Diagnostic verification of hydrometeorological and hydrologic ensembles

Journal:	*Atmospheric Science Letters*
Manuscript ID:	Draft
Wiley - Manuscript type:	Original Manuscript
Date Submitted by the Author:	
Complete List of Authors:	Demargne, Julie; National Oceanic and Atmospheric Administration, National Weather Service, Office of Hydrologic Development; University Corporation for Atmospheric Research Brown, James; National Oceanic and Atmospheric Administration, National Weather Service, Office of Hydrologic Development; University Corporation for Atmospheric Research Liu, Yuqiong; National Oceanic and Atmospheric Administration, National Weather Service, Office of Hydrologic Development; Riverside Technology, Inc. Seo, Dong-Jun; National Oceanic and Atmospheric Administration, National Weather Service, Office of Hydrologic Development; University Corporation for Atmospheric Research Wu, Limin; National Oceanic and Atmospheric Administration, National Weather Service, Office of Hydrologic Development; Wyle Information Systems Toff, Zoltan; National Oceanic and Atmospheric Administration, Earth System Research Laboratory Zhu, Yuejian; National Oceanic and Atmospheric Administration, National Weather Service, National Centers for Environmental Prediction
Keywords:	ensembles, hydrological forecasting, probabilistic verification, uncertainty

Diagnostic verification of hydrometeorological and hydrologic ensembles

Julie Demargne [1,2], James Brown[1,2], Yuqiong Liu[1,3], Dong-Jun Seo[1,2], Limin Wu[1,4], Zoltan Toth[5] and Yuejian Zhu[6]

[1] *National Oceanic and Atmospheric Administration, National Weather Service, Office of Hydrologic Development*
[2] *University Corporation for Atmospheric Research*
[3] *Riverside Technology, Inc.*
[4] *Wyle Information Systems*
[5] *National Oceanic and Atmospheric Administration, Earth System Research Laboratory*
[6] *National Oceanic and Atmospheric Administration, National Weather Service, National Centers for Environmental Prediction*

Keywords: ensembles, hydrological forecasting, probabilistic verification, uncertainty

Abstract

This paper presents a diagnostic verification case study of experimental precipitation and streamflow ensemble reforecasts over a 24-year period, using the Ensemble Verification System (EVS). The results show the improvement in forecast skill, and more significantly in forecast reliability, by using Numerical Weather Prediction (NWP) single-valued forecasts in an ensemble preprocessor for ensemble streamflow prediction. Results also yield insight into the relative contribution of hydrologic uncertainty in comparison to the atmospheric uncertainty. The EVS is proposed as a flexible and modular tool for the HEPEX verification test-bed to evaluate existing and emerging verification methods that are appropriate for hydrologic applications.

1. Introduction

Atmospheric and hydrologic forecasts are subject to uncertainty, which needs to be systematically quantified and effectively communicated to users (NRC, 2006). A common approach to provide such information in an operational setting is to generate ensemble forecasts from which probabilistic statements are issued (e.g., see examples of operational flood forecasting systems based on weather ensemble inputs in (Cloke and Pappenberger, 2009)). Hydrologic ensembles and their corresponding hydrometeorological forecasts need to be *routinely* verified to improve both research and operations (Welles *et al.*, 2007). However forecast verification in hydrology has been limited to date, although a number of verification case studies with hydrologic ensembles have been published (see references quoted in Cloke and Pappenberger, 2009). Furthermore the meteorology and hydrology communities need to closely collaborate to define verification metrics and practices that are appropriate for hydrological applications (Pappenberger *et al.*, 2008). Such forecast verification needs to include two activities (Demargne *et al.*, 2009): 1) diagnostic verification performed by scientists and forecasters to monitor forecast quality over time, analyze the different sources of uncertainty and skill across the entire river forecasting process, and evaluate forecast skill

improvement from new science and technology; 2) real-time verification, which aims to communicate along with real-time forecasts (and before the corresponding observations occur), verification information relative to historical analogue forecasts to assist operational forecasters and end-users in their decision making.

The Office of Hydrologic Development (OHD) of the National Oceanic and Atmospheric Administration (NOAA) National Weather Service (NWS) has been developing various capabilities for the Hydrologic Ensemble Forecast Service (HEFS) to provide river ensemble forecasts for a wide range of spatio-temporal scales, from hours for flash flood forecasts at local scale, to months for water supply forecasts at regional scale. The Ensemble Verification System (EVS) developed by Brown *et al.* (2009) is the diagnostic verification component for the HEFS to verify ensemble forecasts of any continuous numeric variables, produced at discrete locations and for any forecast horizon and time step. The OHD and the National Centers for Environmental Prediction (NCEP) are currently collaborating to improve the climate, weather and river forecasts at the catchment scale for the HEFS and define standard verification metrics and products that are meaningful for water applications.

In this paper, the metrics and the EVS software used in a diagnostic verification case study of hydrologic ensemble forecasts are introduced. Verification results are presented for the experimental HEFS precipitation and streamflow ensembles from a 24-year period to analyze the impact of the atmospheric uncertainty on the quality of the hydrologic ensembles. Finally, future work and on-going collaborations to advance ensemble verification in operational river forecasting are described.

2. Diagnostic verification of hydrologic forecasts

The quality of forecast ensembles includes several attributes (Wilks, 2006), such as reliability, resolution, discrimination, and skill. Therefore, a variety of verification metrics need to be concurrently analyzed in hydrologic forecast verification, as it is reported for atmospheric forecast verification. The main metrics presented in the verification case study are briefly described hereafter (see further details in Jolliffe and Stephenson, 2003, Wilks, 2006, and Brown *et al.*, 2009).

