISFLOOD-FP; TIGGE

Received 13 October 2008; Revised 8 January 2009; Accepted 9 January 2009

1. Introduction

As the severity of floods increases, possibly due to climate and landuse changes (Reynard *et al*., 2004), there is an urgent need for more effective and reliable warning systems. The incorporation of numerical weather prediction (NWP) into a flood warning system can increase forecast lead times from a few hours to a few days. However, a single NWP forecast is not sufficient to quantify the inherent uncertainties in the forecasts.

An ensemble of weather forecasts from one Ensemble Prediction System (EPS), when used on catchment hydrology, can provide improved early flood warning as some of the uncertainties can be quantified (Cloke and Pappenberger, 2008). Recent research has been conducted in real-time flood forecasting using such EPSs in conjunction with spatially distributed catchment models. De Roo *et al*. (2003) developed a European-scale flood forecasting system (EFFS) by coupling the European Centre for Medium-Range Weather Forecasts (ECWMF) EPS forecasts/analysis with the LISFLOOD suite of models, namely the LISFLOOD-RR (5 km grid resolution) and LISFLOOD-FP (50 m grid resolution) (the full cascade is presented by Pappenberger *et al*. (2005)). The results showed most ensemble members tended to underestimate the river flow but the overall performance was encouraging. Bartholmes and Todini (2005) applied the TOPKAPI model at 1 km grid resolution to the River Po (catchment size *ca.* 37 000 km^2) using two European limited area models and the ECMWF EPS. The quantitative precipitation forecasts were reported as unreliable and the forecast discharge was negatively biased in both time and magnitude. Cluckie *et al*. (2006) employed a distributed rainfall-runoff model (GBDM) (500 m grid resolution) driven by the ECMWF EPS which was first resolved to a 2 km grid resolution using a dynamical downscaling technique, then spatially corrected and finally temporally adjusted. The study ascertained the potential values of using EPS in flood forecasting. This study also revealed a common underestimation of precipitation at finer spatial

* Correspondence to: Y. He, Department of Geography, King's College London, Strand, WC2R 2LS, London, UK. E-mail: yi.he@kcl.ac.uk

scales which resulted in biased (low) river flow forecasts. Considerable uncertainties were propagated from weather domains to catchment domains, which lead to a wide spread of river flow ensembles and resultant flood inundation extents.

EPS forecasts from a single weather centre only account for part of the uncertainties originating from initial conditions and stochastic physics (Roulin, 2006). Other sources of uncertainties, including numerical implementations and/or data assimilation, can only be assessed if a grand ensemble (GE) of EPS from different weather centres are combined (Goswami *et al.*, 2007). When various models that produce EPS from different weather centres are aggregated, the probabilistic nature of the ensemble precipitation forecasts can be better retained and accounted for. The availability of 12 global EPSs (a combined total of 232 weather forecasts) through the 'THORPEX Interactive Grand Global Ensemble' (TIGGE) (Shapiro and Thorpe, 2004; Park *et al.*, 2007) offers a new opportunity for the design of an improved probabilistic flood forecasting framework. A prototype of such a framework has been successfully demonstrated in Pappenberger *et al.* (2008a) using the Lisflood-RR model in the European Flood Alert System (EFAS). GE forecasts from seven weather centres were incorporated to hindcast the October 2007 flood event that took place in the Danube basin in Romania.

This paper presents a case study using the TIGGE database for flood warning on a meso-scale catchment. The upper reach of River Severn catchment was selected due to the large quantity of observational data available and its relatively small size (4062 km^2) (compared to the resolution of the NWPs). This choice was deliberate as the hypothesis is that the uncertainty in the forcing of smaller catchments cannot be represented by a single EPS with a very limited number of ensemble members, but only through the variance given by a large number of ensembles. A coupled atmospheric-hydrologic-hydraulic cascade system driven by the TIGGE ensemble forecasts was set up to study the potential benefits of using the TIGGE database in early flood warning. Physically based and fully distributed LISFLOOD-RR and LISFLOOD-FP models were selected to simulate discharge and flood inundation consecutively. The study was embedded in an uncertainty cascade recognising the uncertainties in the numerical weather predictions as well as the hydrological and hydraulic component (similar to Beven *et al.* (2005) and Pappenberger *et al.* (2005)). A probabilistic discharge and flood inundation forecast was provided as the end product of the cascade system.

2. The study area and data

The River Severn collects water from the Welsh mountainous area and its tributaries after passing through many steep sided valleys in Wales. It then flows eastwards and contributes significantly high levels of discharge to Montford (Figure 1). Between Montford and Buildwas, the

