ATMOSPHERIC SCIENCE LETTERS
Atmos. Sci. Let. **11**: 123–131 (2010)
Published online 24 March 2010 in Wiley InterScience
(www.interscience.wiley.com) **DOI:** 10.1002/asl.266

Ensemble flood forecasting in Africa: a feasibility study in the Juba–Shabelle river basin†

Vera Thiemig,[1]* Florian Pappenberger,[2] Jutta Thielen,[1] Hussein Gadain,[3] Ad de Roo,[1] Katalin Bodis,[1] Mauro Del Medico[1] and Flavian Muthusi[3]

[1] *Land Management and Natural Hazards Unit, Institute for Environment and Sustainability, DG Joint Research Centre, European Commission, Via E. Fermi 1, TP 261, 21020 Ispra, Varese, Italy*
[2] *European Centre for Medium-Range Weather Forecasts, Reading RG1 9AX, UK*
[3] *Food and Agriculture Organization of the United Nations, Somalia Water and Land Information Management – SWALIM Project, P. O. Box 30470-00100, Nairobi, Kenya*

**Correspondence to: Vera Thiemig, Land Management and Natural Hazards Unit, Institute for Environment and Sustainability, DG Joint Research Centre, European Commission, Via E. Fermi 1, TP 261, 21020 Ispra, Varese, Italy. E-mail: vera.thiemig@jrc.ec.europa.eu*

† This article was published online 24 March 2010. An error was subsequently identified. This notice is included in the online and print versions to indicate that both have been corrected 20 April 2010.

Received: 8 October 2009
Revised: 1 February 2010
Accepted: 3 February 2010

Abstract

The European flood alert system (EFAS) achieves early flood warnings for large to medium-size river basins with lead times of 10 days. This is based on probabilistic weather forecasts, the exceedance of alert thresholds and persistence. The methodologies have been tested for different events and time scales in mid-latitude basins in Europe. In this article, the transferability of the EFAS-methodologies to equatorial African basins is assessed with the analysis of the Juba–Shabelle river basin as an example using a variety of different meteorological data sources. In this context, ERA-40 and CHARM have been used for the calculation of climatologies; re-forecasts of the current operational European Centre for Medium-Range Weather Forecasts model provided hindcasts of historic flood events. The results show that flood events have been detected successfully in more than 85% of all cases, with a high accuracy in terms of timing and magnitude. Copyright © 2010 Royal Meteorological Society

Keywords: **flood warning; probabilistic flood forecasting; European flood alert system; ensemble prediction system; Africa; LISFLOOD; hydrological ensemble prediction systems; Juba-Shabelle river basin**

1. Introduction

Over the last few years, Africa has increasingly experienced severe transnational floods that have caused substantial socio-economic losses and put enormous pressure on countries across the continent (Theron, 2007; Dartmouth Flood Observatory, 2009). Planning, coordination and realisation of flood prevention, protection and mitigation measures require a certain amount of time prior to the flood event, and this can be provided through early flood prediction.

The forecasting skills of medium-range flood prediction systems depend largely on the quality of the meteorological forecast and on the performance of the hydrological model. In recent years the skill of meteorological forecasts has noticeably improved (McBride and Ebert, 2000; Hamill *et al*., 2007), whereas the skill of hydro-meteorological forecast systems has improved by 1–2 days over the last 10 years (Pappenberger and Thielen, 2009, personal communication). Thus it is not surprising that the application of probabilistic rainfall forecasts from ensemble prediction systems (EPSs) has gained in importance within the hydrological community leading to the development of hydrological ensemble prediction systems (HEPS) (Buizza, 2008; Pappenberger *et al*., 2008a,b; Cloke and Pappenberger, 2009; Cloke *et al*., 2009). This increased use of EPS is due to its ability to account for uncertainties related to future meteorological conditions, with lead times of up to 2 weeks, which is well beyond the limits of possibility of deterministic models, and results in more skilful forecasts (Pappenberger *et al*., 2005; Thielen *et al*., 2009b).

The European flood alert system (EFAS) is an advanced prototype of a continental flood warning system. It uses multiple weather forecasts from EPS to produce probabilistic flood alerts with lead times up to 10 days. EFAS information is complementary to national systems and provides a unique European overview for the coordination of aid during flood events that require international assistance (Thielen *et al*., 2009a).

In this article, we have studied to what extent the methodologies of EFAS can be transferred to semi-arid environments in Africa and outlines the

