University of Reading

New dimensions in early flood warning across the globe using grand-ensemble weather predictions

Article

Published Version

Pappenberger, F., Bartholmes, J., Thielen, J., Cloke, H. L. ORCID: https://orcid.org/0000-0002-1472-868X, Buizza, R. and de Roo, A. (2008) New dimensions in early flood warning across the globe using grand-ensemble weather predictions. Geophysical Research Letters, 35. L10404. ISSN 1944–8007 doi: 10.1029/2008GL033837 Available at https://centaur.reading.ac.uk/29719/

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

To link to this article DOI: http://dx.doi.org/10.1029/2008GL033837

Publisher: American Geophysical Union

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the End User Agreement.

www.reading.ac.uk/centaur

University of Reading

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

New dimensions in early flood warning across the globe using grand-ensemble weather predictions

Florian Pappenberger,[1] Jens Bartholmes,[2] Jutta Thielen,[2] Hannah L. Cloke,[3] Roberto Buizza,[1] and Ad de Roo[2]

Received 5 March 2008; revised 9 April 2008; accepted 11 April 2008; published 24 May 2008.

[1] Early and effective flood warning is essential to initiate timely measures to reduce loss of life and economic damage. The availability of several global ensemble weather prediction systems through the "THORPEX Interactive Grand Global Ensemble" (TIGGE) archive provides an opportunity to explore new dimensions in early flood forecasting and warning. TIGGE data has been used as meteorological input to the European Flood Alert System (EFAS) for a case study of a flood event in Romania in October 2007. Results illustrate that awareness for this case of flooding could have been raised as early as 8 days before the event and how the subsequent forecasts provide increasing insight into the range of possible flood conditions. This first assessment of one flood event illustrates the potential value of the TIGGE archive and the grand-ensembles approach to raise preparedness and thus to reduce the socio-economic impact of floods. **Citation:** Pappenberger, F., J. Bartholmes, J. Thielen, H. L. Cloke, R. Buizza, and A. de Roo (2008), New dimensions in early flood warning across the globe using grand-ensemble weather predictions, *Geophys. Res. Lett.*, *35*, L10404, doi:10.1029/2008GL033837.

1. Introduction

[2] A major research challenge of the 21st century is to provide early warning for floods with potentially disastrous consequences. Floods killed 8349 people, affected 164 million and caused damage in the excess of 21 billion US$ in 2007 (EM-DAT, The OFDA/CRED International Disaster Database, 2007, available at http://www.em-dat.net, Université Catholique de Louvain-Brussels-Belgium). Early warning provides civil protection authorities and the public with vital preparation time and can reduce the socio-economic impacts of flooding, for example by opening polders ahead of time. In addition, early warning of flood disasters can enable the preparation of international assistance actions.

[3] Although the incorporation of numerical weather predictions (NWP) into a flood warning system can significantly increase forecast lead time [*Ahrens and Jaun*, 2007; *Gourley and Vieux*, 2005; *Krzysztofowicz*, 2002; *Pappenberger et al.*, 2005; *Verbunt et al.*, 2006], many hydrological services do not include them as they introduce a non-predictable degree of uncertainty into their forecasts and thus into their decision making process [*Demeritt et al.*, 2007]. In contrast, forecasts from Ensemble Prediction Systems (EPS) improve upon single deterministic forecasts as they quantify some of these uncertainties in the production of multiple weather forecasts for the same period, and, used with a hydrological model, have the potential to provide valuable early flood warning [*Roulin*, 2007]. Recently there has been a move to integrate EPS into flood forecasting systems around the world, for example: the Georgia-Tech/Bangladesh project (T. Hopson and P. Webster, Three tier flood and precipitation forecasting scheme for South-East Asia, 2008, available at http://cfab2.eas.gatech.edu/); the Finnish Hydrological Service (B. Vehvilainen and M. Huttunen, Hydrological forecasting and real time monitoring in Finland: The watershed simulation and forecasting system, 2002, available at http://www.environment.fi/waterforecast); the Swedish Hydro-Meteorological Service [*Olsson and Lindström*, 2007]; the Joint Research Centre (JRC) of the European Commission (EC) with the European Flood Alert System (EFAS) [*Thielen et al.*, 2008]; and the Advanced Hydrologic Prediction Services of NOAA [*Mcenery et al.*, 2005].

