ATMOSPHERIC SCIENCE LETTERS
Atmos. Sci. Let. **11**: 59–63 (2010)
Published online in Wiley InterScience
(www.interscience.wiley.com) **DOI:** 10.1002/asl.267

RMetS
Royal Meteorological Society

Summary of recommendations of the first workshop on Postprocessing and Downscaling Atmospheric Forecasts for Hydrologic Applications held at Météo-France, Toulouse, France, 15–18 June 2009

John Schaake,[1]* Jean Pailleux,[2] Jutta Thielen,[3] Ray Arritt,[4] Tom Hamill,[5] Lifeng Luo,[6] Eric Martin,[2] Doug McCollor[7] and Florian Pappenberger[8]

[1] *NOAA, NWS Office of Hydrologic Development, Silver Spring, MD, USA*
[2] *CNRM-GAME (Météo-France, CNRS), Toulouse, France*
[3] *EU Joint Research Center, Ispra, Italy*
[4] *University of Iowa, Iowa, IA, USA*
[5] *NOAA, Earth System Research Lab, Boulder, CO, USA*
[6] *Princeton University, Princeton, NJ, USA*
[7] *BC Hydro, Vancouver, B.C.*
[8] *ECMWF, Reading, UK*

**Correspondence to: John Schaake, NOAA, NWS Office of Hydrologic Development, Silver Spring, MD, USA. E-mail: john.schaake@noaa.gov*

Received: 1 February 2010
Accepted: 3 February 2010

Abstract

Hydrologists are increasingly using numerical weather forecasting products as an input to their hydrological models. These products are often generated on relatively coarse scales compared with hydrologically relevant basin units and suffer systematic biases that may have considerable impact when passed through the nonlinear hydrological filters. Therefore, the data need processing before they can be used in hydrological applications. This manuscript summarises discussions and recommendations of the first workshop on Postprocessing and Downscaling Atmospheric Forecasts for Hydrologic Applications held at Meteo France, Toulouse, France, 15–18 June 2008. The recommendations were developed by work groups that considered the following three areas of ensemble prediction: (1) short range (0–2 days), (2) medium range (3 days to 2 weeks), and (3) sub-seasonal and seasonal (beyond 2 weeks).

Keywords: **HEPEX; postprocessing; hydrological applications**

1. Introduction

Hydrologic Ensemble Prediction EXperiment (HEPEX) is an international initiative of hydrologists, meteorologist and end users to advance probabilistic hydrological forecast techniques for flood, drought and water management applications (Schaake *et al.*, 2006, 2007; Thielen *et al.*, 2008). HEPEX is sharing information and experience effectively through collaborative test-beds and workshops (http://www.hepex.org, for a literature review on this topic see Cloke and Pappenberger, 2009).

Ensemble prediction systems (EPS) are subject to forecast bias and dispersion errors that arise from limitations in the description of the physical processes in the meteorological models, errors in the subgrid-scale parameterisations and uncertainties in the initial state (Wilks, 2006). Consequently, they may not display enough variability, and thus lead to an underestimation of the uncertainty (Buizza, 2010; Buizza *et al.*, 1999, 2005). Achieving good probabilistic forecasts has become a driving force in meteorological research (Gneiting *et al.*, 2007). Various approaches are proposed to postprocess, or calibrate, the meteorological outputs so that the probabilistic forecasts remain reliable while becoming sharper (Wilks, 2006).

Therefore, a topical HEPEX workshop on Postprocessing and Downscaling Atmospheric Forecasts for Hydrologic Applications was proposed. It was held in Toulouse, France, from 15 to 18 June 2009. It was hosted by Meteo-France and SCHAPI and co-sponsored by CEMAGREF, COSMOLEPS, COST731, ECMWF, Electricite de France (EDF), European Commission DG Joint Research Centre (JRC), GEO, IAHS, MSC, NCAR, NOAA, World Climate Research Program (WCRP)/Global Energy and Water Cycle Experiment (GEWEX)/hydrologic applications project (HAP) and WWRP/THORPEX. More than 100 participants from 21 countries and several International and European organisations were present. The workshop was held jointly with the second COST731 (propagation of uncertainty in advanced meteo-hydrological forecast systems) workshop (http://cost731.bafg.de). The workshop included researchers from the fields of hydrology and meteorology, operational hydrological and meteorological forecasters and end users from reservoir operations and electric power utilities.