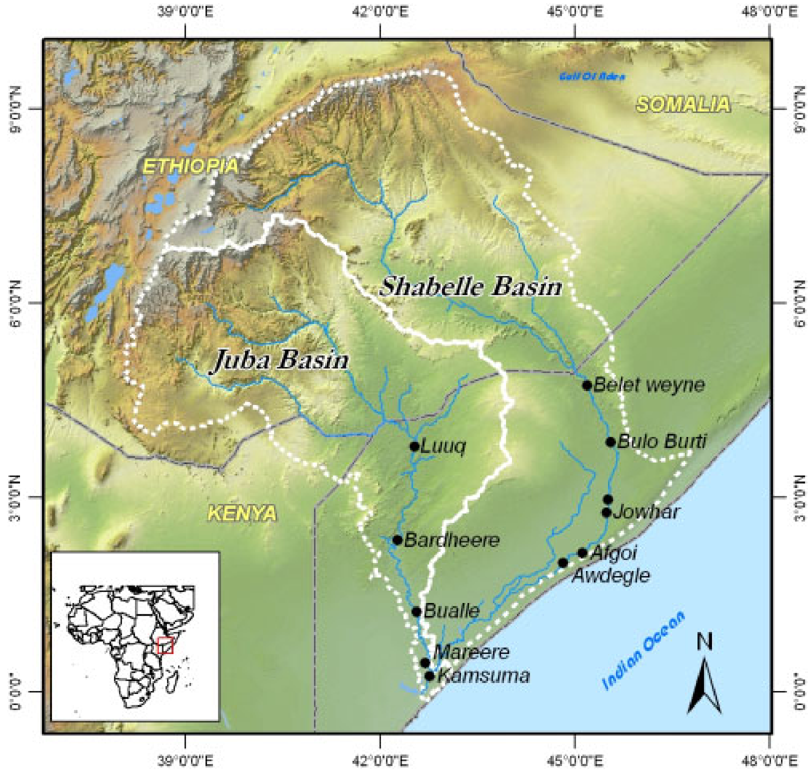

Figure 1. Juba–Shabelle river basin.

quality of the different climatological data. Such a coupled system also enables evaluation of the performance of meteorological forecasts and integrated re-analysis over time and space (Pappenberger and Buizza, 2009).

The remainder of this article is organised in four sections. In Section 2, a short description of the study area, the model set-up, the input data and the EFAS-methodologies are given. This is followed by a presentation of the evaluation methods in Section 3, an analysis of the results in Section 4 and conclusions in Section 5.

2. Description of the study area, input data, model set-up and EFAS-methodology

2.1. The study area: Juba–Shabelle river basin

The Juba–Shabelle river basin is shared between Somalia, Ethiopia and Kenya, and covers an area of 783,000 km^2 (Figure 1). The Juba and Shabelle rivers are 1100 and 1700 km long, respectively. The altitude of both rivers ranges from more than 3000 m above mean sea level (AMSL) in the eastern Ethiopian highlands to sea level.

The climate is determined by the north-easterly and south-easterly winds of the Intertropical Convergence Zone over the Ethiopian highlands resulting in tropical arid to dry and sub-humid conditions, with two distinctive annual rain seasons: the so-called Gu' (long rains) from April to June, and the Deyr (short rains) from October to November (Artan *et al.*, 2007).

The two rain seasons are reflected in the hydrograph by a twice-yearly flooding period (Figure 2). The mean annual runoff at Luuq is more than 2.5 times the one measured at Belet Weyne, although the catchment area upstream of the gauging station Belet Weyne is about one-fifth larger than the catchment area upstream of the gauging station Luuq (Table I). This is partly due to higher annual rainfall in the headwaters of the Juba and due to the different nature of the geological formations underlying the two rivers. Both rivers show a progressive discharge reduction towards the outlet due to the lack of any significant flow contribution in Somalia, major natural losses (e.g. evaporation and infiltration) and human-driven withdrawals for irrigational purposes (Artan *et al.*, 2007).

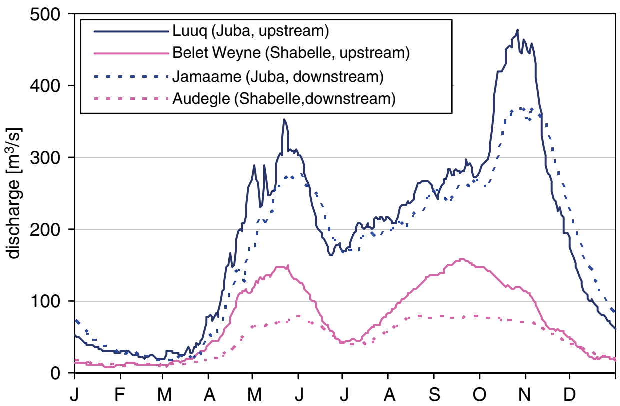

Figure 2. Long-term average (1970–1989) flows in m^3/s.

Table I. Characteristics of gauging stations (locations are also marked in Figure 1).

Station/stream	Altitude (m AMSL)	100-yr return period (m^3/s)	Mean annual discharge (m^3/s)	Upstream area (km^2)	Sampling frequency	Data		
						Start	End	Missing
Luuq/Juba	141.4	2031	191.0	166000	Daily	1951	To date	1968, 1991–2000
Belet Weyne/Shabelle	176.1	576	74.3	207000	Daily	1951	To date	1953, 1991–2001

Table II. Meteorological input data.

	ERA-40 (Uppala *et al.*, 2004)	CHARM (Funk *et al.*, 2003)	ERA-interim (Berrisford *et al.*, 2004)	RFE (NOAA-CPC, 2002)	VAREPS (Buizza, 2009)
Released by	ECMWF	Geography Department, University of California in Santa Barbra (UCSB) and United States Geological Survey (USGS)	ECMWF	National Oceanic and Atmospheric Administration, Climate Prediction Center (NOAA-CPC)	ECMWF
Data type	Re-analysis of past observations using NWP models	Blended gauge-satellite product, re-analysis fields and orographic model output	Re-analysis of past observations using NWP models	Blended gauge-satellite product	Probabilistic re-forecasts using EPS
Temporal resolution	6 h	24 h	6 h	24 h	Staggered, 3 h (days 1–4) and 6 h (days 5–15)
Spatial resolution	120 km	10 km	80 km	0.1°	Staggered, 50 km (days 1–4)* and 80 km (days 5–15)
Time periods provided	1 January 1959 to 31 August 2002	1 January 1961 to 31 December 1996	1 September 2002 to 31 December 2007	1 January 2001 to 31 December 2008	1 October 1977 to 25 November 1977, 16 March 1981 to 3 May 1981

* [Correction made here after initial online publication.]