[4] EPS forecasts from a single forecast centre only address some of the uncertainties inherent in NWP, namely the initial conditions and stochastic physics [*Roulin*, 2007]. By contrast, a grand-ensemble, incorporating the EPS forecasts from multiple forecast centres can improve the simulation of uncertainties arising from numerical implementations and/or data assimilation [*Goswami et al.*, 2007]. This can have a substantial impact in generating the full spectrum of possible solutions, especially in the tails of the distributions where the extreme rainfall events that cause flooding are captured. Therefore, multi-model systems can potentially better represent the true probability distribution of predictions, as seen in other fields such as climate change [*Hagedorn et al.*, 2005].

2. The TIGGE Grand-Ensemble and the European Flood Alert System

[5] In October 2007 the full set of 7 global ensembles in the THORPEX Interactive Grand Global Ensemble (TIGGE) data archive [*Richardson*, 2005; *Park et al.*, 2008] became available. In this paper we demonstrate how grand-ensembles can be used for early flood warning with the example of the first major flooding event since the TIGGE archive became operational; namely the October 2007 floods on several tributaries to the Danube in Romania (Siret, Jiu, Olt and Arges). Table 1 shows the ensemble forecasts used in this study from which we form a TIGGE

[1]European Centre for Medium-Range Weather Forecasts, Reading, UK.
[2]Institute for Environment and Sustainability, DG Joint Research Centre of the European Commission, Ispra, Italy.
[3]Department of Geography, King's College London, London, UK.

Copyright 2008 by the American Geophysical Union.
0094-8276/08/2008GL033837

Table 1. Meteorological Forecast Centres and the Data Used in This Study[a]

Centre	Country/Domain	Ensemble Members	Horizontal Resolution	Vertical Levels	Forecast Length
Bureau of Meteorology	Australia	33	TL119	19	10
China Meteorological Administration	China	15	T213	31	10
National Centre for Environmental Predictions	United States	21	T126	28	16
UK MetOffice	United Kingdom	24	1.25 × 0.83 deg	38	15
Canadian Meteorological Centre	Canada	21	T254 (up to 3.5 days) then T170	64 (up to 3.5 days) then 42	16
Japan Meteorological Agency	Japan	51	TL159	40	9
European Centre for Medium-Range Weather Forecasts	Europe	51	TL399 (up to day 10)	62	15

[a]For the hydrological forecasts only the first 10 days of lead time were used.

grand-ensemble forecast (216 forecast members) by merging the single forecasts with equal weights.

[6] We use TIGGE weather forecasts within the European Flood Alert System (EFAS) [*de Roo et al.*, 2003], a flood prediction system developed by the EC which aims to increase preparedness in trans-national European river basins [*Bartholmes et al.*, 2008; *Thielen et al.*, 2008] by providing early flood warning information on a catchment scale. EFAS uses the hydrological model LISFLOOD [*van der Knijff and de Roo*, 2007] applied on a 5 km grid across Europe, and provides local water authorities with probabilistic flood forecasting information up to 10 days in advance based on 4 warning thresholds (low, medium, high and severe) (as defined by *Bartholmes et al.* [2008] and *Kalas et al.* [2008]).

[7] We make several assumptions in this analysis: all ensemble members are equally likely, ensemble size is irrelevant and NWP are the dominant source of uncertainty. Also, our analysis concentrates on the combined forecast performance of the TIGGE grand-ensemble and not on the forecasts of the single EPS.

3. Methodology

3.1. Rainfall

[8] Comparisons between observed and modelled were made with Root Mean Squared Error (RMS error) between the ensemble mean and the observations; and the Rank Probability Score (RPS) and Brier Score. The threshold categories for the rainfall score are derived from the observed distribution in 10th percentile steps. RMS error of ensemble mean and RPS results were ranked for each lead time and averaged. The observed rainfall fields are calculated by interpolating synoptic station data extracted from the meteorological database of the JRC Institute for the Security and Protection of the Citizen [http://www.agrifish.jrc.it] onto the EFAS grid.

3.2. River Discharge

[9] The thresholds of the river discharge score for RPS (and Brier score) are given by the EFAS warning thresholds. Since spatially distributed observed discharge data are not available, discharge observations are proxies, derived from routing observed rainfall through the hydrological model. As the same model is used for the observations and the predictions, this allows us to control for a number of other uncertainties, which will be examined further in future research.