This workshop was proposed at the third HEPEX workshop held in Stresa, Italy, 27–29 June 2007 (Thielen *et al.*, 2008), to focus on scientific questions concerning how atmospheric prediction systems can be used for hydrological streamflow applications and probabilistic quantitative precipitation forecasting. Special attention was given on how to produce skillful and reliable ensemble forcing for hydrological applications using both single-value and ensemble atmospheric forecasts. Topics for the workshop included: (1) diagnosis of problems with atmospheric weather and climate ensemble forecast systems (e.g. bias and spread deficiencies) and how they affect hydrological predictions; (2) comparisons with precipitation estimation techniques not based on ensemble forecasting (e.g. based on deterministic predictions or statistical interpretation; (3) methods for statistically correcting the ensemble forecasts to provide reliable input to hydrological models; (4) novel techniques for using ensemble input in hydrological models; (5) methods of verifying ensemble forecasts and evaluating their effect on hydrological prediction systems; (6) requirements for new databases and collaborative activities that will support the hydrological prediction system development. The workshop was interdisciplinary, fostering communication between meteorologists, hydrologists and end users to advance the HEPEX goal of helping the science community develop hydrological EPS that can be used with confidence by emergency and water resource managers.

The first part of the workshop was dedicated to oral and poster presentations on its specific topics of interest to serve as a basis for subsequent work group discussions. Three work groups were formed to discuss science issues and identify possible future collaborative activities in the following three areas of ensemble prediction: (1) short range (0–2 days), (2) medium range (3 days to 2 weeks), and (3) intraseasonal and seasonal (beyond 2 weeks). This article presents a summary of the findings and recommendations from these work groups on a range of topics. Because most of these topics were discussed by most of the work groups, the rest of the article is arranged by these topics.

The workshop report is available online at http://www.hepex.org.

2. Requirements for postprocessing and downscaling

Meteorological ensembles may include systematic biases that need to be identified and if possible corrected before useful products can be produced. Hydrological applications may require specific techniques for postprocessing and downscaling that may differ from meteorological applications. Processes and techniques that need to be addressed are

- Production of time series of ensemble future precipitation and temperature forcing for input to hydrological forecast models. These must be at the space scale of each model subarea and at the models computational time step. The ensemble members must have the same climatological characteristics as the precipitation and temperature analysis data that were used to calibrate the hydrological forecast model.
- Removal of systematic biases.
- Production of forecasts that are as sharp as possible.
- Production of reliable estimates of uncertainty.
- Extraction of as much relevant information as possible from the weather and climate forecasts for hydrological applications on different spatial and temporal scales.
- To the extent possible, design of forecasts that are to be consistent in time, i.e. they do not bounce around from one day to the next.
- Development of calibration techniques for EPS forecasts that respect the relationship between variables such as temperature and precipitation. Methods on how to validate this property need also to be developed.

3. Role of dynamic *versus* statistical downscaling

Various techniques are used for downscaling precipitation forecasts from numerical weather prediction (NWP) model outputs (which are representative of the model mesh size and which can be either deterministic or coming from an ensemble). Some techniques are statistical, some are dynamical, i.e. using a somewhat higher resolution model on some aspects. The optimum algorithm is probably a combination of both techniques. Some studies are needed for intercomparing various statistical and dynamical techniques and for working out an optimal algorithm.

Another crucial aspect which has to be studied for downscaling precipitation using a limited area model (LAM) or a regional climate model (RCM) is the relative importance of local small scales *versus* larger scales for the uncertainty on the precipitation product. One typical example (which has already been studied by several groups) is the importance of the boundary conditions in a LAM compared with the importance of scales which are assimilated in the LAM. This relative importance is dependent on both the weather situation and the geographical area considered (orography, etc.).