Two metrics from single-valued forecast verification are used to verify the ensemble means: the mean error to measure how the "best single-valued estimate" from the ensemble forecast agrees with the observed outcome on average, and the correlation coefficient between the ensemble mean and the corresponding observation. The Continuous Ranked Probability Score (CRPS) quantifies the overall forecast quality as the expected squared error of the forecast probabilities for all possible events and it is averaged across the observed-forecast pairs. Its associated skill score, the Continuous Ranked Probability Skill Score (CRPSS), measures the forecast skill (in terms of CRPS) above a given reference forecast to show the usefulness of the forecasting system. It ranges from $-\infty$ to 1, with perfect skill of 1 and negative value when the forecast has worse CRPS than the reference. The CRPS decomposition (Hersbach, 2000) is performed similarly to the Brier Score decomposition (Murphy, 1973) to provide further details

about the forecast performance. The reliability component of the CRPS measures the average reliability of the ensemble forecasts similarly to the rank histogram. Specifically it tests whether the fraction of observations that fall below the *k*-th of *n* ranked ensemble members is equal to *k*/*n* on average. The second component of the CRPS, called the Potential CRPS, represents the CRPS one would obtain when the forecasting system would become perfectly reliable (i.e., Reliability of CRPS = 0). It is sensitive to the average ensemble spread and the frequency and magnitude of the outliers. For best potential CRPS, the forecasting system needs narrow ensemble spread on average without too many and too high ensemble outliers (Hersbach, 2000). The CRPS, the Reliability of CRPS, and the Potential CRPS are all negatively oriented, with perfect score of 0. Finally, the Relative Operating Characteristic (ROC) score is used to describe the ability of the forecasts to discriminate between events and non-events, on average. The ROC curve plots the probability of detection against the probability of false detection for a range of probability levels and for a given event (such as flooding). The ROC score is defined as the area below the ROC curve and above the diagonal, with a perfect score of 1, measuring the average gain in discrimination over climatological forecasts for all probability levels. All these verification metrics were computed in this work for the hydrologic forecasts and their corresponding atmospheric forecasts to describe the different aspects of forecast quality. More detailed statistics (e.g., reliability diagrams) were also examined but discussions of these are not included in this paper.

Hydrologic ensemble forecasts need to account for the atmospheric uncertainty and the hydrologic uncertainty, which includes uncertainty in the initial conditions, the model parameters and the model structure. To analyze the relative importance of the two sources of uncertainty, streamflow ensemble forecasts are verified with the observed flows and with the simulated flows that are produced from the observed hydrometeorological inputs using the same model and the same initial conditions. The verification of streamflow ensembles with observed flows leads to the computation of the total error, including the contribution of the atmospheric uncertainty and the hydrologic uncertainty. The verification with simulated flows allows for the contribution of the atmospheric uncertainty (in the hydrometeorological forecasts) to be diagnosed, assuming that uncertainties in the observed hydrometeorological inputs are much smaller than the hydrologic uncertainty.

Regarding the EVS software used in this diagnostic verification analysis, the main features are summarized below; a detailed description is provided in (Brown *et al.*, 2009). EVS can perform temporal aggregation (e.g., daily total flows aggregated from 6-hourly instantaneous flows) and data stratification to define subsets of forecast-observed pairs depending on the time of interest (e.g., winter months) and/or conditions defined from the variables being verified (e.g., exceedance thresholds). EVS can aggregate the verification statistics produced across different locations based on a user-defined weighted average, in order to easily report forecast quality on larger areas. Finally EVS produces graphics and numerical results of the verification statistics, including graphics with modified box-and-whisker plots of errors in the ensemble members. The EVS software is developed within the NOAA's Community Hydrologic Prediction System to allow cost-effective collaborative research and development with academic and private institutions and rapid

research-to-operations transition of scientific advances. EVS is intended to be flexible, modular and open to accommodate enhancements for both research and operational forecasting purposes. It is planned to become available on line to support collaborative work such as the Hydrological Ensemble Prediction Experiment (HEPEX) verification test-bed project (http://hydis8.eng.uci.edu/hepex/testbeds/Verification.htm).

3. Verification case study

The verification case study concerns experimental ensemble hindcasts of precipitation and streamflow generated with the current HEFS prototype. The precipitation ensembles (as well as temperature ensembles) are generated from single-valued forecasts by the NWS Ensemble Preprocessor (EPP) (Schaake *et al.*, 2007). The EPP aims to remove the bias in the NWP single-valued forecasts while capturing the skill and uncertainty therein. The EPP estimates the joint distribution of single-valued forecasts and observations based on historical pairs. Ensemble members are sampled from the conditional probability distribution of the observations given a particular single-valued forecast. The Schaake Shuffle technique (Clarke *et al.*, 2004) is applied to approximately reconstruct the space-time statistical properties of the precipitation and temperature variables for multiple lead times and locations based on historical observations. When no single-valued forecast is available, EPP estimates the climatological distribution from the historical observations and applies the Schaake Shuffle to the values sampled from the distribution. The resulting ensembles, called resampled climatological ensembles, are used as reference forecasts to analyze the skill in the ensembles derived from the NWP single-valued forecasts.