4. Rainfall Forecasts: Performance of the Grand-Ensemble Versus Single EPS

[10] Examples of the probability forecasts of rainfall (in excess of 25 mm over a 48 hour period) per grid cell can be seen in Figure 1. There is a high spatial variability in the rainfall prediction between the single EPS forecasts, and the spatial distribution of the grand-ensemble is closest to observations. The Brier score for rainfall exceeding the median observed rainfall in October in each EFAS grid cell over the whole of Europe has also been analysed: the average score drops from a Brier Score of 0.8 at a lead time of 1 day to 0.55 at a lead time of 10 days. All ensembles perform similarly, with TIGGE in the top 3 performances (all of which were nearly identical). In the Romanian case study all ensemble forecasts overpredict high rainfalls (above 80th percentile of observed) at lead times up to 4 days [*Pappenberger et al.*, 2008]. By contrast, at longer lead times, 4 EPS and the TIGGE forecasts underpredict high rainfalls. However, the RMS error of TIGGE forecasts is superior to 5 single EPS forecasts especially at long lead times, and overall, TIGGE forecasts and 2 single EPS forecasts show top performances across all lead times with an average performance rank of 2.8, 3.0 and 2.9 respectively (out of 8, lowest being best). At all lead times the error of observed to forecasted mean rainfall for TIGGE forecasts remains below 20%, and the TIGGE grand-ensemble forecasts are always within the top three performing systems according to the RPS.

5. River Discharge: Grand-Ensemble Predictions Are Within the Range of Measurement Uncertainty

[11] For all forecasts, the error in the high river discharge predictions of EFAS is lower than for the corresponding rainfall forecasts [*Pappenberger et al.*, 2008], because the hydrological model acts as a non-linear filter which reduces uncertainties. The maximum average uncertainty for river discharge measurements is often assumed to be on the order of 8.5% [*Pappenberger et al.*, 2006]. All EPS forecasts are within these uncertainty bounds for short lead times, and TIGGE and 2 of the single EPS forecasts are within these bounds for all lead

