An information-theoretical perspective on weighted ensemble forecasts[☆]

Steven V. Weijs[1,a,*], Nick van de Giesen[b]

[a]*Environmental Fluid Mechanics Laboratory, EFLUM/IEE/ENAC, EPFL, Lausanne, Switzerland*
[b]*Water Resources Management, TU Delft, Delft, The Netherlands*

Abstract

This paper presents an information–theoretical method for weighting ensemble forecasts with new information. Weighted ensemble forecasts can be used to adjust the distribution that an existing ensemble of time series represents, without modifying the values in the ensemble itself. The weighting can, for example, add new seasonal forecast information in an existing ensemble of historically measured time series that represents climatic uncertainty. A recent article in this journal compared several methods to determine the weights for the ensemble members and introduced the pdf-ratio method. In this article, a new method, the minimum relative entropy update (MRE-update), is presented. Based on the principle of minimum discrimination information, an extension of the principle of maximum entropy (POME), the method ensures that no more information is added to the ensemble than is present in the forecast. This is achieved by minimizing relative entropy, with the forecast information imposed as constraints. From this same perspective, an information–theoretical view on the various weighting methods is presented. The MRE-update is compared with the existing methods and the parallels with the pdf-ratio method are analysed. The paper provides a new, information–theoretical justification for one version of the pdf-ratio method that turns out to be equivalent to the MRE-update. All other methods result in sets of ensemble weights that, seen from the information–theoretical perspective, add either too little or too much (i.e. fictitious) information to the ensemble.

Keywords: weighted ensemble forecast, minimum relative entropy, maximum entropy, information theory

1. Introduction

> *"Probability is relative, in part to [our] ignorance, in part to our knowledge."* - Pierre-Simon Laplace, 1825

A recent article in this journal by Stedinger and Kim (2010) addressed methods to produce weighted ensemble forecasts. In this type of forecasts, weights for the individual ensemble members are updated to deviate from the usual equal weights, to reflect information that was not included in the generation of the ensemble. The method of weighting ensemble forecasts has great potential to combine information from different sources at different scales and weighted ensembles are readily applicable in water resources planning and reservoir optimization. A possible application is the generation of seasonal ensemble forecasts by weighting historically observed time series based new forecast information from large scale circulation patterns. Two existing methods for weighting ensembles are the pdf-ratio method of Stedinger and Kim (2010), which adjusts the weights proportional to the ratio of an original and a new pdf, and the nonparametric and parametric updates of Croley (1996, 2003), which find weights by minimizing the mean squared difference in probabilities, with new information imposed as constraints.

Although the updating of the weights constitutes information added to the ensemble, the problem of weighting ensembles has never before been analysed from the perspective of information theory. Applying information–theoretical principles can be essential to ensure that the reduction in uncertainty caused by the weighting is in balance with the information provided by the forecast.

In this paper we argue that the problem of weighting ensembles is amenable to the information–theoretical principle of maximum entropy (POME). We present the minimum relative entropy update (MRE-update), a method for finding weights based on an extension of POME. The objective of the paper is to present the theoretical basis for the MRE-update and to demonstrate its use. Furthermore, we present an information–theoretical view on previous weighting methods and compare them with the MRE-update, specifically focusing on information flows that result from the different methods.

Using the information–theoretical concepts, we show that the Croley parametric update is a second order approximation of the MRE-update (Weijs, 2011). The parallels with the pdf-ratio method of Stedinger and Kim (2010) are also explored. We conclude that when the Gaussian version of the pdf-ratio method is forced to exactly match the prescribed conditional

mean and variance, the results are identical to the MRE-update. Firstly, this is an information–theoretical justification for this version of the pdf-ratio method. Secondly, it indicates the pdf-ratio method as a fast way to solve the MRE-update in case the forecast information consists of a conditional mean and variance. We only give a short introduction to weighted ensemble forecasts here. For more background and references, the reader is referred to Stedinger and Kim (2010).

1.1. Use of ensembles in water resources

Decision making about water resources systems often requires uncertainty to be taken into account. Ensembles are a common method to describe uncertainty in forecasts, such as future inflows to a reservoir system. An ensemble consists of several scenarios (also called members or traces), which represent the possible future development of the variables of interest. An ensemble of past measured streamflows can, for example, be used as a stochastic description of the inputs to a system of reservoirs (Kelman et al., 1990). Ideally, such a historical ensemble represents the climatic uncertainty about the inter-annual variability, and at the same time contains a realistic stochastic description of spatial and temporal variability and correlations at smaller timescales. Using ensembles directly has the advantage that no statistical models have to be assumed. Ensemble members can be multivariate, e.g. yearly sets of daily time series of various hydrological variables on various locations (Faber and Stedinger, 2001). Other examples of ensembles are the ensemble weather forecasts of European Centre for Medium range Weather Forecasts (ECMWF), real-time hydrological forecasts based on these ensemble weather forecasts (Reggiani et al., 2009; Pagano et al., 2013), and the extended streamflow predictions (ESP) that are used throughout the USA as an input for reservoir optimization models or decision making about flood protection.

To make the remainder of this introduction not too abstract, it will focus on ESP ensembles, even though the methods presented are more generally applicable to any ensemble. ESP forecasts reduce uncertainty by conditioning predicted streamflow on initial basin conditions. The ESP forecast is produced by feeding a distributed hydrological model, which has an initial state consistent with actual basin conditions, with past observed weather patterns (Day, 1985; Wood and Lettenmaier, 2006). The result is an ensemble of streamflows with one trace for each historical weather pattern. The ensemble reflects the climatic uncertainty, but also the information that is in the actual basin conditions (Wood and Lettenmaier, 2008). This information reduces the climatic uncertainty in the flows.

Apart from the initial basin conditions, information about the streamflows might be present in climatic indexes that characterize long term persistence in the atmospheric and oceanic circulation (Piechota et al., 1998). For example, the phase of the El Niño Southern Oscillation (ENSO) and the Pacific Decadal Oscillation (PDO) gives information about the precipitation in the Pacific Northwest of the USA (Hamlet and Lettenmaier, 1999). Hamlet et al. (2002) proposed a method to select only the one third of ESP traces that match the ENSO conditions of the actual year. They calculated that this information, in combination with a more flexible reservoir operation strategy, could lead to "an increase of non-firm energy production from the major Columbia River hydropower dams, ... resulting in an average increase in annual revenue of approximately $153 million per year in comparison with the status quo." This paper presents a method to include this type of information into ensemble forecasts by weighting rather than selecting ensemble traces.

1.2. Weighted ensembles

Ensemble traces are often stated to be "equally likely", see e.g. Cloke and Pappenberger (2009). This should not be taken too literally. For example, streamflows close to the inter-annual mean are more likely than streamflows of the most extreme ensemble members in terms of probability density. Also, the probability of any scenario exactly occurring with the same real number(s) as outcome is zero. Ideally ensemble forecasts are produced in such a way that all ensemble members can be considered to represent equally probable ranges of the possible outcomes "closest" to the scenario. This is reflected by the fact that scenarios usually lie closer to each other around the mean value. Each scenario represents the same probability, but a different region of the space (or line in the univariate case) of the outcome and therefore a different probability density; see Fig. 1. In that way, the ensemble is a discrete representation of the underlying probability density function and can be used in Monte-Carlo approaches, risk analysis and decision making, see e.g. Georgakakos and Krzysztofowicz (2001) or Van Overloop et al. (2008). It also means that we can treat the ensemble weights as probabilities, even though strictly speaking they are not.

Often, long-term forecasts based on, for example, ENSO do not contain information at a high spatial and temporal detail level, but rather contain information about averages in time and space, e.g. the total flow at the outlet of a river basin, averaged over several months. Yet, risk analysis or optimal operation may depend on events and sequences at shorter timescales and smaller spatial scales. One could attempt to shift the time series in the ensemble to match the long term forecast (bottom right Fig 1), but this could destroy the realism of the traces if the shifting or scaling procedure is not sophisticated enough.

A reasonable alternative to combine detailed information in the ensemble of historical time series with forecast information is to update the probabilities of individual ensemble members based on the averages the series are labeled with, while leaving the time series they contain intact (Werner et al., 2004; Stedinger and Kim, 2010). This has the advantage of preserving high-resolution stochastic structure within the ensembles. The update of the weights is thus based on averages in space and time, derived from the time series, compared to information on these same quantities in the seasonal forecast; see top right of Fig. 1.