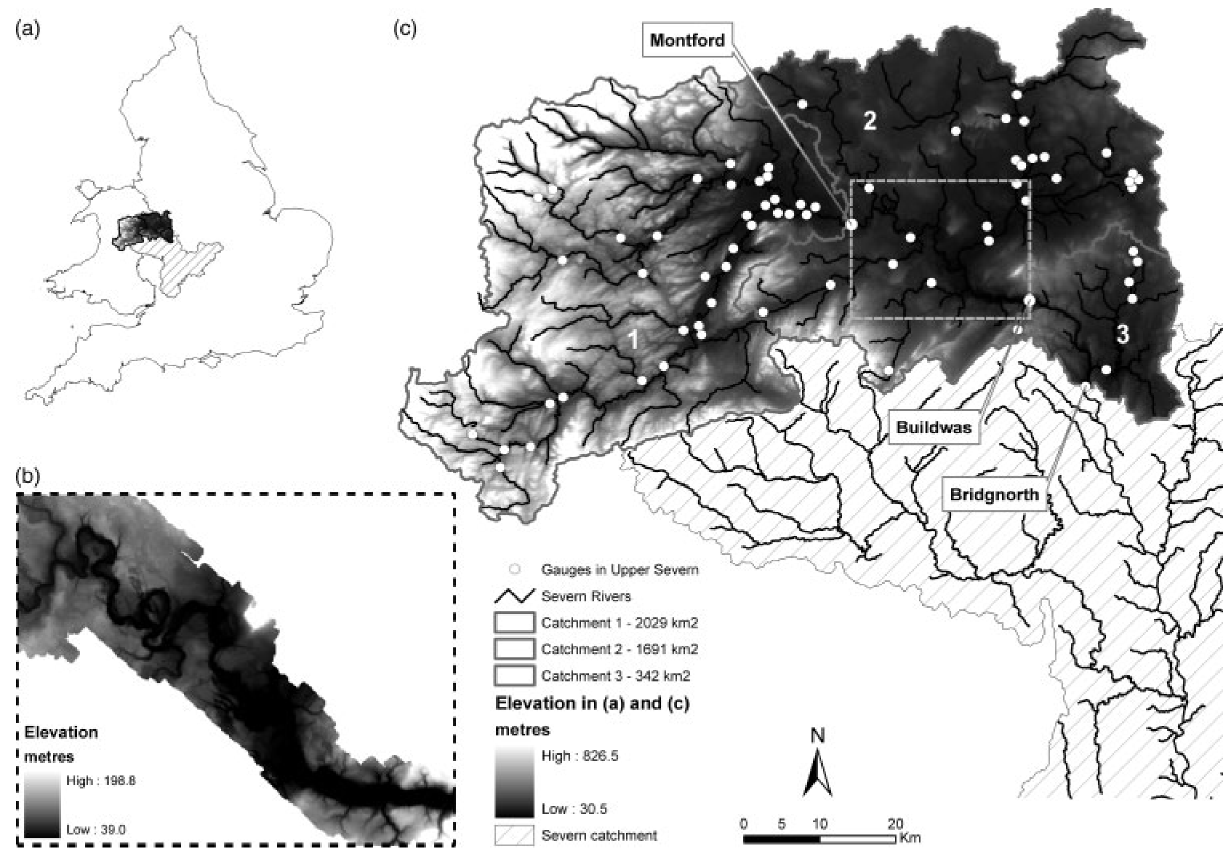

Figure 1. The Upper Severn catchment. (a) Severen Catchment located in the Midlands region of England; (b) Montford-Buildwas river section; (c) Digital Elevation Model (DEM) of the Upper Severn Catchment.

river channel meanders through a low-lying flood plain. Elevation on both sides of Buildwas rises above the flood plain forcing the river to cut through and form a gorge behind Buildwas. The river diverts its flow southwards just before passing through Bridgnorth. The drainage area up until Bridgnorth is referred to as the Upper Severn catchment in this study. It is approximately 4062 km^2 with urban or sealed areas, forest and agricultural land accounting for 3, 7.1 and 48.5% of the area respectively. Loosely packed peat soil dominates the Upper Severn catchment. Flow levels are generally high in autumn and winter and low in summer.

While the discharge simulation was performed for the entire Upper Severn catchment, the study of flood inundation focused on the river section between Montford Bridge (UK Ordnance Survey grid reference SJ412144) and Buildwas (SJ644044). The Shropshire town of Shrewsbury is situated within the incised meander loop of this reach and vulnerable to floods.

The digital elevation model of the Upper Severn catchment was obtained from the NEXTMap Britain dataset through the UK NERC Earth Observation Data Centre. The observed discharge data were provided by the UK Environment Agency (EA) Midlands region. Observed precipitation and temperature station data were provided from the UK Meteorological Office (2008). When the study was conducted, EPS data were available from seven centres in the TIGGE database with the majority delivering EPS from October 2007 onwards. The flood event took place in January 2008 was hence selected as the first high flow event in the study area. Table I lists the seven weather centres and their numbers of ensemble forecasts. Each centre provides one 'central' unperturbed analysis generated by a data-assimilation procedure and a number of forecasts with perturbed initial conditions. All forecast members were assigned equal weights (recommended by Park *et al.*, 2007) in this study. Along with the forecast precipitation, observed temperature of January 2008 was used as driving inputs to simulate discharge. The latter is due to (1) the flood event that took place in January 2008, a precipitation driven event with temperature playing a minor role, and (2) the study focused on uncertainties propagated from the forecast precipitation to the catchment models.

3. Methodology

3.1. Discharge simulation

The LISFLOOD-RR model (LFRR) has been used within European Flood Alert System (EFAS) and thus selected for this study. LFRR is a fully distributed and semi-physically based rainfall-runoff model. The model is described in detail by De Roo (1999) and van der Knijff and De Roo (2008). The model was set up at 1 × 1 km resolution for the Upper Severn catchment. Uncertainty analysis of the LFRR model has been presented by Pappenberger and Beven (2004), Pappenberger *et al.* (2004) and Mo *et al.* (2006). In this study 1000 Sobol quasi-random sets (Sobol, 1967) of five model parameters were generated within *a priori* parameter space recommended by Feyen *et al.* (2007). The Nash-Sutcliffe efficiency (Nash and Sutcliffe, 1970) was used as a performance indicator with emphasis given to the peak flow:

$$R^2 = 1 - \frac{\sum_{t=1}^{T} Q_t(\hat{Q}_t - Q_t)^2}{\sum_{t=1}^{T} Q_t(Q_t - \overline{Q})^2} \quad (1)$$

where t is the time step ($t = 1, \ldots, T$), Q_t and $\hat{Q}_t$ are the observed and simulated discharge at time step t respectively, and $\overline{Q}$ is the mean observed discharge. Due to the complex interaction among parameters, several sets of parameters can lead to equally good model performance at various temporal or spatial resolutions and in different time periods. To take into consideration the phenomenon also know as equifinality (Beven, 1993; Freer *et al.*, 1996) and to account for parameter uncertainties, the best six sets of parameters were selected and combined with 216 forecast members to form an ensemble of simulated discharges. In other words, 6 × 216 = 1296 forecast discharges were obtained at each time step. An implicit assumption was made that the six behavioural parameter sets selected from the 1000 Sobol simulations could represent the best possible model configuration whilst disregarding model sensitivity to different spatial resolution.

Table I. List of the meteorological forecast centres used in the case study.

Country/Region	Weather centre	Centre Abbriviation	Centre Code	Ensemble members
Australia	Bureau of Meteorology	BMRC	AMMC	32 + 1
Canada	Meteorological Service of Canada	MSC	BABJ	14 + 1
China	China Meteorological Administration	CMA	CWAO	20 + 1
Europe	European Centre for Medium-Range Weather Forecasts	ECWMF	ECMF	50 + 1
Japan	Japan Meteorological Agency	JMA	RJTD	50 + 1
UK	UK Met Office	UKMO	EGRR	23 + 1
USA	National Centres for Environmental Prediction	NCEP	KWBC	20 + 1

DOI: 10.1002/met

3.2. Precipitation map correction

The LFRR was originally driven by the UK Met Office raingauge data interpolated to a 1×1 km grid, while the TIGGE precipitation data was disaggregated from approximately $100 \times 100 - 1 \times 1$ km. There remain concerns with respect to negative impacts on the forecast discharges using the TIGGE data, due to (1) LFRR not being adjusted to respond to inputs at a different spatial resolution, and (2) high elevation areas that contribute to high level of discharge not being reflected in the TIGGE data due to its rather coarse spatial resolution. It was therefore necessary to apply a spatial correction to the input maps before they were run through the LFRR. The corrected precipitation was calculated on a grid-by-grid basis in comparison with observed precipitation with the aim of retaining the mass balance over the catchment area:

$$\widehat{p}_{ij}^{f} = \frac{p_{ij}^{f} c_{ij}^{f}}{\overline{C}}, \; where \quad c_{ij}^{f} = \frac{\overline{P}^{f} \overline{p}_{ij}^{o}}{\overline{P}^{o} \overline{p}_{ij}^{f}} \tag{2}$$

where c is the correction factor and p is the precipitation at location i and j, either forecasted (f) or observed (o), and capitals denote mean values over the entire catchment. The correction factors were calculated with observed and forecasted values for the preceding month (December 2007 in this study). The corrected patterns inherit some of the forecast pattern but most of the spatial variability still originates from the observed precipitation field (Figure 2).