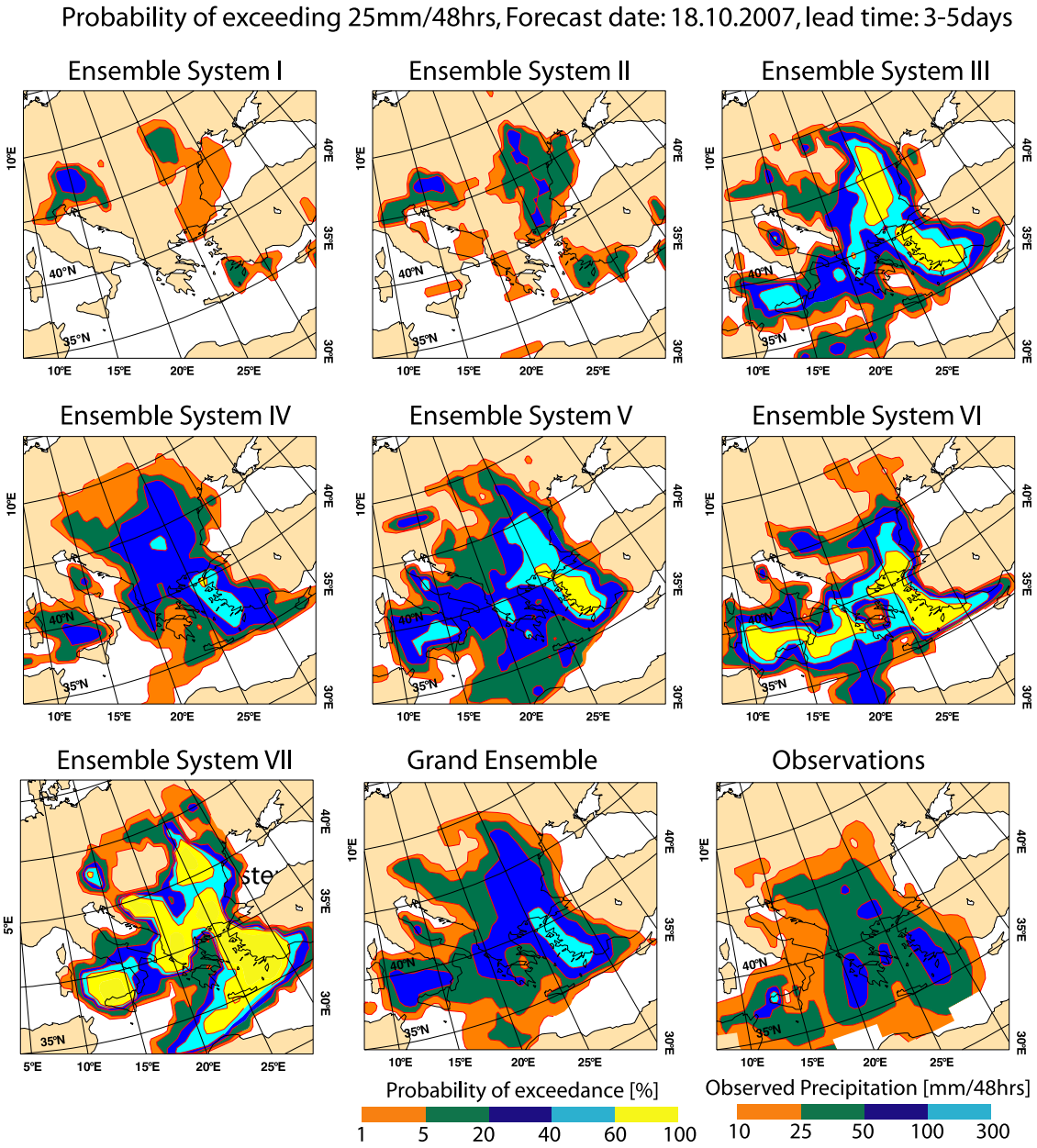

Figure 1. Probability of exceeding 25 mm of rainfall in 48 hrs for the TIGGE grand-ensemble and the 7 single EPS forecasts issued at 12 UTC of 18 October 2007 valid for 20 to 22 October (i.e., with a 48-to-96 hour lead time). A map of observations is also shown.

times and thus are suitable for predicting this flood within the range of uncertainty of the measurements.

6. River Discharge Predictions: Performance of the Grand-Ensemble Versus Single EPS

[12] The Brier Score of the forecasts exceeding the median discharge in October 2007 for the whole of Europe has been analysed: the score drops from 0.6 at a lead time of 1 day to roughly 0.4 at a lead time of 10 days. No clear favourite ensemble emerges and all of the forecast systems perform similarly, including TIGGE. In the Romanian case study, the RMS error of the ensemble mean for all EPS forecasts rises by a factor of between 3 and 6 for lead times of 10 days compared to a 1 day lead time. The RMS error of TIGGE forecasts is lower than 5 of the single EPS forecasts, and is as good as the remaining 2, especially for long lead times (more than 4 days). In the RPS of river discharge, TIGGE forecasts are only outperformed by 2 other EPS forecasts (Figure 2). It should also be noted that the high performance of one particular EPS forecast is actually due to its persistent underprediction of the event (limitation of the RPS methodology). Overall, for the prediction of flood discharges for the Romanian floods of October 2007, the TIGGE grand-ensemble forecasts show a high performance across all lead times with the best average performance rank of 2.9. The other single EPS forecasts show mixed results again, particularly for the high river discharges which indicate flooding. Therefore in this case study the TIGGE

