Probabilistic Quantitative Precipitation Forecast for Flood Prediction: An Application

P. REGGIANI* AND A. H. WEERTS

Foundation Delft Hydraulics, Delft, Netherlands

(Manuscript received 6 December 2006, in final form 31 May 2007)

ABSTRACT

This paper outlines a methodology to produce probabilistic quantitative precipitation forecasts by means of a dedicated uncertainty processor for weather model output. The uncertainty processor is developed as a component of a Bayesian forecasting system for river flow prediction. In this context the quantitative precipitation forecast is envisaged as a mixed binary–continuous predictand. The processor is applied to the quantitative precipitation forecasts and to precipitation observations covering a 5-yr period, whereby the forecasted and observed relative air humidity are used as ancillary meteorological indicators. The application of the processor to the selected dataset highlights a significantly larger skill of the quantitative precipitation forecast in predicting event occurrence rather than event depth and provides an objective quantification of the forecast uncertainty. The methodology applied here remains restricted to small basins, in which spatial variability of precipitation can be considered negligible. The need for processing the uncertainty induced by spatial variability of rainfall is briefly addressed.

1. Introduction

Recent large floods in Europe, such as those that have occurred in the Meuse and Rhine basins in 1995, over large areas of the United Kingdom in 1998 and 2000, in the Elbe basin in 2002 and in the Danube basin in 2005 and 2006, have led to increased interest in research on upgrading existing operational flood forecasting systems. In this context special emphasis is put on the quantification of the uncertainty associated with a forecast. A variety of operational users are getting progressively accustomed to the reality that typical predicants such as precipitation, river stages, discharges, or volumes at critical locations need to be associated with an accurate measure of the predictive uncertainty. Several approaches [Beven and Binley 1992; Lardet and Obled 1994; Schaake and Larson 1998; Krzysztofowicz 1999; Liu et al. 2005; the Hydrological Ensemble Prediction Experiment (HEPEX) hosted at the National Weather Service and the European Centre for Medium-Range Weather Forecasts; Schaake et al. 2006] have been pursued with the aim of quantifying predictive uncertainty due to different sources.

The sources of predictive uncertainty can be manifold. They may be attributable to uncertainty of precipitation forecasts and observations, errors in the conceptualization of the system, model initialization, or process parameterization. Whatever the source of the uncertainty, the final probabilistic forecast must be accompanied by an adequate measure of the total uncertainty in the form of probability density functions, which are revised on the basis of a priori knowledge from historical information.

For the particular area of flood forecasting it is desirable to detach the responsibility of the forecaster, which consists in predicting a critical river level by means of operational models, from those of civil protection services in charge of making the decision of whether an evacuation or controlled flooding is necessary. This is only feasible if a reliable measure of the uncertainty associated with a discharge or river stage prediction is available and communicated by the forecaster to the decision maker.

There is thus a general need to provide robust uncertainty assessment tools, which can be implemented and used in operational flood forecasting systems. These tools should be internally consistent, statistically self-calibrating (Krzysztofowicz 1999), and require

* Additional affiliation: Section of Hydraulic Structures and Probabilistic Design, Delft University of Technology, Delft, Netherlands.

Corresponding author address: P. Reggiani, Foundation Delft Hydraulics, P.O. Box 177, 2600 MH, Delft, Netherlands.
E-mail: paolo.reggiani@wldelft.nl

DOI: 10.1175/2007JHM858.1

modest computational effort by performing the bulk of the computations offline. The latter attribute is especially relevant, because in emergency situations the time available between performing a forecast, issuing a warning, and organizing response actions is generally limited. This fact is exacerbated in river basins with typically short response times.

The general framework for a Bayesian forecasting system (BFS) presented by Krzysztofowicz (1999) constitutes a major effort in formalizing the quantification of uncertainty in the flood forecasting process, given arbitrary deterministic models. The BFS consists of three components: (i) an input uncertainty processor (IUP), (ii) a hydrologic uncertainty processor (HUP), and (iii) an integrator of uncertainty (IU). The BFS is based on a complete and consistent theory that identifies and separates relevant sources of uncertainty. It provides a general reference framework for building the uncertainty processors and the integrator, components that can be readily implemented in a real-time flood forecasting system. Sequel papers have implemented various BFS components as a one-branch processor (precipitation-independent model; Kelly and Krzysztofowicz 2000; Krzysztofowicz and Kelly 2000) or two-branch processor (precipitation-dependent model; Krzysztofowicz and Herr 2001), analyzed their statistical properties and carried out preliminary verifications. One of the most salient features of the BFS theory is the provision of a formal basis for the development and successive revision of its components, while preserving internal consistency and statistical properties, which make it appealing for operational use.

Operational experience has shown that the most prominent source of uncertainty in the flood forecasting process is due to the quantitative precipitation forecast (QPF). The aim of this paper is to present an application of a Bayesian processor for QPF output (BPO), which is to be used as an IUP for the BFS. The proposed processor quantifies the uncertainty of the QPF and is suited for offline execution. In contrast to the BFS introduced in Krzysztofowicz (1999), we suggest modifying the system whereby the originally proposed IUP, formulated as a precipitation uncertainty processor (Kelly and Krzysztofowicz 2000), is substituted by a Bayesian processor of output (BPO) (Krzysztofowicz 2004; Maranzano and Krzysztofowicz 2004).

The original IUP (Krzysztofowicz 1999) is formulated in terms of probability distributions of hydrological model output, conditional on measured precipitation. Here we propose to separate the hydrological model output from QPFs obtained from numerical weather prediction (NWP) in the Bayesian forecasting system. The aim is to enhance the segregation of sources of uncertainty by means of dedicated processors that are designed for processing hydrological and numerical weather prediction model output individually. In this manner the uncertainty on precipitation, given a numerical weather forecast, is treated explicitly in the BFS. Thus we propose in this paper an input uncertainty processor based on probability of precipitation and conditional on two variates: 1) the QPF and 2) the relative air humidity acting as ancillary indicator on the state of the atmosphere. Further below we implement the processor.

Given the large attention of the scientific and operational community on the ensemble techniques, the present application needs to be put into perspective in this context (Schaake et al. 2006). The IUP presented here is used for a deterministic precipitation forecast. An extension of the BPO for use with the ensemble prediction system is straightforward. An ensemble weather forecast represents an auxiliary randomization of the future precipitation and can be handled naturally within the BFS theory (Krzysztofowicz 2001a). The restriction of the present application to deterministic forecasts has been chosen to simplify the exposure. Finally we observe that attention needs to be paid to the fact that the present application, as well as the applications by Krzysztofowicz (1999), Krzysztofowicz and Kelly (2000), and Krzysztofowicz and Herr (2001) are restricted to typically small river basins, where the impact of spatial and temporal distribution of precipitation on the overall forecast uncertainty can be ignored. For larger basins however, this aspect needs to be addressed explicitly, as spatial and temporal variability of precipitation may have significant impact on the predictive uncertainty of the basin outflow forecasts. In such cases the predictive uncertainty on precipitation at individual locations within the basin and over a range of temporal scales needs to be integrated into the procedure. However, this aspect remains at present matter of further investigation. The following exposure is structured into five parts: Section 2 gives a brief overview of the BFS and introduces the BPO as an IUP in the BFS. Section 3 summarizes the theoretical concepts of the theory underlying the BPO, section 4 describes the actual implementation of the BPO, and section 5 is dedicated to a discussion of the results and draws the conclusions.

2. The Bayesian forecasting system

a. Assumptions

The BFS proposed by Krzysztofowicz (1999) rests on the general assumption that the principal sources of

uncertainty in the flood forecasting process can be split into two groups: 1) the *input uncertainty* and 2) the *hydrologic uncertainty*.

The first group is attributable to all the forcing inputs of the hydrologic models (e.g., precipitation, temperature, and evaporation), assuming the model to be perfect and the internal model states to be well known.

The second group is due to the conceptualization errors in the description of the hydrological and hydraulic processes, the uncertainty in the choice of the parameters in the governing equations, the uncertainty of the knowledge of the internal model states, and the measurement errors of environmental variables, assuming that the precipitation is perfectly known and certain.

In this fashion the two principal sources of uncertainty are isolated from each other and treated separately. With particular attention to short-term flood forecasting (in the order of several hours), evaporation or temperature (their impact on the hydrologic response can be considered medium term) are assumed to have negligible contribution to the forecast, in contrast to precipitation. For this reason input uncertainty is considered attributable to precipitation only. The uncertainty due to the remaining meteorological variables is lumped into the hydrological uncertainty. However, the general structure of the BFS allows a relaxation of this hypothesis if a separate treatment of such sources of uncertainty is deemed necessary.

b. Variates

Adopting the notation by Krzysztofowicz (1999), the following variates are considered:

$$\mathbf{W} = [W_1, \ldots, W_J]' \tag{1}$$

is an ensemble of precipitation depth observations at time intervals $1, \ldots, J$;

$$\mathbf{S} = [S_1, \ldots, S_J]' \tag{2}$$

is an ensemble of relative air humidity observations at different locations and times $1, \ldots, J$. Next,

$$\mathbf{Z} = [Z_1, \ldots, Z_J]' \tag{3}$$

are forecasted estimates of precipitation at a given location and at time steps at which the forecasts become available, while

$$\mathbf{T} = [T_1, \ldots, T_J]' \tag{4}$$

are forecasted estimates of relative air humidity at the same locations and times. The ensemble

$$\mathbf{H} = [H_1, \ldots, H_N]' \tag{5}$$

called the predictand, are discharges (or river stages) recorded at the basin outlet at forecasting times $1, \ldots, N$. These quantities lie in the future with respect to those observed at the same locations, at an arbitrary number of past times $1, \ldots, L$ up to the time 0 at which the forecasts are started:

$$\mathbf{H}_0 = [H_0, \ldots, H_{0-L}]'. \tag{6}$$

Last,

$$\mathbf{R} = [R_1, \ldots, R_N]' \tag{7}$$

is an ensemble of sets of modeled discharges (or river stages) at the same location and times of the observations $\mathbf{H}$. The realizations of the variates $\mathbf{W}$, $\mathbf{T}$, $\mathbf{S}$, $\mathbf{Z}$, $\mathbf{H}$, $\mathbf{H}_0$, and $\mathbf{R}$ are denoted with the lowercase letters $\mathbf{w}$, $\mathbf{t}$, $\mathbf{s}$, $\mathbf{z}$, $\mathbf{h}$, $\mathbf{h}_0$, and $\mathbf{r}$, respectively.

c. Predictive uncertainty

Next, two marginal probability densities $\psi(\mathbf{w}|\mathbf{z}, \mathbf{y})$ and $\theta(\mathbf{h}|\mathbf{r}, \mathbf{x})$ are introduced, whose estimates constitute the IUP and the HUP, respectively. The vector $\mathbf{y}$ includes forecasted atmospheric state variables, such as relative air humidity, which are predictors of the predictand $\mathbf{W}$, and initial and boundary conditions of the weather model; $\mathbf{x}$ is a vector of random variables, which are the predictors of the predictand $\mathbf{H}$ and are considered fixed for a given forecast. The probabilistic quantitative flow forecast (PQFF) conditional on the combined vector $(\mathbf{x} \cup \mathbf{y} \cup \mathbf{z})$ for a selected location and a characteristic forecast time n is thus expressed in terms of the following total probability law:

$$\chi(\mathbf{h}|\mathbf{x}, \mathbf{z}, \mathbf{y}) = \int_{-\infty}^{\infty} \theta(\mathbf{h}|\mathbf{r},\mathbf{x})\psi(\mathbf{w}|\mathbf{z},\mathbf{y})\, d\mathbf{r}\, d\mathbf{w}. \tag{8}$$

The actual composition of $\mathbf{x}$ and $\mathbf{y}$ needs to be tailored to the specific problem. For the present application we assume $\mathbf{y} = \mathbf{t}$, namely, equal to a vector $\mathbf{t}$ of observed relative air humidity, while the initial and boundary conditions of the meteorological model are assumed fixed and thus omitted from the probabilistic formulation. The pair $(\mathbf{z}, \mathbf{t})$ is said to be a sufficient predictor for $\mathbf{W}$.