From the information–theoretical perspective presented in this paper, updating ensemble probabilities to deviate from equal probabilities implies adding information to the ensemble. That information is measured by the relative entropy between the original and the new probabilities (Cover and Thomas, 2006);
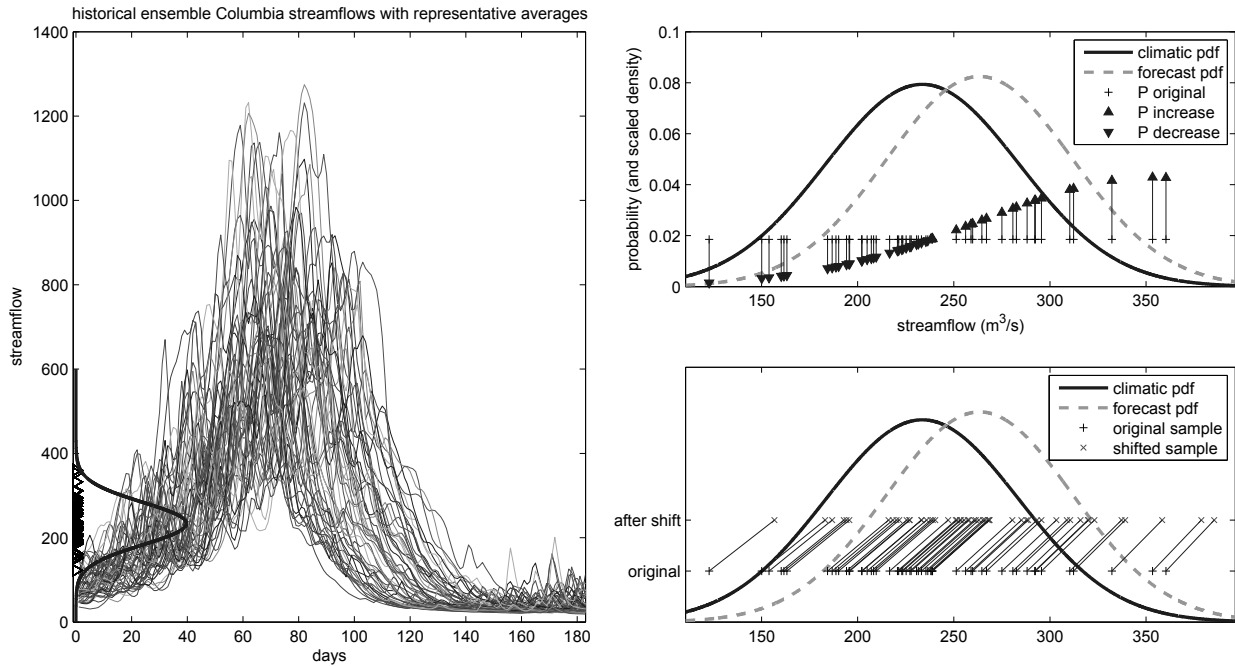
Figure 1: Equally weighted ensemble members can represent a nonuniform density. This density can be changed by shifting or by weighting the ensemble traces. The left panel in the figure shows historical time series together with their temporal means used to label them, so they can be updated with information about those means. The right panels show how these means can represent a probability density function, and how an update of this probability density function can be translated back to be represented in the ensemble. In the upper panel this is done by weighting the initially uniform weights, shown on the vertical axis. In the lower panel, the update is done by shifting the means of the time series to other streamflow values, after which these shifts should be represented by changing the values in the corresponding time series. Note that the vertical axis in the lower panel just represents the transition from the original to the new location of the points, and does not indicate weights; the weights remain unchanged and uniform.

see section 2. To prevent adding more information to the ensemble than is justified by the forecast, the relative entropy should be minimized, constrained by the information in the forecast. In this paper, this method to derive the weights is referred to as the minimum relative entropy update (MRE-update). The method is consistent with and an extension of POME applied to the ensemble weights.

Apart from weighting climatic ensembles with additional forecast information, the MRE-update method is generally applicable whenever information is added to an ensemble by adjusting probabilities; see e.g. Another possible application could be a bias correction or variance adjustment for ensembles generated by Monte Carlo simulations with models. In finance, the concept of minimum relative entropy has been used to include price information in a Weighted Monte Carlo simulation, which is mathematically equivalent to the proposed bias correction application of the MRE-update (Avellaneda et al., 2001).

1.3. Previous work on adding forecast information to climatic ensembles by weighting

Croley (1996) presents a method for updating ensemble member probabilities, assuming forecast information is given by a third party in the form of conditional tercile probabilities (probability triplets). These are the probabilities of below normal, normal or above normal conditions, which have equal probabilities of $\frac{1}{3}$ in the climatic distribution. Croley presents a nonparametric probability adjustment procedure based on minimization

of the sum of squared deviations of the probabilities from the uniform distribution. The result is a block adjustment of probabilities, in which all ensemble members within one tercile get assigned the same weight. This is in line with the literal interpretation of the probability triplets as considered by Wilks (2000, 2002). The method can also deal with multiple forecasts, including deterministic "most probable event" forecasts (Croley, 1997). A procedure that one by one eliminates constraints according to user priorities helps to reach a feasible solution for the probability weights.

Croley (2003) presents an alternative parametric approach, in which sample moments of the forecast distribution can be imposed as equality constraints on corresponding moments of the weighted ensemble. A problem with this method is that often many of the probabilities become zero, so only part of the original ensemble is used. The cause for this partly lies in the objective function that Croley proposes. Although it seems reasonable to minimize the adjustment to the probabilities, there is no clear rationale for using minimum squared deviations as objective function; see page 345 of Jaynes (2003). Among the two methods, the parametric one leads to more reasonable, smoother adjustments than the block adjustment, avoiding sudden jumps in probability between adjacent ensemble members (Stedinger and Kim, 2010).

Stedinger and Kim (2002, 2007, 2010) introduce the pdf-ratio method, also focusing on obtaining a weighted ensemble, but now assuming a forecast is given by the third party as a

target conditional distribution. They also argue that forecast information in the form of probability triplets should not be taken literally, but as a representation of a smooth underlying target distribution. They propose that probability triplets should be converted to a likely target distribution that can subsequently be used in the pdf-ratio method. The pdf-ratio method adjusts the probability of each ensemble member with the ratio between marginal (climatic) and conditional (forecast) probability density functions (pdf) at each sample point. The pdf-ratio method then normalizes the probabilities to make them sum to one. Although the method does not seem to be very sensitive to distribution type, one still has the problem of assuming a distribution from only two tercile probabilities or moments. Another problem is that for relatively large deviations from the climatic distribution, significant deviations of the resulting moments from the target moments occur, but this can be solved by an additional adjustment step.

In this paper we analyse the problem of updating ensemble probabilities with forecast information from an information–theoretical viewpoint. We present a new method to include forecast information in a historical ensemble, based on minimizing relative entropy. In a comparison between our method and the existing methods, we explicitly show the assumptions in the existing methods and the differences between various ways of presenting forecast information. Before introducing the MRE-update, a short review of relevant information–theoretical concepts and principles is given.

2. Information, assumptions and entropy

In information theory, information and uncertainty are quantities, measurable in "bits". Uncertainty, entropy and missing information are all equivalent properties of a probability distribution. The central measure for uncertainty, information entropy, as defined by Shannon (1948), specifies the amount of extra information necessary to obtain certainty. For a discrete probability distribution the Shannon-entropy is defined as:

$$H(P) = -\sum_{x \in X} p(x) \log p(x) \tag{1}$$

in which X is the support set of the random variable with discrete distribution P and $p(x)$ is the probability of a certain outcome x. When using a base 2 logarithm, the unit of the resulting entropy is bits. Shannon (1948) derived his entropy measure starting from very basic desired properties of a measure of uncertainty, which uniquely define (1) as measure for uncertainty (Jaynes, 1957). Shannon's entropy measure has been proven useful in many fields of science (see Cover and Thomas, 2006).