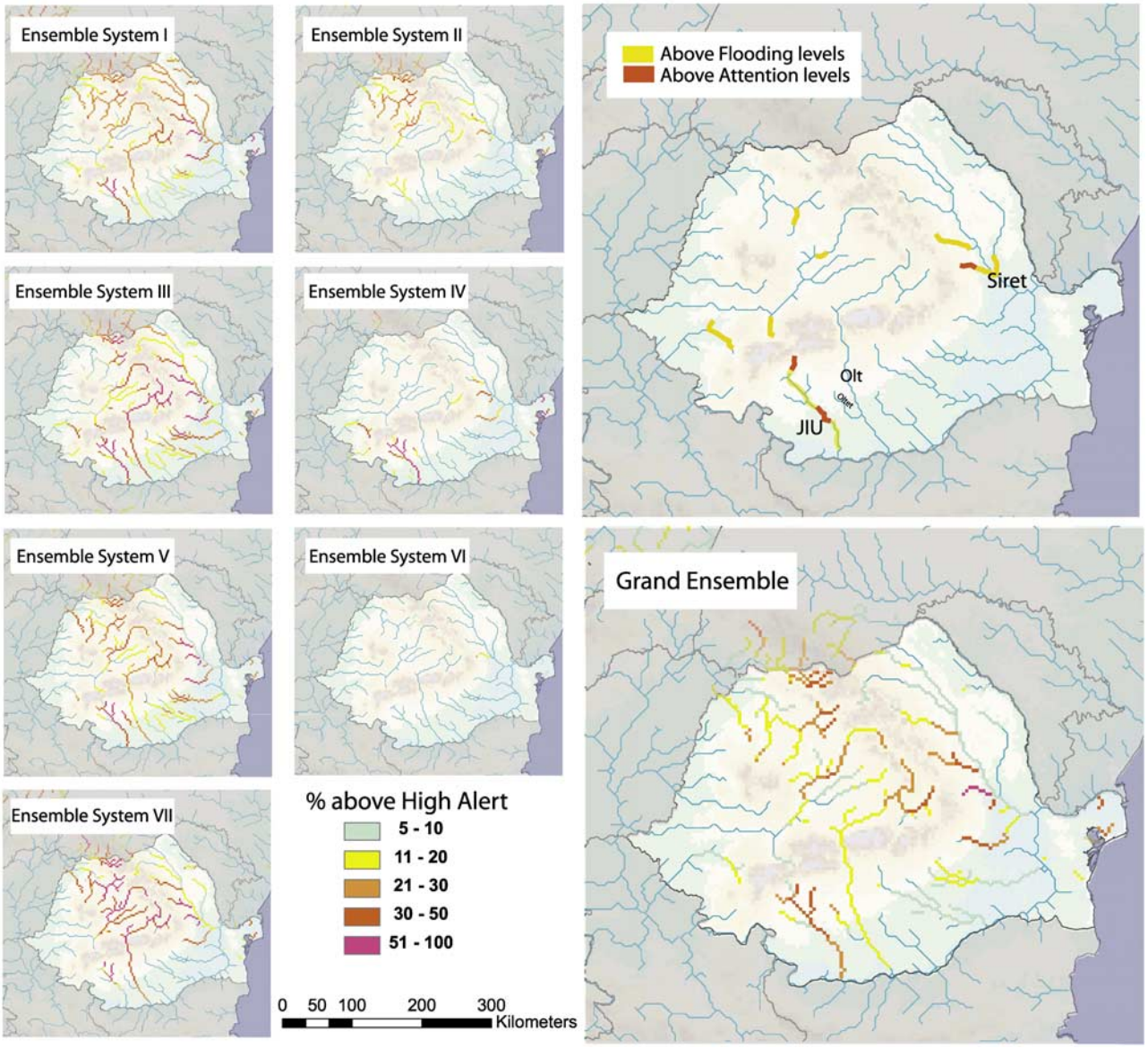

Figure 2. Probability of exceeding the high alert threshold of river discharge for TIGGE grand-ensemble and the single EPS forecasts issued at 12 UTC of 19 October 2007 valid for 24 of October. A map of observations is also shown.

grand-ensemble provides better river discharge predictions than any of the single EPS forecasts.

7. Characterisation of Important Hydrograph Features

[13] Flood warnings are often based on point predictions. In Figure 3, forecasts for a 5 day lead time are shown for a point on the river Jiu where flooding was observed. The distribution is large and can bracket flows both below and above the warning thresholds. TIGGE and 6 single EPS forecasts predict the onset of the rising limb correctly in terms of timing and river discharge thresholds, which is the most essential part of the flood hydrograph to represent correctly in terms of flood preparedness and disaster mitigation. TIGGE forecasts also bracket the flood peak correctly, as do 2 other single EPS forecasts. None of the forecasts perform very well for the lower end of the recession limb. The ensemble spread widens with lead time and thus more observations are bracketed. The widening distribution also means that a lower percentage of river discharge predictions are above the warning thresholds. This has important implications for issuing warnings, as at long lead times there will be fewer ensemble members that will be able to trigger such warnings and this should be taken into account. Overall, using the TIGGE grand-ensemble consistently gives a good prediction of the flood hydrograph for the Romanian floods of October 2007, apart from the falling recession limb.

8. Flood Warnings for Single Locations

[14] In Figure 4, the probability of exceedance of the high warning threshold for each forecast centre for 13 consecutive forecast dates is shown for a location on the Jiu. Figure 4 concentrates on the onset of the flood (24 October) and

