



HAL
open science

The reduction continuous rank probability score for evaluating discharge forecasts from hydrological ensemble prediction systems

Bao N. Trinh, Jutta Thielen, Guillaume Thirel

► **To cite this version:**

Bao N. Trinh, Jutta Thielen, Guillaume Thirel. The reduction continuous rank probability score for evaluating discharge forecasts from hydrological ensemble prediction systems. *Atmospheric Science Letters*, 2013, 14 (2), pp.61-65. 10.1002/asl2.417. hal-01493230

HAL Id: hal-01493230

<https://hal.science/hal-01493230v1>

Submitted on 21 Mar 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

The reduction continuous rank probability score for evaluating discharge forecasts from hydrological ensemble prediction systems

B. N. Trinh,¹ J. Thielen-del Pozo^{1*} and G. Thirel^{1,2}

¹European Commission, Joint Research Centre, Institute for Environment and Sustainability, 21020, Ispra, Italy

²IRSTEA, 92761, Antony Cedex, France

*Correspondence to:

J. Thielen-del Pozo, European Commission, DG Joint Research Centre, Institute for Environment and Sustainability, 21020 Ispra, Italy. E-mail: jutta.thielen@jrc.ec.europa.eu

Abstract

Ensemble prediction system (EPS) meteorological forecasts are increasingly being used as input to hydrological models in order to extend flood-warning lead-times and to improve forecasts and the knowledge of their uncertainty. Probabilistic skill scores classically used in meteorology are usually also utilized for assessing the quality of hydrological ensemble forecasts. However, the different river discharge magnitudes can make difficult the interpretation and comparisons of these scores, as it is for the continuous rank probability score (CRPS). In this letter, a novel 'Reduction' CRPS (RCRPS), which takes into account the different river discharge magnitudes, is proposed, and its usefulness is exhibited. Copyright © 2013 Royal Meteorological Society

Received: 16 May 2012
Revised: 24 September 2012
Accepted: 15 October 2012

Keywords: flood forecasting; skill scores; discharge; probabilistic forecasting; ensemble prediction system

1. Introduction

Hydrological forecasting systems increasingly use meteorological forecasts from ensemble prediction systems (EPS) as inputs (Schaake *et al.*, 2007; Bogner and Kalas, 2008; Pappenberger *et al.*, 2008; Thielen *et al.*, 2008; Cloke and Pappenberger, 2009; Cloke *et al.*, 2009; Pappenberger and Buizza, 2009). Such hydrological ensemble prediction system (HEPS) forecasts often have both a deterministic and probabilistic component (Thielen *et al.*, 2009a, 2009b). The deterministic forecast provides a single high resolution estimate, most useful for accurate spatial prediction at short lead-times. The probabilistic forecast provides additional information, i.e. the indication of the possibility of an extreme event, an estimate of forecast uncertainty, possibility of extending flood warning lead-times to the medium-range, and the probabilistic information for the transformation into cost-loss and hedging options in decision making (Gouweleeuw *et al.*, 2005; Laio and Tamea, 2007; Roulin, 2007; Verbunt *et al.*, 2007).

However, evaluating probabilistic HEPS can be a problem due to the different magnitude order of rivers discharges. A similar problem of the evaluating probabilistic forecast by the Spread of ensemble and root mean squared error (RMSE) scores has already been pointed out by Randrianasolo *et al.* (2010) and Thirel *et al.* (2010). Regarding the continuous rank probability score (CRPS), a probabilistic skill-score often applied in meteorology, Laio and Tamea (2007) pointed out this problem while they used it on different catchments.