The hydrometeorological ensemble hindcasts produced by the EPP are ingested into the Hydrologic Ensemble Hindcaster (HEH) (Demargne *et al.*, 2007) to produce corresponding streamflow ensemble hindcasts based on various hydrological models. The HEH retrospectively generates the initial conditions of the hydrological models for each hindcast date. These retrospective initial conditions may not reflect the initial conditions used in real-time forecasting, which are usually modified by the forecasters based on their expertise, or by data assimilation techniques (for which further evaluation is under way). However, this hindcast process supports the analysis of the impact of the atmospheric ensembles on the quality of hydrologic ensembles. Two sets of streamflow ensembles are generated: one using the EPP ensembles derived from the NWP single-valued forecasts, the other using the EPP resampled climatological ensembles, to analyze the skill in the streamflow forecasts when incorporating information from the NWP single-valued forecasts. These two sets of hydrologic ensembles account only for the atmospheric uncertainty, the hydrologic uncertainty being quantified by other components of the HEFS.

The verification study was performed for the North Fork of the American River above the North Fork Dam (USGS stream gauge station ID 11427000) near Sacramento in California. The NWP single-valued forecasts were obtained from the ensemble means of the precipitation and temperature reforecasts from the frozen version (circa 1998) of the NCEP's Global Forecast System (GFS) for 14 days into the future (Hamill *et al.*, 2006). The EPP produced 6-hourly mean areal precipitation and mean areal temperature

ensemble hindcasts at 12:00 UTC, from which the HEH generated 6-hourly streamflow ensembles for 14 days of forecast horizon to mimic the operational forecasting process. These hindcasts were produced for a period of almost 24 years from 1 January 1979 to 30 September 2002, each hindcast containing 55 ensemble members. The EPP resampled climatological ensembles and the corresponding climatology-based streamflow ensembles were also produced as reference forecasts. The EPP was calibrated using the forecasts and observations from the same period; independent verification analysis is currently being conducted. The precipitation forecasts were aggregated in EVS to be verified as daily totals using precipitation observations. The precipitation verification statistics were also aggregated across two precipitation sub-areas. The 6-hourly flow forecasts were aggregated to daily averages to be verified with the USGS streamflow measurements that were available only at daily time step. To assess the relative contribution of the atmospheric and hydrologic uncertainties in the streamflow forecasts, the 6-hourly flow forecasts were also verified with the 6-hourly flow simulations generated from the observed hydrometeorological inputs using the same model and the same initial conditions.

Verification statistics were computed using the whole 24-year period to verify, with sufficiently large sample sizes, the forecast performance for high events (defined by thresholds on the observed sample), which is critical for operational forecasting. Work is under way to estimate the confidence intervals of the verification metrics based on a bootstrapping approach to account for the sampling uncertainty. A preliminary assessment of confidence intervals for this case study (not shown) showed that sampling uncertainty becomes significant after Day 10 (especially for the higher thresholds), rendering it difficult to draw any meaningful conclusions regarding the differences in forecast quality between the climatology-based ensembles and the GFS-based ensembles for these long forecast horizons.

4. Results

The daily precipitation totals are verified for all the forecast-observed pairs (8660 pairs for the first 24-hour lead time) and for different subsets of pairs defined by the observation exceeding 0 mm (i.e., probability of precipitation, PoP), 1 mm, 5 mm, 12.5 mm, 25 mm, and 50 mm. The last three thresholds correspond to non-exceedance probabilities of approximately 0.9, 0.94 and 0.98, respectively.

In Fig. 1, the mean error and the correlation coefficient of the ensemble means, as well as the CRPS reflect the decreasing forecast quality with increasing lead time and with increasing observed precipitation amount for the GFS-based precipitation ensembles. Regarding the CRPSS, the GFS-based ensembles have more skill than the resampled climatological ensembles at all lead times, with a larger gain for high precipitation events (above 12.5 mm) compared to low precipitation events. The skill score is slightly negative for the lower thresholds (when excluding the no-rain events) after Day 9, showing that the GFS-based ensembles are not skillful for the small precipitation events beyond this forecast horizon. However the GFS-based ensembles clearly outperform the resampled climatological ensembles for the prediction of PoP at all lead times.

The box-and-whisker plot given in Fig. 2 for the 24-hour lead time gives the distribution of the errors in the ensemble members by increasing observed precipitation amount. The forecast error (ensemble member – observation) is represented with a box-and-whisker diagram for the 0, 20, 40, 60, 80 and 100 percentiles of the forecast error distribution, the box corresponding to the 20-80 percentiles. The GFS-based precipitation ensembles exhibit a large conditional bias that increases with forecast lead time, as the mean error on Fig. 1 also indicates: they tend to over-forecast small precipitation amounts and under-forecast large precipitation amounts.

In Fig. 3, verification statistics for the GFS-based precipitation ensembles and the resampled climatological ensembles are compared against each other with respect to the Reliability of CRPS, the Potential CRPS, and the ROC Score. The statistics are also presented for two subsets of forecast-observed pairs. The GFS-based ensembles exhibit very good reliability at all lead times for all the forecast-observed pairs and when excluding the no-rain events. The reliability component accounts for most of the CRPS after Day 5 at the > 25 mm threshold; it steadily degrades with increasing lead time. The GFS-based ensembles significantly improve the forecast reliability compared to the resampled climatological ensembles. For the intermittency threshold, the reliability is improved up to Day 12. For the > 25 mm precipitation threshold, this relative improvement starts from 75% at Day 1 to reach 17% at Day 14 (positive improvement at all lead times can also be seen from the > 5 mm threshold to the > 50 mm threshold). The GFS-based ensembles exhibit a Potential CRPS that degrades with lead times for the lower precipitation events. They significantly outperform the resampled climatological ensembles at all lead times for all the forecast-observed pairs, and until Day 7 when excluding the no-rain events, due to their narrower spread for small precipitation events. However for the > 25 mm precipitation threshold (as well as the >12.5 mm and > 50 mm thresholds), the GFS-based ensembles have worse Potential CRPS due to their larger ensemble spread. Therefore, for the high precipitation events, the GFS-based ensembles exhibit better CRPS than climatology-based ensembles due to significantly improved reliability. Regarding the ROC Score, the forecast discrimination is very significantly improved with the GFS-based ensembles compared to resampled climatological ensembles, especially for the probability of precipitation event (> 0 mm); this gain decreases with lead time, as expected.