Next, the vector $\mathbf{x}$ can be cast into a subvector, which is recognized as sufficient predictor of $\mathbf{H}$. We assume $\mathbf{x} = (\mathbf{h}_0, \mathbf{u}, \mathbf{w})$, where $\mathbf{u}$ is the internal state vector of the hydrological model. Should $(\mathbf{r}, \mathbf{x})$ prove at a later stage to be insufficient predictors of $\mathbf{h}$, and $(\mathbf{z}, \mathbf{t})$ insufficient predictors of $\mathbf{W}$, this assumption can be relaxed by including additional variables. In an operational forecasting system, for which the hydrological model has been calibrated over a sufficiently long time period, the model parameters can be considered static. However, in principle different parameter sets that have been calibrated specifically for the high and low flow seasons

or for different months of the year can be adopted. In such case the parameter vector could be considered as a potential source of uncertainty in the forecast and explicitly included in the Bayesian formulation. Equation (8) can now be expanded to

$$\chi(\mathbf{h}|\mathbf{h}_0, \mathbf{u},\mathbf{z},\mathbf{t}) = \int_{-\infty}^{\infty} \theta(\mathbf{h}|\mathbf{r},\mathbf{h}_0, \mathbf{u}, \mathbf{w})\psi(\mathbf{w}|\mathbf{z},\mathbf{t})\, d\mathbf{r}\, d\mathbf{w}, \quad (9)$$

where $\psi(\cdot|\mathbf{z}, \mathbf{t})$ is the density of the actual precipitation conditional on the QPF $\mathbf{z}$ produced by the NWP model and on forecasted relative air humidity; $\theta(\cdot|\mathbf{r}, \mathbf{h}_0, \mathbf{u}, \mathbf{w})$ is the density of the predictand $\mathbf{h}$, conditional on the predictand estimate $\mathbf{r}$ (the modeled discharge) and the predictand up to the forecast time, $\mathbf{h}_0$, the internal state vector $\mathbf{u}$, and the actual precipitation $\mathbf{w}$, given that $\mathbf{w}$ is certain.

The integration operator (IU) in Eq. (9) combines the hydrological uncertainty and the uncertainty on the precipitation, yielding the total predictive uncertainty on $\mathbf{H}$. The resulting probability density $\chi(\cdot|\mathbf{h}_0, \mathbf{u}, \mathbf{z}, \mathbf{t})$, the PQFF, expresses the predictive uncertainty on the predictand $\mathbf{H}$, conditional on the set $(\mathbf{h}_0, \mathbf{u}, \mathbf{z}, \mathbf{t})$ and marginalized with respect to the actual precipitation $\mathbf{w}$ and the predictand estimate $\mathbf{r}$.

3. Bayesian processor

a. Operational requirements

The implementation of the BFS based on Eq. (9) requires the estimation of the marginal densities $\psi(\cdot|\mathbf{z}, \mathbf{t})$, and $\theta(\cdot|\mathbf{s}, \mathbf{h}_0, \mathbf{u}, \mathbf{w})$, which represent the two groups of uncertainties. This constitutes a major challenge and is the key to a successful application of the processor. The distribution estimators IUP and HUP are both formulated as Bayesian processors, in which one revises a prior density by means of a likelihood function, which statistically represents the performance of the forecasting model over a historical period. For an application in a real-time forecasting system the operational requirements impose clear limitations on the methods that can be used in specifying the two conditional densities. Carrying out a large number of simulations, as for instance required when putting Monte Carlo–sampled data though a deterministic model (Lardet and Obled 1994), should be possibly avoided in a real-time forecasting situation because of time constraints. This issue becomes even more important when using computationally intensive process-based hydrological models. The method should also meet the criteria of robustness and protect the user automatically from producing an unrealistic forecast (Krzysztofowicz 2001b), if for instance model predictions turn out to be noninformative. Thus modeling the densities with the aid of analytical-numerical expressions, which can be prepared offline but evaluated and updated by processing a limited amount of information at the time of the forecast, is a valuable solution.

The IUP formulated as precipitation uncertainty processor (PUP) in Kelly and Krzysztofowicz (2000) and the HUP (Krzysztofowicz and Kelly (2000)) meet both criteria. These requirements, as will become clear later on, are also met by the BPO (Krzysztofowicz 2004), which is implemented hereunder and proposed as IUP in (9). The application of the HUP will not be addressed further in this paper and remains a matter for future investigations.

b. Input uncertainty processor

The IUP quantifies the uncertainty on the actual precipitation, given a precipitation forecast. Precipitation can be modeled as an intermittent process, in which the variate assumes either the value zero in the absence of rain or a positive value if an event occurs. Stochastically, precipitation is represented (Krzysztofowicz 2004) by the predictand $\mathbf{W}$, a mixed (binary–continuous) variate combining two processes: (i) the binary process of event occurrence/nonoccurrence, expressed through the binary predictand $V = 1$ for an event, and $V = 0$ in absence of an event, and (ii) a continuous process w representing the precipitation depth and conditional on event occurrence. The predictand $\mathbf{W}$ has a predictor $\mathbf{Z}$, which is analogously a mixed (binary–continuous) variate. A forecasted event at a location and at a point in time can thus be represented stochastically by the joint realization (v, z) of the binary and the continuous predictands. The precipitation depth and its prediction are represented by the joint realization (w, z). A revised posterior distribution on the model input conditional on a forecast is obtained from an assumed prior by means of Bayesian inference as proposed by Krzysztofowicz (2004) and reported herein.

1) Bayesian revision of the binary process

For the binary process we identify a prior probability, which is revised in the Bayesian formulation on the basis of the likelihood function. The likelihood incorporates the statistics of the system behavior on the basis of historical observations. The revision leads to a posterior distribution. The information content of the posterior distribution strongly depends on that of the prior, which needs therefore to be carefully selected. The prior probability of event occurrence is denoted with $g = P(V = 1)$ and the density of the forecasted precipitation $\mathbf{Z}$ at a location, conditional on the assumption

that the event happened ($v = 1$) or did not happen ($v = 0$), with $f_v(\mathbf{z}) = p(\mathbf{Z} = \mathbf{z}, |V = v)$, $v \in \{0, 1\}$. Given $\mathbf{Z} = \mathbf{z}$, the pair (f_0, f_1) represents the likelihoods of the event $v = 0$ and $v = 1$, respectively. If the prior probability g on event occurrence is combined with the likelihood functions (f_0, f_1) in the Bayesian revision process, one obtains the posterior probability $\Pi_\mathbf{z} = P(V = 1|\mathbf{Z} = \mathbf{z})$ of event $V = 1$ conditional on the realization of the predictor $\mathbf{Z} = \mathbf{z}$ as follows:

$$\Pi_\mathbf{z} = \frac{f_1(\mathbf{z})g}{k(\mathbf{z})}, \tag{10}$$

where the expectation k at the denominator is given by the weighted sum:

$$k(\mathbf{z}) = f_0(\mathbf{z})(1 - g) + f1(\mathbf{z})g. \tag{11}$$

Combining (10) and (11), an analytical expression of the posterior probability of event occurrence $V = 1$, conditional on the realization of the predictor $\mathbf{Z} = \mathbf{z}$ is obtained:

$$\Pi_\mathbf{z} = \left[1 + \frac{(1 - g)}{g}\frac{f_0(\mathbf{z})}{f_1(\mathbf{z})}\right]^{-1}. \tag{12}$$

Equation (12) defines the theoretical structure of the BPO for a binary predictand.

2) Bayesian revision of the continuous process

The Bayesian revision of the continuous process requires an appropriate prior distribution on precipitation depth $\mathbf{Z} = \mathbf{z}$ and a corresponding likelihood function. Both distributions are combined in the Bayesian formulation to obtain a revised posterior distribution. If we assume that $g(\mathbf{w}|\mathbf{s}) = p(\mathbf{w}|\mathbf{s})$ represents a prior density of the predictand $\mathbf{W}$ conditional on observed relative air humidity $\mathbf{S} = \mathbf{s}$, then $f(\mathbf{z}|\mathbf{w}, \mathbf{t}) = p(\mathbf{z}|\mathbf{W} = \mathbf{w}, \mathbf{T} = \mathbf{t})$ is the density function of the forecast $\mathbf{Z}$, given the measured precipitation $\mathbf{W} = \mathbf{w}$ and the forecasted air humidity $\mathbf{T} = \mathbf{t}$. For a precipitation forecast $\mathbf{Z} = \mathbf{z}$, $f(\mathbf{z}|\cdot)$ expresses the likelihood of $\mathbf{W}$, denoted with $f(\mathbf{z}|\mathbf{w}, \mathbf{t})$. It embodies the statistical properties of the system based on historical weather model performance and climatic records. We apply once more Bayesian inference on the prior density $g(\mathbf{w}|\mathbf{s})$, obtaining the revised posterior density function $\phi(\mathbf{w}|\mathbf{z}, \mathbf{t}, \mathbf{s})$ of predictand $\mathbf{W}$ conditional on the realizations $\mathbf{Z} = \mathbf{z}$, $\mathbf{T} = \mathbf{t}$ and $\mathbf{S} = \mathbf{s}$:

$$\phi(\mathbf{w}|\mathbf{z}, \mathbf{t}, \mathbf{s}) = \frac{f(\mathbf{z}|\mathbf{w}, \mathbf{t})g(\mathbf{w}|\mathbf{s})}{k(\mathbf{z}|\mathbf{t}, \mathbf{s})}, \tag{13}$$

where

$$k(\mathbf{z}|\mathbf{t}, \mathbf{s}) = \int_{-\infty}^{\infty} f(\mathbf{z}|\mathbf{w}, \mathbf{t})g(\mathbf{w}|\mathbf{s})\, d\mathbf{w}. \tag{14}$$

The posterior cumulative distribution function $\Phi(\mathbf{w}|\mathbf{z}, \mathbf{t}, \mathbf{s}) = P(\mathbf{W} \le \mathbf{w}|\mathbf{Z} = \mathbf{z}, \mathbf{T} = \mathbf{t}, \mathbf{S} = \mathbf{s})$ is obtained by integrating (13) as follows:

$$\Phi(\mathbf{w}|\mathbf{z}, \mathbf{t}, \mathbf{s}) = \frac{1}{k(\mathbf{z}|\mathbf{t}, \mathbf{s})}\int_{-\infty}^{\mathrm{w}} f(\mathbf{z}|\zeta, \mathbf{t})g(\zeta|\mathbf{s})\, d\zeta. \tag{15}$$