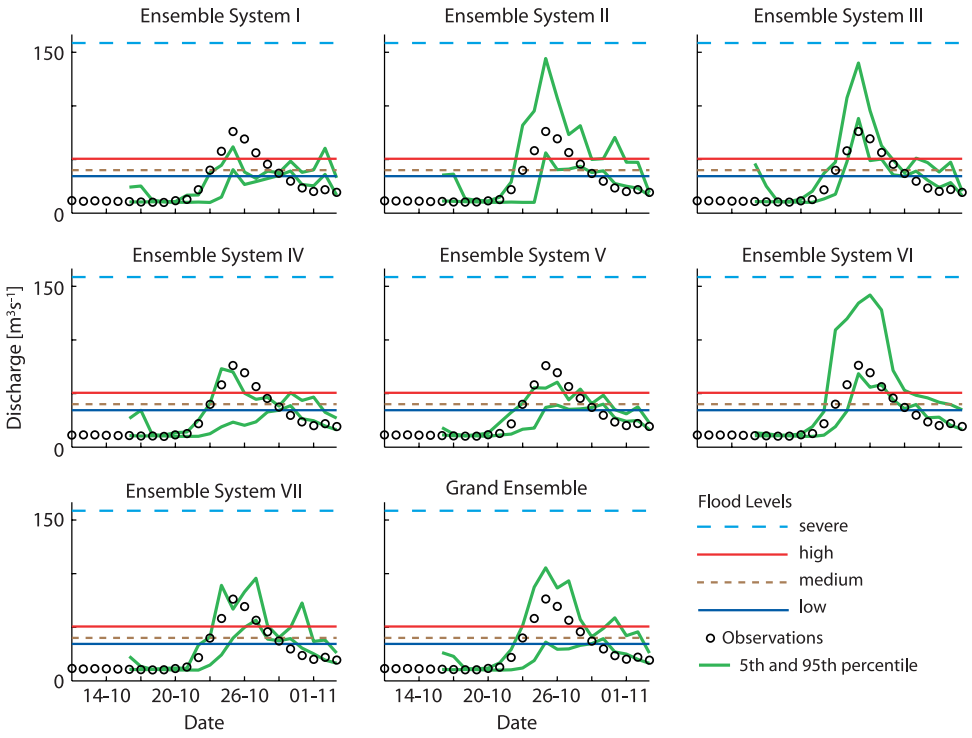

Figure 3. For a point on the river Jiu where flooding was observed, the 5th and 95th percentile of river discharge predictions are shown for the different forecasts with a 5 day lead time. The dashed horizontal lines show the four EFAS warning thresholds. "Observed" discharges refer to simulations based on observed meteorological input.

therefore only shows the forecasts for 11 October to the 23 October. The exceedance levels indicate that most EPS forecasts start to predict the forthcoming flooding in the forecasts issued from 14 October to 17 October. The signal persists from forecast to forecast, which provides the necessary reassurance. This type of persistency is one method used by EFAS to decide whether warnings will be issued [*Thielen et al.*, 2008]. From 19 October onwards, the signal is very strong, although initially the flooding is predicted one day too early. This means that there is an efficient flood warning 5 days in advance and a possible warning 8 days in advance.

[15] Clearly, in this case study the success of the flood warning based on an single EPS forecast would have depended a great deal on the choice of EPS. Some EPS forecasts missed the event almost entirely and others showed inconsistent results from one forecast to another (a lack of persistency which would lead to lower flood probabilities being assigned in EFAS [*Bartholmes et al.*, 2008]). In terms of early flood warning, however, it is the missed events that have more weight: false alarms and hits are identified as the events draw nearer, while missed events lead to late preparations and can initially result in doubts over the short-term forecast results, thus reducing preparedness even more. The grand-ensemble which is composed of several EPS and thus includes a large range of possible weather forecasts, is less likely to miss an event entirely and therefore more useful for early flood warning. We have also performed the above analysis for locations where no flooding was observed during the Romanian floods, and results indicated that in this particular case study the flood forecast based on multiple weather forecasts could reduce the false alarm rate (see *Pappenberger et al.* [2008] for more information).

9. Conclusions on the TIGGE Grand-Ensemble: A New Dimension in Flood Forecasting?

[16] The task of evaluating meteorological forecasts for hydrological applications is extremely difficult because of the very low sample size (flood events are rare; catchment characteristics may change between consecutive flood events; verification on medium size flows has nothing to do with the performance of models at flood flows; floods are seasonal and so cannot be averaged over traditional meteorological time lengths; antecedent conditions are important and thus not every extreme rainfall leads to a flood). However, this paper has demonstrated on a single available test case the utility of grand ensemble systems for flood forecasting.