2.2. Input data

2.2.1. Meteorological data

Meteorological observations available for this study area are sparse, derived from an unevenly distributed gauge network, and time series are often interrupted by missing records. Therefore, the study has instead been based on four re-analysis datasets, ERA-40, CHARM, ERA-interim and RFE 2.0 (Table II). In fact, this article focusses exclusively on ERA-40 and CHARM, because the temporal period of both presented hindcasts is only captured by these datasets.

For the probabilistic flood forecast, the European Centre for Medium-Range Weather Forecasts (ECMWF)- VAREPS data (Table II; Buizza *et al.*, 2007; Vitart *et al.*, 2008) were used, which were specifically recalculated for this case study by ECMWF, using the latest model version (cycle 33R3, Buizza, 2009).

2.2.2. Reference discharge data

Within Somalia, the two uppermost gauging stations, Luuq and Belet Weyne, have been chosen as reference gauges for the hydrological calculations. These locations allow the expansion of research from earlier studies (the *Somalia Flood Forecasting Model,* Artan *et al.*, 2007), which provide short-range inflow forecasts based on river routing only without hydrological components and the use of NWP data. Table I gives an overview of the characteristics of these stations.

2.3. Model set-up and EFAS-methodology

In EFAS, the hydrological calculations are accomplished using LISFLOOD – a distributed grid-based hydrological rainfall-runoff model (Van der Knijff and de Roo, 2008). LISFLOOD is a hybrid model consisting of physical and conceptual components. As such, most model parameters are physically based and thus derivable from literature, whereas the remaining parameter values have to be adjusted through model calibration.

LISFLOOD was calibrated based on meteorological and hydrological data from the last five decades. Parameter sensitivity has been established by using a global sensitivity analysis (Cloke *et al.*, 2008; Pappenberger *et al.*, 2008a,b). Sensitive parameters have then been optimised in a two-step approach: first manually using the trial-and-error method and then automatically applying the Shuffled Complex

Evolution (SCE-UA) algorithm (Duan *et al.*, 1994). No hydrological parameter uncertainty has been considered under the assumption that meteorological uncertainty will dominate the hydrological response (He *et al.*, 2009).

Due to large differences in the spatial and temporal distribution of the rainfall fields within the different meteorological datasets (ERA-40, CHARM, ERA-interim and RFE 2.0) all datasets were applied independently from each other during the calibration and calculation of initial conditions.

For the actual flood forecasting (i.e. hindcasting), 15-day ensemble weather forecast data were fed into LISFLOOD, resulting in an ensemble of simulated hydrographs.

In EFAS, a flood alert is emitted if the discharges exceed critical thresholds (low, medium, high and severe) and if the exceedance has been sufficiently persistent over previous forecasts (the EFAS decision support rules; Ramos *et al.*, 2007; Bartholmes *et al.*, 2009; Thielen *et al.*, 2009). For EPS-based forecasts, an event is defined if a certain number of EPS (i.e. at least 10 out of 51) have persistently exceeded the critical thresholds (Bartholmes *et al.*, 2008).

The determination of the thresholds thus plays a crucial role for the system. For this study, the alert thresholds were derived based on return-periods. The 2-year, 5-year, and 10-year return-periods of the hydrological time series were defined as the low, medium and high alert threshold, whereas the severe alert threshold corresponds to the highest recorded discharge value.

3. Evaluation method

Hindcasts for the historical flood events of Deyr 1977 and Gu' 1981 were calculated. A long-term run has been computed over a time period of 21 years (1 January 1961 to 31 December 1981) using the meteorological datasets CHARM and ERA-40 as well as the respective calibration setting. The initial conditions were extracted from the simulated discharge series on a daily basis during the flood event. Based on these initial settings, the hindcasts using the probabilistic precipitation re-forecasts of the ECMWF EPS are calculated on a daily time step. Hence, every flood event has been hindcasted twice, once for each initialisation setting.

This study addresses the potential use of VAREPS and re-analysis data for flood forecasting in Africa. Influences from insufficient model calibration or determination of initial conditions are therefore excluded and ensemble forecasts are compared against simulated discharges driven with ERA-40 and CHARM, the so-called proxy hydrological record.

The applicability of the hindcasting methodology is assessed visually and by multiple statistical methods (as suggested by e.g. Cloke *et al.*, 2008). The visual assessment of the model performance is facilitated by the compilation of colour-coded threshold exceedance diagrams; whereas the statistical reasoning is based on various skill scores such as the continuous rank probability score (CRPS), relative operating characteristics (ROC), spread–skill relationship, ranked histogram and the Pearson product-moment correlation coefficient (R).