Studying statistical downscaling methods should lead to a better understanding of the different sources of errors. Schematically, the sources of errors can be classified in the following way: (1) systematic drifts in the atmospheric models (e.g. spin-up problem as encountered in data assimilation); (2) model truncations (both space resolution and lack of some physical processes which cannot be resolved or parameterised) and (3) forecast errors due to initial errors. The statistical methods should be able to filter out error types (1) and (2), but probably not (3) which is too dependent on weather situation and observation availability. The size of the ensemble is important to consider for

(1), not for (2) as all the members will suffer from the same model truncations. In the implementation of a statistical method an *a priori* estimate of the climatology is always useful as a reference. In general statistical methods are limited, they cannot replace the modelling of a process whose physics is understood and can be put into equations. This is why the use of statistical methods should be limited, or made 'weather-regime dependent', or combined with physical or dynamical downscaling techniques.

The meteorological and hydrological community should rigorously compare statistical downscaling (i.e. numerical forecasts postprocessed using re-forecasts) against dynamical downscaling (high-resolution modelling systems) and multi-model ensemble approaches. The relative merits of the various approaches are still not yet well understood.

4. Requirements for re-forecasts

The availability of re-forecasts has opened new possibilities in dealing with EPS. Hamill *et al*. (2004) and Hamill and Whitaker (2006) have declared re-forecasting as an efficient means of calibrating EPS and argue that the benefits of re-forecasts are so large that they should become an integral part of the NWP process. Hagedorn *et al*. (2007) point out that a growing body of literature indicates the potential utility of re-forecast methodology for improving operational ensemble predictions. Unfortunately, re-forecasts are costly to produce and are not systematically available for all NWP EPS.

Re-forecasts were discussed at length during the workshop. There was consensus among the participants that they are useful for several reasons:

- They provide a large training sample for calibrating statistical postprocessing algorithms.
- They are needed for validating postprocessing algorithms over a large number of cases, thereby spanning many rare events.
- They are needed by users to calibrate and test decision support systems (DSS).

The hydrological community strongly recommends that operational weather-climate predictions should be supplemented with the regular computation of re-forecast data sets using the same model and assimilation systems. These data sets are useful both for training of statistical postprocessing algorithms, and they allow the testing of hydrological systems over a wider number of cases in the past. Accompanying time series of observations and/or re-analyses are also required.

Topics for future research on the use of re-forecasts include:

- Considering that re-forecasts are computationally expensive, best practices for configurating re-forecasts are required. Whether they should be done regularly and/or for special meteorological events only may depend on the purpose of the re-forecasts. Equally, the necessary frequency and the number of necessary members need to be evaluated. If dates of re-forecasts are not uniform and specially include extreme events, what effect may this limited training sample have on the the postprocessing statistics? If re-forecasts are computed irregularly for special events, what procedure should be followed to decide upon the events?
- The possible effects of climate-change signals on the accuracy of postprocessed products produced with long re-forecast training data sets need to be explored further.
- Observations/analysis data must be synthesised for the same time span as the re-forecasts, and consistent with the ones used for hydrological validation. High-resolution precipitation analyses are of particular importance.
- What are the advantages and disadvantages of re-forecast-based calibration (statistical downscaling) relative to dynamical downscaling?

5. Atmospheric modelling issues

It is important to find a good trade off between NWP model resolution and the number of members in the ensemble. This is one particular example of choices which have to be made on several parameters for the general configuration of a NWP ensemble suite in an operational context. Other parameters which have to be chosen are, e.g. the parameters defining the computation of the perturbations. One needs to consider both the initial condition uncertainty and the model uncertainty.

Other issues include:

- Do we need multiple climate models, multiple hydrological models, multiple initial conditions to produce ensembles?
- How many ensemble members are sufficient to capture as much as possible the full range of possible weather outcomes?
- What is the most effective use of the computer resources used to produce atmospheric ensemble forecasts?