A similar measure, relative entropy, expresses divergence between two probability distributions or the information gain that is associated with going from one distribution to another. Relative entropy, or Kullback-Leibler divergence, is defined as:

$$D_{KL}(Q \| P) = \sum_{x \in X} q(x) \log \left(\frac{q(x)}{p(x)} \right) \tag{2}$$

in which P and Q are the original and updated distributions respectively and $q(x)$ is the new probability estimate of event x. Relative entropy has been applied in meteorology in the study of atmospheric predictability (Leung and North, 1990; Kleeman, 2002) and as measure for forecast quality (e.g. Roulston and Smith, 2002; Ahrens and Walser, 2008; Weijs et al., 2010b; Weijs and Van de Giesen, 2011).

A reduction of entropy implies that information is added about the uncertain event the distribution describes. Information can be added in the form of data or knowledge, but can also enter implicitly by unwarranted assumptions, all reducing the entropy of the distribution. When new information in, for example, a forecast motivates a revision of the probability distribution from $P(X)$ to $Q(X)$, the relative entropy $D_{KL}(Q \| P)$ is an exact measure of the amount of information added by that specific forecast. The expectation of $D_{KL}(Q \| P)$ over all possible forecasts is equal to the mutual information between the forecasts and the random variable X. A good overview of information-theoretic concepts reviewed in this section can be found in Cover and Thomas (2006).

2.1. The principle of maximum entropy

Among all discrete probability distributions, the uniform distribution, in which all outcomes in the support set are equally likely, maximizes entropy, i.e. has maximum missing information. So without any information available except for the support set, it is rational to assume a uniform distribution. Assuming any other distribution leads to less uncertainty without having the information to justify that reduction. This idea was already formulated by Laplace (1825) .

Jaynes (1957) first formulated the principle of maximum entropy, which is in fact a generalization of Laplace's principle. It states that when making inferences based on incomplete information, one should choose the probability distribution that maximizes uncertainty (entropy), subject to the constraints provided by the available information. Applying this principle leads to a distribution with maximum uncertainty, but bounded by what is known. This automatically implies that no false certainty is created and only truly existing information is added. The continuous variant of principle of maximum entropy has been widely applied for derivation of continuous prior distributions and parameter estimation; see e.g. Singh and Singh (1985); Singh and Rajagopal (1987); Singh (1997) for relevant references.

Along the same lines of reasoning, the principle of minimum relative entropy or principle of minimum discrimination information (Kullback, 1997, first edition 1959) states that given new facts, a new distribution should be chosen that is consistent with those facts, but apart from that minimizes the information gain with respect to the original distribution. This principle ensures that not more new information is included than is justified by the new facts. The principle leads to results identical to those of POME, but generalizes to non-uniform prior distributions. For the discrete form, this means minimum relative entropy takes into account the initial probabilities in adjusting a distribution to new facts.

3. Methods

3.1. The Minimum Relative Entropy update

We propose to apply the principle of minimum relative entropy to adjust the probabilities of a climatic ensemble to reflect new forecast information. This method is referred to as the minimum relative entropy update (MRE-update). The MRE-update is a constrained minimization of relative entropy to optimally combine new information in the form of constraints with an existing ensemble. The method can be used for updating a climatic ensemble, whose members may contain high resolution spatial and temporal patterns of several variables. New added information concerns some averaged quantities that characterize the traces in the climatic ensemble. This new information can, for example, be expressed in the form of conditional moments of those averaged quantities. The information is added by adjusting the weights of the ensemble members in such a way that the weighted moments match the forecast.

3.1.1. Rationale of the method

The amount of new information added by the forecast is the relative entropy between the original uniform distribution of probabilities and the updated probabilities assigned to the ensemble. Minimizing this relative entropy, constrained by the information contained in the forecast, will find weights that exactly use all information in the forecast, without adding information that is not in the forecast. Consequently, the new ensemble is consistent with the forecast, but does not deviate more than necessary from the observed climatic distribution. The MRE-update optimally combines new forecast information with climatic information already present in an ensemble.

3.1.2. Formulation of the method

In the MRE-update, we try to find updated probabilities q_i by minimizing relative entropy to the original uniform distribution of probabilities p_i from the updated probabilities q_i assigned to the n samples x_i, given the general constraints of probabilities and the constraints posed by the forecast information. This results in a nonlinear optimization problem with objective function:

$$\min_Q D_{KL}(Q\|P) = \min_{q_i\ldots q_n}\{\sum_{i=1}^{n} q_i \log(\frac{q_i}{p_i})\} \tag{3}$$

Because in this case we start from a uniform distribution of equiprobable ensemble members (p_i is constant), which has maximum entropy, minimizing relative entropy is equivalent to maximizing the entropy of the distribution of q_i:

$$\max_Q H(Q) = \max_{q_i\ldots q_n}\{-\sum_{i=1}^{n} q_i \log(q_i)\} \tag{4}$$

In the remainder of the paper, we focus on MRE (Eq. 3), because POME (Eq. 4), although in this case equivalent, is less flexible to take into account possible prior weights assigned to

the ensemble. Also, MRE makes the analogy with the Croley methods that we will compare the method to more evident.

The optimization is subject to the constraint that the probabilities sum to one

$$\sum_{i=1}^{n} q_i = 1 \tag{5}$$

and that all probabilities are non-negative:

$$q_i \geq 0 \tag{6}$$

This last constraint is never binding in the MRE-update, because the objective function (Eq. 3) already ensures positive weights. Without any additional forecast information, no extra constraints are added. Objective function (3) minimizes the divergence from the original uniform distribution. Because constraints (5) and (6) are already satisfied by the original distribution, no adjustment is made.

When forecast information is available, it can be introduced by additional constraints to the minimization problem. In case the forecast information is given as probability triplets of below (p_b) and above (p_a) normal conditions, the following constraints are added:

$$\sum_{i\in S_b} q_i = p_b \tag{7}$$

$$\sum_{i\in S_a} q_i = p_a \tag{8}$$

In which S_b and S_a are the sets of i for which $x_i \leq x_b$ and $x_i \geq x_a$ respectively. With x_b and x_a being the lower and upper terciles of the climatic distribution for X.

In case the forecast information is given as the conditional mean μ_1 and standard deviation σ_1, the following constraints are imposed:

$$\sum_{i=1}^{n} q_i x_i = \mu_1 \tag{9}$$

$$\sum_{i=1}^{n} q_i(x_i - \mu_1)^2 = \sigma_1^2 \tag{10}$$

The resulting constrained convex optimization problem is subsequently solved using a standard gradient search.

To show the flexibility of the MRE-update method to include other forms of information, an example of including a prescribed skew will be tested by adding the additional constraint

$$\sum_{i=1}^{n} q_i(x_i - \mu_1)^3 = \gamma_1 \sigma_1^3$$

where γ_1 is the target skew.

5

3.2. Theoretical test case on a smooth sample and comparison to existing methods

In this experiment, the results of the MRE-update are compared with the results of the Croley nonparametric adjustment (Croley, 1996), the Croley parametric adjustment (Croley, 2003) and the pdf-ratio method Stedinger and Kim (2010). The same example as the univariate case in Stedinger and Kim (2010) was used. In this example, an artificially generated smooth climatic sample of $n = 50$ scalar values x_i is updated with forecast information of the previously mentioned forms. In a real-world application, the sample would represent an ensemble of, for example, time series. The sample is created by evaluating the inverse cumulative distribution function of the prescribed original distribution at the Hazen plotting positions $((i - 0.5)/n)$; see Stedinger and Kim (2010). The sample is drawn from a normal distribution with mean 3 and standard deviation 1. In absence of extra information, all 50 samples are considered equiprobable, with probability $1/50$.

The challenge is now to update the probabilities of the sample values in such a way that the ensemble reflects the forecast distribution, given the climate information, conditioned on forecast information. It must be noted that when we compare the methods, we implicitly also compare forecast information of three different types:

(TC)   The conditional tercile probabilities p_b, p_n and p_a

(N)   An assumed forecast normal distribution with given parameters.

(M)   The mean μ_1 and standard deviation σ_1 of the forecast distribution

Table 1 gives an overview of the methods and forecast types, indicating which combinations are compared and which abbreviations are used for the results.