To overcome this problem and to have a probabilistic skill-score independent on the magnitude of the studied variable, the Reduction CRPS (RCRPS) is introduced in this letter. This score is based on a normalization of the CRPS by the standard deviation of river discharges and preserves the noteworthy properties of the CRPS (Hersbach, 2000; Candille *et al.*, 2007; Laio and Tamea, 2007; Candille, 2009), namely:

1. It is sensitive to the entire permissible range of the parameter of interest;
2. Its definition does not require the introduction of a number of predefined classes, on which results may depend;
3. It can be interpreted as an integral over all possible Brier scores (Brier, 1950);
4. For a deterministic forecast, the CRPS is equal to the mean absolute error (MAE) and therefore, has a clear interpretation, and thus this score permits the comparison of the deterministic and the probabilistic forecast of an HEPS. This quality is necessary for decision making for earlier flood alerts, when the weighting of different components of an HEPS needs to be optimized. In the case of the RCRPS, it is equal to Ratio-MAE for deterministic forecasts;
5. It can be interpreted as an average value of Expected Cost (the cost/loss ratio score is mentioned in Laio and Tamea, 2007);
6. It permits the comparison of two (H)EPSs.

In Section 2 of this letter, we present a resume of the CRPS and introduce the definition of Reduction CRPS. In Section 3, the test bed and HEPS datasets

used for this study are described. In Section 4, we carry out the evaluation of the HEPS discharge forecast in order to demonstrate best practice in the application of skill scores and discuss the results. Conclusions and a proposal for future work are drawn in Section 5.

2. Continuous rank probability score and its adaptation for hydrological applications, the reduction CRPS

The CRPS (Stanki *et al.*, 1989) measures the distance between the predicted and the observed cumulative distribution functions (CDFs) of scalar variables. Hersbach (2000) proposed a reliability-resolution-uncertainty decomposition of the CRPS. This decomposition is convenient for practical computations, especially for a short observational period, and gives objective criteria for comparing two EPS forecasts, which is necessary when deciding whether to implement a new or revised EPS (Candille *et al.*, 2007; Candille, 2009). Also, for hydrological applications the benefit of decomposition for the most appropriate forecast evaluation has been demonstrated (Bartholmes *et al.*, 2009).

Thus while very useful, the dependence on magnitude of the data reference makes spatial comparisons across catchments difficult. To overcome the disadvantage of the magnitude order dependence of the CRPS, we propose a variant of this score, the 'Reduction' CRPS (RCRPS). The RCRPS is defined by the reduction of the standard deviation σ_0 (or statistical normalization) obtained for each hydrological station over the study time period:

$$\overline{\text{RCRPS}} = \frac{\overline{\text{CRPS}}}{\sigma_0} \quad (1)$$

This equation is demonstrated by the normalization of the variable and the calculation of the integration on the probability spaces (Ω, ρ_X) and (Ω_Y, ρ_Y) .

$$y = \frac{x - m_0}{\sigma_0}; \rho_Y(y) = \frac{1}{\sigma_0} \rho_X\left(\frac{x - m_0}{\sigma_0}\right);$$

$$P(x) = \int_{-\infty}^x \rho(z) dz$$

$$P_a(x) = H(x - x_a) = \begin{cases} 0 & \text{for } x < x_a \\ 1 & \text{for } x \geq x_a \end{cases}$$

Then $P_a(y) = H(y - y_a) = H(x - x_a) = P_a(x)$

$$E \left\{ \int_{\Omega_Y} [P(y) - P_a(y)]^2 dy \right\}$$

$$= E \left\{ \int_{\Omega_Y} \left[P\left(\frac{x - m_0}{\sigma_0}\right) - P_a\left(\frac{x - m_0}{\sigma_0}\right) \right]^2 \left(\frac{dx}{\sigma_0}\right) \right\}$$

$$= \frac{1}{\sigma_0} E \left\{ \int_{\Omega} [P(x) - P_a(x)]^2 dx \right\} \quad (2)$$

where m_0 is the mean and σ_0 is the standard deviation of all observational hydrological data.

We recognize that this normalization is a numerical normalization, and therefore preserves all properties of the CRPS mentioned in the introduction and the decomposition of Hersbach (2000). And we note also that this score RCRPS is dimensionless and the error made in the larger a bin width (magnitude order) of CRPS is normalized by the standard deviation of the observational hydrological data.