4. Results

4.1. Quality of meteorological data

The quality of the two different long-term meteorological datasets, ERA-40 and CHARM, is determined based on mutual comparison. In principle, as both datasets try to reproduce a close approximation of the meteorological situation, they should be in agreement with each other. However, the long-term annual average precipitation reveals a significant difference between both datasets. Table III shows that the discrepancy between ERA-40 and CHARM remains independent of the reference time period almost constant with CHARM corresponding approximately to 1.5 times ERA-40.

In contrast, the spatial distribution of the long-term annual average precipitation (Figure 3) shows that despite the different resolution of the datasets, the rough spatial distribution of the precipitation fields is in good agreement. Both datasets show an increased amount of precipitation over the Ethiopian highlands and in the coastal area. Quantitatively, however, the

Table III. Comparison of the different meteorological datasets.

		Luuq			Belet Weyne		
		ERA-40		**CHARM**	**ERA-40**		**CHARM**
Long term	Annual areal average upstream of the gauging station (mm/year)	520		814	505		776
	Correlation		0.38			0.36	
1976–1981 (calibration period)	Annual areal average upstream of the gauging station (mm/year)	617		903	561		777
	Correlation		0.39			0.41	
1982–1987 (validation period)	Annual areal average upstream of the gauging station (mm/year)	477		663	449		684
	Correlation		0.47			0.38	

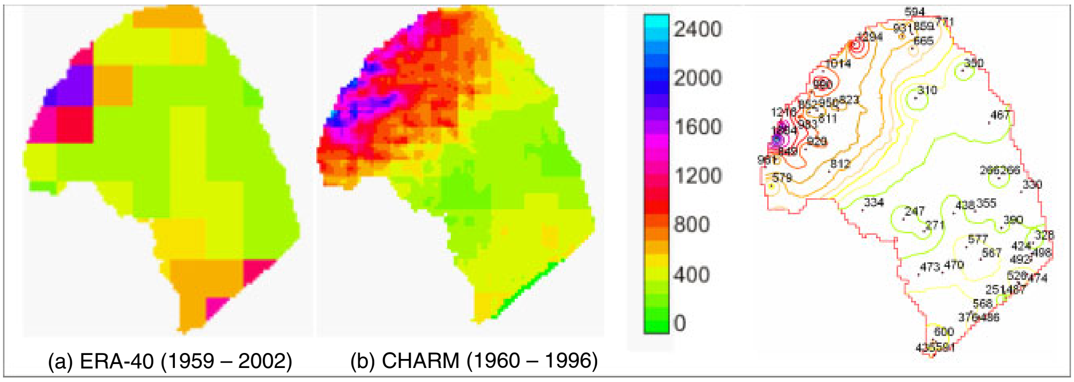

Figure 3. Comparison of the long-term annual average precipitation (mm/a) of the different long-term meteorological datasets ERA-40 (a) and CHARM (b) with the observations (mm/a) released within the FAO CLIMWAT database.

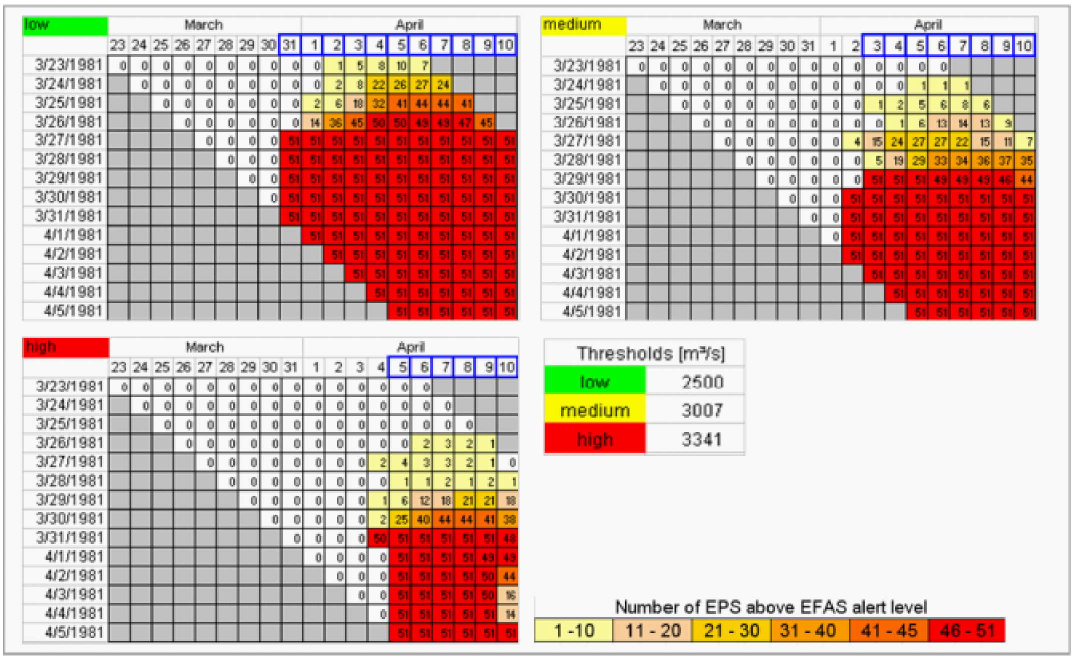

Figure 4. Colour-coded threshold exceedance diagram for the hindcast in Gu' 1981 for Belet Weyne, showing the persistence on the number of EPS-based simulations exceeding the low, medium and high threshold; initial conditions based on CHARM data. The *x*-axis shows the day of each forecast. The *y*-axis indicates the forecasted day. The time period, in which the proxy hydrological record exceeds the equivalent alert threshold, is indicated using a dark blue cell border at the heading of the matrix.

datasets are clearly different. In order to assess the uncertainty, the datasets are visually compared with the long-term observations that have been released within the FAO CLIMWAT database (Smith, 1993). Comparing those, ERA-40 shows a slight overestimation at the coast, whereas the CHARM data exhibits an overestimation in the northern part of the catchment.