For the pdf-ratio method, Stedinger and Kim (2010) considered normal, lognormal and gamma type distributions. This paper focuses on their results using the assumption of a normal distribution for both climatic and forecast distribution, that has some parallels with the MRE-update using mean and variance as constraints. In Stedinger and Kim (2010), the forecast information of type TC is converted to type N, using the assumption of a normal distribution. The rationale behind this is that the forecast of type TC is likely to represent a smooth underlying distribution. They compare results of the pdf-ratio method using forecast N with the Croley nonparametric adjustment using forecast TC and to CP-M. Since the pdf-ratio method does not always match the target moments, we also use a variant that changes the parameters of the target distribution of pdf-N in an optimization, until the resulting moments after normalization exactly match the targets.

The Croley methods are based on an optimization using the same constraints as the MRE-update for both the forecast TC (results: CP-TC) and the forecast M (results: CP-M). The objective function that Croley (1996, 2003) proposed for both methods, however, is different from the MRE-objective (Eq. 3); instead of minimizing relative entropy, the Croley methods find the weights by minimizing the squared adjustment in probabilities (Eq. 11), which is in fact the unique difference between the CP and MRE methods.

$$\min_{q_i}\{\sum_{i=1}^{n}(q_i - p_i)^2\} \qquad (11)$$

We compare results of the MRE-update with both tercile probability constraints (MRE-TC) and constraints on mean and standard deviation (MRE-M) to the results of the Croley nonparametric adjustment, the Croley parametric adjustment, and the pdf-ratio method with and without the optimization to exactly match the target moments. The comparison is made for a hypothetical case, with the various combinations of μ_1 and σ_1 as described in Stedinger and Kim (2010). For the case of tercile probability constraints (TC), target tercile probabilities are derived from μ_1 and σ_1, assuming a normal distribution.

3.3. Information-uncertainty trade-off

To visualise the trade off between uncertainty and information from the forecast we can plot information-theoretical measures of the resulting weight sets in a two dimensional space. A Pareto front can be obtained by formulating a multi-objective optimization problem and solving it by using the fast global optimization algorithm AMALGAM, developed by Vrugt and Robinson (2007). The problem consists of minimizing two objectives by finding Pareto-optimal vectors Q of 50 weights for the ensembles. The first objective is the maximization of the entropy of the weights. When formulated as the minimization of the difference with the entropy of uniform weights (Eq. 12) it can be interpreted as a measure of original uncertainty lost.

$$\min_{Q}\{H_{\text{uniform}} - H(Q)\} \qquad (12)$$

The second objective is minimization of the Kullback-Leibler divergence of the sought distribution from the closest distribution that exactly matches the target moments. This objective measures the information loss with respect to the exact forecast.

$$\min_{Q}\{D_{KL}(P_{\text{target}}\|Q)\} \qquad (13)$$

where

$$P_{\text{target}} = \arg\min_{P}\{D_{KL}(P\|Q)\} \qquad (14)$$

and P_{target} is subject to the constraints on sum, mean and variance in equations 5, 9 and 10. The position of results from different methods relative to the Pareto front will give insight in how they balance reduction of uncertainty with information to justify it.

3.4. Multivariate MRE-update

An important extension is the update of ensemble probabilities to reflect forecast information on multiple variables (Stedinger and Kim, 2010). For more background on the importance of the multivariate case; see Stedinger and Kim (2007,

Table 1: An overview of the methods and types of forecasts that are compared in this paper.

Adjustment method	Forecast used		
	Tercile constraints (TC) p_b, p_n, p_a	Conditional distribution (N) $N(\mu_1, \sigma_1)$	Conditional mean and variance (M) μ_1, σ_1^2
pdf-ratio method		(pdf-N)	
Croley non-parametric adjustment	(CP-TC)		
Croley parametric adjustment			(CP-M)
Minimum relative entropy update	(MRE-TC)		(MRE-M)

2010). For all variables, constraints on mean and variance can be specified separately. Because the size of the ensemble stays the same, the dimensionality of the optimization problem does not increase. The only difference is the addition of more constraints, which results in a slightly higher risk of the optimization problem becoming infeasible. However, tests show that for most practical problems, enough degrees of freedom exist to find a solution. An important issue for the multivariate case is the preservation of cross-correlations (Stedinger and Kim, 2010), especially in cases where risk depends on the joint occurrence of for example high water temperatures and low flows. Preservation of the cross-correlations can be ensured by imposing additional equality constraints on the weighted cross-correlation of the adjusted sample.

Although in this paper we concentrate on the univariate case, we will briefly show some results to demonstrate the potential of the MRE-update also for multivariate updates. We consider the theoretical example from (Stedinger and Kim, 2010) for comparison, using the exact same data. We use the same bubble plots to show the resulting weights as a function of both variables.

4. Results

The results are presented as graphs showing the weights of the individual traces as a function of the value they represent and the cumulative weights, which form an empirical cumulative distribution function (CDF). Next to the graphs, a number of tables shows the resulting tercile probabilities, moments and relative entropies

The codes in Tables 2-4 and the legends of figures 2-5 correspond to the different methods given in Table 1. For the case of tercile probability constraints (forecast TC), the MRE-update (with objective eq. 3) always results in exactly the same block adjustment as for the Croley nonparametric adjustment (objective eq. 11). For tercile constraints, the minima of the objective functions thus coincide. The identical results for these two methods are indicated with TC. Pdf-N indicates pdf-ratio method, using normal climatic and normal forecast distributions (forecast N).

4.1. Resulting weights

Figures 2-5 show the assigned probabilities for individual ensemble members against their x-values. The right graphs in these figures show the corresponding discrete approximations

for the cumulative distribution functions (CDF), using Hazen plotting positions, following Stedinger and Kim (2010). The original smooth sample with uniform weights, drawn from the climatic normal distribution is also shown (climatic). In the cumulative distribution plots, also the cumulative distribution function of the target distribution, used in the pdf-ratio method is plotted.

From these results, it becomes clear that there is a large difference between forecasts in the form of probability triplets and forecasts in the form of moments. When the deviations from the original moments are small, the results for the methods using moments and the pdf-ratio method are similar. When deviations become larger, the Croley parametric method shows clearly different behavior, while the MRE-update and the pdf-ratio method show very similar results in many cases. .

4.2. Resulting moments and tercile probabilities

Table 2 shows the target and resulting tercile probabilities for below and above normal conditions (p_b and p_a) for the methods, while Table 3 shows resulting means and standard deviations. Small rounding errors occur due to the limited number of ensemble members and the way the original sample is drawn. The effects become apparent for the case $\mu_1 = \mu_0 = 3$ and $\sigma_1 = \sigma_0 = 1$ (no new information). Firstly, the discrete approximation of outer tercile probabilities with 17 of the 50 members, results in probabilities of 0.34 rather then $\frac{1}{3}$. Secondly, the standard deviation of the original sample is not exactly one, but 0.987.

When ignoring small differences due to these numerical effects, Table 2 shows that for all methods p_b and p_a match the assumed target reasonably well for cases with $\sigma_1 \leq 1$. For the cases with increased variance, the methods using moment constraints show somewhat larger deviation from target probabilities p_a and p_b.

Results for the mean-variance forecast (type M) show a difference between MRE-update (MRE-M) and Croley parametric adjustment (CP-M). The latter tends to result in more ensemble member probabilities set to zero; see Fig. 2-5. Although results are different, both satisfy the constraints given by the forecast. The difference in results is purely due to the difference in objective function. Naturally, because the moments are imposed as constraints, both methods will exactly match the target mean and standard deviation (Table 3), like the results for methods using tercile constraints (TC) exactly match p_b and p_a (Table 2).