Daily average flow ensembles are verified for all forecast-observed pairs (8660 pairs for the first 24-hour lead time) and subsets of pairs based on the following non-exceedance probability thresholds (defined from the 24-year observation record): 0.25 (2 $m^3 s^{-1}$), 0.5 (7 $m^3 s^{-1}$), 0.75 (30 $m^3 s^{-1}$), 0.9 (60 $m^3 s^{-1}$), 0.95 (84 $m^3 s^{-1}$), and 0.99 (210 $m^3 s^{-1}$). As indicated in Fig. 4 by the mean error, the coefficient of correlation and the CRPS, the forecast quality decreases significantly with increasing flow thresholds and with lead time. The GFS-based flow ensembles exhibit a conditional bias consistent with the conditional bias of the precipitation ensembles (Fig. 1 and 2): over-forecasting of small events and under-forecasting of large events. Regarding the CRPSS (Fig. 4) in reference to the climatology-based flow ensembles, the GFS-based flow ensembles are more skillful at all forecast horizons and their skill at individual lead times increases with the

flow thresholds until Day 10. The sharp increase in skill between Day 1 and Day 2 is due to the basin response time to precipitation amount. The influence of the atmospheric ensembles on the flow forecasts is more pronounced after Day 1; the two sets of flow ensembles have more similar verification statistics on Day 1 as indicated in Fig. 5. Furthermore, since these flow ensembles do not capture any hydrologic uncertainty, both sets of flow ensembles are significantly less reliable at Day 1. As shown in Fig. 5, forecast reliability degrades with lead time especially for very high flow. However, the GFS-based flow ensembles outperform the climatology-based ensembles in terms of the Reliability of CRPS, the Potential CRPS, and the ROC Score for all lead times and all flow thresholds, except the Potential CRPS for the > 0.95 non-exceedance probability threshold. This is similar to the pattern in the precipitation results (Fig. 3). For the very high flows, the GFS-based flow ensembles have better CRPS than the climatology-based flow ensembles due to significantly improved reliability (for the > 0.95 non-exceedance probability threshold, the relative improvement in reliability varies from 86% at Day 1 to 32% at Day 14).

Regarding the relative contribution of the atmospheric and hydrologic uncertainties, verification statistics are presented in Fig. 6 for the GFS-based flow ensembles verified with observed flows (solid lines) and with simulated flows (dashed lines). The forecasts verified with flows simulated from observed hydrometeorological inputs exclude the hydrologic uncertainty (and the observed hydrometeorological input uncertainty) whereas the verification with observed flows includes all sources of uncertainty. Note that the 0.75 and 0.95 non-exceedance probability thresholds correspond to similar flow values for both the observations and the simulations. All four verification statistics indicate that the hydrologic uncertainty is significant for short lead times and depends on the flow values: for example, for the CRPS, it significantly degrades the score up to Day 7 for all flows and up to Day 2 for very high flows. This indicates that uncertainty in hydrologic initial conditions is a major source of the hydrologic error. However, because of the other hydrologic uncertainty sources (e.g., the model structure and parameters), the hydrologic error tends to degrade the forecast reliability, the Potential CRPS, and the ROC Score at all lead times.

5. Conclusions and future work

Diagnostic verification is carried out with EVS on experimental ensembles from the HEFS to quantify potential forecast improvement. This case study quantifies the improvement of using the NWP single-valued forecasts in the HEFS ensemble preprocessor (versus climatological inputs) for ensemble streamflow prediction. The improvement is due mostly to improved reliability for very high events. The relative impact of the hydrologic uncertainty is significant for short lead times due to the uncertainty in hydrologic initial conditions. Additional verification studies are under way to include ensembles produced from all the HEFS components (e.g., ensemble post-processor, data assimilation), reducing and accounting for the hydrologic uncertainty and using additional weather and climate forecast information, to help target future improvements of the forecasting system and show the value of such improvements to forecasters and users. These verification studies also include more detailed verification

statistics (including statistics conditioned on the forecast) and more user-oriented verification statistics for operational forecasting. Planned enhancements to EVS include the ability to separate the timing (phase) and amplitude errors in hydrologic forecasts, and the capability to derive additional measures of skill using other reference forecasts. The OHD, the NCEP, and the NWS forecasters are also working together and with users to develop meaningful verification products and capabilities to effectively help forecasters and external users in their decision making.

This paper aims to motivate the meteorological and hydrologic research and operations communities for collaborative research and development of verification capabilities and services to generate and communicate verification information for weather, climate and water forecasts at the catchment scale. One such initiative is the cross-cutting HEPEX verification test-bed, for which EVS is proposed as one of the verification tools. This verification test-bed aims to address the following challenges in hydrologic ensemble verification: verification of rare events, characterization of the timing error, definition of an optimal set of reference forecasts for skill evaluation, definition of quality measures to be easily integrated in forecasters' and end users' decision process, and development of methods which are appropriate for multivariate forecasts (e.g., forecasts issued for multiple locations and time steps) and methods to analyze forecast predictability on multiple spatial and temporal scales.

References

Brown JD, Demargne J, Seo DJ, Liu Y. 2009. The Ensemble Verification System (EVS): a software tool for verifying ensemble forecasts of hydrometeorological and hydrologic variables at discrete locations. Submitted to *Environmental Modelling and Software*.