3) The probabilistic quantitative precipitation forecast

For a complete probabilistic description of the predictand $\mathbf{W}$, which is a mixed (binary–continuous) variate, one needs to combine two processes: the probability of precipitation occurrence

$$\Pi_\mathbf{z} = P(V = 1|\mathbf{Z} = \mathbf{z}), \tag{16}$$

which is given by Eq. (10), and the probability of precipitation depth, conditional on forecasted precipitation depth, $\mathbf{z}$; the forecasted and observed humidity, $\mathbf{t}$ and $\mathbf{s}$; and precipitation occurrence, $V = 1$. The cumulative distribution function

$$\Phi(\mathbf{w}|\mathbf{z}, \mathbf{t}, \mathbf{s}) = P(\mathbf{W} \le \mathbf{w}|\mathbf{Z} = \mathbf{z}, \mathbf{T} = \mathbf{t}, \mathbf{S} = \mathbf{s}, V = 1) \tag{17}$$

is given by Eq. (15). The total probabilistic quantitative precipitation forecast (PQPF) Ψ is then expressed in terms of the cumulative distribution function Ψ of $\mathbf{w}$, conditional on $\mathbf{z}$, $\mathbf{t}$, and $\mathbf{s}$:

$$\begin{aligned}\Psi(\mathbf{w}|\mathbf{z}, \mathbf{t}, \mathbf{s}) &= P(\mathbf{W} \le \mathbf{w}|\mathbf{Z} = \mathbf{z}, \mathbf{T} = \mathbf{t}, \mathbf{S} = \mathbf{s}) \\ &= \Pi_\mathbf{z}\Phi(\mathbf{w}|\mathbf{z}, \mathbf{t}, \mathbf{s}) + (1 - \Pi_\mathbf{z}), \quad \mathbf{w} \ge 0.\end{aligned} \tag{18}$$

The corresponding probability density function conditional on $\mathbf{z}$, $\mathbf{t}$, and $\mathbf{s}$ is obtained by derivation of (18) with respect to $\mathbf{w}$:

$$\psi(\mathbf{w}|\mathbf{z}, \mathbf{t}, \mathbf{s}) = \Pi_\mathbf{z}\phi(\mathbf{w}|\mathbf{z}, \mathbf{t}, \mathbf{s}). \tag{19}$$

4. Application

The processor theory is applied to a complete dataset for the province of Limburg, Netherlands. For the experiment proposed in this paper we use precipitation and temperature–dewpoint temperature difference measurements observed at Maastricht-Aachen airport, and the QPF from the High-Resolution Limited-Area Model (HIRLAM) (Undén et al. 2002) operated with a model cell size of 22 km × 22 km. The temperature–dewpoint temperature difference is directly related to the relative air humidity S and thus constitutes a surrogate indicator for the latter:

$$S = S(T - T_d). \tag{20}$$

Because of the lack of temperature and dewpoint temperature forecasts, observations are used instead in the sense of a pseudoforecast. This choice can be justified by the fact that we limit ourselves here to showing the working principles of the BPO. In this context we note that an implementation of the present theory is hampered by the difficulties in accessing continuous records of observations and archived forecasts that have a consistent spatial and temporal resolution, especially because of the rapid evolution of the NWP models.

The HIRLAM model delivers hourly precipitation depths for each model cell, which we aggregated to 6-hourly cumulative values. The forecasts are carried out every 6 h with base times 0000, 0600, 1200, and 1800 UTC + 1. The measured precipitation, recorded at hourly intervals, is accumulated into 6-hourly cumulative precipitation amounts. The hourly temperature–dewpoint temperature differences observations are averaged over 6-h periods. The length of the total investigation period (NWP forecasts and measurements) ranges from 1 January 2000 to 31 October 2005. The Bayesian revision is performed with the 6-hourly QPFs, whereby lead times of 24 and 48 h are chosen. Since the observed and forecast precipitation variables are mixed (binary–continuous), the prior density functions, as well as the posterior probability and posterior distribution functions, need to be evaluated separately for the binary and continuous processes. The binary variable is assumed 0 in the absence of a precipitation event, and 1 otherwise. The distinction between an event and a nonevent is made on the basis of a minimum precipitation threshold of 1 mm in 6 h.

a. Preliminary analysis of forecasted and measured precipitation

First the 6-h cumulative forecasted and measured precipitation variables Z and W, including the temperature differences T, are grouped into four seasons: winter (1 January–31 March), spring (1 April–30 June), summer (1 July–30 September), and autumn (1 October–31 December). This grouping is necessary because the precipitation process is nonstationary and it is not recommended to carry out a stationary statistical analysis, as precipitation patterns change during the annual cycle. In principle, to properly account for effects of nonstationarity, a data analysis on a monthly basis should be performed for applications in many parts of the world. However, given the climate of the study area, which is typically rather stationary over longer periods of time, we consider only seasonal variations. The separation into seasons yields separate statistics that pay tribute to the particular climatic characteristics of the season. Next, the predictor realizations z produced by the NWP model for the two selected lead times (24 or 48 h) are sorted into four series, conditional on the actual event occurrence $v = 1$ or nonoccurrence $v = 0$, indicated with z_{0n} and z_{1n}, where $n = \{24, 48\}$ indexes the 24- or 48-h forecast, respectively.

Figure 1 shows the empirical cumulative distributions conditional on $v = 0$ and $v = 1$ for winter and summer, including the distribution of the actually measured precipitation w. The left plot (winter) shows five distributions. The red dotted curve represents the measured precipitation. The green (24-h forecast) and blue dotted curve (48-h forecast) represent the distribution of the forecasted precipitation depth for events, at which precipitation has actually been observed ($v = 1$). The remaining two curves (crosses) show the probabilities of precipitation depth for those forecasts, in which no actual precipitation has been recorded ($v = 0$) and thus represent the lowest skill predictions. The right plot shows the same empirical probability distributions for the summer season. It is evident that the model has a higher skill in the summer season, yielding a probability distribution on precipitation depth conditional on $v = 1$ more similar to the one of the observations than for the winter season.

b. Normal quantile transform

The next procedural step is to map the respective variates into the normal space through the application of the normal quantile transform (NQT) (e.g., Kelly and Krzysztofowicz 1994). The application of the NQT yields Gaussian surrogate variates in the normal space. In the normal space, expressions for moments of distributions are obtained analytically and relationships between statistical variables, as necessary for deriving the likelihood, can be fitted via (multi)linear regressions. Parametric expressions for conditional marginal densities such as a likelihood, a priori, and expected or posterior density are obtained. The subsequent transformation back into the original space yields the sought-after parametric expressions for the conditional densities and distributions required for the Bayesian formulation of the uncertainty processors.

The approach of employing the NQT, applicable only to absolutely continuous positive non-Gaussian random variables (Herr and Krzysztofowicz 2005), is pursued here to obtain parametric expressions for the prior and posterior densities as well as the likelihood for the IUP. Gaussian variates are obtained by matching an empirical probability distribution $\Gamma[(\cdot)]$ with a Gaussian distribution $Q[(\cdot)]$ and then performing an inversion $Q^{-1}[(\cdot)]$. The NQT, thus indicated with

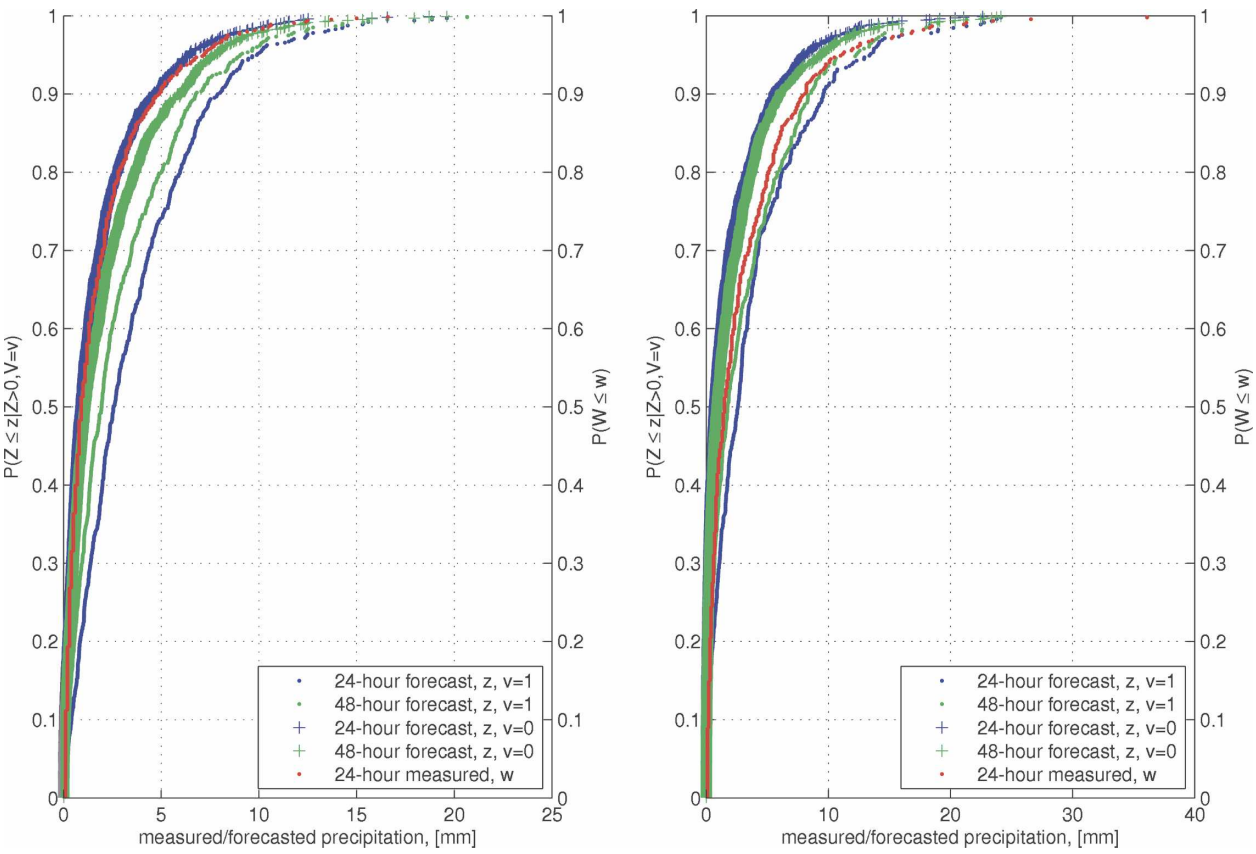

FIG. 1. Cumulative probability of precipitation distributions at Maastricht: (a) winter and (b) summer.

$Q^{-1}[(\cdot)]$, of a non-Gaussian distribution yields thus a series of corresponding transformed normal variables:

$$\omega = Q^{-1}[\Gamma(w)], \tag{21}$$

$$\zeta = Q^{-1}[\Delta(z)], \tag{22}$$

$$\tau = Q^{-1}[\Lambda(t)], \tag{23}$$

$$\sigma = Q^{-1}[\Upsilon(s)], \tag{24}$$

where Γ, Δ, Λ, and Υ are the empirical counterparts of corresponding modeled distributions. The inverse $\Gamma^{-1}[Q(\cdot)]$, $\Delta^{-1}[Q(\cdot)]$, $\Lambda^{-1}[Q(\cdot)]$, or $\Upsilon^{-1}[Q(\cdot)]$ yield again the non-Gaussian variate in the original space.