Table 2: Resulting tercile probabilities for the four methods compared

True moments		Target probabilities		Estimated probabilities							
				TC		pdf-N		MRE-M		CP-M	
μ_1	σ_1	p_b	p_a	p_b	p_a	p_b	p_a	p_b	p_a	p_b	p_a
3.00	0.25	0.043	0.043	0.043	0.043	0.049	0.049	0.050	0.050	0.047	0.047
2.00	0.50	0.873	0.002	0.873	0.002	0.880	0.002	0.881	0.002	0.886	0.000
3.00	0.50	0.195	0.195	0.195	0.195	0.205	0.205	0.205	0.205	0.227	0.227
4.00	0.50	0.002	0.873	0.002	0.873	0.002	0.880	0.002	0.881	0.000	0.885
4.50	0.50	0.000	0.984	0.000	0.984	0.000	0.986	n.a.	n.a.	n.a.	n.a.
5.00	0.50	0.000	0.999	0.000	0.999	0.000	0.999	n.a.	n.a.	n.a.	n.a.
3.00	1.00	0.333	0.333	0.333	0.333	0.340	0.340	0.342	0.342	0.342	0.342
3.00	1.20	0.360	0.360	0.360	0.360	0.364	0.364	0.375	0.374	0.382	0.382
4.00	1.20	0.117	0.682	0.117	0.682	0.127	0.669	0.142	0.712	0.110	0.840
4.50	1.20	0.054	0.814	0.054	0.814	0.064	0.791	0.097	0.843	0.081	0.919
5.00	1.20	0.021	0.905	0.021	0.905	0.029	0.877	0.071	0.931	0.071	0.931

Table 3: Resulting mean and standard deviation for the various methods

Target assuming normal				Mean				Standard deviation			
p_b	p_a	μ_1	σ_1	TC	pdf-N	CP-M	MRE-M	TC	pdf-N	CP-M	MRE-M
0.043	0.043	3.00	0.25	3.00	3.00	3.00	3.00	0.41	0.25	0.25	0.25
0.873	0.002	2.00	0.50	2.07	2.00	2.00	2.00	0.61	0.50	0.50	0.50
0.195	0.195	3.00	0.50	3.00	3.00	3.00	3.00	0.76	0.50	0.50	0.50
0.002	0.873	4.00	0.50	3.93	4.00	4.00	4.00	0.61	0.50	0.50	0.50
0.000	0.984	4.50	0.50	4.05	4.52	n.a.	n.a.	0.52	0.50	n.a.	n.a.
0.000	0.999	5.00	0.50	4.07	4.96	n.a.	n.a.	0.51	0.41	n.a.	n.a.
0.333	0.333	3.00	1.00	3.00	3.00	3.00	3.00	0.98	0.99	1.00	1.00
0.360	0.360	3.00	1.20	3.00	3.00	3.00	3.00	1.01	1.13	1.20	1.20
0.117	0.682	4.00	1.20	3.61	3.84	4.00	4.00	0.88	1.04	1.20	1.20
0.054	0.814	4.50	1.20	3.81	4.19	4.50	4.50	0.75	0.95	1.20	1.20
0.021	0.905	5.00	1.20	3.95	4.47	5.00	5.00	0.64	0.84	1.20	1.20

Table 4: Resulting relative entropy for the different methods

Target assuming normal distribution				Resulting divergence (relative entropy) from original distribution (bits)			
μ_1	σ_1	p_b	p_a	TC	pdf-N	CP-M	MRE-M
3.00	0.25	0.043	0.043	1.132	1.324	1.379	1.324
2.00	0.50	0.873	0.002	1.002	1.174	1.251	1.181
3.00	0.50	0.195	0.195	0.257	0.459	0.500	0.459
4.00	0.50	0.002	0.873	1.002	1.174	1.251	1.178
4.50	0.50	0.000	0.984	1.438	2.081	n.a.	n.a.
5.00	0.50	0.000	0.999	1.547	3.356	n.a.	n.a.
3.00	1.00	0.333	0.333	0.001	0.000	0.000	0.000
3.00	1.20	0.360	0.360	0.005	0.036	0.081	0.078
4.00	1.20	0.117	0.682	0.371	0.561	1.333	0.949
4.50	1.20	0.054	0.814	0.712	1.105	2.870	2.283
5.00	1.20	0.021	0.905	1.035	1.709	5.274	5.274

4.3. Information contained in forecasts

While the relative entropy is on the one hand used to find the weights for the MRE-update, by minimizing it in an optimization, it can also be used as a measure to analyse the results from all methods. Table 4 shows the resulting relative entropy for the set of ensemble member probabilities, relative to the original uniform distribution. The "n.a." entries correspond to combinations of μ_1 and σ_1 constraints for which the optimization based methods (CP-M) and (MRE-M) were not able to find a solution. This means that those μ_1 and σ_1 combinations are not achievable with the given sample. If a solution is needed, it can only be found when relaxing the constraints. In all cases, the result for TC has the lowest relative entropy, followed by pdf-N, MRE-M and CP-M. The relative entropy resulting from the MRE-update is the uncertainty reduction by the information in the forecast. Hence we can see that the forecast of type TC is less informative than type M (compare TC and MRE-M). The entropies for (MRE-M) in Table 4 also show that larger shifts in mean result in larger relative entropy. This corresponds to the intuition that forecasts add more information when they deviate more from climatology.

Because the pdf-N and CP-M methods have no information-theoretical founding, the relative entropy resulting from those methods does not say much about the amount of information in the forecast, but does indicate the uncertainty reduction in the ensemble. Because CP-M has higher relative entropy than MRE-M, we can say that this first method introduces information that is not present in the forecast. This will be further explained in the discussion in section 5.1. For cases with reduced variance and not too large a shift in mean, results of pdf-N very closely resemble MRE-M. Apparently, the information contained in the mean and variance constraints is the same information contained in a normal distribution. This is related to the fact that the maximal entropy distribution for a given mean and variance is a normal distribution; see Appendix A. In general, there is a duality between sufficient statistics and the constraints of a maximum entropy distribution; see Jaynes (2003), page 520.

If forecast information is given as constraints on conditional tercile probabilities (TC), there are infinitely many adjustments possible to satisfy those constraints. However, the adjustment that minimizes relative entropy is a block adjustment. Generally, when information about a distribution is given in form of constraints on quantile probabilities, the maximum entropy distribution is piecewise uniform. This also holds for the distribution of probabilities over the discrete scenarios. For the case of tercile constraints, the objective of minimum squared deviations of the Croley method has its minimum in the same location, leading to identical results as MRE.

It might seem strange that results for the MRE-update depend on how the forecast information is presented. Forecast TC and M give completely different results. However, considering the tercile probabilities (TC) and moments (M) as different ways to present the same information implies that some underlying information is present, i.e. that we know more about the information than what is presented. In this example forecast

M contains the extra knowledge that the forecast distribution is normal. Taking the information-theoretical viewpoint, forecast TC and M contain different information. If we do not know anything but these forecasts, we have to take them literally and thus we obtain different results for both forecasts. Forecasts of type TC appear to be less informative than forecasts of type M; see Table 4. Moreover, forecast TC does not seem an appropriate way to represent a smooth forecast distribution. Deriving a mean and standard deviation (forecast M) as given variables from tercile probabilities (forecast TC), as was done in Stedinger and Kim (2010), implicitly introduces the assumption that the forecast distribution is normal and hence results in a smooth update. In an ideal situation, the forecast information is already in a format that does not need additional assumptions to be interpreted (Weijs et al., 2010a).

4.4. MRE compared to Croley

The MRE-M and CP-M methods both use a forecast of type M. Although the same information is used by both methods, they lead to different results. Logically, because it is the objective, the MRE-update results in a smaller divergence or relative entropy than CP-M (Table 4). This means that the MRE-update retains more of the climatic uncertainty. However, the Croley parametric method uses the same constraints as the MRE-update, so the amount of forecast information used is the same. Table 3 shows that both results are equally consistent with the forecast, because mean and standard deviation are exactly reproduced in both cases.

Consequently, from an information-theoretical point of view, we can say that the Croley method makes an unnecessary extra deviation from climatology, not motivated by the forecast. The higher relative entropy means uncertainty is reduced by artificially introduced information. The minimum squared deviation objective therefore results in an over-confident adjustment. This is also demonstrated by the fact that several probabilities are set to zero, without having information that explicitly rules out those scenarios as representing possible future conditions.

4.5. MRE compared to pdf-ratio, using equivalent forecasts M and N

Because the maximum entropy distribution for given mean and variance is a normal distribution, forecast M implies a normal forecast distribution. When a forecast used in the pdf-ratio method is a normal distribution (forecast type N) with mean and standard deviation (of forecast type M) as parameters, forecast N and M are equivalent and add the same information. Therefore, the differences in the resulting weights for (pdf-N) and (MRE-M) are purely due to the methods.