Clarke M, Gangopadhyay S, Hay L, Rajagopalan B, Wilby R. 2004. The Schaake Shuffle: a method for reconstructing space-time variability in forecasted precipitation and temperature fields. *Journal of Hydrometeorology* 5: 243–262.

Cloke HL, Pappenberger F. 2009. Ensemble flood forecasting: a review. *Journal of Hydrology* 375: 613-626.

Demargne J, Mullusky M, Werner K, Adams T, Lindsey S, Schwein N, Marosi W, Welles E. 2009. Application of Forecast Verification Science to Operational River Forecasting in the U.S. National Weather Service. *Bulletin of the American Meteorological Society* 90 (6): 779-784.

Demargne J, Wu L, Seo DJ, Schaake J. 2007. Experimental hydrometeorological and hydrologic ensemble forecasts and their verification in the U.S. National Weather Service. *Quantification and Reduction of Predictive Uncertainty for Sustainable Water Resources Management (Proceedings of Symposium HS2004 at IUGG2007, Perugia, July 2007). IAHS Publication* 313: 177-187.

Hamill TM, Whittaker JS, Mullen SL. 2006. Reforecasts: an important data set for improving weather predictions. *Bulletin of the American Meteorological Society* 87(1): 33-46.

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

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

Murphy AH. 1973. A new vector partition of the probability score. *Journal of Applied Meteorology* 12 (4) 124: 595-600.

National Research Council of the National Academies (NRC). 2006. Completing the Forecast: Characterizing and Communicating Uncertainty for Better Decisions Using Weather and Climate Forecasts [Available at: http://www.nap.edu/, accessed 09/30/09].

Pappenberger F, Scipal K, Buizza R. 2008. Hydrological aspects of meteorological verification. *Atmospheric Science Letters* 9: 43-52.

Schaake J, Demargne J, Hartman R, Mullusky M, Welles E, Wu L, Herr H, Fan X, Seo DJ. 2007. Precipitation and temperature ensemble forecasts from single-value forecasts. *Hydrology and Earth Systems Sciences Discussions* 4: 655-717.

Welles E, Sorooshian S, Carter G, Olsen B. 2007. Hydrologic Verification: A Call for Action and Collaboration. *Bulletin of the American Meteorological Society* 88: 503–511.

Wilks DS. 2006: *Statistical Methods in Atmospheric Sciences, 2nd ed.* Academic Press: San Diego, California.

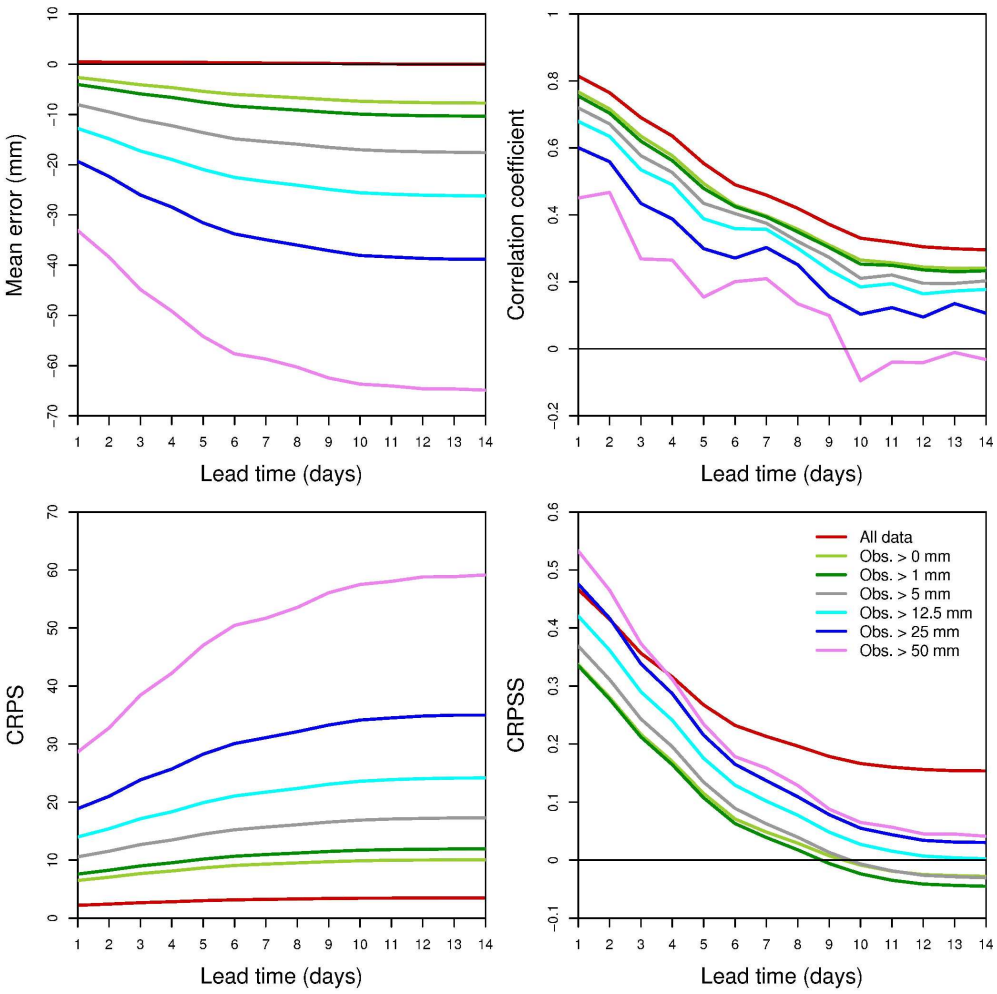

Figure 1: Mean Error and Correlation Coefficient of the ensembles means, as well as CRPS and CRPSS (in reference to resampled climatological ensembles) for the GFS-based precipitation ensembles
177x177mm (600 x 600 DPI)