Consequently, the RCRPS allows the averaging over all hydrological stations in the same catchment and to compare the skill score of different catchments for the deterministic and probabilistic forecast across space and time.

3. Description of the HEPS discharge dataset

3.1. Test bed

The Danube catchment has been selected as a study area in order to demonstrate best practice in the application of the skill scores for HEPS discharge predictions. It has an area of 817 000 km² and is the 25th biggest drainage area of the Earth. It is located in the southeast of Europe, originates from the Alps and Carpathians, and drains from 19 different countries.

The Danube can be divided into three sub-regions (Figure 1): the upper basin (upstream of Bratislava station, fast responding, and snow melting), the middle basin (upstream of Pancevo station) including the Tisza and Sava, and the lower basin (including the Danube Delta: Ceatal Izmail station, slow responding). Figure 1 shows the relevant sub-basins of the Danube Catchment and the eight hydrological stations used in this study.

3.2. Case study

The hydrological model setup for the simulations and forecasts is obtained from the European Flood Awareness System (EFAS), which is described in detail in Thielen *et al.* (2009b) and Bartholmes *et al.* (2009). EFAS has been used in the past for case studies in the Danube (Kalas *et al.*, 2008; Pappenberger *et al.*, 2008; Thielen *et al.*, 2009a). The hydrological model of EFAS is Lisflood, a semi-physically based rainfall-runoff-routing model designed to simulate hydrological processes that occur at medium to large scales (Van Der Knijff *et al.*, 2010).

The period studied is 01 November 2007 to 31 October 2008, representing one hydrological year. This is too short to draw definite conclusions about the long-term quality of any HEPS discharge forecasts (see discussion in Cloke and Pappenberger, 2009), however, several studies on the HEPS have already been realized on similar periods (19 months for Thirel *et al.*, 2008 and 16 months for Brochero *et al.*, 2011) and we use this as a first step in the comparison of