3.3. Precipitation map evaluation

It is important that the forcing of the model cascade is evaluated in a hydrologically meaningful way (Cloke and Pappenberger, 2008; Pappenberger *et al*., 2008b). In this case study the input precipitation fields were evaluated on a monthly basis using continuous ranked probability scores (CRPS) (Hersbach, 2000). CRPS is a verification tool that compares how the distribution of an ensemble of forecasts to the observed value. It is sensitive to bias in the forecast values as well as variability. Probability scores are commonly used to evaluate forecasts (Toth *et al*., 2003):

$$\mathrm{CRPS} = \frac{1}{N} \sum_{n=1}^{N} \int_{-\infty}^{\infty} [F(x) - H(x - x_o)]^2 \mathrm{d}x \tag{3}$$

where N is the number of forecasts, x is the forecast variable, x_o is the observed variable, $F(x)$ is the cumulative distribution function (c.d.f.) $F(x) = \int_{-\infty}^{x} \rho(y)\mathrm{d}y$ and $H(x - x_0)$ is the Heaveside function, which has the value 0 when $(x - x_o) < 0$ and 1 otherwise. In order to quantify the skill of the probability score, the skill score was calculated as:

$$\mathrm{SS_{CRPS}} = 1 - \frac{\mathrm{CRPS}_F}{\mathrm{CRPS}_R} \tag{4}$$

where CRPS_F denotes the forecast score and CRPS_R is the score of a reference forecast of the same predictand. A skill score close to unity means a successful simulation. A negative score indicates a worse performance than the reference forecast. The reference forecast in this study was taken as the mean precipitation on the calendar day over the years 1951–2000.

3.4. Forecast procedure and performance evaluation

The forecast discharges were evaluated using a contingency table, where the observed flood events were compared with the forecasted. The possible outcomes in a contingency table (von Storch and Zwiers, 1999) are: (1) hit (H), the observed flood is correctly forecast, (2) miss (M), the observed flood is not forecast, (3) false alarm (FA), a flood event is wrongly forecast and (4) correct negative (CN), a non-event. The relative operating characteristic (ROC) is computed from the contingency table by plotting H against FA for different percentiles of the observed discharge. ROC is a measure that demonstrates forecast ability to separate events from non-events. A hit table was also used, showing H as a percentage of the total number of forecasts over a certain flood alert level for each weather centre. A flow rate above a specific flood alert level is labelled as a

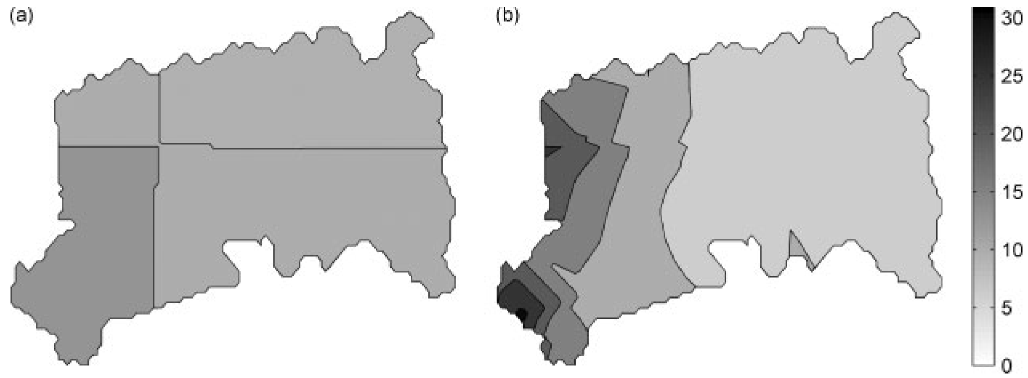

Figure 2. Precipitation (mm) forecast at 12 : 00 UTC of January 16 by ECMWF for ensemble member 1 with a lead time of 132 h (a) precipitation before correction, (b) precipitation after correction.

flood event. The flood alert level uses the UK EA recommended warning level converted to flow rate using the corresponding rating curve (source: West Gauge Board Data, UK EA). A water level of 6.1, 6.9, 7.1 and 7.3 m above gaugeboard datum at Montford are regarded as 'Flood Warning', 'Severe Flood Warning', 'Severe Warning Update' and 'Breaking point' respectively. The first flood alert level - 'Flood Warning' corresponding to a discharge of 331 m^3 s^{-1} was used in this study.

The forecast procedure was tested in an iterative way such that the disaggregated precipitation input maps could display satisfactory spatial patterns and produce comparable discharge level as that driven by raingauge observation. Once the ensemble of forecast discharges was obtained, a hydraulic model was coupled to the end of the cascade to forecast flood water depth and inundation extent. A schematic flowchart of the test procedure is shown in Figure 3. The time used in this paper is UTC.

3.5. Flood inundation

The LISFLOOD-FP model (LFFP) has been selected for simulating flood inundation both in the channel and on the flood plain. LFFP uses the Saint-Venant equations of open channel flow within the channel, and a 2D raster model based on the Manning equation for the floodplain. A detailed description of the model can be found in Bates and De Roo (2000) and Hunter *et al.* (2007). An uncertainty analysis of the LFFP model have been presented in Pappenberger *et al.* (2006) and Pappenberger *et al.* (2007).

4. Results

4.1. Precipitation skill scores

All EPSs show decreasing skills with increasing lead times for the forecast precipitation (Figure 4). Both the overall performance and deterioration vary depending on the month of the year and seem to be model independent. Dry months had higher skill scores than wet months. September 2007 was a relatively dry month compared to January 2008 for the study area, consequently, SS_{CRPS} for September 2007 shows higher skills, even with increasing lead times.

The GE consists of all the seven EPSs on a given day. The advantage of using GE over an EPS from one centre is that GE is more stable and overall also performs well (Figure 4). Some EPSs affected GE negatively and it could be argued that the ensemble components in the GE ought to be weighted according to performance rather than lumped together with equal weights.