The upper plot in Fig. 2a shows the forecasted precipitation z versus the measured precipitation w at Maastricht, during winter for lead times of 24 and 48 h, respectively. Table 1 reports the number of events ($\upsilon = 1$) and nonevents ($\upsilon = 0$) for the predictand W and the predictor Z, grouped per seasons. The lower plot presents the normal quantile transforms $Q^{-1}[\Gamma(w)]$ against $Q^{-1}[\Delta(z)]$ of the same data series in the normal space. The data for the 24-h forecast are indicated with dots, while those relative to the 48-h forecast are indicated with diamonds. The linear regressions of the data in the normal space are indicated with a dashed line for the 24-h forecasts and with a dashed–dotted line for the 48-h period. Figure 2b shows the same data for the summer season. It is evident that the correlation between the forecasted and the measured 6-hourly cumulative precipitation is generally poor. The evidently poor correlation between measured and forecasted indicates a low forecasting skill as far as the precipitation depth is concerned. As will become clear further below, this uncertainty on precipitation depth propagates through the IUP into the revised posterior probability through the Bayesian processor.

1) PRIOR DENSITIES

The prior densities need to be derived separately for the binary and the continuous process. For the binary process the prior density reduces to a numerical value, while for the continuous process it is given by an actual density. The necessary procedural steps are given next.

(i) Binary process

The prior probability of precipitation occurrence g in Eq. (10) is derived directly from the historical sample w

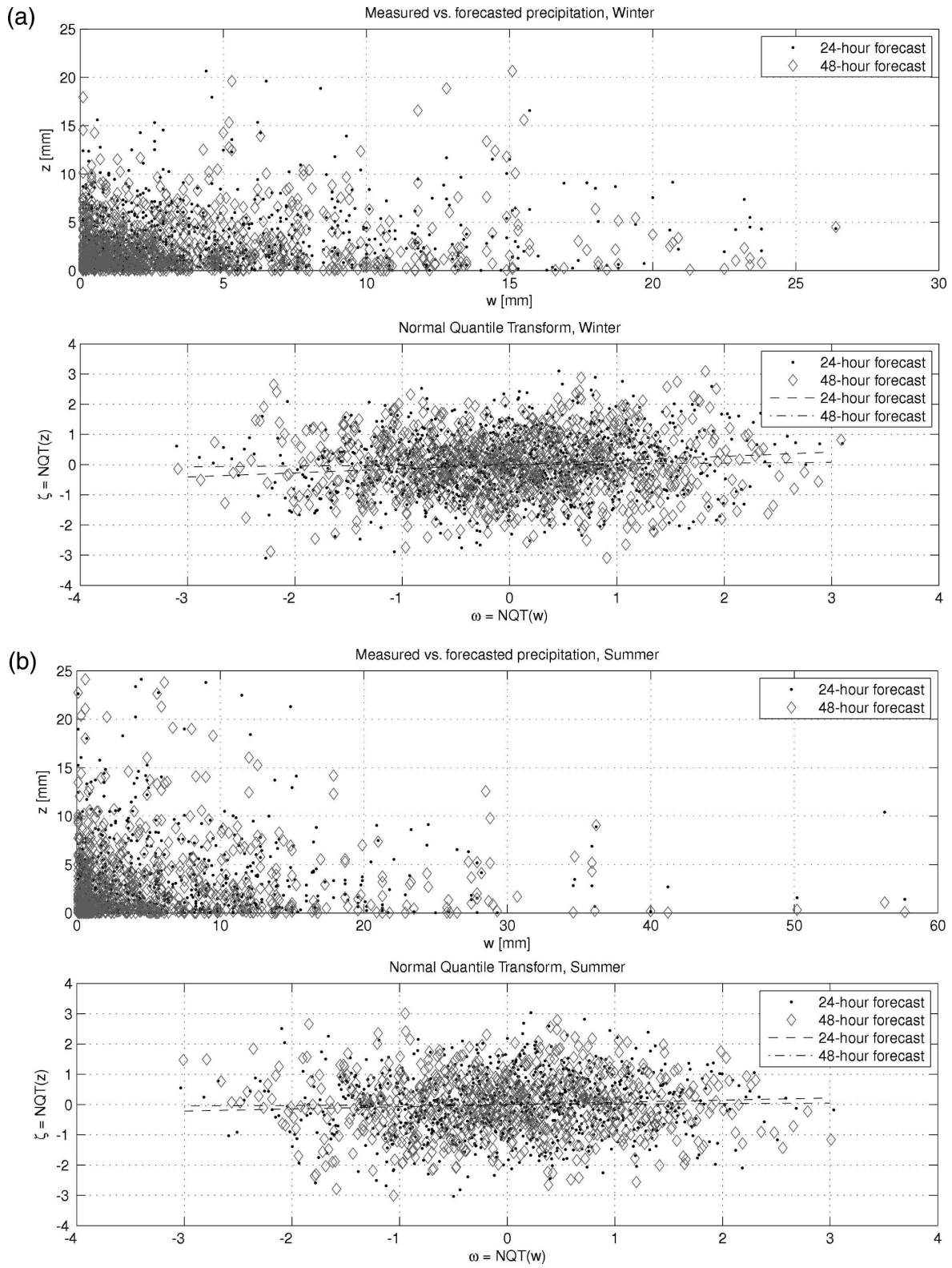

FIG. 2. (a) (top) Forecasted vs measured precipitation and (bottom) normal quantile transform, 24- and 48-h forecasting lead time, winter. (b) (top) Forecasted vs measured precipitation and (bottom) normal quantile transform, 24- and 48-h forecasting lead time, summer.

TABLE 1. Number of events for **W** and **Z**, per season, Maastricht.

Variate	V	Winter	Spring	Summer	Autumn
W	0	1245	1050	1066	1059
	1	923	1154	1142	885
Z	0	1642	1370	1522	1470
	1	526	834	686	474

TABLE 2. Prior probabilities of precipitation occurrence per season, Maastricht.

	Winter	Spring	Summer	Autumn
Sample size	2168	2204	2208	1944
Prior probability g	0.4257	0.5236	0.5172	0.4578

of precipitation recorded at Maastricht, grouped into the four seasons. Table 2 summarizes the prior probabilities of precipitation occurrence over the chosen historical period 1 January 2000–31 October 2005. At Maastricht the probability of precipitation occurrence is slightly higher in spring and summer than in autumn and winter.

(ii) Continuous process

The prior probability density function $g(w|s)$ of the precipitation depth in (13) is also obtained from the historical record. First a two-parameter Weibull distributions are fitted to the empirical distributions $\Gamma(w)$ and $\Upsilon(s)$ of recorded precipitation depths w and the relative humidity s. Once the fitting parameters have been optimized, the probability densities functions are also available in parametric form. Figures 3a and 3b show the empirical cumulative distributions, the fitted two-parameter Weibull distributions, and the derived parametric densities for observed precipitation and temperature–dewpoint temperature differences, respectively, on a seasonal basis. Alternative fitted parametric distribution functions such as the log-logistic and the log-Weibull functions have also been tested. However, these latter models provide suboptimal results for the data series used in this analysis and are not reported here. Next we apply the NQT to the measured precipitation W and the observed humidity S, obtaining the transformed normal variates $\omega = Q^{-1}[\Delta(w)]$ and $\sigma = Q^{-1}[\Upsilon(s)]$. Then, we perform a linear regression, applicable to absolutely continuous random variables:

$$\Omega = d\Sigma + e + \Xi, \tag{25}$$

where d and e are regression constants and Ξ is a variate stochastically independent of Σ and normally distributed with mean zero and variance μ^2. Figure 4 reports a validation of the normal-linear model for Ω on Σ by showing the homoscedasticity of the dependence and the normality of residuals Ξ for the winter season. The conditional mean and variance consequently are

$$E(\Omega|\mathcal{T} = \tau) = d\sigma + e, \tag{26}$$

$$\mathrm{Var}(\Omega|\mathcal{T} = \tau) = \mu^2, \tag{27}$$

while the conditional prior density is

$$g_Q(\omega|\sigma) = \frac{1}{\mu} q\left(\frac{\omega - d\sigma - e}{\mu}\right). \tag{28}$$

Table 3 reports the values of the standard deviation μ and the correlation coefficient R (coinciding with the slope of the linear regression d) for the two lead times and the four seasons. Figures 5a and 5b show the scatterplot of S versus W and σ versus ω and the linear regressions for the four seasons that exhibit a clear correlation between the two variables. The meta-Gaussian prior density in the original space is then obtained by applying the inverse operator of the NQT (e.g., Kelly and Krzysztofowicz 1994):

$$g(w|s) = \frac{\gamma(\omega)}{\mu q\{Q^{-1}[\Gamma(w)]\}} q\left\{\frac{Q^{-1}[\Gamma(w)] - dQ^{-1}[\Upsilon(s)] - e}{\mu}\right\}, \tag{29}$$

while the meta-Gaussian prior cumulative probability function is

$$G(w|s) = Q\left\{\frac{Q^{-1}[\Gamma(w)] - dQ^{-1}[\Upsilon(s)] - e}{\mu}\right\}. \tag{30}$$

For the actual evaluation of (29) and (30) the modeled Weibull fits for $\Gamma(w)$, $\gamma(w)$, and $\Upsilon(s)$ are used (e.g., refer to Figs. 3a and 3b).

2) LIKELIHOOD

The likelihood functions need also to be established separately for the binary and the continuous process, whereby the binary process leads to a likelihood that is parametric with respect to $v = \{0, 1\}$, while the likelihood function of the continuous process results in family of densities, as shown below. Maranzano and Krzysztofowicz (2004) presented the equations for the binary process [Eq. (10)] based on two cases for the predictor Z ($Z = 0$ and $Z > 0$) and performed preliminary verifications. Here we extend the application to the combined binary–continuous process.