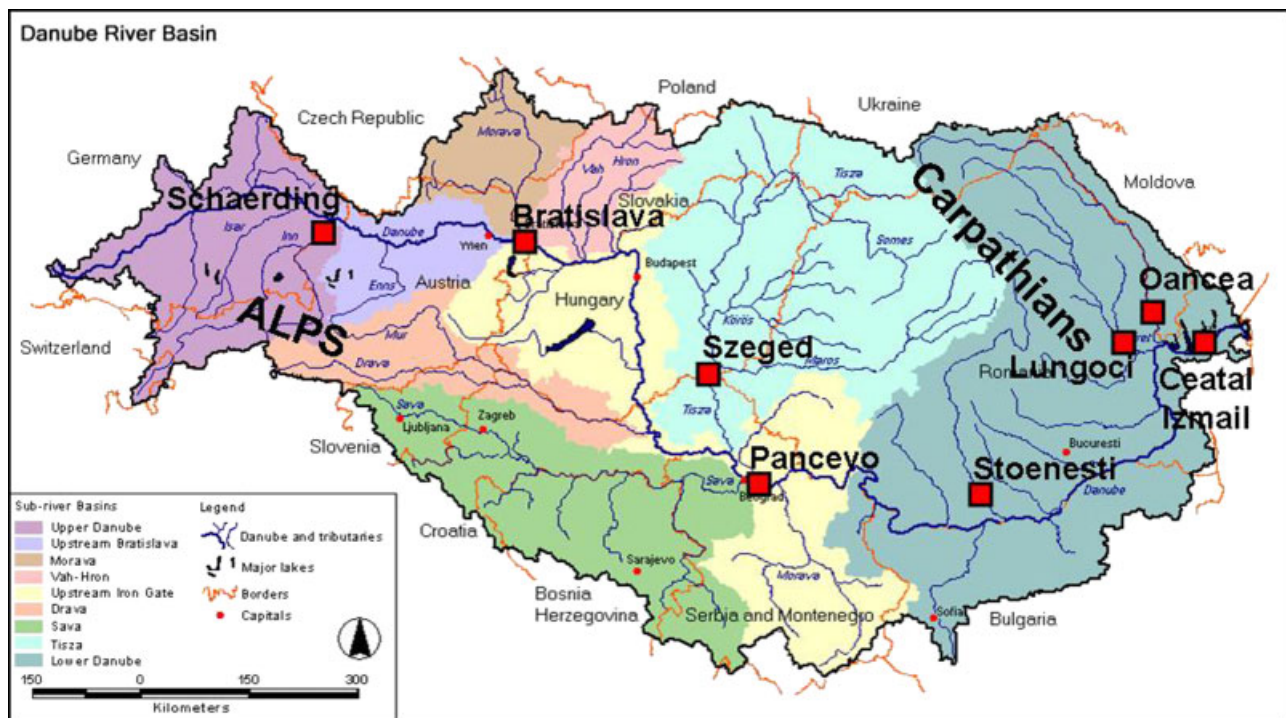


Figure 1. Danube catchment with eight hydrological stations: Schaerding (Inn River), Bratislava (upper part of the main Danube), Szeged (Tisza River), Pancevo (middle part of the main Danube), Stoenesti (Olt River), Lungoci (Siret River), Oancea (Prut River), and Ceatal Izmail (close to the outlet of the Danube) [based on IDPCR (http://www.icpdr.org/icpdr-pages/river_basin.htm)].

skill forecast scores at various stations representing different hydrologic regimes.

4. Evaluation of the HEPS discharge forecasts by RCRPS and CRPS

Table I shows information about the eight hydrological stations selected above (Figure 1) and their corresponding highest simulated discharge (Q Max), highest CRPS over the complete lead-time (CRPS Max), and highest RCRPS over the complete lead-time (RCRPS Max).

The correlation coefficient between the Q Max and the CRPS Max is $R = 0.9355$, whereas between the Q Max and the RCRPS Max it is $R = -0.7167$, for the eight hydrological stations presented in Table I. This shows the RCRPS Max is less dependent on Q Max than the CRPS Max. To show how the dependency of RCRPS Max on Q Max is reduced, we used

211 hydrological stations available in this study area (not shown in Figure 1). The correlation coefficient between the Q Max and the CRPS Max is $R = 0.9432$ and between the Q Max and the RCRPS Max is $R = -0.5832$ for these 211 hydrological stations. And Figure 2 shows the linear dependency of the CRPS Max and RCRPS Max on the Q Max (Maximum of the discharge) of all hydrological stations in this study area.

That demonstrates the lower dependence of the Reduction CRPS on the magnitude order of reference discharge (given here through Q Max). It makes the RCRPS more suitable than the CRPS for HEPS discharge forecasts evaluation.

The comparison of the CRPS and RCRPS is presented in Figure 3 in more details for four selected stations in the Danube catchment. Their evolution with the lead-time (days) of the forecasts is given. Note that both panels of Figure 3 are not using the same y-scale.

Table I. The river basin, upstream area, maximum discharge, CRPS and RCRPS of eight hydrological stations.

Hydrological stations	River basin	Upstream area (km ²)	Q Max (m ³ s ⁻¹)	CRPS Max	RCRPS Max
Stoenesti	Olt	24 125	440	35	0.45
Lungoci	Siret	36 025	595	43	0.46
Schaerding	Inn	26 625	950	50	0.22
Oancea	Prut	27 300	1 120	46	0.32
Bratislava	Upper Danube	132 775	2 685	160	0.36
Szeged	Tisa	139 475	3 070	90	0.18
Pancevo	Middle Danube	519 450	7 985	280	0.19
Ceatal Izmail	Lower Danube	791 400	11 040	260	0.14