4.2. Hydrograph ensemble statistics

The LFRR model was run with the 1000 Sobol quasi-random parameter sets for the time period between January 2006 and January 2008 with 6-h time step. The best and worst Nash-Sutcliffe efficiency at Montford of the six selected behavioural parameter sets are 0.8841 and 0.8744 respectively. A total number of six behavioural LFRR parameter sets along with 216 TIGGE forecast members were run for the entire January 2008. An example is shown in Figure 5 with the mean precipitation forecasts issued by ECMF on 18 January, 2008 over the upstream area of Montford and the resulting ensemble

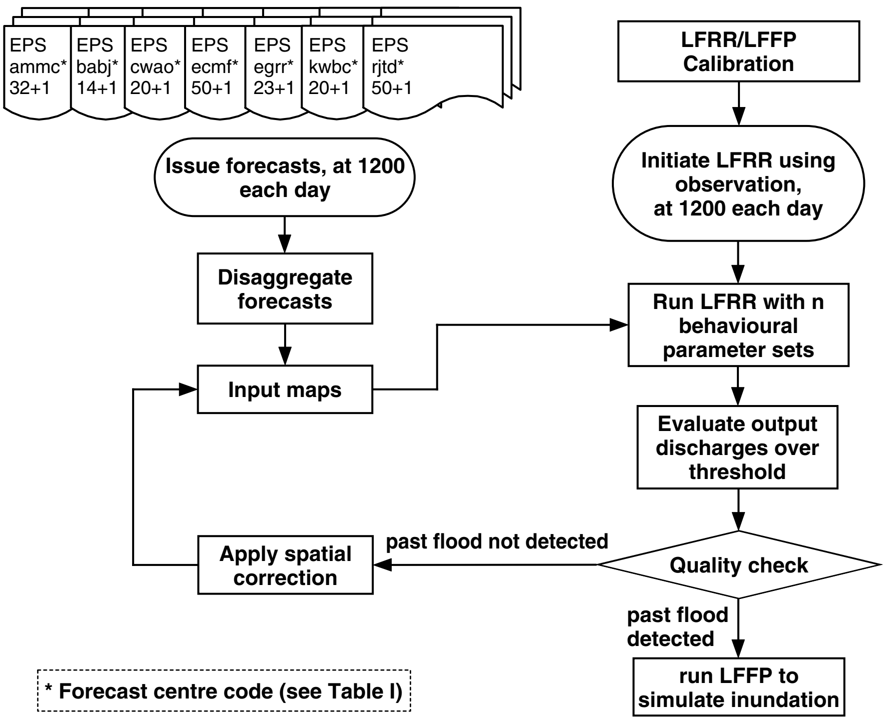

Figure 3. Schematic flowchart showing the GE flood forecast procedure in a test mode.

DOI: 10.1002/met

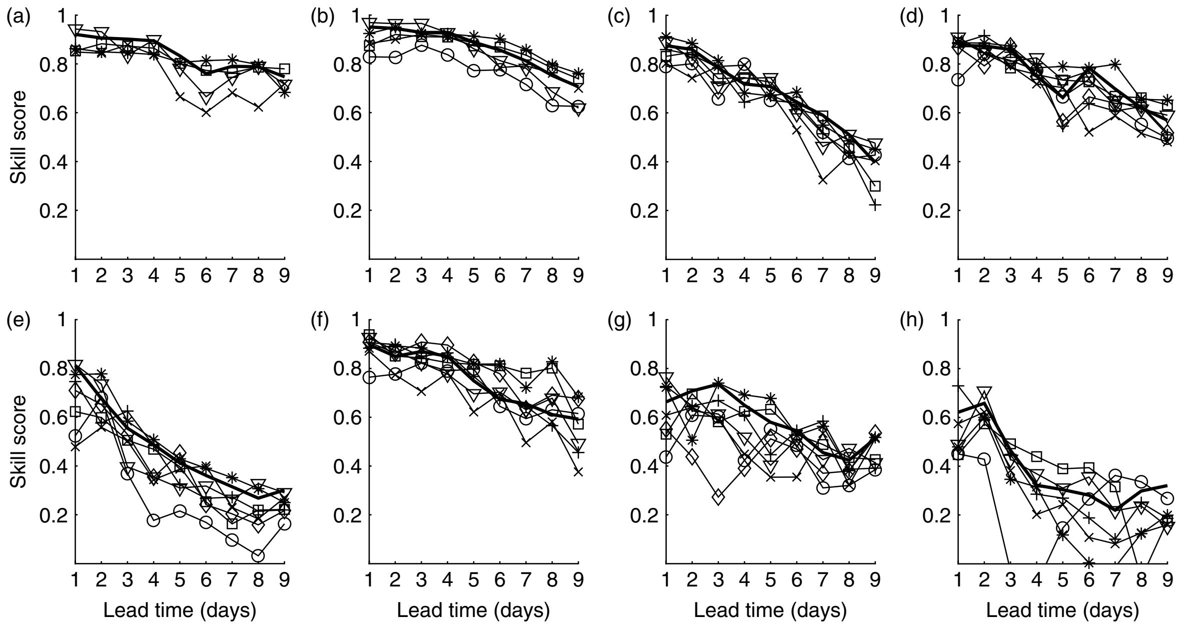

Figure 4. SS_{CRPS} of monthly average for the seven EPSs over the study area. The Figures (a)–(h) are for the months from September 2007 to April 2008. The symbols denote: AMMC (circles), BABJ (crosses), CWAO (plus signs), ECMF (stars), EGRR (boxes), KWBC (diamonds) and RJTD (triangles). GE is the bold line.

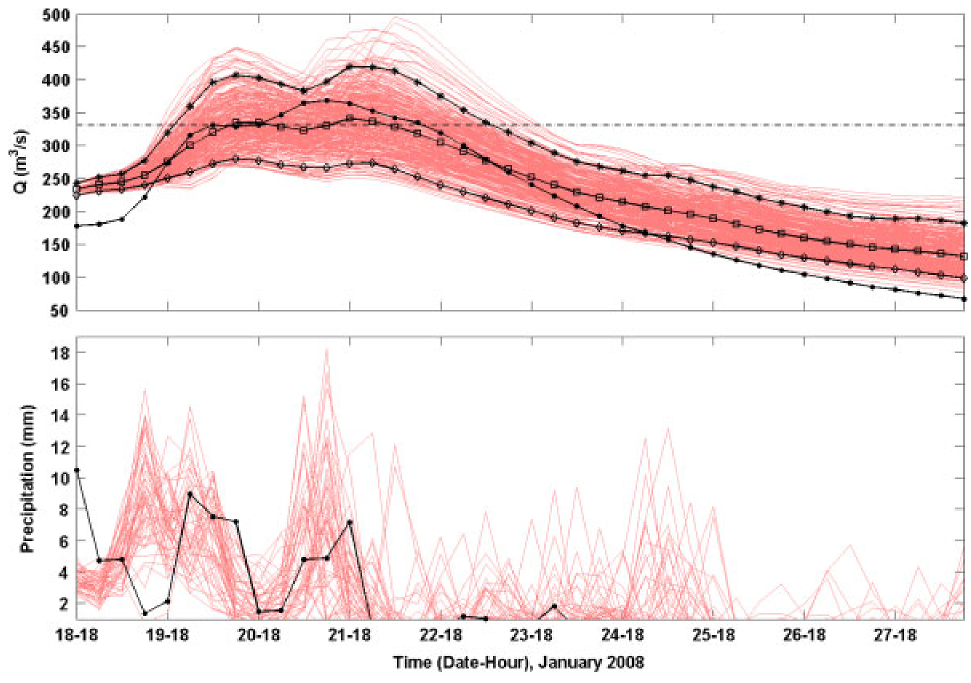

Figure 5. Ensemble precipitation forecasts issued on 18 January, 2008 by 'ECMF' (upper) and ensemble forecast discharges (lower) in comparison with observation. The horizontal dashed line is the EA warning level. Lines marked with diamond, squares and stars represent the 5th, 50th, 95th percentile of the forecast discharges respectively. The lines marked with circles and the solid thin represent the observed and the forecast values respectively. This figure is available in colour online at www.interscience.wiley.com/ma