Obviously this is a very coarse verification and one should bear in mind that due to the non-linear filter of the hydrological system small but systematic differences in rainfalls may have significant consequences for the subsequent discharges.

Furthermore the fairly low correlation between ERA-40 and CHARM (from 0.36 to 0.47) in Table III confirms a high degree of uncertainty in the meteorological datasets.

Similar results have been reported in the literature. Funk *et al.* (2003) stress that the temporal variability of precipitation is in general very well captured by the CHARM algorithm. As the literature evaluates the performance of CHARM with reference to two case studies (one in Kenya and the other in Mali) no direct statement about the performance for the region examined in this study can be given here. However, Funk *et al.* (2003) show that in mountainous regions, that are well represented within the Global Historical Climate Network (GHCN) dataset, the long-term average shows considerable positive bias, indicating an overestimation of the precipitation. As the Ethiopian highlands are well represented by the GHCN dataset it can be presumed that the precipitation is uncertain.

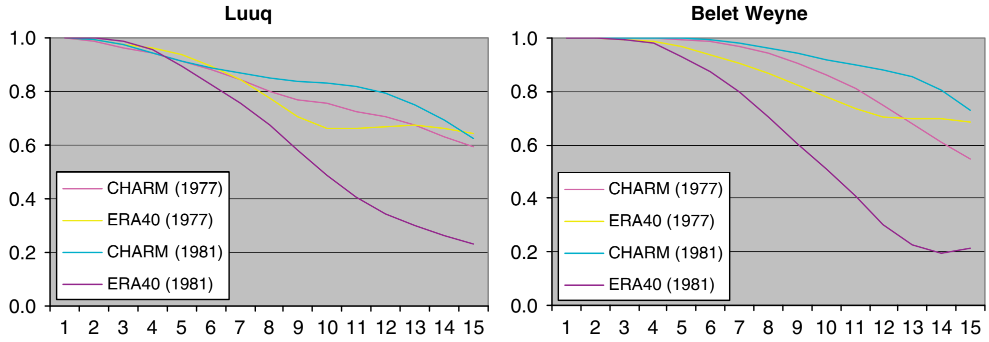

Figure 5. Development of the Pearson product-moment correlation between the proxy hydrological record and the EPS-based simulations over the 15-day lead time period.

Concerning the quality of the ERA-40 data Diro *et al*. (2009) state that the spatial and seasonal variability of the rainfall climatology are in general well captured in Ethiopia, whereas the mean amount of rainfall over the Ethiopian highlands is underestimated.

4.2. Quality of hydrological predictions

Figure 4 summarises the results from the forecasts starting on 23 March 1981. Threshold exceedances are represented in a calendrical matrix; each row holds a 15-day EPS-based hydrological hindcast for the specific day the meteorological forecast has been processed; each column represents the consecutive development of the hindcast for a specific calendar day. Therefore, the number of EPS-based simulations exceeding the flood alert thresholds is shown within the matrix, whereas the time period, in which the proxy hydrological record exceeds the equivalent alert threshold, is indicated using a dark blue cell border at the heading of the matrix. Hence, if the colour-coded rows and columns overlap each other then the flood has been hindcasted successfully (Ramos *et al*., 2007).

The threshold exceedance diagram for the flood hindcast of Gu' 1981 for the gauging station Belet Weyne, shown in Figure 4, serves as a representative example of the general performance of all hindcasts. Keeping the decision support rules of EFAS in mind (Section 2.3), the alarm for the flood in Gu' 1981 at the gauging station Belet Weyne would have been primarily processed on 25 March indicating a low flood with a medium probability (18 out of 51 EPS-based simulations) starting on 3 April. Due to consecutive daily monitoring of the expected situation the starting date of the flood event is progressively adjusted. Hence, on 27 March the flood event is predicted to start from 31 March, which is also the date at which the proxy hydrological record exceeds this threshold. The same accuracy can be seen for the medium and high alert threshold. Despite the loss in lead time caused by the criteria of persistence, the flood event has been forecasted successfully 4 days in advance. That the magnitude of the flood event is well captured can be determined by the fact that both EPS-based simulations and proxy hydrological record exceed the same thresholds. Hence from a visual point of view, the hindcasting methodology results in a flood forecast that is characterised by a very high accuracy of both the timing and magnitude.

Considering the visual evaluation of all hindcasts (not shown), the flood signal could be detected successfully in seven out of eight cases. The accuracy in lead times varies from case to case and alert level significantly from 1 up to 8 days, whereas the accuracy concerning the magnitude of flooding was captured correctly in six out of eight events.