6. Verification

Precipitation forecast information needed for hydrological applications is space and time scale dependent. As a result, standard verification methods normally used in NWP are not sufficient. Verification techniques and scores have to be worked out that are adequate for hydrology. It is not sufficient to verify precipitation at a set of independent observing points, as the space and time structure of precipitating events have to be considered and atmospheric models do not predict precipitation at a point. Moreover, relationships between physically dependent variables like, e.g. precipitation and temperature should be respected.

Ensemble forecasts are often verified through two scores which are the reliability and the 'resolution' (resolution used with its ensemble meaning, which is different from the model mesh size). The optimal tuning between resolution and reliability is necessarily a compromise which is dependent on the application. For catching extreme events at short range, resolution might be more important than reliability.

More techniques need to be developed for the diagnosis and validation of the time series properties of forcing inputs and the hydrological outputs. Among other things these need to test if the timing of onset, offset, duration, and intensity are correct. Techniques should also be developed that evaluate the consistency of the forecasts from one day to the next. Metrics are needed that are meaningful for both hydro and meteorological communities.

7. Intercomparison studies

One of the motivations for having this workshop was to explore the possibility of conducting a set of intercomparison studies over the next several years to test ensemble precipitation postprocessing and downscaling techniques. Some of the suggestions for such studies are listed as follows without being ordered according to importance:

- Intercompare dynamical and statistical downscaling approaches.
- Specify common outputs for comparison (common time scales, etc.).
- Assess predictability as a function of location, time scale, space scale, regime (e.g. drought, ENSO phase). Evaluate these in a controlled way.
- Look at variability within and between seasons.
- Consider the effect of data limitations.
- Assess the information content of the model atmospheric forecast model. Best relationships may be with broad-scale model fields rather than nearest grid point.
- Decide on a few catchments/basins and apply methods to each location.
- Look at a broad range of spatial scales. For example, in the USA we can look at different climate regions (data are compiled for 102 regions), then smaller or larger areas.
- When setting up intercomparison studies, specific applications must be considered, some with small cost/loss ratios, some with medium cost/loss ratios.

Regarding re-forecasts, several existing potential sources of re-forecasts were identified during the workshop, that might be used to initiate intercomparison studies. These include:

- NCEP GFS,
- NCEP CFS,
- DEMETER,
- ENSEMBLES,
- ECMWF System 3.

HEPEX test-beds should be invited to participate in intercomparison studies that might be led by the different groups or organisations. Intercomparison studies for intra-seasonal to inter-annual (ISI) time scales should be coordinated with GEWEX/HAP. Intercomparison studies for time scales out to about 2 weeks should be coordinated with WWRP.

8. Application issues

A fair amount of time of the workshop revolved around practical application issues. For example, who is the best to do the statistical postprocessing of meteorological forecasts? Should this be done by meteorologists, who may understand the meteorology of the problem, or rather by hydrologists, who understand better the hydrological issues and what is needed as inputs to their models? Should not statisticians be more involved or rather develop the methodologies? It is very likely that the multi-disciplinary approach would yield best results. In this case, however, it is very important that the different communities learn to understand the terminology and requirements of the others. For example, a typical request of meteorologist in this context is that hydrologists should describe more concretely their requirements.

It may be preferable to have centrally computed hydrological forecast products so that local hydrologists are freed from the burden of producing numerical forecasts themselves. The lack of centralisation of computations may result in discrepancies from one basin and one forecast office to the next.

Organisationally, how can we cooperate better between hydrological and weather services? How can hydrologists in return help the weather modellers? Possible examples include through providing data sets that help us validate near-surface processes in our models.

9. Future cooperation

For stimulating or triggering cooperation on the above-mentioned topics, it was suggested to use already existing groups and programmes in the World Meteorological Organisation (WMO). Strong candidates here include: the WCRP, GEWEX; the World Weather Research Program (WWRP); the Commission for Atmospheric Sciences (CAS) research commission, and the Commission for Basic Systems (CBS) for operations in meteorology. Also, at the European level opportunities exist for fostering collaboration, e.g. with organisations like EUMETNET, or with the EU research projects. Several temporary projects which include field experiments have taken place recently in Europe or will take place soon: Convective orographically forced precipitation systems (COPS); MAP D-Phase, etc. For the coming years, an important potential focus for some of these studies should be the Mediterranean projects, especially Hydrological cycle in the Mediterranean Experiment (HyMeX).