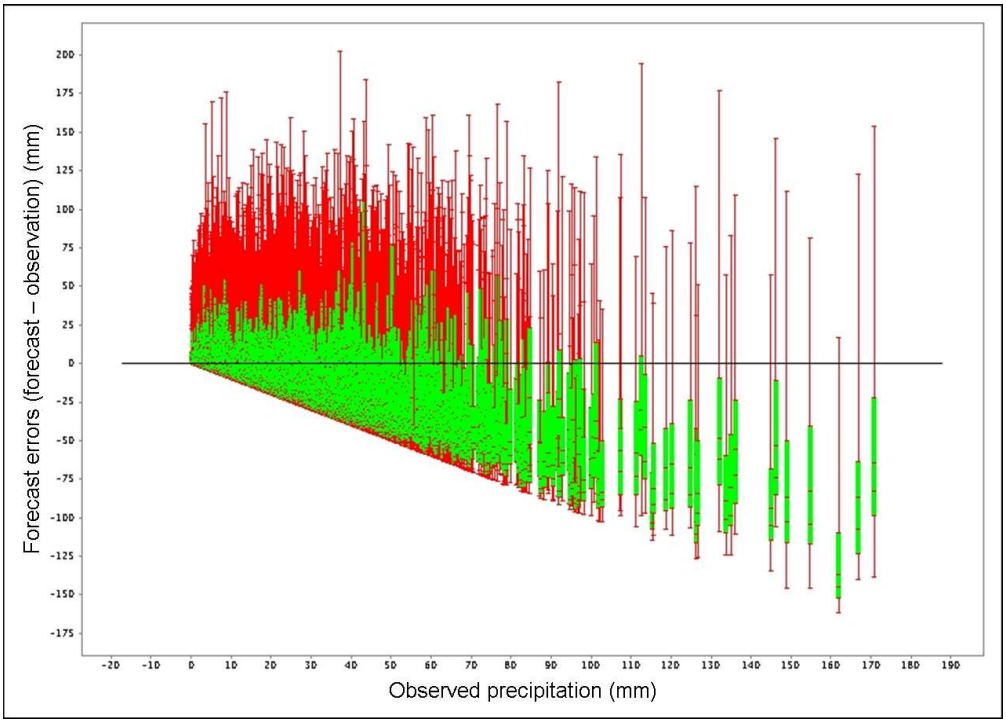

Figure 2: Box-and-whisker plot for the 0, 20, 40, 60, 80 and 100 percentiles of the forecast error distribution for the GFS-based precipitation ensembles and for the first 24-hour lead time

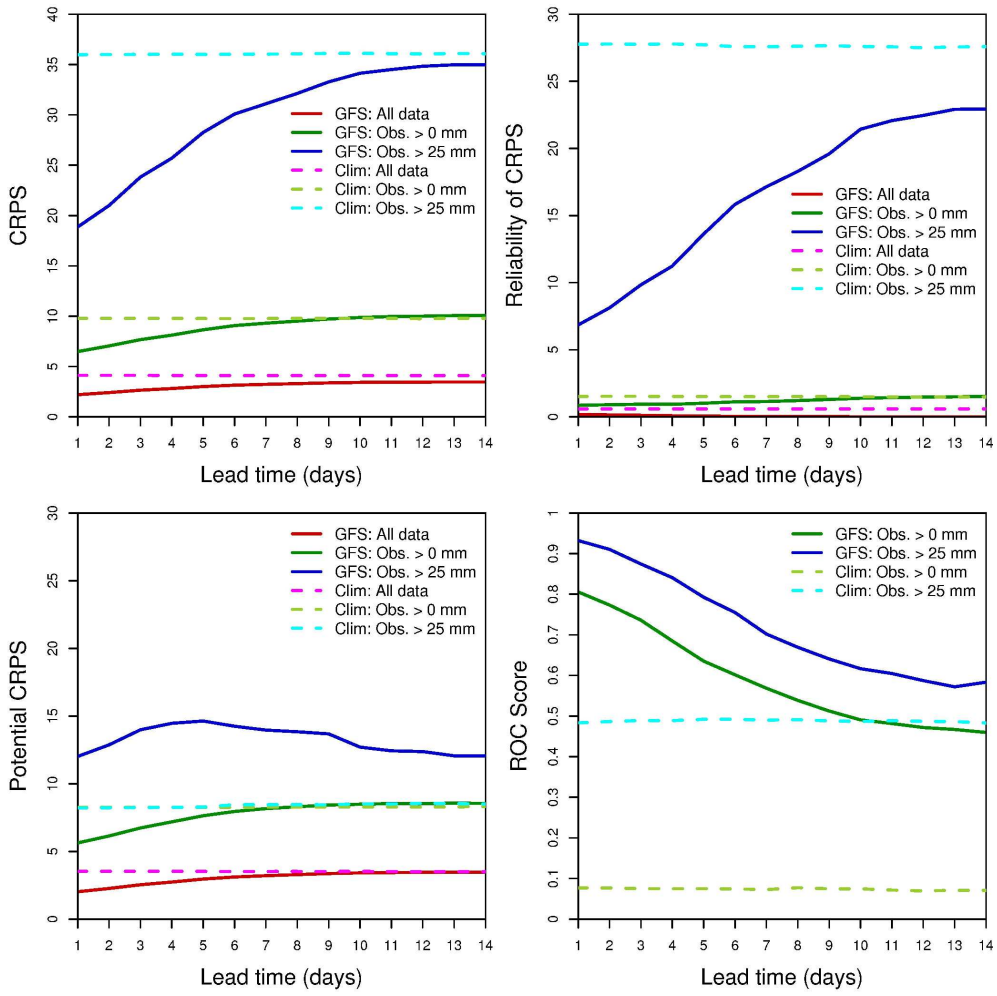

Figure 3: CRPS, Reliability component of CRPS, Potential CRPS, and ROC Score for the GFS-based precipitation ensembles ("GFS") and the resampled climatological ensembles ("Clim")
177x177mm (600 x 600 DPI)