Statistical analysis supports the visual findings. The correlation plots (Figure 5) indicate a very high association ($R > 0.9$) for the first few lead days and a fairly good overall association ($R > 0.6$). Only the correlation for the ERA-40-based hindcast (1981) decreases significantly to a very low level of 0.2 indicating major deviations from the proxy observations. Further, comparing both plots it can be seen that Belet Weyne shows the tendency to perform slightly better than Luuq.

The spread–skill relationship assesses the relationship between mean standard deviation and standard deviation of the mean absolute error. A perfect ensemble implies that the statistical properties of the true value of discharge (here: of the proxy hydrological record) are identical to the statistical properties of the ensemble, in which case the mean standard deviation is equal to the standard deviation of the mean absolute error (Palmer *et al*., 2005) [Correction made here after initial online publication]. Based on this, the ERA-40 exhibits a very good spread–skill relationship as the curves are located very close together (Figure 6(a)). The CHARM relationship performs less well indicating an ensemble which is significantly under-dispersive. It is probably not a surprise that the ECMWF-VAREPS forecasts provide a better spread–skill relationship with respect to the ERA-40 than CHARM as there is a larger similarity in the model generation. The comparison with CHARM suggests an under-dispersive model, which does not exhibit enough spread to capture the proxy hydrological record sufficiently.

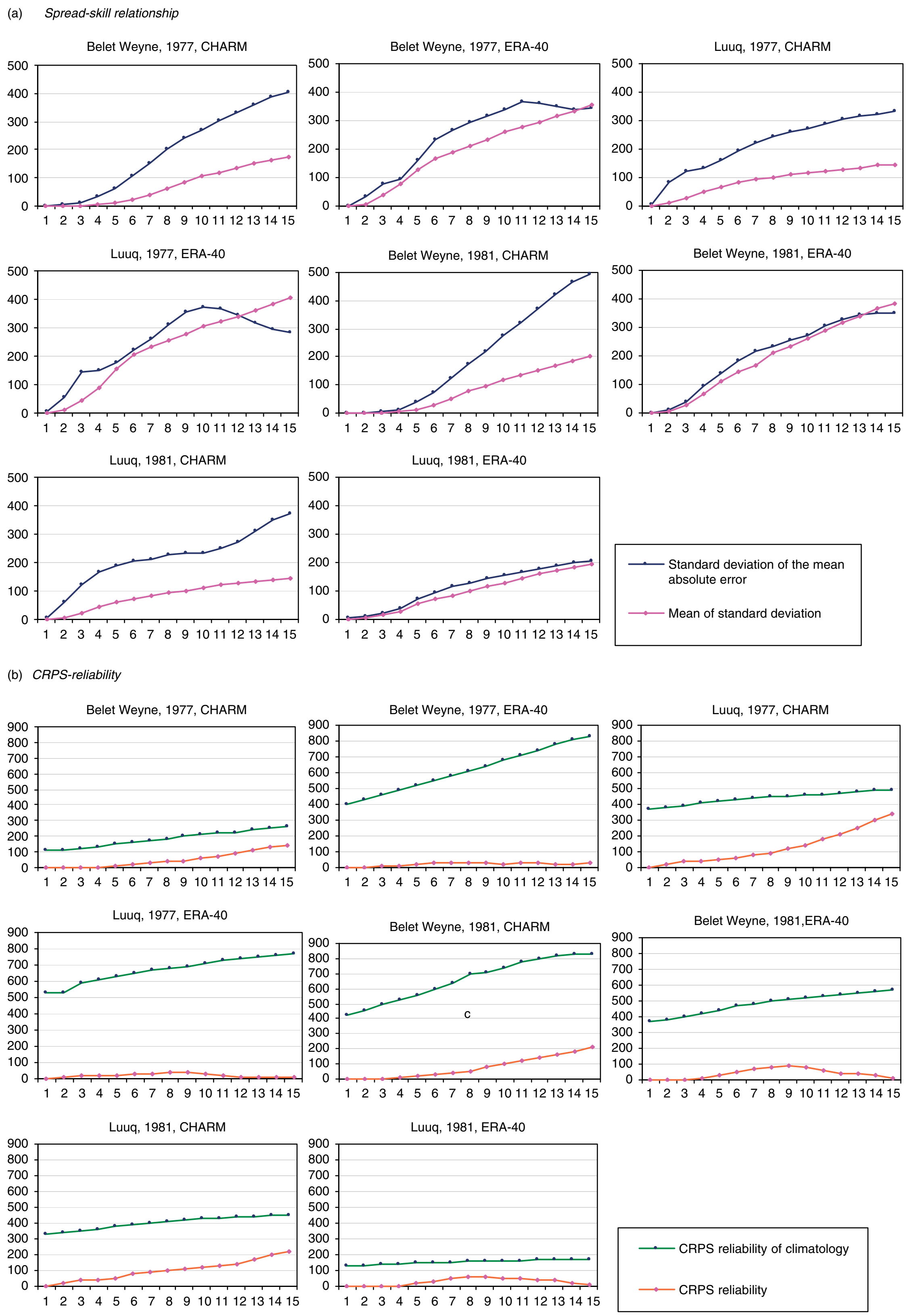

Figure 6. Spread–skill relationship (a) and CRPS reliability (b) (both in m^3/s) for the flood hindcasts in Gu' 1977 and Deyr 1981 over the 15-day lead time.