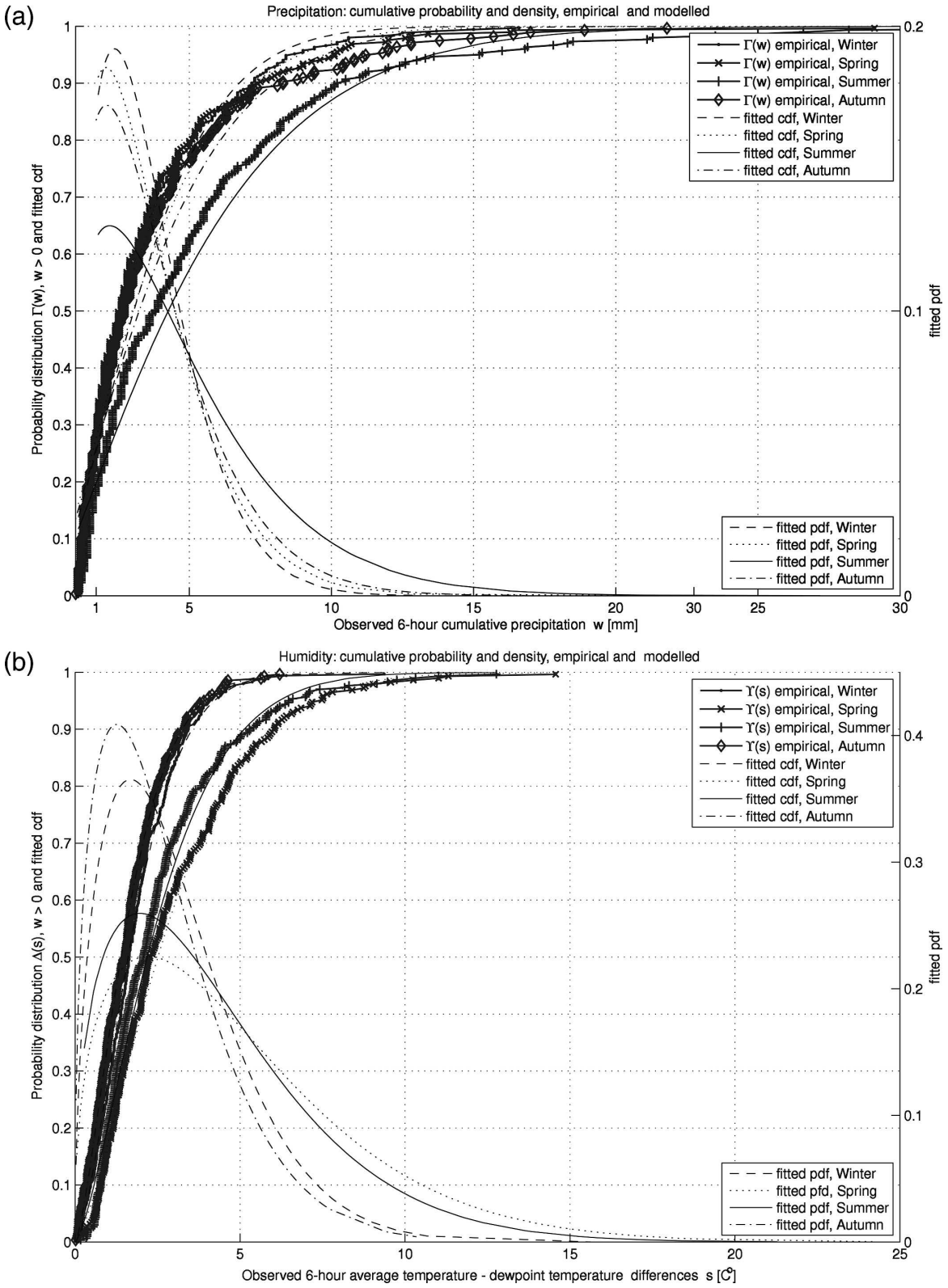

FIG. 3. (a) Empirical and modeled cdf, modeled pdf of *w*, all seasons. (b) Empirical and modeled cdf, modeled pdf of *s*, all seasons.

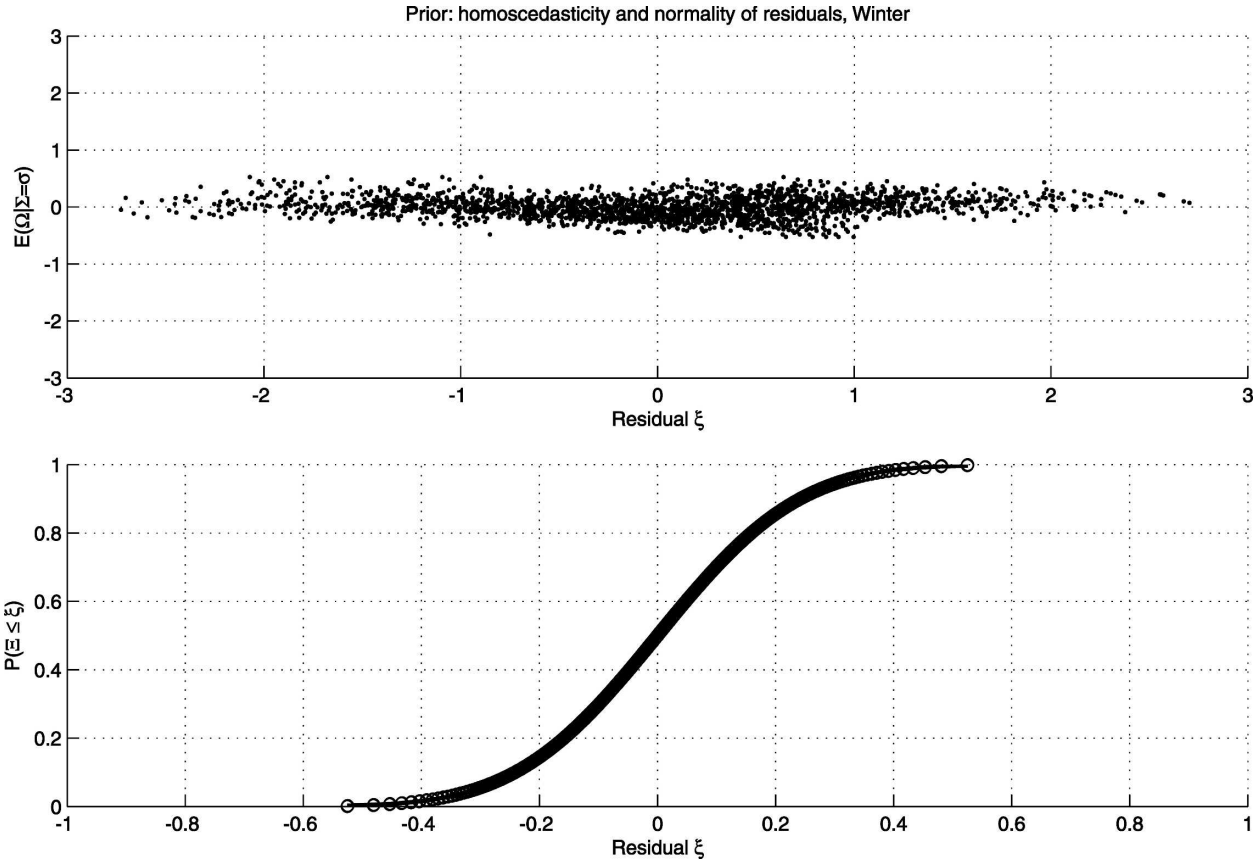

FIG. 4. Prior: homoscedasticity and normality of residuals.

(i) Binary process

The cumulative distribution functions $F_{vn} = P(Z < z|Z > 0, V = v)$, $v = \{0, 1\}$, $n = \{24, 48\}$ have the corresponding density functions f_{vn}. The distributions F_{vn} are derived from historic data for each season and for the two forecasting lead times. Subsequently, a parametric cumulative distribution function is fitted to the data by searching optimal scale and shape parameters. The parametric density functions are obtained from the fitted cumulative distribution functions by derivation. Figure 6a shows the empirical distributions of the forecasted precipitation depths z versus precipitation depth for the two lead times already reproduced in Fig. 1, and the respective Weibull fits, for $v = \{0, 1\}$ and $n = \{24, 48\}$. Once again this parametric model has proven to give the best fit to the particular set of data. Figure 6b shows the corresponding parametric densities f_{vn}. Also, these densities are monotonically decreasing functions indicating a higher probability concentration for low precipitation depths.

TABLE 3. Prior parameters, with d and e the regression constants, μ the std dev of residuals, and R the correlation coefficients.

Regression constant	d	e	μ	R
Winter	−0.1892	0.0	0.1866	0.1892
Spring	−0.2764	0.0	0.2726	0.2764
Summer	−0.3245	0.0	0.3201	0.3245
Autumn	−0.3220	0.0	0.3172	0.3220

(ii) Continuous process

For the continuous process the likelihood function is parameterized by mapping the forecasted precipitation into the standard normal variable $\zeta_n = Q^{-1}[\Delta(z_n)]$, the observed precipitation into $\omega_n = \mathrm{Q}^{-1}[\Delta(w_n)]$, and forecasted relative air humidity into $\tau_n = Q^{-1}[\Lambda(t_n)]$ via the NQT. We recall that the NQT is applied only for values $Z > 0$ and $W > 0$. A linear regression is fitted between the transformed standard normal variates of the precipitation, Ω_n, the forecast, $\mathcal{Z}_n$, and the surrogate humidity $\mathcal{T}_n$ for the forecasting lead times $n = \{24, 48\}$:

$$\Omega_n = a_n\mathcal{Z}_n + b_n\mathcal{T}_n + c_n + X_n, \tag{31}$$

with a_n, b_n, and c_n regression constants and X_n a Gaussian variable, stochastically independent of $(\mathcal{Z}_n, \mathcal{T}_n)$ with

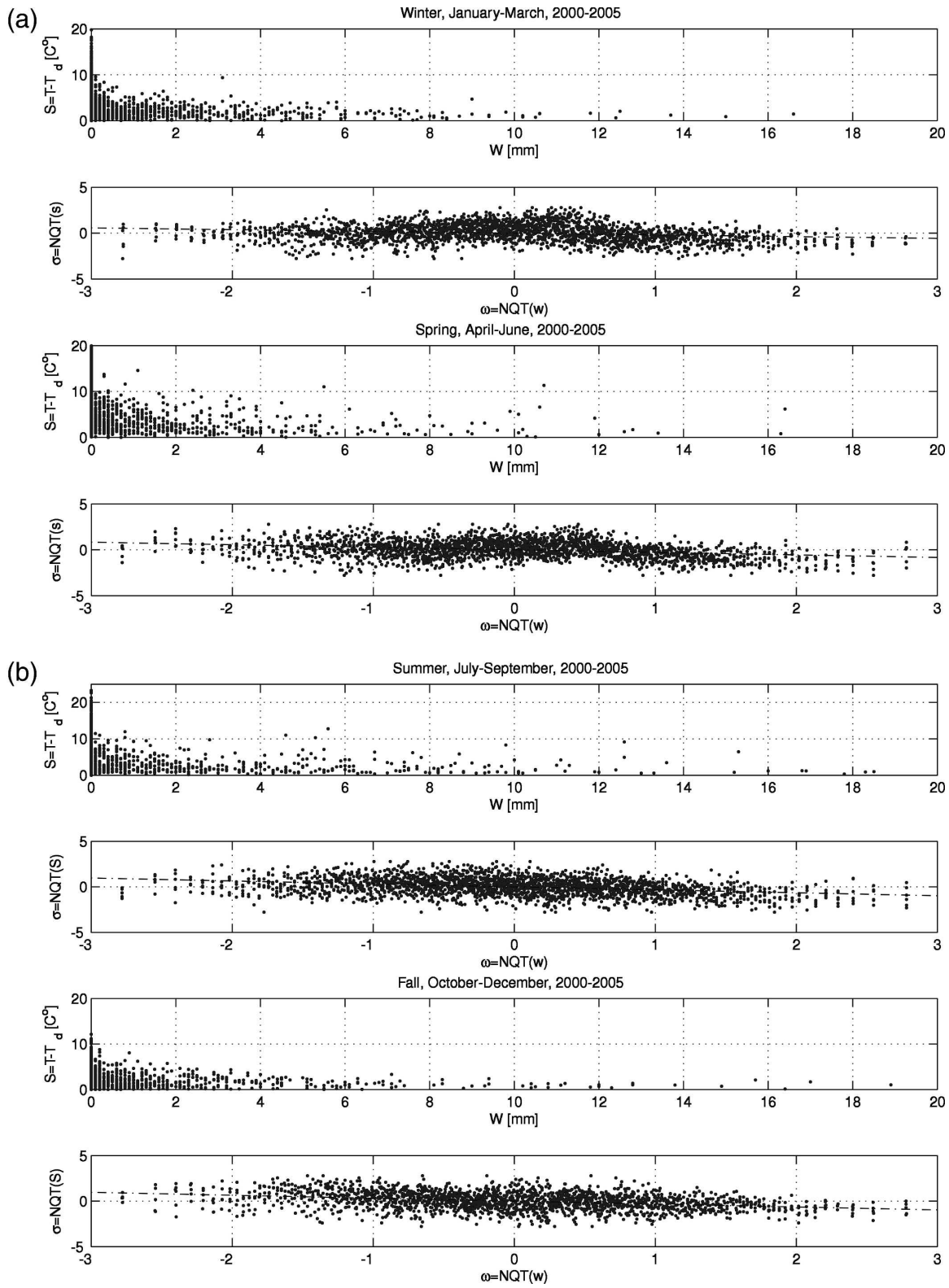

FIG. 5. (a) Scatterplots of observed humidity S vs observed precipitation W and of NQT of S vs NQT of W, winter and spring. (b) Scatterplots of observed humidity S vs observed precipitation W and of NQT of S vs NQT of W, summer and autumn.