MRE-M gives results similar to the pdf-ratio method (pdf-N) for adjustments that are not too large and where the ensemble is sufficiently dense; see fig. 1. In cases where the adjustment is large and the sample values do not cover the entire forecast distribution range, the approximation of the forecast distribution needs probability mass in the range outside the sample values. In the pdf-ratio method, this results in some "missing" probability.
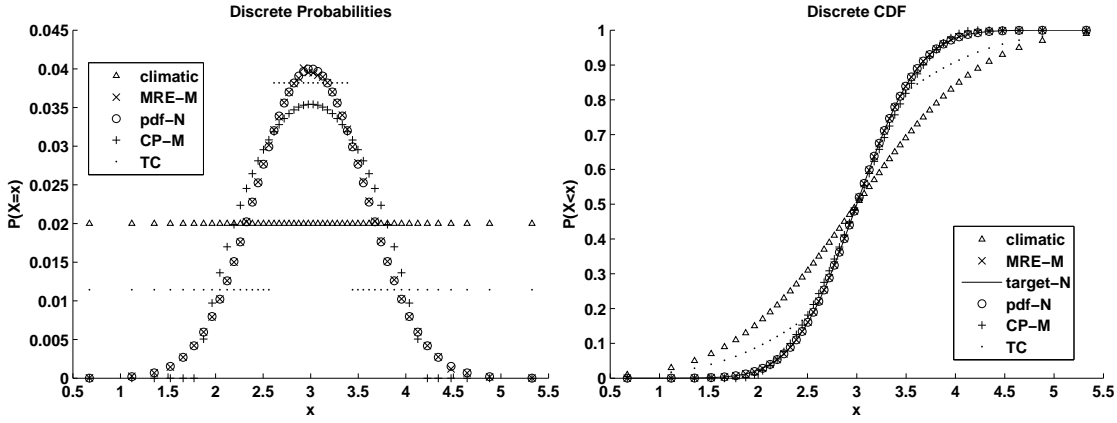
Figure 2: Resulting ensemble member probabilities for μ_1=3, σ_1=0.5 and resulting empirical cumulative distribution. The decreased variance is accounted for by giving less weight to the outer ensemble members.
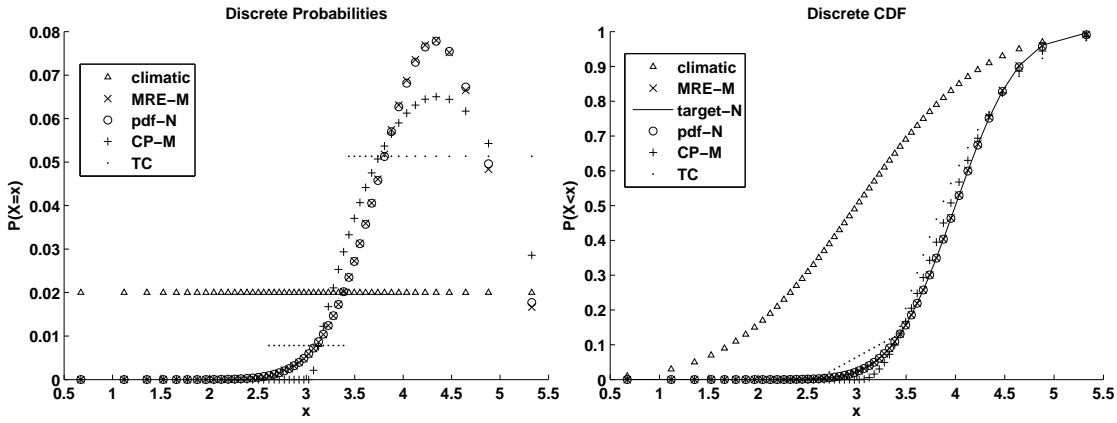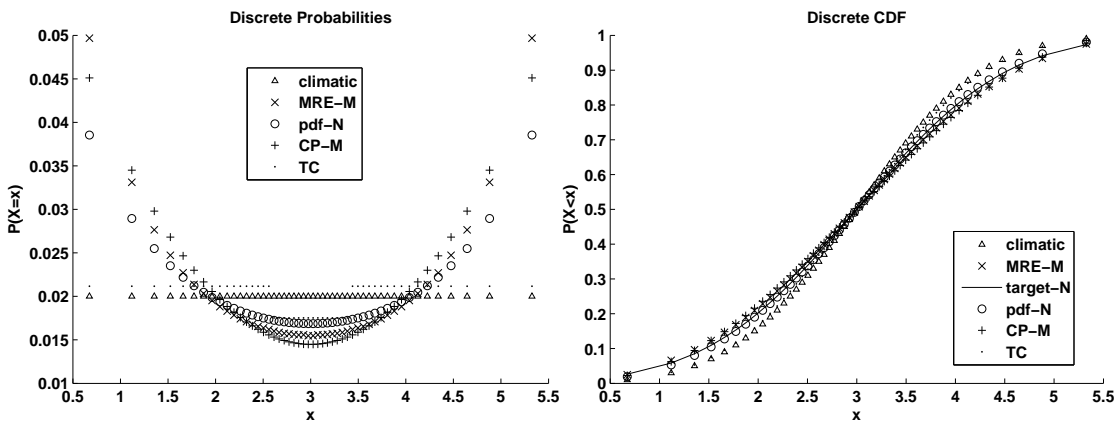

Figure 3: Resulting ensemble member probabilities for μ_1=4, σ_1=0.5 and resulting empirical cumulative distribution. The Croley parametric method (CP-M) gives zero weight to the ensemble members below a value of 3, while the MRE-update and pdf-ratio methods give small but non-zero weights.

Figure 4: Resulting ensemble member probabilities for μ_1=3, σ_1=1.2 and resulting empirical cumulative distribution. The inflated variance is accounted for by giving more weight to the outer ensemble members.
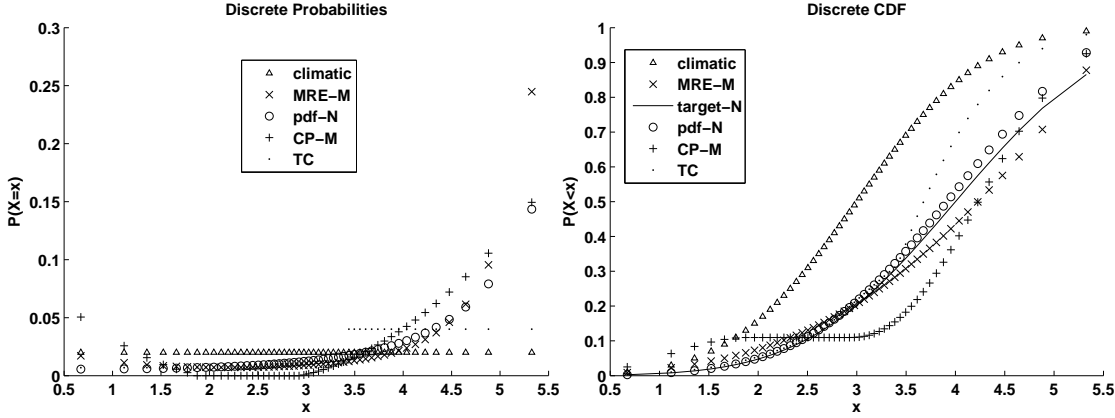
10

Figure 5: Resulting ensemble member probabilities for μ_1=4, σ_1=1.2 and resulting empirical cumulative distribution.
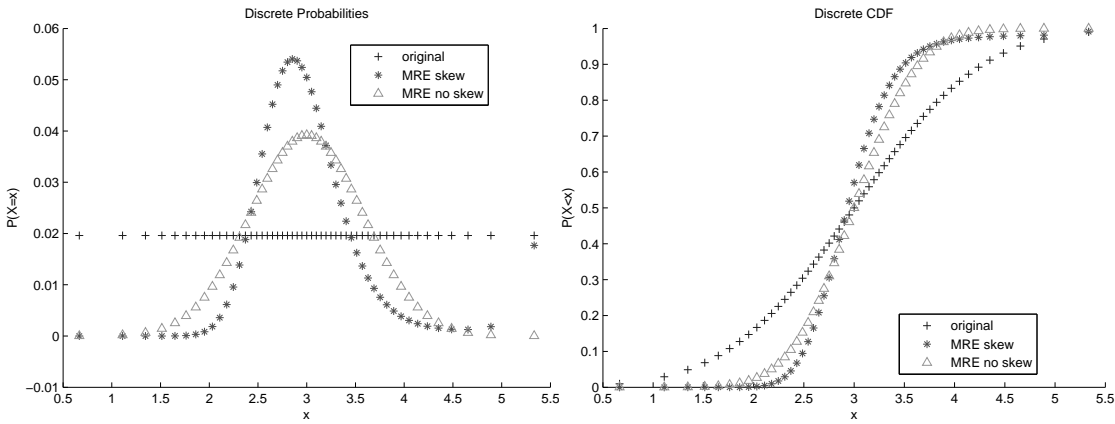

Figure 6: Resulting weights (left) and CDF (right) when an extra constraint on skewness is imposed in the MRE-update. The result is shown for $\mu_1 = 3$, $\sigma_1 = 0.5$ and a target skewness of 2.