[17] In particular, we have shown that although ensemble systems are designed to theoretically capture all possible weather developments, and although this might be true on average, severe events can be entirely missed by single EPS. In contrast a grand-ensemble (instead of single EPS) has been shown to produce more reliable results of a flooding event and therefore can have significant added value for an operational flood forecasting system. Results are based on weather ensemble forecasts from the TIGGE archive, and flood predictions generated using EFAS for the example of the severe flooding on the river Danube in Romania in October 2007. It is extremely difficult to carry out a statistically significant evaluation for flood events, but we believe that this work gives encouraging indications that a

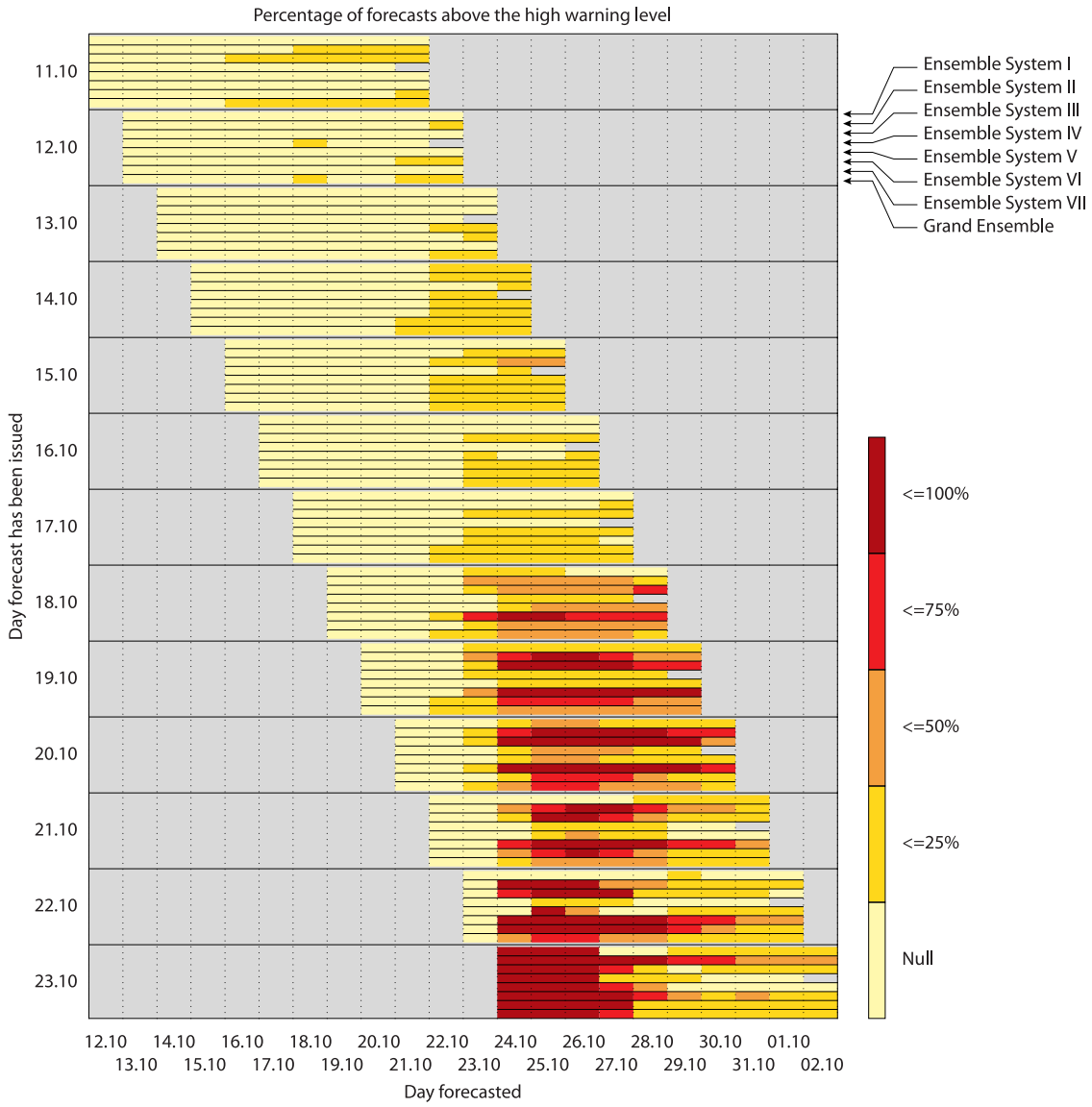

Figure 4. For a location on the river Jiu where flooding was observed, percentage of forecasts exceeding the high thresholds from 11 October to 23 October for all forecast systems.

multi-system grand-ensemble can provide more valuable forecasts than a single ensemble in predicting extreme flood events.