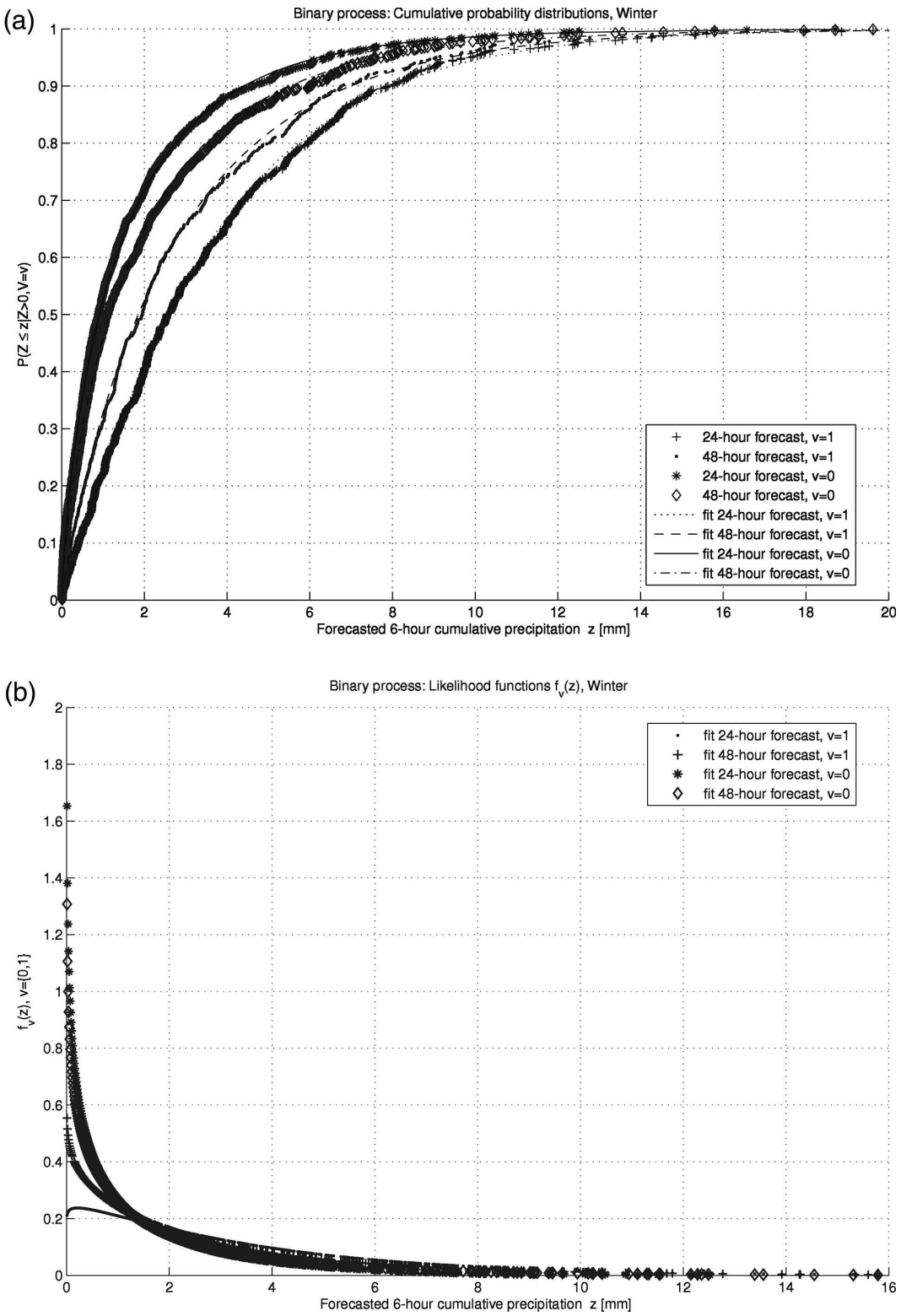

FIG. 6. (a) Binary process: parametric cdfs. (b) Binary process: parametric pdfs.

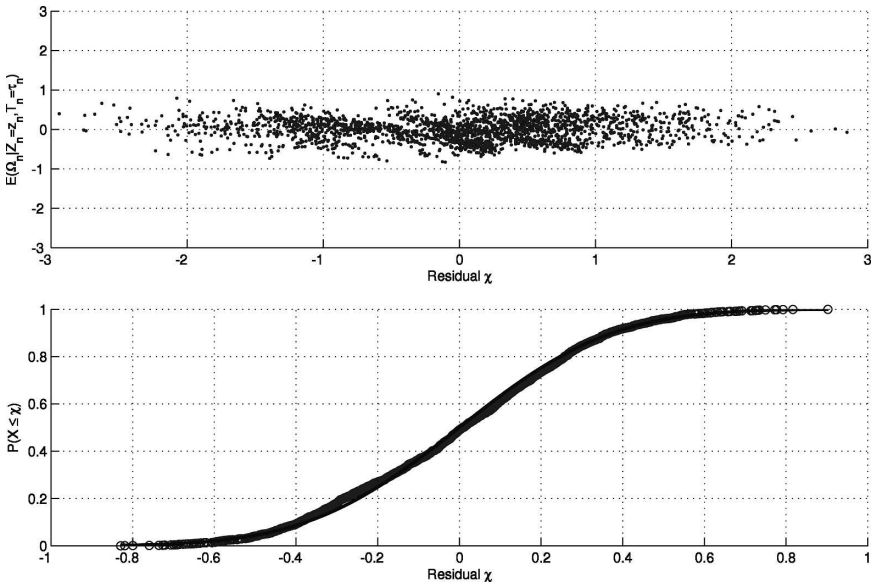

FIG. 7. Likelihood: homoscedasticity and normality of residuals, lead time 24 h.

zero mean and variance ν_n^2. Figure 7 gives a validation of the normal-linear model (31) by proving the homoscedasticity of the dependence and the normality of the residuals X_n for the winter season, lead time 24 h. The conditional mean and variance are

$$E(\Omega_n|\mathcal{Z}_n = \zeta_n, \mathcal{T}_n = \tau_n) = a_n\zeta_n + b_n\tau_n + c_n, \quad (32)$$

$$\mathrm{Var}(\Omega_n|\mathcal{Z}_n = \zeta_n, \mathcal{T}_n = \tau_n) = \nu_n^2, \quad (33)$$

respectively. The transformed conditional density $\phi_{Qn}(\omega_n|\zeta_n, \tau_n)$ of $\phi_n(w_n|z_n, t_n)$ in (13) is given by the expression

$$\phi_{Qn}(\omega_n|\zeta_n, \tau_n) = \frac{1}{\nu_n} q\left(\frac{\omega_n - a_n\zeta_n - b_n\tau_n - c_n}{\nu_n}\right), \quad (34)$$

where q is the normal density. An overview of the values of the regression parameters and the variance of the residuals for the four seasons is given in Table 4. The projection of the linear regression (31) on the $\zeta - \omega$ plane for the winter and summer data mapped to the normal space and two lead times is visible in Figs. 2a and 2b.

TABLE 4. Likelihood parameters, with a_n, b_n, and c_n the regression constants, ν_n the std dev of residuals, and R_n correlation coefficients.

	Lead time n = 24, 48 h	a_n	b_n	c_n	ν_n	R_n
Winter	n = 24 h	0.2333	−0.1428	0.0	0.0857	0.2968
Spring		0.2927	−0.1915	0.0	0.1506	0.3935
Summer		0.2593	−0.2197	0.0	0.1572	0.4019
Autumn		0.2557	−0.2575	0.0	0.1600	0.4060
Winter	n = 48 h	0.2408	−0.1225	0.0	0.2936	0.2977
Spring		0.1467	−0.2186	0.0	0.2926	0.2966
Summer		0.0790	−0.3048	0.0	0.3326	0.3372
Autumn		0.2460	−0.2772	0.0	0.3918	0.3978

3) POSTERIOR PROBABILITY

The posterior revised densities for the binary and the continuous process are derived next. The continuous process requires an inverse transformation from the Gaussian to the original space.

(i) Binary process

Given the family of likelihood functions, and given a prior probability g, the posterior probability Π is given by appropriate substitutions in Eq. (12). Figure 8 shows the probabilities of actual precipitation occurrence during the winter season for a forecasted precipitation depth, given different values of the prior probability g and forecasting lead times 24 and 48 h, respectively. We observe that the probabilities of precipitation occurrence are monotonically increasing for increasing forecasted precipitation in the 24-h forecasts, while the same probabilities are increasing and successively decreasing in the 48-h forecasts. This implies that for 48-h forecasts with high precipitation depths an event occurrence is less likely than for forecasted depths that lie in the medium range. The implications of this behavior of Π on the BPO will be discussed in more depth in section 5.

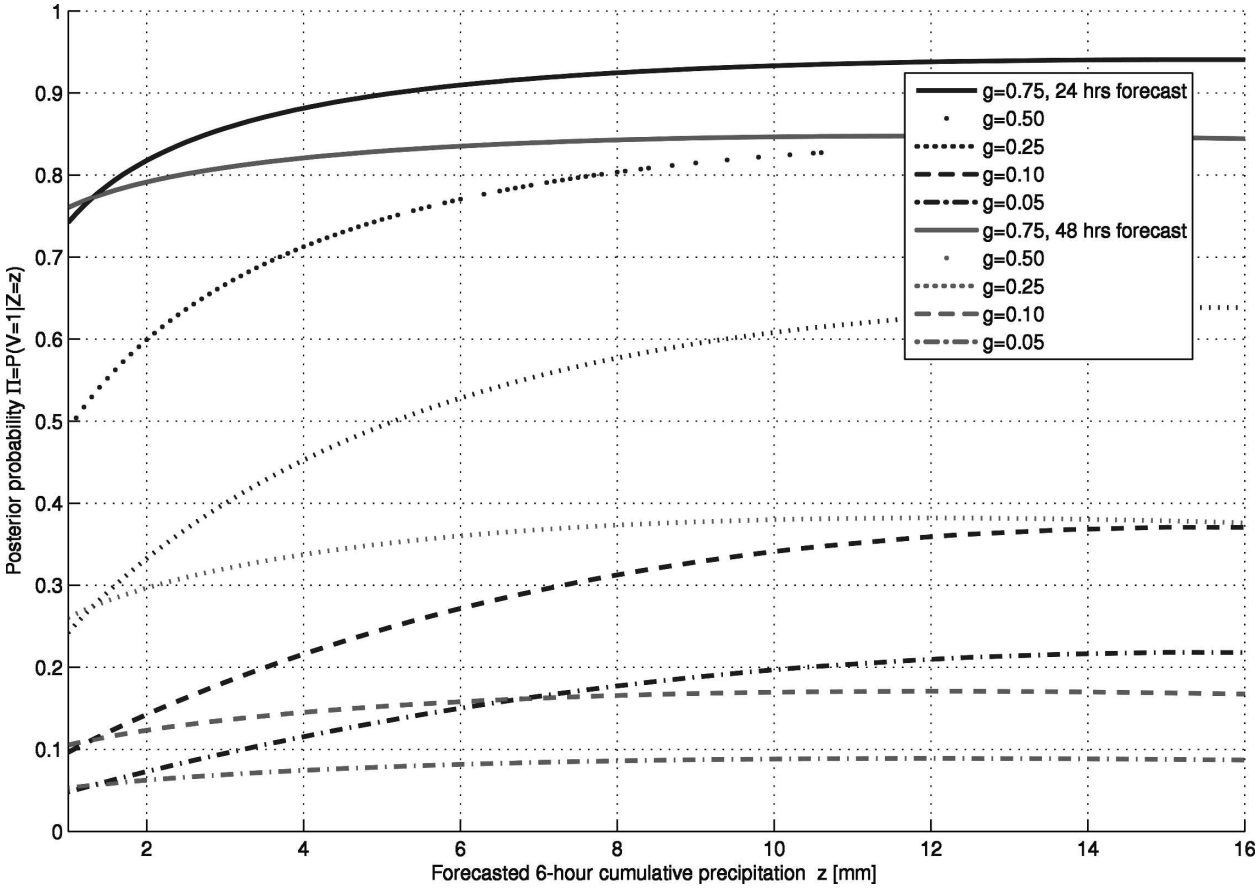

FIG. 8. Posterior probability of precipitation occurrence for a range of values of the prior probability, 24- and 48-h forecast, winter.