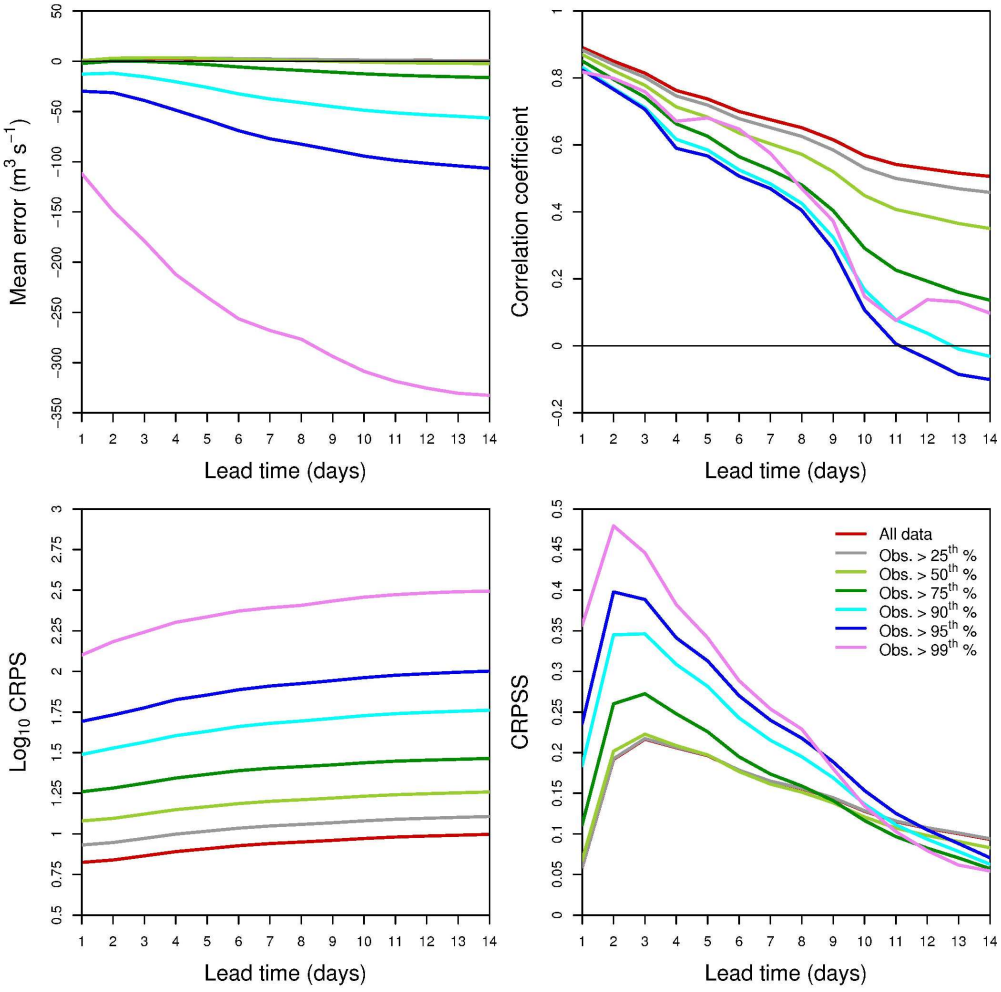

Figure 4: Mean Error and Correlation Coefficient of the ensembles means, as well as CRPS and CRPSS (in reference to climatology-based flow ensembles) for the GFS-based flow ensembles
177x177mm (600 x 600 DPI)