discharges at Montford. Three major clusters of peaks can be observed in the precipitation forecasts (Figure 5, Time: 19-18, 21-18 and 24-18). The first two possibly resemble the second and third peaks of the observed precipitation with a noticeable delay, which in turn resulted in the delay of the forecasted flow peaks. No precipitation was observed for the last half of this forecast period and the observed hydrograph transited from peak to recession (Figure 5, Time: 23-18 to the end), however, a third cluster is seen in the precipitation forecasts which possibly lead to an over-estimation on the recession segment of the ensemble hydrograph. Overall, the peak flow was well captured and encompassed in the ensemble hydrograph.

The ensemble spread is measured by standard deviation normalized by the corresponding mean value (Figure 6).

DOI: 10.1002/met

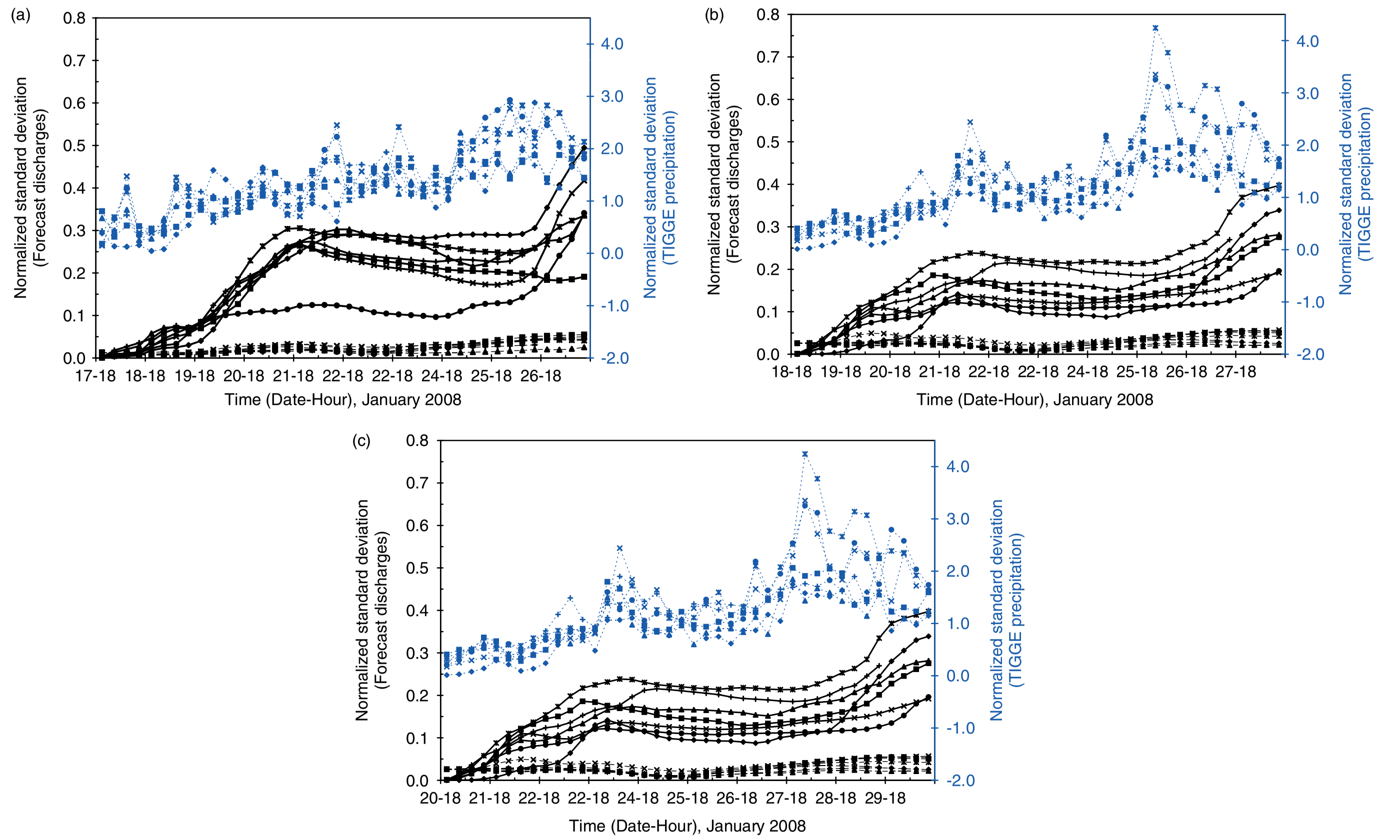

Figure 6. Ensemble statistics associated with forecasts issued on (a) 17, (b) 18 and (c) 20 January, 2008 with 10-day lead time except 'RJTD' with 9-day lead time. Y-axis on the left is for the spread of discharges. Y-axis on the right is for the spread of forecast precipitation. The markers denote: AMMC (diamonds), BABJ (squares), CWAO (triangles), ECMF (crosses), EGRR (stars), KWBC (circles) and RJTD (plus signs). The dotted, solid, dashed lines represent the spread of the TIGGE forecast precipitation, spread of discharges reflected in the forecast members and spread of discharges reflected in the parameter sets respectively. This figure is available in colour online at www.interscience.wiley.com/ma

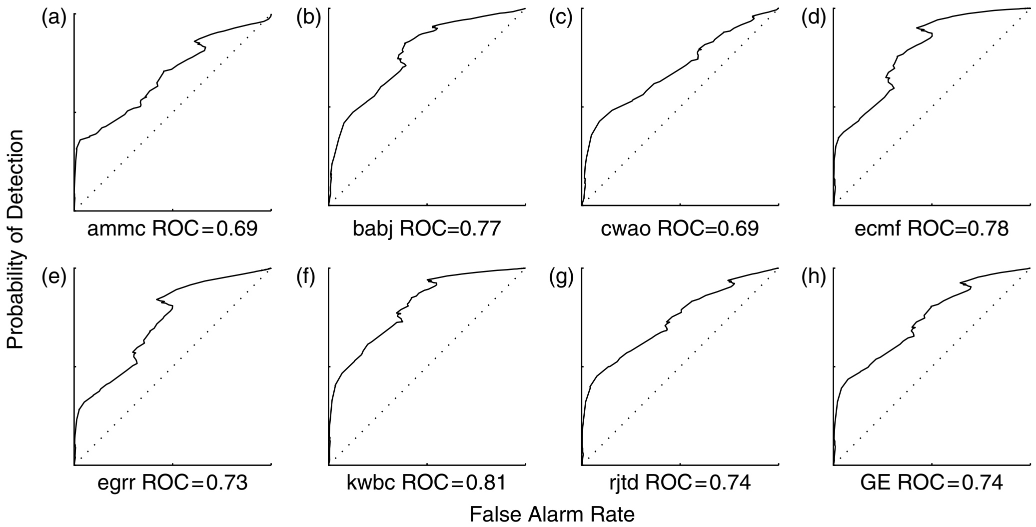

Figure 7. Relative operating characteristic for the seven weather centres and the GE for the January 2008 flood event. A value of 0.5 denotes no skill.