(ii) Continuous process

The meta-Gaussian posterior density is obtained by applying the theory of conjugate families of distributions (DeGroot 1970), owing to the normal-linear form of the likelihood (34) and the prior density in the normal space $g_Q(\omega|\sigma) = q\{Q^{-1}[\Gamma(w|s)]\}$, as shown in Krzysztofowicz and Kelly (2000). The expected density at the denominator of (13) becomes

$$k_{Qn}(\zeta_n|\tau_n, \sigma) = \frac{1}{(\mu^2 + \nu_n^2)^{1/2}} q\left[\frac{-a_n\zeta_n - b_n\tau_n - c_n + d\sigma + e}{(\mu^2 + \nu_n^2)^{1/2}}\right]. \quad (35)$$

In practice the posterior density in the normal space is derived by combining (35) with (28) and (34) and by applying the relationship between conjugate normal prior and posterior distributions:

$$\phi_{Qn}(\omega_n|\zeta_n, \tau_n, \sigma) = \frac{1}{K_n} q\left(\frac{\omega_n - A_n\zeta_n - B_n\tau_n - C_n\sigma - D_n}{K_n}\right). \quad (36)$$

The constants A_n, B_n, C_n, D_n, and K_n^2 have the following algebraic expressions:

$$A_n = \frac{\mu^2 a_n}{\mu^2 + \nu_n^2}, \quad (37)$$

$$B_n = \frac{\mu^2 b_n}{\mu^2 + \nu_n^2}, \quad (38)$$

$$C_n = \frac{\nu_n^2 d}{\mu^2 + \nu_n^2}, \quad (39)$$

$$D_n = \frac{\mu^2 c_n + \nu_n^2 e}{\mu^2 + \nu_n^2}, \quad (40)$$

$$K_n^2 = \frac{\mu^2 \nu_n^2}{\mu^2 + \nu_n^2}. \quad (41)$$

The numerical values of the constants, evaluated for the selected dataset, are summarized in Table 5. Equation (36) represents a family of distributions of the normal variable ω_n, transform of the measured precipita-

TABLE 5. Coefficients of the posterior density in the normal space.

Coefficient		A_n	B_n	C_n	D_n	K_n
Lead time n = 24, 48 h		$(\mu^2 a_n/\mu^2 + \nu_n^2)$	$(\mu^2 b_n/\mu^2 + \nu_n^2)$	$(\nu_n^2 d/\mu^2 + \nu_n^2)$	$(\mu^2 c_n + \nu_n^2 e/\mu^2 + \nu_n^2)$	$(\mu^2\nu_n^2/\mu^2 + \nu_n^2)$
Winter	n = 24 h	0.0858	−0.0525	−0.1196	0.0	0.0220
Spring		0.1345	−0.0880	−0.1493	0.0	0.0402
Summer		0.1563	−0.1324	−0.1289	0.0	0.0407
Autumn		0.1311	−0.1321	−0.1569	0.0	0.0490
Winter	n = 48 h	0.0926	−0.0471	−0.1165	0.0	0.0214
Spring		0.0882	−0.1315	−0.1102	0.0	0.0296
Summer		0.0438	−0.1690	−0.1446	0.0	0.0456
Autumn		0.1158	−0.1305	−0.1703	0.0	0.0532

tion w_n, parameterized with respect to all possible ζ_n, transforms of the forecasted precipitation z_n, and τ_n, and σ transforms of t_n and s, respectively. The meta-Gaussian posterior density in the original space is then obtained by applying the inverse operator of the NQT [e.g., Kelly and Krzysztofowicz (1994)]:

$$\phi(w_n|z_n, t_n, s) = \frac{\gamma(w_n)}{K_n q\{Q^{-1}[\Gamma(w_n)]\}} \cdot q\left\{\frac{Q^{-1}[\Gamma(w_n)] - A_n Q^{-1}[\Delta(z_n)] - B_n Q^{-1}[\Lambda(t_n)] - C_n Q^{-1}[\Upsilon(s)] - D_n}{K_n}\right\} \tag{42}$$

with $\gamma(\cdot)$ the normal density, $\Gamma(\cdot)$ the cumulative distribution function, and

$$J(w_n) = \frac{\gamma(w_n)}{q\{Q^{-1}[\Gamma(w_n)]\}} \tag{43}$$

the Jacobian of the transformation. The meta-Gaussian posterior cumulative distribution finally reads as follows:

$$\Phi(w_n|z_n, t_s, s) = Q\left\{\frac{Q^{-1}[\Gamma(w_n)] - A_n Q^{-1}[\Delta(z_n)] - B_n Q^{-1}[\Lambda(t_n)] - C_n Q^{-1}[\Upsilon(s)] - D_n}{K_n}\right\}. \tag{44}$$

For the evaluation of expressions (42) and (44) the modeled Weibull fits for $\gamma(w_n)$, $\Gamma(w_n)$, $\Delta(z_n)$, $\Lambda(t_n)$, and $\Upsilon(s)$ are used.

4) THE PROBABILISTIC QUANTITATIVE PRECIPITATION FORECAST

For the PQPF, the posterior density function ψ given by (19) and the posterior distribution function Ψ given by (18) are obtained by combining Eqs. (42) and (44) with the posterior probability Π_z given by (12). Figures 9a and 9b show the posterior conditional density $\psi(w_n|z_n, t_n, s)$ given by the BPO and corresponding distribution functions Ψ for (a) winter and (b) summer, for forecasted values of z_n and t_n and an observed value of s_0 at the begin of the forecast. Equation (19) represents the input uncertainty processor $\psi(\cdot|\mathbf{z}, \mathbf{t}, \mathbf{s})$ (IUP), which then needs to be integrated with the hydrologic uncertainty processor $\theta(\cdot|\mathbf{r}, \mathbf{h}_0, \mathbf{u}, \mathbf{w})$ (HUP) to obtain the total predictive uncertainty of the forecast [Eq. (9)].

5. Discussion

The precedent exposure shows the necessary procedural steps for the implementation of an IUP as a Bayesian processor of output for operational use. In this section we discuss some of the most relevant properties of the processor for operational flood forecasting. A series of general properties of the proposed IUP are highlighted first:

1) The processor is based exclusively on QPFs and the measured historical precipitation time series. It does not require output information from the hydrological model. As such, the BPO strictly separates the precipitation uncertainty due to numerical weather prediction from uncertainty attributable to the hydrological component of the forecast.
2) The computations required for operating the processor can be carried out offline. From the statistical analysis of observed and forecasted precipitation,

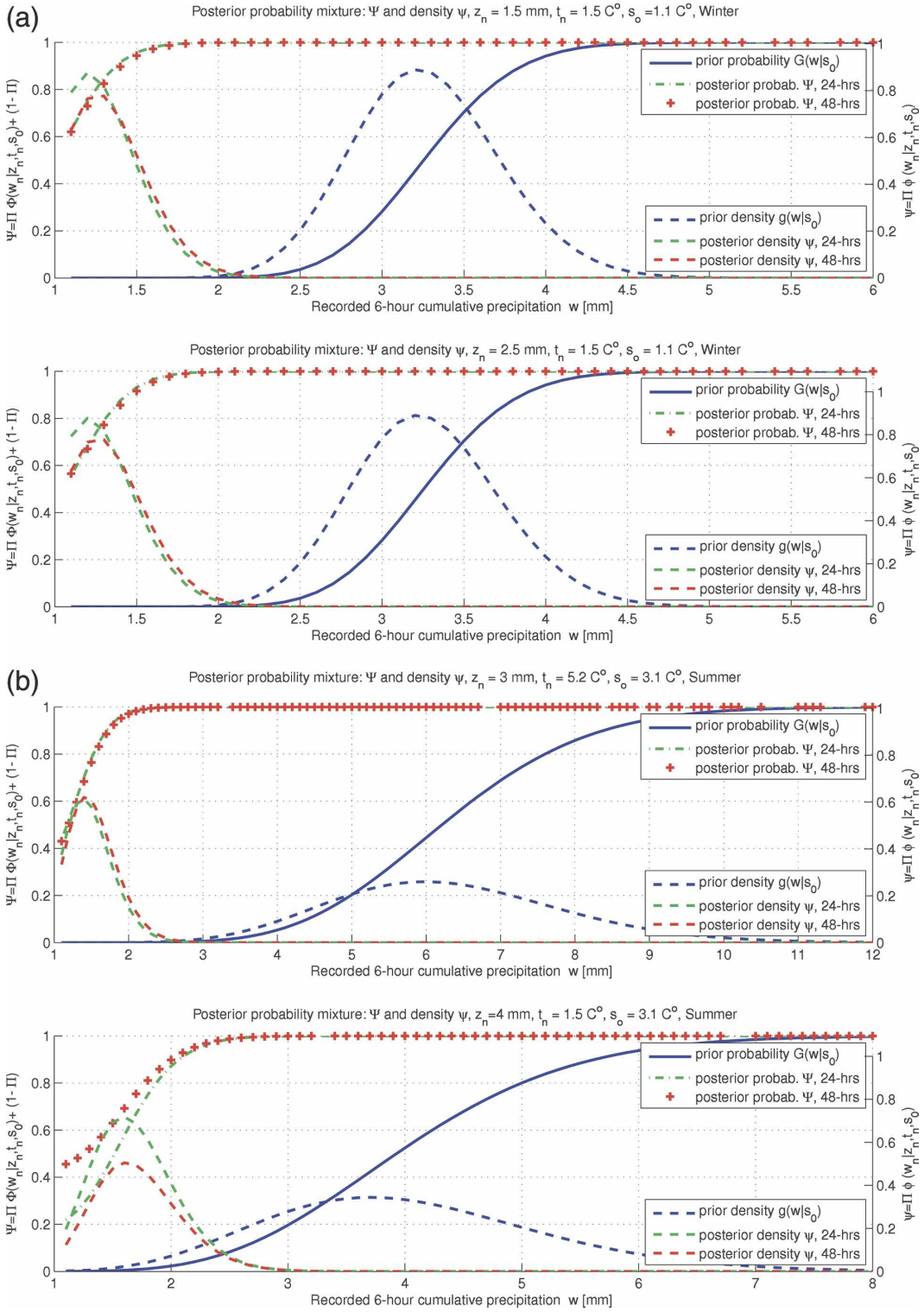

FIG. 9. (a) Prior probability distribution and posterior mix of revised distributions, prior probability density functions, and posterior mix of revised density functions, winter. (b) Prior probability distribution and posterior mix of revised distributions, prior probability density functions, and posterior mix of revised density functions, summer.

the parameters A_n, B_n, C_n, D_n, and K_n^2 in Table 5 are fitted for the four seasons. These parameters need to be updated periodically as the historical forecasted and measured series expand in time. In this sense the processor is well suited for an application in an operational context, where as little as possible computational effort should be spent at forecasting time. The execution of the processor consists in evaluating first expressions (12), (42), and (44) for a given forecasted precipitation depth **z** and the (surrogate) relative air humidities **t** and **s** and then the mixed cumulative distributions and densities (18) and (19). The latter two expressions are plotted in Figs. 9a and 9b for selected forecasted precipitation depths and relative air humidity (i.e., temperature–dewpoint temperature differences).