The ROC allows to assess the performance of a forecasting system that distinguishes between the intrinsic discrimination capacity and the decision threshold of the system. In this context, discrimination means the ability of the forecast system to distinguish between occurrences of the event and non-occurrences, by forecasting a different set of probabilities before the occurrences of the event than before the non-occurrences. A forecast is skilful if the ROC-graph lies above the 1 : 1 line (Jolliffe and Stephenson, 2003).

Although it was observed that the graph decreased with increasing lead time, it always remained well above the 1 : 1 line indicating a skilful forecast with a high discriminatory power. This is representative for all hindcasts (except flood event 1981, Luuq, ERA-40).

In a last step, it was assessed that the forecast (hindcast) performs better than guessing or a benchmark. The ROC score has already established that there is some skill in the system. An additional way to evaluate this is to compute the CRPS of a climatology, i.e. benchmark. In this study, the benchmark is set to an average performance for this particular date. The results (Figure 6(b)) show that the CRPS of this benchmark (green graph) is always significantly worse than the one of the forecast (red graph), which indicates that the forecast system is skilful.

5. Conclusions

The potentials of a hydrological ensemble prediction system for flood forecasts using the EFAS-methodologies in combination with medium-range probabilistic weather forecasts have been investigated for Africa, i.e. the Juba–Shabelle river basin. The applied flood forecasting approach relies on the criteria of threshold exceedance (minimum 10 out of 51 EPS-members) and the persistence of this signal.

In order to assess the pure performance of the hindcasting methodology the model outcome was compared against the proxy hydrological record. In this case, the performance of the hindcast depends solely on the quality of the meteorological data and the capabilities of the hydrological model; possible after-effects from the calibration are excluded.

The results show that in seven out of eight cases the flood signal has been detected successfully. The visual inspection depicts a high accuracy in terms of timing and magnitude. Focussing on the processing of flood forecasts that exceed the high alert threshold the lead time is on average around 6–8 days. The statistical evaluation applying a number of appropriate skill scores, such as the Pearson product-moment correlation coefficient (R), the spread–skill relationship, ROC and CRPS, revealed the capability of the forecasting system to perform better than guessing.

Future model adaptations focussing on the implementation of hydrological model components that account for semi-arid areas and significant water management issues (e.g. water withdrawal) promise to improve the accuracy of the hydrological calculations.

Overall, the article has shown that the EFAS-methodologies are transferable to African basins to produce probabilistic flood forecasts to assist national authorities in improved decision making, provided that there is good-quality meteorological data.

Acknowledgements

The authors gratefully acknowledge the support of all staff of the JRC's Institute for Environment and Sustainability who contributed to this pilot study, especially, to Peter Salamon and Niall McCormick for their detailed review of this paper. The authors also wish to thank the European Centre for Medium-Range Weather Forecast, in particular, Roberto Buizza, and SWALIM for providing data. Special acknowledgement also to the commission of the Ruhrverband that supported this work with a scholarship. Florian Pappenberger is funded by the SAFER (EC, Framework 7) project and a direct contract with the DG JRC, which are both gratefully acknowledged.

References

Artan G, Gadain H, Muthusi F, Muchiri P. 2007. *Improving flood forecasting and early warning in Somalia – feasibility study*. FAO-SWALIM (GCP/SOM/EC045) Project Publication N°W-10, Nairobi, Kenya.

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

Berrisford P, Dee D, Fielding K, Fuentes M, Kallberg P, Kobayahi S, Uppala S. 2009. *The ERA-Interim archive*. Technical report. European Centre for Medium-Range Weather Forecasts (ECMWF): Reading, UK. http://www.ecmwf.int/publications/library/ecpublications/_pdf/era/era_report_series/rs_1.pdf.

Buizza R. 2008. The value of probabilistic prediction. *Atmospheric Science Letters* **9**: 36–42.

Buizza, R., 2009. Current status and future developments of the ECMWF EPS. EFAS Workshop, 29–30 January 2009, European Centre for Medium-Range Weather Forecast (ECMWF), Reading, United Kingdom, available online on http://www.ecmwf.int/newsevents/meetings/workshops/2009/EFAS/presentations/04_RB_2009_01_EFAS_CurrStat_FutDev_EPS.pdf.

Buizza R, Bidlot J-R, Wedi N, Fuentes M, Hamrud M, Holt G, Vitart F. 2007. The new ECMWF VAREPS (Variable Resolution Ensemble Prediction System). *The Quarterly Journal of the Royal Meteorological Society* **133**: 681–695.

Cloke HL, Pappenberger F. 2008. Evaluating forecasts of extreme events for hydrological applications: an approach for screening unfamiliar performance measures. *Meteorological Applications* **15**: 181–19.

Cloke HL, Pappenberger F. 2009. Operational flood forecasting: review of ensemble techniques. *Journal of Hydrology* **375**: 613–626.

Cloke HL, Pappenberger F, Renaud J-P. 2008. Multi-Method Global Sensitivity Analysis (MMGSA) for modelling floodplain hydrological processes. *Hydrological Processes* **22**: 1660–167.