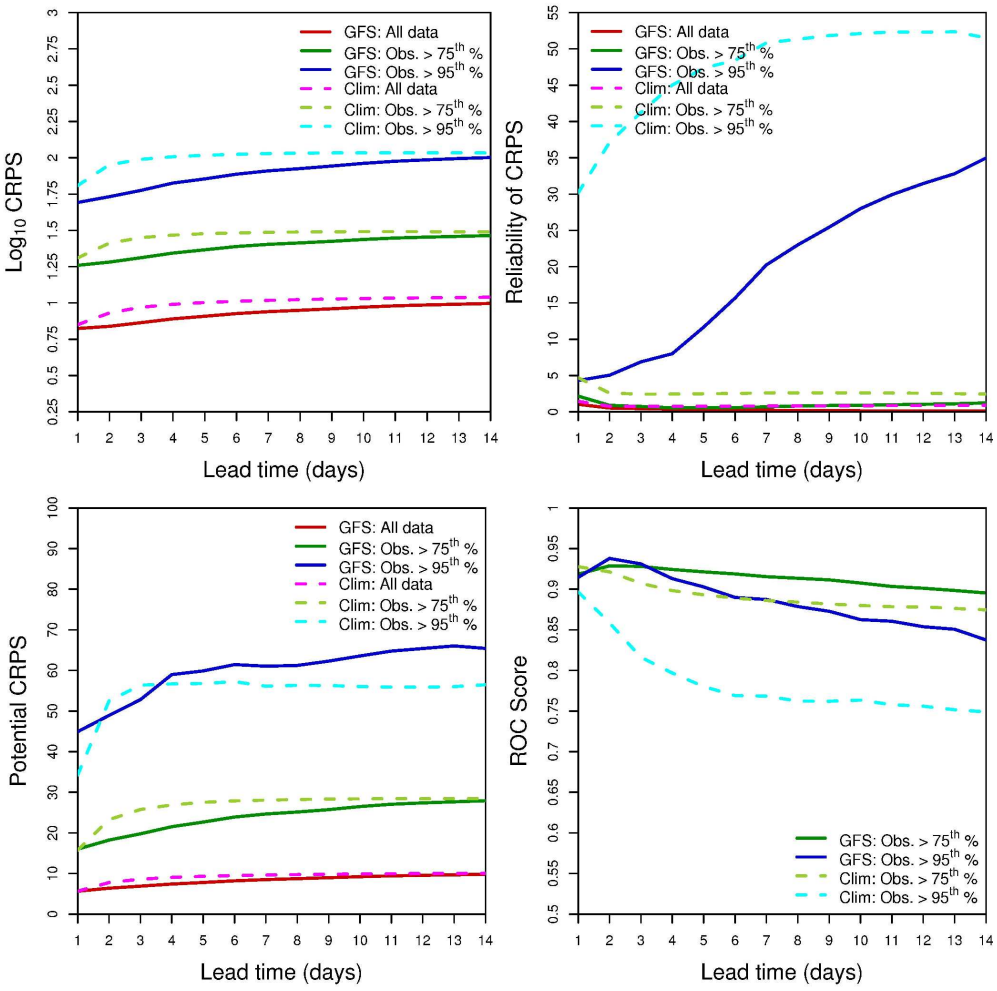

Figure 5: CRPS, Reliability component of CRPS, Potential CRPS, and ROC Score for the GFS-based flow ensembles ("GFS") and the climatology-based flow ensembles ("Clim")
177x177mm (600 x 600 DPI)

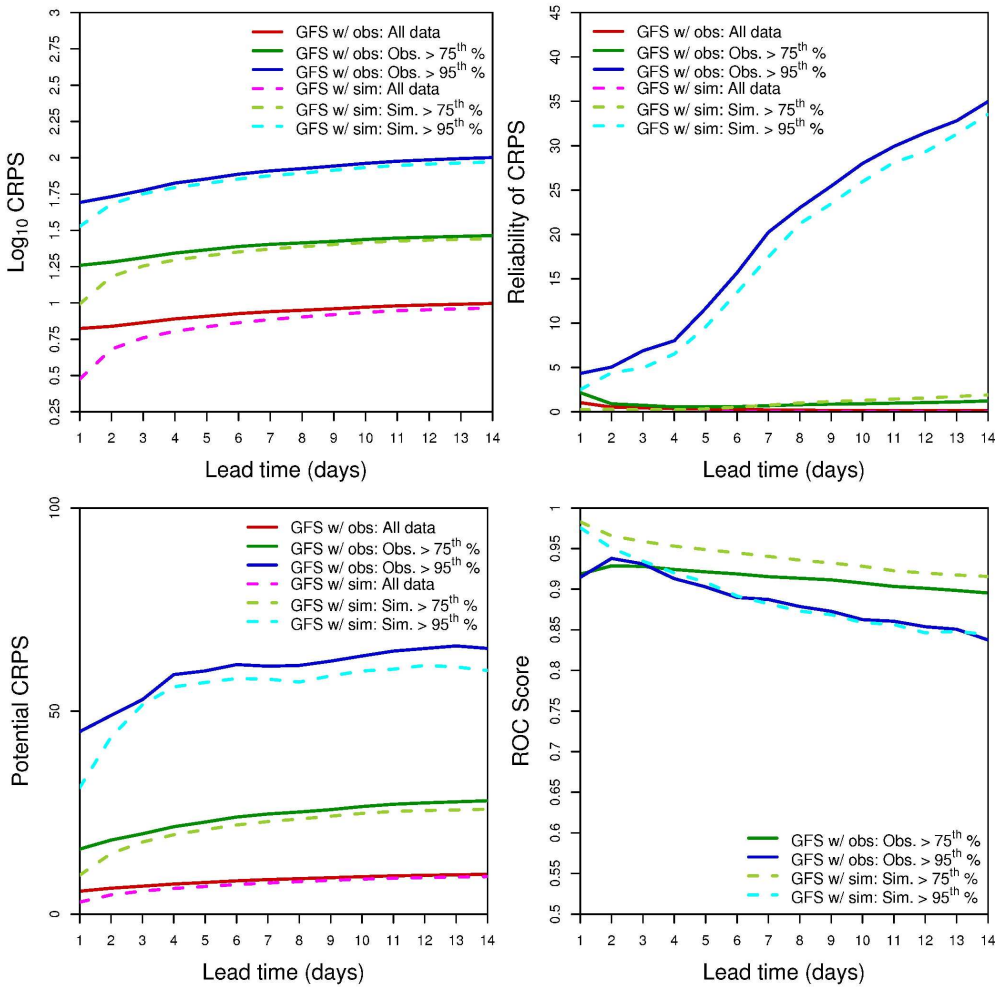

Figure 6: CRPS, Reliability component of CRPS, Potential CRPS, and ROC Score for the GFS-based flow ensembles verified with observed flows ("GFS w/ obs") and simulated flows ("GFS w/ sim") 177x177mm (600 x 600 DPI)