The three dates show similar behaviour in that the spread of the precipitation forecasts (SPrec) is large than the spread of the hydrograph as a result of different forecast members (SQforc) which is again large than the spread of the hydrograph as a result of the use of six different parameter sets (SQpara). In general, SPrec and SQforc increase with the lead times, but SQpara is fairly stable along the lead times, which indicates precipitation input uncertainties dominate and propagate through the cascade chain and the level of uncertainties is dampened to a certain extent after running through the LFRR model. However, SQpara can be dominant within the first time steps. The LFRR model acts as a low-pass filter that removes the high-frequency noise in the system and gives out a less noisy signal.

The ROC diagram for the flood forecast shows that all individual centres have forecast skill, and the GE performs as the median (Figure 7). According to the ROC diagrams, the best model for the January 2008 flood event was KWBC. This is because of KWBC's low false alarm rate rather than its ability to correctly model the flood event.

The use of more than one forecast centre gives the forecaster more information at an earlier stage and potentially increases the ability to detect a flood (Figure 8). The flood is visible in some centres as early as 10 days ahead, although only on the 25%-level, but the signal is not persistent until 4 days ahead. It is difficult to issue a warning given the afore-mentioned information, but a forecaster could better prepare for the possibility of a flood at an early stage, and then issue the real warning if the signal strengthens over time. It is still difficult to identify the best forecast centre(s), but the advantage of using a multi-centre approach is the gained lead time. It should be noted that the warning level used in this study is the EA warning level determined from long-term observed river flows while most of the other studies have used threshold level obtained from proxy discharges simulated by rainfall-runoff models (e.g., Thielen *et al.*, 2008; Bartholmes *et al.*, 2008).

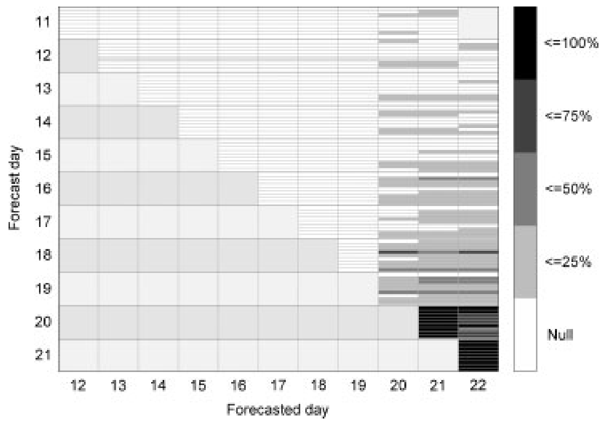

Figure 8. Percentages of ensemble discharges above the threshold at Montford for the January 2008 flood event. The horizontal bars represent the seven centres in the same order as in Table I. The bottom bar for each forecast is the GE.

4.3. Flood inundation

The LFFP model was set up at 50 m grid resolution for the Montford-Buildwas river section and calibrated against the flood event that took place on 30 October 2000 with a peak flow of 435 $m^3 s^{-1}$. The model calibration was performed in the non-adaptive time mode (Hunter *et al.*, 2006, Hunter *et al.*, 2007). The calibrated Manning coefficient values are 0.035 and 0.05 in the channel and floodplain respectively. The ensemble hydrograph at Montford were taken as input and routed downstream to Buildwas.

An example of the hydrograph and inundation ensembles is shown in Figure 9. The forecasts were issued on 17 January, 2008 at 1200 at a 6-hour time step for the next 10 days, with the exception of the 'RJTD' centre providing a 9-day lead time. The first peak was reached