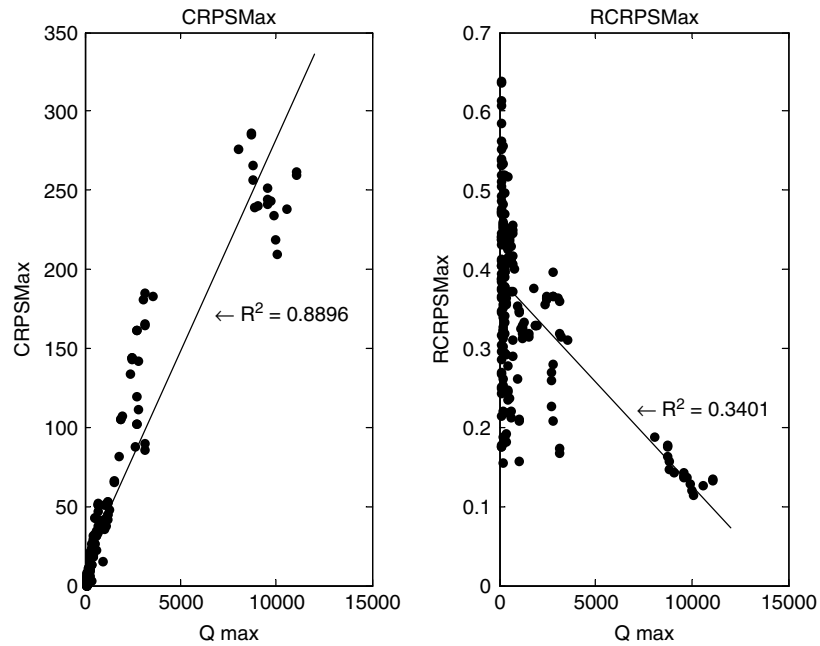


Figure 2. The linear dependency of the CRPS Max and RCRPS Max on the Q Max (maximum of the discharge).

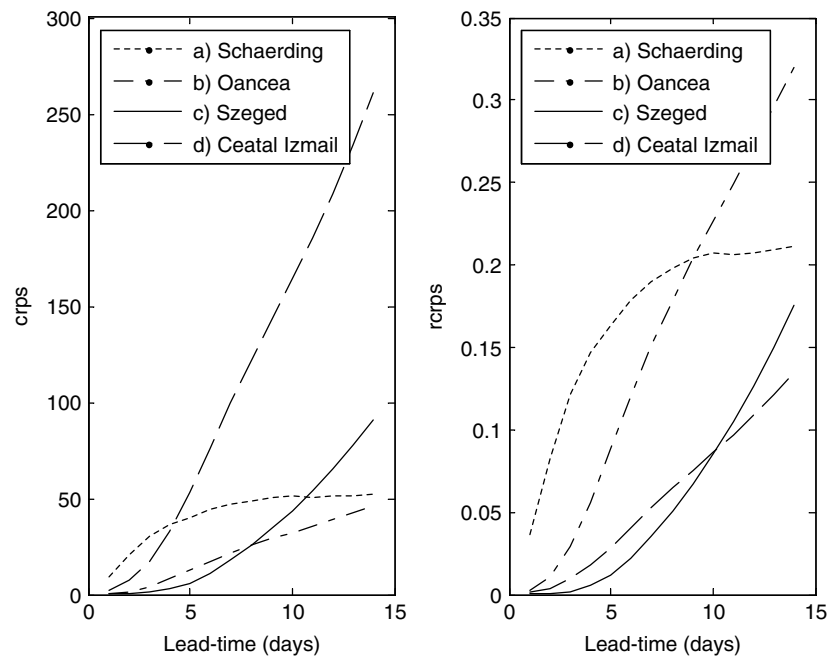


Figure 3. Curves of the score CRPS (left panel) and RCRPS (right panel) at stations: Schaerding (dotted line), Oancea (dash-dot line), Szeged (solid), and Ceatal Izmail (dashed).