This can clearly be seen in the Fig. 5 (right), where the value of the target CDF at the highest sample value is still far from one. The pdf-ratio method (pdf-N) needs a large normalization factor in these cases. All ensemble member probabilities are multiplied by the same factor, to make them sum to one. This results in deviations from the target mean and variance, because the missing probability outside the sample range is divided equally over the traces. Although this leads to smooth adjustments, it also results in weighted ensembles that do not conserve the mean and variance of the new information, possibly biasing results of planning and risk analysis.

The (MRE-M) method distributes the missing probability to the sample values in a way to match exactly the target mean and variance, as long as the constraints do not become over-restrictive. Especially when a high adjustment in the mean and a small variance are required, the problem might become infeasible (the "n.a." entries in the tables). An infeasible MRE-update indicates that the new information is conflicting with the historical ensemble and use of a weighted ensemble may be questionable.

When the MRE-update is forced to match the resulting moments of the weighted ensemble resulting from pdf-N, the results of MRE-M and pdf-N are identical.

Conversely, when pdf-N is forced to exactly match the target moments from the forecast, it will yield a result identical to MRE-M with the original target moments. In Appendix A, it is shown analytically why the methods yield the same results. It is also shown that the Croley parametric method (CP-M) results in a second order approximation of the MRE-result.

4.6. Results for skew and multivariate constraints

The following results are meant to show the flexibility of the MRE-update to incorporate more information by additional constraints. Fig. 6 shows for example how an extra constraint on skewness results in a different update that represents skew. For the multivariate MRE-update, the top plot in Fig. 7 shows the results for the MRE-update exactly matching mean and variance of both variables, but without explicitly preserving the initial cross-correlation by including it as a constraint. The middle plot shows the MRE-update result when a constraint enforced the preservation of the initial cross-correlation of 0.8. The bottom plot in Fig. 7 shows the resulting weights when the MRE-update is asked to exactly match the means, variances and cross correlation resulting from the pdf-ratio method with a bivariate normal distribution ($\sigma_{1x} = 1.145$ $\sigma_{1y} = 1.292$ $\rho_1 = 0.751$). Also for the multivariate case, it turned out that the MRE-update using means, variances, and cross-correlation

11

is equivalent to the pdf-ratio method with a bivariate normal distribution, when its moments and cross-correlation would be forced to exactly match the targets.

5. Discussion

5.1. An information-theoretical view

The previous results showed that the pdf-ratio method does not exactly match the target moments in the case of large shifts. Stedinger and Kim (2010) discussed whether it is desirable to exactly match target moments, arguing that if the moments in the forecast can not be trusted completely, it might be better to not exactly match them. The question can then be asked what justifies that deviation and in what way the resulting moments should deviate from the forecast.

In the information-theoretical framework, the information is the reduction in uncertainty. A requirement of the distribution of updated weights is therefore that the uncertainty is maximum, given a quantity of information that is added. If less information is taken from the forecast because it is not completely trusted, the maximum permissible reduction in uncertainty will also be less, and vice versa. This can be visualized as a tradeoff between forecast information lost and uncertainty lost. Fig. 8 shows the tradeoff as a Pareto front. Points below the Pareto front are not attainable, because the change in weights to include a given portion of the forecast information inevitably leads to a given minimum loss of uncertainty. Solutions above the Pareto front, however, lose more uncertainty than justified by the information. In other words, these weighted ensembles incorporate a gain in information that did not come from the forecast. The solutions on the Pareto front can also be useful when the given ensemble cannot represent the new information in the forecast; e.g. the n.a. entries in Tables 2,3 and 4. In that case some forecast information must be lost in order to find a solution, which should be sought on the Pareto front.

From Fig. 8, it can be seen that the climatic distribution, and the MRE-update both lie on the Pareto front. In contrast, the Croley parametric method is not Pareto optimal according to these criteria. Although it exactly matches the forecast and loses no information from the forecast, it does so with more reduction in uncertainty than strictly needed. Also the pdf-ratio method does not reach the Pareto front, although is comes close in many cases. It is conjectured that when the forecast is seen as two separate pieces of information, one about the mean and one about the variance, the pdf-ratio solution would lie on the Pareto front in a 3 dimensional space, where lost information with respect to mean and variance would be plotted on separate axes.

While all solutions on the Pareto front indicate rational sets of weights given varying degrees of trust in the forecast, the MRE-update lies on the unique point that completely trusts the forecast. The onus is therefore on the forecaster to reflect both the information and the uncertainty about the variable under consideration in that forecast. The MRE-update reflects exactly this information and uncertainty in the weighted ensemble, and
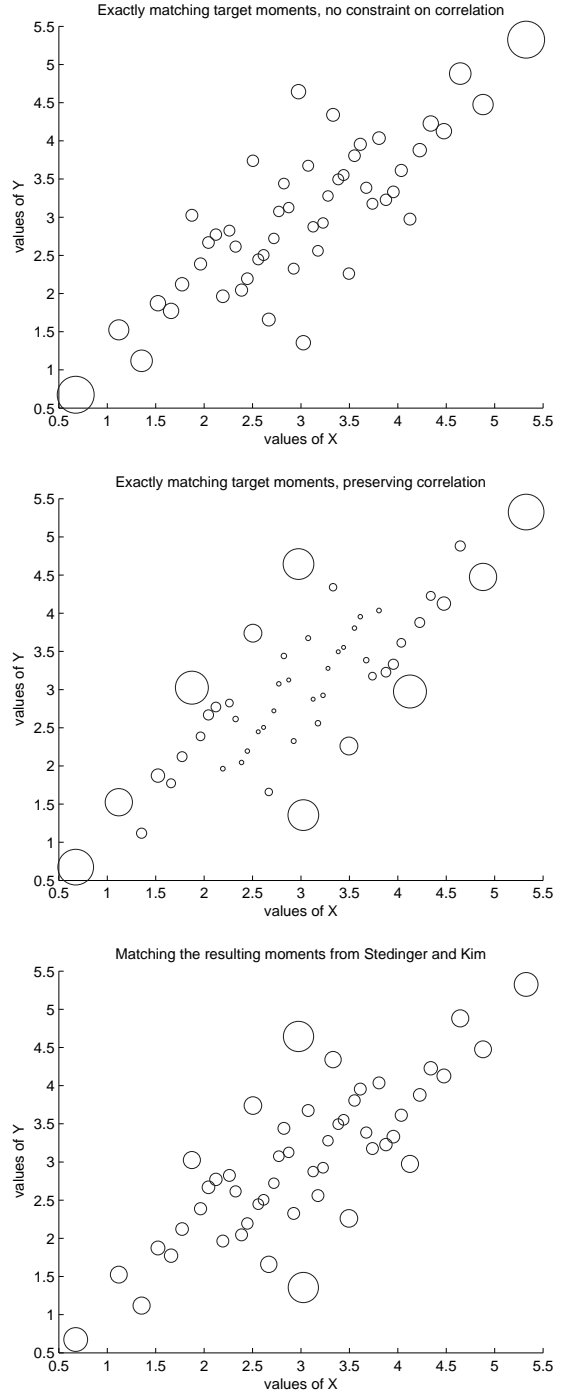