The WCRP/GEWEX project includes an HAP that already is well coordinated with HEPEX. All of HEPEX activities related to ISI hydrological ensemble prediction are joint HEPEX/HAP activities.

The workshop recommended developing stronger relationships with the WWRP perhaps through the THORPEX and TIGGE projects. Accordingly, the following suggestions were made:

- Use TIGGE data sets to evaluate their value for hydro forecasts, and compare against re-forecast-based techniques.
- Use THORPEX–HEPEX interaction as a vehicle for the dynamical–statistical downscaling comparisons that we know need to be done.
- TIGGE data format (grib; netCDF) not easily usable in hydro community, also because of file size; can folks who pioneer their use feed back their algorithms to the TIGGE archive centers to make it easier for others to use?

One particular application that was discussed is electricity power production that is dependent on several meteorological parameters, one being the precipitation at different scales. In addition, the application is relevant (with different tools) in the short range, the medium range and the long range. One possibility could be to design projects around this application in connection with regional hydrometeorological projects such as HYMEX in the Mediterranean basin, and in connection with specific observing networks (Mediterranean observatory in France).

Future studies are needed to understand how uncertainty propagates from weather and climate forecasts through hydrological models. These studies need to make use of both re-forecasts and hydrological post-processing algorithms to fix bias and spread problems in hydrological ensemble forecasts.

10. Future activities

For the coming year, plans will be developed for an intercomparison study of downscaling techniques for the medium range of 0–14-day forecast period. It is expected that this study will focus on one or two target areas in the North America and Europe. This will be coordinated with both the WWRP and WCRP/GEWEX/HAP.

References

Buizza R. 2010. Horizontal resolution impact on short- and long-range forecast error. *Quarterly Journal of the Royal Meteorological Society* under revision.

Buizza R, Houtekamer PL, Toth Z, Pellerin G, Wei MZ, Zhu YJ. 2005. A comparison of the ECMWF, MSC, and NCEP global ensemble prediction systems. *Monthly Weather Review* **133**: 1076–1097.

Buizza R, Miller M, Palmer TN. 1999. Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. *Quarterly Journal of the Royal Meteorological Society* **125**: 2887–2908.

Cloke HL, Pappenberger F. 2009. Ensemble Flood Forecasting: a review. *Journal of Hydrology* **375**: 613–626.

Gneiting T, Balabdaoui F, Raftery, AE, 2007. Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society Series B-Statistical Methodology* **69**: 243–268.

Hagedorn R, Hamill TM, Whitaker JS. 2007. Probabilistic forecast calibration using ECMWF and GFS ensemble forecasts Part I: 2-meter temperature. *Monthly Weather Review* **136**: 2608–2619.

Hamill T, Whitaker J, Wei X. 2004. Ensemble reforecasting: improving medium-range forecast skill using retrospective forecasts. *Monthly Weather Review* **132**: 1434–1447.

Hamill TM, Whitaker JS. 2006. Probabilistic quantitative precipitation forecasts based on reforecast analogs: theory and application. *Monthly Weather Review* **134**: 3209–3229.

Schaake J, Franz K, Bradley A, Buizza R. 2006. The Hydrologic Ensemble Prediction EXperiment (HEPEX). *Hydrology and Earth System Sciences Discussions* **3**: 3321–3332.

Schaake JC, Hamill TH, Buizza R, Clark M. 2007. HEPEX – The hydrological ensemble prediction experiment. *Bulletin of the American Meteorological Society* **88**(10): 1541–1547.

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. *Atmos. Sci. Let.* **9**(2): 29–35.

Wilks DS. 2006. *Statistical Methods in the Atmospheric Sciences*, 2nd ed. Elsevier Academic Press: 627.