Figure 3, (left panel) shows that the lowest CRPS score value is observed at Oancea, the largest is at Ceatal Izmail. However, since this score depends clearly on the magnitude order of discharges and it is not possible to compare these four curves altogether. Therefore, neither is averaging the CRPS across different stations possible, nor the comparison of the CRPS for different stations.

In contrast, on the right panel of Figure 3, the RCRPS is shown: in contrast to the CRPS on the left hand side, the RCRPS cannot be linked easily to the Q Max of the stations. It is in decreasing order

of performance for: Szeged, Ceatal Izmail, Oancea, and Schaerding for the lead-times lower than 10 days. This shows the lower dependence of the RCRPS with regards to Q Max and that this score preserves the shape of the curves with lead-time that were observed with the CRPS, for each station.

5. Conclusions and future work

For hydrological applications where scores for different stations across a river basin are to be compared or

averaged, the CRPS should be replaced with the normalized Reduction CRPS. It permits the comparison of ensemble discharge forecasts for two or more hydrological stations and to allow averaging scores over diverse hydrological stations. The Reduction CRPS maintains the beneficial properties of CRPS and it permits the comparison of deterministic and probabilistic HEPS discharge forecasts with the Reduction MAE and Reduction CRPS. Following this evaluation, the Reduction CRPS will now be used within EFAS in order to make forecast evaluation easier to interpret for the different types of ensemble discharge forecasts used: deterministic (ECMWF and DWD) and probabilistic (ECMWF EPS and COSMO-LEPS). We also note that this probabilistic score has been checked with different distributions and different parameters (not shown here). That means that this probabilistic score has been validated for the different environmental datasets with different magnitude orders.

We suggest that future work should consider observational error (between the reference discharge and the actual observed discharge) using the score reduce centered random variable (RCRV: Candille and Talagrand, 2008) in order to give a skill-score of the use of the discharge reference. We also recommend that future work should attempt to classify a wide range of skill Reduction CRPS forecasts in order to know the level of skill-score forecast by this Reduction CRPS.

Acknowledgements

We thank F. Pappenberger, H. Cloke, and K. Bogner for their substantial contribution to the revision of our manuscript and we also thank the complete EFAS team for help and assistance, especially M. Del Medico, D. Nappo, and H. Sint. The national providers of hydrological and meteorological data for EFAS, and thus this study, are being acknowledged.