DOI: 10.1002/met

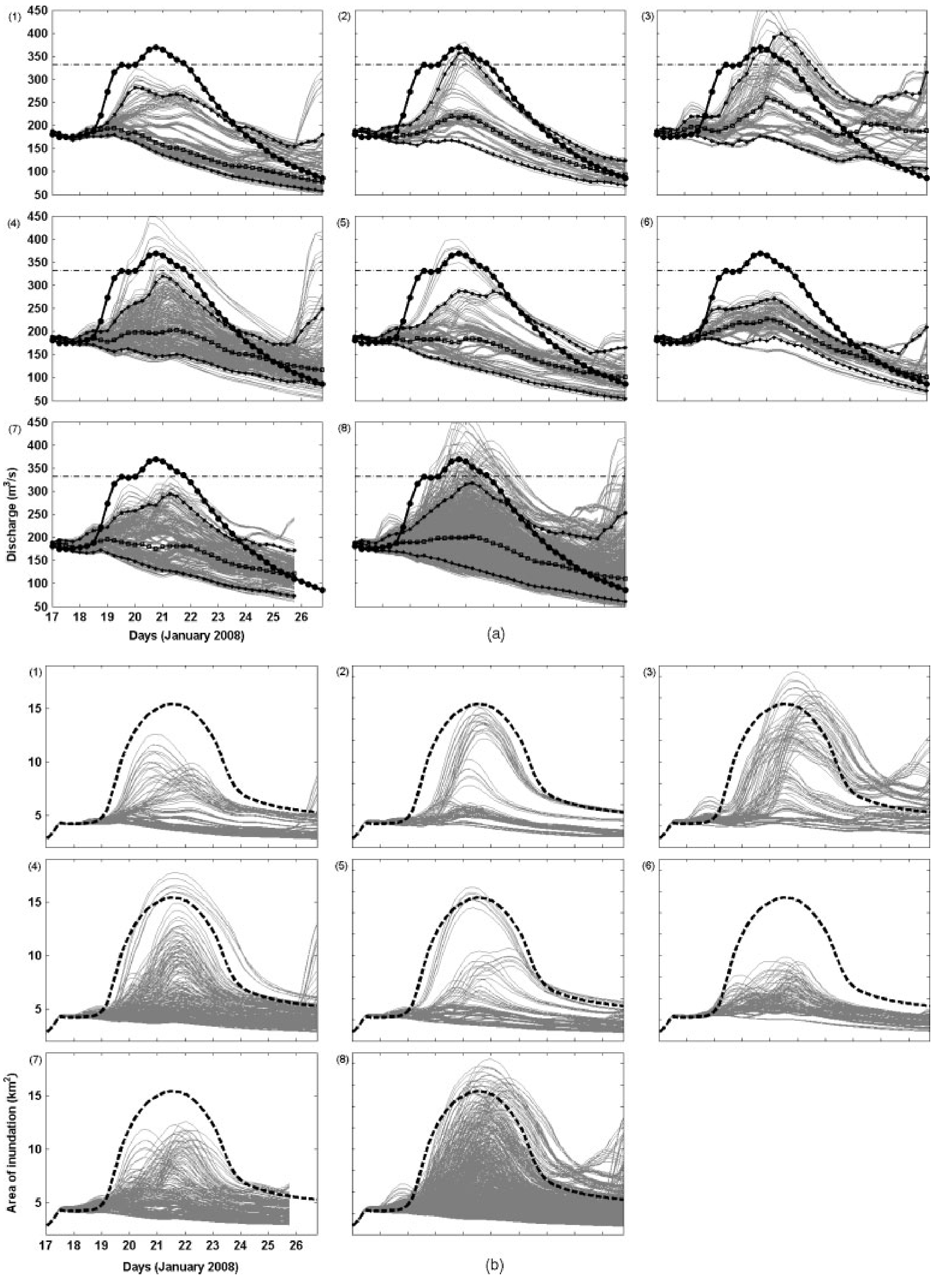

Figure 9. Comparison of (a) ensembles of forecast discharges simulated using EPS from seven weather centres and the six behavioural parameter sets and (b) ensembles of the forecast inundation areas simulated using the discharge ensembles shown in (a) for the Montford-Buildwas river section (simulations were driven by the TIGGE forecasts issued on 17 January, 2008 with a 10-day lead time except 'RJTD' with 9-day lead time only). (1) AMMC; (2) BABJ; (3) CWAO; (4) ECMF; (5) EGRR; (6) KWBC; (7) RJTD; (8) GE.

DOI: 10.1002/met

at around 0600 on the 20 January and followed by the second higher peak around 1200 on the 21 January. This means that the ensemble of forecast discharges has approximately 2.5-day lead time for the first peak and 4-day for the second peak. CPRS drops below 0.6 after the third day (lead time) for January 2008, as can be seen in Figure 4, possibly causing a worse performance propagated through the cascade and reflected as an overall under-estimation of the peak flows (Figure 9(a)). As a consequence, the inundated areas with 9/10-day lead time were largely under-estimated by almost all the seven EPSs (Figure 9(b)), except a number of forecast members from 'CWAO', 'ECMF' and 'EGRR'. With respect to the ensemble of forecast discharges, 'ECMF' and 'EGRR' lead to the best results for this specific set of forecasts and 'KWBC' the worst in terms of the timing and flow magnitude, although the SS_{CRPS} (Figure 4) shows 'AMMC' as the worst in terms of monthly average for January 2008. The spread of flood inundated areas is equivalent to that of forecast discharges, which implies that the uncertainties originated from the ensemble of precipitation input dampened by the LFRR model is being propagated though the LFFP model without much change.

5. Conclusion and outlook

The paper presents a case study using the TIGGE database for flood warning on a meso-scale catchment (4062 km^2) located in the Midlands region of England. A coupled atmospheric-hydrologic-hydraulic cascade system driven by the TIGGE ensemble forecasts is set up to study the potential benefits of using the TIGGE database in early flood warning. A probabilistic discharge and flood inundation forecast is provided as the end product of the cascade system. The results show the TIGGE database is a promising tool for producing forecasts of discharge and flood inundation comparable with the observed discharge and simulated inundation driven by the observed discharge.

The spread of discharge forecasts varies from centre to centre, but it is generally large, implying a significant level of uncertainties. Precipitation input uncertainties dominate and propagate through the cascade chain. The current NWPs fall short of representing the spatial variability of precipitation on a comparatively small catchment. This perhaps indicates the need to improve NWPs resolution and/or disaggregation techniques to narrow down the spatial gap between meteorology and hydrology. The simple correction of the input precipitation forecast employed in this study substantially improved the results. As the resolution of NWPs improves in the future, this will become less of an issue. Nevertheless, this opens opportunities for improved methods to spatially correct the forecasts and enhance forecast reliability in meso/small-scale catchments.

A total of six sets of parameters and 216 ensemble forecasts formed the basis of this uncertainty study. It is worth mentioning that the procedure to select the parameter sets is not rigorous enough to justify a sound configuration of a hydrological model. In the future, other methods such as the robust parameter estimation (ROPE) approach (Bárdossy and Singh, 2008) could be used to improve the model performance whilst reducing the parameter space search domain.

Only one flood event was studied in this paper because most of the TIGGE data became available after October 2007. It is necessary to test the TIGGE ensemble forecasts with other flood events in catchments with different hydrological and climatic regimes before general conclusions can be made on its robustness and applicability. Whilst it is important that a wider range of flood events be studied before a reliable assessment of the technique can be made, it is also important to study non-flood events where the system may forecast flooding (i.e. false alarms).

It is difficult to identify the best forecast centre(s), but in general the chance of detecting floods is increased by the incorporation of the TIGGE database. It is not necessarily true that early flood warning becomes more reliable when more ensemble forecasts are employed. There is no guarantee that models from different weather centres are absolutely uncorrelated. An ensemble approach does not mean errors can be compensated. If a reliable flood warning could be obtained with fewer forecasts, considerable computing time and storage space could be saved.

An atmospheric-hydrologic-hydraulic cascade probabilistic flood warning framework has been established in this study. More effort is required before the performance of the ensemble flood warning system could be labelled as 'satisfactory'. The study, however, provides very encouraging results and has hopefully ushered in a new era.

Acknowledgement

This work was funded by the UK Natural Environment Research Council (NERC), Flood Risk from Extreme Events (FREE) programme with the grant number NE/E002242/1. The TIGGE data were accessed through the European Centre for Medium-Range Weather Forecasts (ECWMF). The Joint Research Centre (JRC) of the European Commission and the School of Geographical Sciences of the University of Bristol are gratefully acknowledged for providing the LISFLOOD-RR and LISFLOOD-FP models respectively as well as relevant technical support. The authors will also like to thank Richard Cross (UK EA) for his valuable comments and inputs.

References

Bárdossy A, Singh S. 2008. Robust estimation of hydrological model parameters. *Hydrology and Earth System Sciences* **12**: 1273–1283.

Bartholmes J, Thielen J, Ramos M, Gentilini S. 2008. The European Flood Alert System EFAS–Part 2: statistical skill assessment of probabilistic and deterministic operational forecasts. *Hydrology and Earth System Sciences Discussions* **5**: 289–322.

Bartholmes J, Todini E. 2005. Coupling meteorological and hydrological models for flood forecasting. *Hydrology and Earth System Sciences* **9**(4): 333–346.

DOI: 10.1002/met

Bates PD, De Roo A. 2000. A simple raster-based model for flood inundation simulation. *Journal of Hydrology* **236**: 54–77.