[18] **Acknowledgments.** Funding has been provided by the PREVIEW project, NERC-FREE (NE/E002242/1) and the FRIEND flood group project (Unesco contract 4500043129). We thank E. Anghel (Romanian National Institute of Hydrology and Water Management), the IPSC of the DG JRC, Mauro del Medico (JRC), and Paul Bates (Bristol University).

References

Ahrens, B., and S. Jaun (2007), On evaluation of ensemble precipitation forecasts with observation-based ensembles, *Adv. Geosci.*, *10*, 139–144.

Bartholmes, J., et al. (2008), The European Flood Alert System EFAS–part 2. Statistical skill assessment of probabilistic and deterministic operational forecasts, *Hydrol. Earth Syst. Sci. Discuss.*, *5*, 289–322.

de Roo, A., et al. (2003), Development of a European flood forecasting system, *Int. J. River Basin Manage.*, *1*, 49–59.

Demeritt, D., et al. (2007), Ensemble predictions and perceptions of risk, uncertainty, and error in flood forecasting, *Environ. Hazards*, *7*(2), 115.

Goswami, M. (2007), et al., Development of regionalisation procedures using a multi-model approach for flow simulation in an ungauged catchment, *J. Hydrol.*, *333*(2–4), 517–531.

Gourley, J. J., and B. E. Vieux (2005), A method for evaluating the accuracy of quantitative precipitation estimates from a hydrologic modeling perspective, *J. Hydrometeorol.*, *6*(2), 115–133.

Hagedorn, R., et al. (2005), The rationale behind the success of multi-model ensembles in seasonal forecasting-I. Basic concept, *Tellus, Ser. A*, *57*(3), 219–233.

Kalas, M., et al. (2008), Evaluation of the medium-range European flood forecasts for the March–April 2006 flood in the Morava River, *J. Hydrol. Hydromech.*, in press.

Krzysztofowicz, R. (2002), Bayesian system for probabilistic river stage forecasting, *J. Hydrol.*, *268*(1–4), 16–40.

Mcenery, J., et al. (2005), NOAA'S Advanced Hydrologic Prediction Service: Building pathways for better science in water forecasting, *Bull. Am. Meteorol. Soc.*, *86*, 375–385.

Olsson, J., and G. Lindström (2007), Evaluation and calibration of operational hydrological ensemble forecasts in Sweden, *J. Hydrol.*, *350*(1–2), 14–24.

Pappenberger, F., et al. (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), *Hydrol. Earth Syst. Sci.*, *9*(4), 381–393.

Pappenberger, F., et al. (2006), Influence of uncertain boundary conditions and model structure on flood inundation predictions, *Adv. Water Resour.*, *29*(10), 1430–1449.

Pappenberger, F., et al. (2008), TIGGE: Medium range multi model weather forecast ensembles in flood forecasting (a case study), *Tech. Memo*. 557, Eur. Centre for Medium-Range Weather Forecasts, Reading, U.K.

Park, Y.-Y., et al. (2008), TIGGE: Preliminary results on comparing and combining ensembles, *Tech. Memo.* 548, Eur. Centre for Medium-Range Weather Forecasts, Reading, U.K.

Richardson, D. (2005), The THORPEX Interactive Grand Global Ensemble (TIGGE), *Geophys. Res. Abstr.*, *7*, Abstract EGU05-A-02815.

Roulin, E. (2007), Skill and relative economic value of medium-range hydrological ensemble prediction, *Hydrol. Earth Syst. Sci.*, *11*, 725–737.

Thielen, J., et al. (2008), The European Flood Alert System–part I. Concept and development, *Hydrol. Earth Syst. Sci. Discuss.*, *5*, 257–287.

van der Knijff, J. and A. de Roo (2007), Lisflood: Distributed water balance and flood simulation model: User manual, Eur. Comm., Brussels, Belgium.

Verbunt, M., et al. (2006), Verification of a coupled hydrometeorological modelling approach for alpine tributaries in the Rhine basin, *J. Hydrol.*, *324*(1–4), 224–238.

J. Bartholmes, A. de Roo, and J. Thielen, Institute for Environment and Sustainability, DG Joint Research Centre of the European Commission, T.P. 261, I-21027, Ispra, Italy.

R. Buizza and F. Pappenberger, European Centre for Medium-Range Weather Forecasts, Reading RG2 9AX, UK. (florian.pappenberger@ecmwf.int)

H. L. Cloke, Department of Geography, King's College London, Strand, London, WC2R 2LS, UK.