References

- Bartholmes JC, Thielen J, Ramos MH, Gentilini S. 2009. The European Flood Alert System EFAS – Part 2: Statistical skill assessment of probabilistic and deterministic operational forecasts. *Hydrology and Earth System Sciences* **13**(2): 141–153.
- Brochero D, Anctil F, Gagne C. 2011. Simplifying a hydrological ensemble prediction system with a backward greedy selection of members – Part 1: optimization criteria. *Hydrology and Earth System Sciences* **8**: 2739–2782.
- Bogner K, Kalas M. 2008. Error-correction methods and evaluation of an ensemble based hydrological forecasting system for the Upper Danube catchment. *Atmospheric Science Letters* **9**(2): 95–102.
- Brier GW. 1950. Verification of forecasts expressed in terms of probability. *Monthly Weather Review* **78**: 1–3.
- Candille G. 2009. The multi-ensemble approach: the NAEFS example. *Monthly Weather Review* **137**(5): 1655–1665.
- Candille G, Talagrand O. 2008. Impact of observational error on the validation of ensemble prediction systems. *Quarterly Journal of the Royal Meteorological Society* **134**(633): 959–971, DOI: 10.1002/qj.268.
- Candille G, Cote C, Houtekamer PL, Pellerin G. 2007. Verification of an ensemble prediction system against observations. *Monthly Weather Review* **135**(7): 2688–2699.
- Cloke HL, Pappenberger F. 2009. Ensemble flood forecasting: a review. *Journal of Hydrology* **375**(3–4): 613–626.
- Cloke HL, Thielen del Pozo J, Pappenberger F, Nobert S, Balint G, Edlund C, Koistinen A, De Saint-Aubin C, Sprokkereef E, Viel C, Salomon 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.
- Gouweleeuw BT, Thielen J, Franchello G, De Roo APJ, Buizza R. 2005. Flood forecasting using medium-range probabilistic weather prediction. *Hydrology and Earth System Sciences* **9**(4): 365–380.
- Hersbach H. 2000. Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather and Forecasting* **15**: 559–570.
- Kalas M, Ramos MH, Thielen del Pozo J, Babiakova G. 2008. Evaluation of the medium-range European flood forecasts for the March–April 2006 flood in the Morava river. *Journal of Hydrology and Hydromechanics* **56**(2): 116–132.
- Laio F, Tamea S. 2007. Verification tools for probabilistic forecasts of continuous hydrological variables. *Hydrology and Earth System Sciences* **11**: 1267–1277.
- Pappenberger F, Buizza R. 2009. The skill of ECMWF predictions for hydrological modelling. *Weather and Forecasting* **24**(3): 749–766.
- Pappenberger F, Scipal K, Buizza R. 2008. Hydrological aspects of meteorological verification. *Atmospheric Science Letters* **9**(2): 43–52.
- Randrianasolo A, Ramos MH, Thirel G, Andreassian V, Martin E. 2010. Comparing the scores of hydrological ensemble forecasts issued by two different hydrological models. *Atmospheric Science Letters* **11**: 100–107.
- Roulin E. 2007. Skill and relative economic value of medium-range hydrological ensemble predictions. *Hydrology and Earth System Sciences* **11**(2): 725–737.
- Schaake JC, Hamill TM, Buizza R, Clark M. 2007. HEPEX: the hydrological ensemble prediction experiment. *Bulletin of the American Meteorological Society* **88**(10): 1541–1547. Retrieved from <http://journals.ametsoc.org/doi/abs/10.1175/BAMS-88-10-1541>.
- Stanki HR, Wilson LJ, Burrows WR. 1989. Survey of common verification methods in meteorology, Technical report No. 8 Geneva: WMO.
- Thielen J, Bogner K, Pappenberger F, Kalas M, del Medico M, de Roo A. 2009a. Monthly-, medium-, and short-range flood warning: testing the limits of predictability. *Meteorological Applications* **16**: 77–90.
- Thielen J, Bartholmes J, Ramos M-H, de Roo A. 2009b. The European Flood Alert System – Part 1: concept and development. *Hydrology and Earth System Sciences* **13**(2): 125–140.
- Thielen J, Schaake J, Hartman R, Buizza R. 2008. Aims, challenges and progress of the Hydrological Ensemble Prediction Experiment (HEPEX) following the third HEPEX workshop held in Stresa 27 to 29 June 2007. *Atmospheric Science Letters* **9**(2): 29–35.
- Thirel G, Rousset-Regimbeau F, Martin E, Habets F. 2008. On the impact of short-range meteorological forecasts for ensemble streamflow predictions. *Journal of Hydrometeorology* **9**: 1301–1317.
- Thirel G, Martin E, Mahfouf J-F, Massart S, Ricci S, Regimbeau F, Habets F. 2010. A past discharge assimilation system for ensemble streamflow over France – Part 2: impact on the ensemble streamflow forecasts. *Hydrology and Earth System Sciences* **14**: 1639–1653.
- Van Der Knijff JM, Younis J, De Roo APJ. 2010. LISFLOOD: a GIS-based distributed model for river basin scale water balance and flood simulation. *International Journal of Geographical Information Science* **24**(2): 189–212.
- Verbunt M, Walser A, Gurtz J, Montani A, Schar C. 2007. Probabilistic flood forecasting with a limited-area ensemble prediction system: selected case studies. *Journal of Hydrometeorology* **8**(4): 897–909.