Cloke HL, Thielen J, Pappenberger F, Norbert S, Balint G, Edlund C, Koistinen A, Saint-Aubin C, Sprookkereef E, Viel C, Salamon P, Buizza R. 2009. Progress in the implementation of Hydrological Ensemble Prediction Systems (HEPS) in Europe for operational flood forecasting. *ECMWF Newsletter* **121**: 20–24.

Dartmouth Flood Observatory. 2009. [website]: Global Active Archive of Large Flood Events. Viewed on March 1, 2009 at http://www.dartmouth.edu/~floods/Archives/index.html.

Diro GT, Grimes DIF, Black E, O'Neill A, Pardo-Iguzquiza E. 2009. Evaluation of reanalysis rainfall estimates over Ethiopia. *International Journal of Climatology* **29**: 67–78.

Duan Q, Sorooshian S, Gupta VK. 1994. Optimal use of the SCE-UA global optimization method for calibrating watershed models. *Journal of Hydrology* **158**: 265–284.

Funk C, Michaelesen J, Verdin J, Artan G, Husak G, Senay G, Gadain H, Magadazire T. 2003. The Collaborative Historical African Rainfall Model: description and evaluation. *International Journal of Climatology* **23**: 47–66.

Hamill TM, Hagedorn R, Whitaker JS. 2007. Probabilistic forecast calibration using ECMWF and GFS ensemble forecasts. Part II: precipitation. *Monthly Weather Review* **136**: 2620–2632.

He HY, Cloke HL, Wetterhall F, Pappenberger F, Freer J, Wilson M. 2009. Tracking the uncertainty in flood alerts driven by grand ensemble weather predictions. *Meteorological Applications* **16**: 91–101.

Jolliffe IT, Stephenson DB. 2003. *Forecast Verification: A Practitioner's Guide in Atmospheric Science*. Wiley and Sons: Chichester; 240.

McBride JL, Ebert EE. 2000. Verification of quantitative precipitation forecasts from operational numerical weather prediction models over Australia. *Weather and Forecasting* **15**: 103–121.

National Oceanic and Atmospheric Administration – Climate Prediction Centre (NOAA-CPC). 2002. [website]. African Rainfall Estimation Algorithm Version 2.0 – Technical description. Viewed on May 7, 2009 at http://www.cpc.ncep.noaa.gov/products/fews/RFE2.0_tech.pdf.

Palmer TN, Buizza R, Hagedorn R, Lawrence A, Leutbecher M, Smith L. 2005. Ensemble prediction – a pedagogical perspective. *ECMWF Newsletter* **106**: 10–17.

Pappenberger F, Bartholmes J, Thielen J, Cloke HL, de Roo A, Buizza R. 2008a. New dimensions in early flood warning across the globe using GRAND ensembles. *Geophysical Research Letters* **35**: L10404.

Pappenberger F, Beven KJ, Hunter N, Gouweleeuw B, Bates P, 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 Science* **9**: 381–393.

Pappenberger F, Beven KJ, Ratto M, Matgen P. 2008b. Multi-method global sensitivity analysis of flood inundation models. *Advances in Water Resources* **31**: 1–14.

Pappenberger F, Buizza R. 2009. The skill of ECMWF predictions for hydrological modelling. *Weather and Forecasting* **24**: 749–766.

Ramos M-H, Bartholmes J, Thielen-del Pozo J. 2007. Development of decision support products based on ensemble forecasts in the European flood alert system. *Atmospheric Science Letters* **8**: 113–119.

Smith, M., 1993. CLIMWAT for CROPWAT. A climatic database for irrigation planning and management. FAO Irrigation and Drainage Paper (0254–5284, no. 49), Rome, Italy.

Theron M. 2007. Climate change and increasing floods in Africa – implications for Africa's development [website]. Consultancy Africa Intelligence/Africa Watch Newsletter. Viewed on March 19, 2009 at http://www.consultancyafrica.com/africa-watch/newsletter/november-2007.

Thielen J, Bartholmes J, Ramos M-H, de Roo A. 2009. The European Flood Alert System – Part 1: concept and development. *Hydrology and Earth System Sciences* **13**: 125–140.

Thielen J, Bogner K, Pappenberger F, Kalas M, Del Medico M, de Roo A. 2009b. Monthly-, medium-, and short-range flood warning: testing the limits of predictability. *Meteorological Applications* **16**: 77–90.

Uppala S, Kallberg PW, Hernandez A, Saarinen S, Fiorino M., Li X, Onogi K, Sokka N, Andrae U, da Costa Brechtold V. 2004. ERA-40: ECMWF 45-year reanalysis oft the global atmospheric and surface conditions 1957–2002. *ECMWF Newsletter* **101**: 2–21.

Van der Knijff, de Roo APJ. 2008. *LISFLOOD. Distributed Water Balance and Flood Simulation Model – Revised User Manual*. European Commission [EUR 22166 EN/2] Joint Research Centre, Institute for Environment and Sustainability: Ispra, Italy.

Vitart F, Buizza R, Alonso Balmasemo M, Balsamo G, Bidlot JR, Bonet A, Fuentes M, Hofstadler A, Molteni F, Palmer TN. 2008. The new VAREPS-monthly forecast system: a first step towards seamless prediction. *The Quarterly Journal of the Royal Meteorological Society* **134**: 1789–1799.