Beven KJ. 1993. Prophecy, reality and uncertainty in distributed hydrological modelling. *Advances in Water Resources* **16**: 41–51.

Beven K, Romaniwicz R, Pappenberger F, Young P, Werner M. 2005. The uncertainty cascade in flood forecasting. In *Advances and Implementation of Flood Forecasting Technology*, Balbanis P, Lambroso D, Samuels P (eds). ACTIF: Tromso.

Cloke HL, Pappenberger F. 2008. Evaluating forecasts for extreme events for hydrological applications: an approach for screening unfamiliar performance measures. *Meteorological Applications* **15**(1): 181–197.

Cluckie ID, Xuan Y, Wang Y. 2006. Uncertainty analysis of hydrological ensemble forecasts in a distributed model utilizing short-range rainfall prediction. *Hydrology and Earth System Sciences Discussions* **3**: 3211–3237.

De Roo A. 1999. *LISFLOOD: A Rainfall-runoff Model for Large River Basins to Assess the Influence of Land use Changes on Flood Risk*, Ribamod: river basin modelling, management and flood mitigation. Concerted action, European Commission, EUR 18287, 349–357.

De Roo A, Gouweleeuw B, Thielen J, Bartholmes J, Bongioannini-Cerlini P, Tonidi E, Bates PD, Beven KJ, Pappenberger F, Heise E, Rivin G, Hils M, Hollingsworth A, Holst B, Kwadijk J, Reggiani P, van Dijk M, Sattler K, Sprokkereef E. 2003. Development of a European flood forecasting system. *International Journal of River Basin Management* **1**(1): 49–59.

Feyen L, Vrugt JA, Nualláin BO, van der Knijff J, De Roo A. 2007. Parameter optimisation and uncertainty assessment for large-scale streamflow simulation with the lisflood model. *Journal of Hydrology* **332**(3–4): 276–289.

Freer J, Beven K, Ambroise B. 1996. Bayesian estimation of uncertainty in runoff prediction and the value of data: an application of the GLUE approach. *Water Resources Research* **32**(7): 2161–2173.

Goswami M, OConnor K, Bhattarai K. 2007. Development of regionalisation procedures using a multi-model approach for flow simulation in an ungauged catchment. *Journal of Hydrology* **333**(2–4): 517–531.

Hersbach H. 2000. Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather and Forecasting* **15**(5): 559–570.

Hunter NM, Bates PD, Horritt MS, Wilson MD. 2006. Improved simulation of flood flows using storage cell models. *Proceedings of the Institution of Civil Engineers-Water Management* **159**: 9–18.

Hunter NM, Bates PD, Horritt MS, Wilson MD. 2007. Simple spatially-distributed models for predicting flood inundation: a review. *Geomorphology* **90**: 208–225, DOI:10.1016/j.geomorph.2006.10.021.

Mo XG, Pappenberger F, Beven K, Liu SX, De Roo A, Lin ZH. 2006. Parameter conditioning and prediction uncertainties of the LISFLOOD-WB distributed hydrological model. *Hydrological Sciences Journal-Journal Des Sciences Hydrologiques* **51**(1): 45–65.

Nash JE, Sutcliffe JV. 1970. River flow forecasting through conceptual models, Part I. A discussion of principles. *Journal of Hydrology* **10**: 282–290.

Pappenberger F, Beven K. 2004. Functional classification and evaluation of hydrographs based on multicomponent mapping. *International Journal of River Basin Management* **2**(2): 89–100.

Pappenberger F, Beven KJ, de Roo A, Thielen J, Gouweleeuw B. 2004. Uncertainty analysis of the rainfall runoff model LisFlood within the Generalized Likelihood Uncertainty Estimation (GLUE). *International Journal of River Basin Management* **2**(2): 123–133.

Pappenberger F, Beven KJ, Hunter N, Gouweleeuw B, Bates PD, de Roo A, Thielen J. 2005. 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). *Hydrology and Earth System Sciences* **9**(4): 381–393.

Pappenberger F, Beven KJ, Frodsham K, Romanovicz R, Matgen P. 2006. Grasping the unavoidable subjectivity in calibration of flood inundation models: a vulnerability weighted approach. *Journal of Hydrology* **333**(2–4): 275–287.

Pappenberger F, Bartholmes J, Thielen J, Cloke H, Buizza R, de Roo A. 2008a. New dimensions in early flood warning across the globe using grand-ensemble weather predictions. *Geophysical Research Letters* **35**: L10404, DOI:10.1029/2008GL033837.

Pappenberger F, Scipal K, Buizza R. 2008b. Hydrological aspects of meteorological verification. *Atmospheric Science Letters* **9**(2): 43–52.

Pappenberger F, Frodsham K, Beven K, Romanowicz R, Matgen P. 2007. Fuzzy set approach to calibrating distributed flood inundation models using remote sensing observations. *Hydrology and Earth System Sciences* **11**(2): 739–752.

Park YY, Buizza R, Leutbecher M. 2007. *TIGGE: Preliminary Results on Comparing and Combining Ensembles*, ECMWF TM 548, European Centre for Medium-Range Weather Forecasts (ECWMF): Reading.

Reynard NS, Prudhomme C, Crooks SM. 2004. The flood characteristics of large UK rivers: potential effects of changing climate and land use. *Climatic Change* **48**(2–3): 343–359.

Roulin E. 2006. Skill and relative economic value of medium-range hydro-logical ensemble predictions. *Hydrology and Earth System Sciences Discussions* **3**: 1369–1406.

Shapiro M, Thorpe A. 2004. *THORPEX International Science Plan*. WMO, WWRP, document.

Sobol IM. 1967. The distribution of points in a cube and the approximate evaluation of integrals. *USSR Computational Mathematics and Mathematical Physics* **7**(4): 86–112.

Thielen J, Bartholmes J, Ramos M, de Roo A. 2008. The European flood alert system–part 1: concept and development. *Hydrology and Earth System Sciences Discussions* **5**: 257–287.

Toth Z, Talagrand O, Candille G, Zhu Y. 2003. Probability and ensemble forecasts. In *Environmental Forecast Verification: A Practitioners Guide in Atmospheric Science*, Jolliffe IT, Stephenson DB (eds). Wiley: Chichester, UK, pp 137–164.

UK Meteorological Office. 2008. *MIDAS Land Surface Stations data (1853-current)*, British Atmospheric Data Centre, 2006, Available from http://badc.nerc.ac.uk/data/ukmo-midas, Accessed 2008-09-01.

van der Knijff J, De Roo A. 2008. *LISFLOOD, Distributed Water Balance and Flood Simulation Model*, Revised User Manual. Vol. EUR 22166 EN/2. Scientific and Technical Reports, Joint Research Centre, European Commission: Ispra (Va), Italy.

von Storch H, Zwiers F. 1999. *Statistical Analysis in Climate Research*, Cambridge University Press: New York.