On the present application of the BPO to a chosen set of measured and forecasted data, the following observations can be made:

3) The correlations between the 6-hourly accumulated forecast data for either lead time with the measured precipitation are very poor. To improve these correlations, areal averages of measured precipitation over larger areas need to be compared with corresponding forecasts that are averaged over more adjacent model cells. After the actual data have been transformed into the normal space (Figs. 2a and 2b) a linear regression between the transformed variables yields very low correlation coefficients as evident from Table 4. As a result the informational added value of the QPF with respect to precipitation depth is low, resulting in negligible added value of the Bayesian revision of the prior probability density $g(\mathbf{w}|\mathbf{s})$. This fact underscores that the contribution to the uncertainty of a flood forecast that can be attributed to the numerical weather forecast is significant. This uncertainty needs to be quantified systematically when QPFs are used to drive hydrological models in predicting critical discharges at a basin outlet. The BPO provides such a quantification tool.
4) Similarly, the correlation between observed (surrogate) relative air humidity and forecasted precipitation is also low (see Table 4) and would be even lower if forecasted humidity would have been used in estimating the likelihood instead of observations. From this we deduce that the inclusion of additional auxiliary indicators on the state of the atmosphere, such as relative air humidity on the ground or ground air pressure, adds little informational value to the Bayesian revision process of the prior probability $g(\mathbf{w}|\mathbf{s})$ because of the inherently limited skill of the model in predicting precipitation depth.
5) Graphs in Figs. 9a and 9b show the mixed posterior distribution function given by (18) and posterior density function given by (19), which are plotted for winter and summer and selected values of the forecasted precipitation depth z, the (pseudo) forecasted humidity t, and the humidity $\mathbf{s_0}$ observed at the begin of the forecast. The posterior mix Ψ is significantly influenced by the weighting with the probability Π of precipitation occurrence. The weighting is responsible to a large extent for the shift of the posterior mix Ψ with respect to the prior distribution $G(w|s_0)$. The posterior probability Φ and density ϕ of precipitation depth (not shown graphically) resemble in shape the respective priors. This similarity between prior and posterior is explained by the fact that the QPF has significantly more skill in predicting precipitation occurrence/nonoccurrence rather than predicting precipitation depth at a location (when comparing the QPF with one precipitation measurement at a single location). The utilization of an uncertainty processor like the BPO allows us to identify and quantify to what extent the model adds value to a forecast. With decreasing information content of the QPF with respect to precipitation depth and probability of event occurrence, the posterior mix of probabilities collapses onto the prior distributions. The forecaster will not be able to deliver a posterior probability, which is better than what he could have predicted without a new model run, solely on the basis of information on historical model performance expressed by the likelihood function.
6) To effectively assess the validity of the BPO procedure and the benefits of using it in producing PQPFs a sound validation is required. In the present case a series of actual observations in the January 2000–October 2005 period could be compared against probabilities of occurrence produced by the processor. However, given the low skill of the forecasts for the available dataset, we have refrained from performing such validation. Further validation can be done by applying the BPO to a set of forecasts at different locations. Systematic comparisons with other common methods for probability of precipitation determination, for example, the model output statistics (MOS) technique (Glahn and Lowry 1972), need also to be investigated. Such an analysis however is beyond the scope of this paper and is omitted here.
7) The impact of spatial and temporal variability of precipitation on the predictive uncertainty has not been addressed here. The focus is exclusively on a single small-scale plot. We emphasize however that

this assumption does not hold if the processor is to be used to assess the predictive uncertainty for larger areas. In this case precipitation may vary strongly in space and in time according to the particular spatial (e.g., local, regional, continental) and temporal (e.g., 6-, 12-, 24-, 48-hourly, or longer periods) scale of observation. The wide range of spatiotemporal scales must be taken into account by identifying focus areas of various size and by aggregating forecasts over periods representing typical problem-specific time scales of interest. These time scales typically discern small basins with response times of a few hours from large basins, whose contraction times lie in the order of days.

For the actual implementation of the full BFS the proposed IUP needs to be integrated with an appropriate hydrologic uncertainty processor. The design and implementation of such a processor is given in Krzysztofowicz and Kelly (2000) (precipitation occurrence-independent model) and Krzysztofowicz and Herr (2001) (precipitation-dependent model). A combination of the BPO as IUP with the HUP through an integrator of uncertainty will be matter of sequel work.

6. Summary and conclusions

An implementation of a modified BPO (previously described in Krzysztofowicz 2004) to process output from a numerical weather prediction model has been presented, in which relative air humidity is used as an ancillary indicator of the state of the atmosphere. The purpose of the processor is to quantify the uncertainties associated with an event for both the event occurrence and the precipitation depth, on the basis of quantitative precipitation forecasts (QPFs) produced by the HIRLAM model. The application is carried out for 6-h QPFs and lead times of 24 and 48 h. The processor can be used as a input uncertainty processor (IUP) for the forcing input into hydrological models. Integration of the IUP with a corresponding hydrologic uncertainty processor yields a Bayesian forecasting system (BFS) for quantification of the total predictive uncertainty in flood prediction.

The precipitation process is represented as a binary–continuous mixture of random variables. The processor is Bayesian in the sense that *prior* knowledge on the predictive capability of the QPF over a historical period is used to obtain a *revised* probabilistic quantitative precipitation forecasts expressed in terms of *posterior* revised statistics.

The processor has been tested with forecasted precipitation series produced by the HIRLAM QPF. The forecasts are compared with precipitation measured at one single station. A preliminary analysis on the measured data is performed to establish prior probabilities of precipitation occurrence and depth for the four seasons. From the analysis of the forecasts it is evident that the skill of the model lies mainly in predicting the event occurrence/nonoccurrence (binary process) rather than in predicting the event depth (continuous process).

The Bayesian processor has the underlying property that poor model performance through noninformative forecasts, such as seen here, can be promptly identified, because the revised probabilistic forecasts show little improvement in terms of information content with respect to prior knowledge. From the results it is evident that the mixed binary–continuous posterior probabilities are mainly influenced by the revised probabilities of the binary process, where the larger forecasting skill of the QPF actually lies. The revised posterior probabilities of the continuous process show limited improvement with respect to the prior information and have thus negligible impact on the posterior mix. This is attributable to the rather low skill of the QPF with respect to precipitation depth. This example shows how the input uncertainty processor formulated as a Bayesian processor of output and shown here is suited as instrument to quantify sources of uncertainty in the forecasting process and emphasizes the utility for employing Bayesian theory in uncertainty quantification.

The present investigation ignores spatial and/or temporal variability of rainfall and is therefore only applicable for small basins, in which these characteristics are negligible. The uncertainty due to spatiotemporal variability of the precipitation fields must be adequately addressed and remains a matter for future investigations.

Acknowledgments. We thank the Dutch Royal Meteorological Institute (KNMI) for having provided the HIRLAM model results and the weather station data records. We would also like to thank three anonymous referees for suggestions that have helped improve the manuscript.

APPENDIX

Acronyms

BFS	Bayesian forecasting system
BPO	Bayesian processor of weather model output
HIRLAM	High-Resolution Limited-Area Model
HUP	Hydrologic uncertainty processor
IU	Integrator of uncertainty

IUP	Input uncertainty processor
PQPF	Probabilistic quantitative precipitation forecast
PQFF	Probabilistic quantitative flood forecast
QPF	Quantitative precipitation forecast

REFERENCES

Beven, K., and A. Binley, 1992: The future of distributed models: Model calibration and uncertainty prediction. *Hydrol. Processes,* **6,** 279–298.

De Groot, M. H., 1970: *Optimal Statistical Decisions.* McGraw-Hill, 489 pp.

Glahn, H. R., and D. A. Lowry, 1972: The use of model output statistics (MOS) in objective weather forecasting. *J. Appl. Meteor.,* **11,** 1203–1211.

Herr, H. D., and R. Krzysztofowicz, 2005: Generic probability distribution of rainfall in space: The bivariate model. *J. Hydrol.,* **306,** 234–263.

Kelly, K. S., and R. Krzysztofowicz, 1994: Probability distributions for flood warning systems. *Water Resour. Res.,* **30,** 1145–1152.

——, and ——, 2000: Precipitation uncertainty processor for probabilistic river stage forecasting. *Water Resour. Res.,* **36,** 2643–2653.

Krzysztofowicz, R., 1999: Bayesian theory of probabilistic forecasting via deterministic hydrologic model. *Water Resour. Res.,* **35,** 2739–2750.

——, 2001a: Reply. *Water Resour. Res.,* **37,** 441–442.

——, 2001b: The case for probabilistic forecasting in hydrology. *J. Hydrol.,* **249,** 2–9.

——, 2004: Bayesian processor of output: A new technique for probabilistic weather forecasting. Preprints, *17th Conf. on Probability and Statistics in the Atmospheric Sciences,* Seattle, WA, Amer. Meteor. Soc., 4.2.

——, and K. S. Kelly, 2000: Hydrologic uncertainty processor for probabilistic river stage forecasting. *Water Resour. Res.,* **36,** 3265–3277.

——, and H. D. Herr, 2001: Hydrologic uncertainty processor for probabilistic river stage forecasting: Precipitation-dependent model. *J. Hydrol.,* **249,** 46–68.

Lardet, P., and C. Obled, 1994: Real-time flood forecasting using a stochastic rainfall generator. *J. Hydrol.,* **162,** 391–408.

Liu, Z., M. L. V. Martina, and E. Todini, 2005: Flood forecasting using a fully distributed model: Application of the TOPKAPI model to the Upper Xixian Catchment. *Hydrol. Earth Syst. Sci.,* **1** (9), 69–86.

Maranzano, C. J., and R. Krzysztofowicz, 2004: Bayesian processor of output for probabilistic forecasting of precipitation occurrence. Preprints, *17th Conf. on Probability and Statistics in the Atmospheric Sciences,* Seattle, WA, Amer. Meteor. Soc., 4.3.

Schaake, J., and L. Larson, 1998: Ensemble streamflow prediction (ESP): Progress and research needs. Preprints, *Special Symp. on Hydrology,* Amer. Meteor. Soc., Phoenix, AZ, J19–J24.

——, K. Franz, A. Bradley, and R. Buizza, 2006: The Hydrologic Ensemble Prediction Experiment (HEPEX). *Hydrol. Earth Syst. Sci. Discuss.,* **3,** 3321–3332.

Undén, P., and Coauthors, 2002: HIRLAM-5 scientific documentation, December 2002. HIRLAM Tech. Rep., 132 pp.+appendix.