Figure 7: Results for the bivariate update of a sample, used in Stedinger and Kim (2010), with initial means $\{\mu_{0x}, \mu_{0y}\} = \{3, 3\}$ and initial standard deviations $\{\sigma_{0x}, \sigma_{0y}\} = \{1, 1\}$, the new means and standard deviations are $\{\mu_{1x}, \mu_{1y}\} = \{3, 3\}$ and $\{\sigma_{1x}, \sigma_{1y}\} = \{1.5, 1.5\}$. The area of the circles represent the weights. The above two graphs result from the MRE-update and match these targets exactly. The middle plot maintains the original cross-correlation $\rho_0 = 0.8$, while the upper plot results in a cross-correlation of 0.934. The bottom plot shows the result of the MRE-update for the target moments that are the resulting moments of the pdf-ratio method with bivariate normal target. The resulting weights are identical to the pdf-ratio solution.
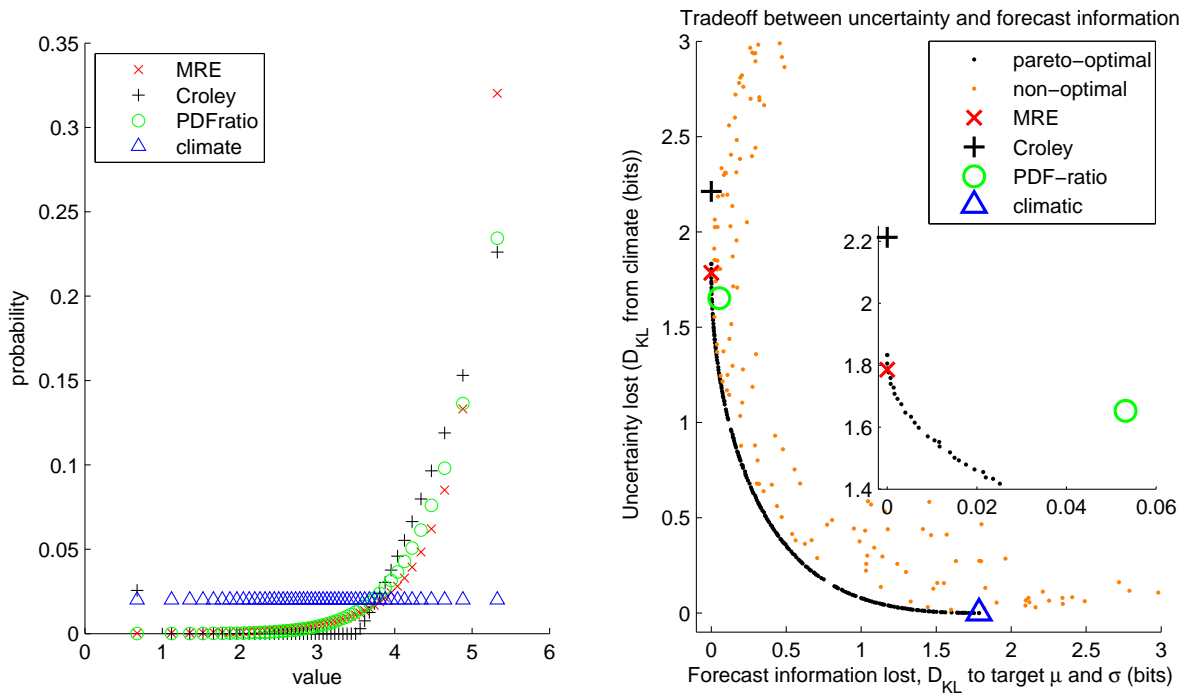
Figure 8: Pareto front showing the trade-off between losing as little uncertainty as possible and losing as little information in the forecast as possible, using Kullback-Leibler divergences from Eqs. 12 and 13 to measure information loss. The different points of the front represent different levels of trust in the forecast. The result is shown here for $\mu_1 = 4.5$ and $\sigma_1 = 0.8$. On the left, the weight sets belonging to 3 solutions and climatology are shown. As can be seen in the information-space on the right, the MRE- and climatic weights are on the extremes on the Pareto front, while the Croley method loses more uncertainty than justified by the forecast and the pdf-ratio approaches the Pareto front at a point where not all forecast information is used.

does not further increase or decrease the uncertainty. A discussion on why forecasters should communicate carefully chosen summary statistics or preferably their entire probability estimates can be found in Weijs et al. (2010a).

While we realize that this analysis can be perceived as biased because the different weighting methods are evaluated using the same information-theoretical principles as are used as the basis for the MRE-update, we think it is useful as a further illustration of the information-theoretical perspective, which is the theoretical basis for the MRE-update.

5.2. About the choice of the weighting methods in practice

In general, weighting is most useful when an ensemble is available that cannot easily be regenerated, and a mild adjustment is needed to reflect new information. For stronger adjustments, a weighted ensemble cannot represent the new state of knowledge anymore and resampling or stochastic time series generation would be better approaches than weighting. For weighting, the following considerations are important when choosing the method.

When forecast TC is received, additional information should be gathered about the moments, support set, or distribution types to assume. If really no other information is available, the MRE-update can be used directly with the forecast, resulting in block adjustment. When information about moments or other appropriately summarized statistics of the forecast distribution is available, the MRE-update is the most suitable method, as it ex-

actly uses the available information and does not make implicit assumptions.

The use of the optimization based adjustments proposed by Croley (1996; 1997; 2003) are a second order approximation of the MRE-update, but can introduce a reduction of uncertainty that is not supported by the forecast information. The MRE objective function should be preferred over the quadratic objective on information-theoretical grounds. For implementing the MRE-update in places where the Croley method is applied, it suffices to replace the quadratic objective function by relative entropy. This also resolves the problem of many probabilities set to zero.

Because the pdf-ratio method does not need to solve an optimization problem, it is easier to apply and faster then the MRE-update. Another advantage of the pdf-ratio method is that it is relatively easy to include a large amount of information, included in estimated climatic and forecast distributions. In many practical cases, the forecast distribution lies well within the climatic distribution, and large normalization is not required in the pdf-ratio method. In those cases, the pdf-ratio method provides a fast and correct adjustment, given that no unfounded assumptions are introduced in the estimation of climatic and forecast distributions. When an extra optimization is done to exactly match the target moments, it can be used as a fast solver for the MRE-update.

The MRE-update uses the full information from the forecast, provided the information contained in the forecast distri-

13

bution can be converted into mathematical constraints for the optimization problem. In principle the MRE-update offers possibilities to include constraints on for example skew, variance of the log-transformed variable, other quantiles or correlations in a multivariate setting. Many known parametric distributions are in fact maximum entropy distributions for combinations of these types of constraints; see e.g. Singh and Singh (1985); Singh and Guo (1995). This offers the possibility to reformulate pdf-ratio problems as MRE-update problems.

Conversely, it allows fast parametric solution of the MRE-update by using the pdf-ratio method which is forced to exactly match the constraints. This offers an opportunity to significantly reduce the dimensionality of the optimization problem for the MRE-update in case of a mean-variance forecast. Instead of seeking values for all individual weights, it suffices to optimize the 2 parameters of the target normal distribution and the normalization factor. Appendix A shows that this amounts to finding the 3 Lagrange multipliers in the analytical solution to the MRE-update.

When we have more information available about the forecast distribution than only mean and variance, like the complete time series of the predictors and responses, it is possible to estimate a joint pdf for them. Bivariate kernel density estimators, as applied by Sharma (2000), would then be a good way to derive continuous climatic and target distributions for the pdf-ratio method. Once one has the joint pdf, the marginal climatic and conditional forecast pdfs can be derived from it and used in the pdf-ratio method. If the conditional distribution from the kernel density estimate can be summarized in a number of constraints, it can also be used in the MRE-update.

6. Conclusions and recommendations

In this paper, we introduced the minimum relative entropy update (MRE-update) as an approach to update ensemble member probabilities. Our method is based on the minimization of relative entropy, with forecast information imposed as constraints. The main advantage of the method is that it optimally combines available climatic and forecast information, without introducing extra assumptions. Results were compared with three existing methods to make probability-adjustments to ensembles, based on different type of forecast information. We considered forecast information given as conditional tercile probabilities, a normal distribution, and a given mean and variance. Analysis of the results from an information-theoretical viewpoint explicitly revealed the differences in information contained in these different types of forecasts and the way existing methods interpret them.

The block adjustment that results from the Croley nonparametric method may be undesirable due to the discontinuities in weights at the arbitrarily selected quantiles. However, the result is in line with the literal information-theoretical interpretation of the probability triplets. When interpreting the forecasts in any other way, information is added. It is important to be aware of this and think carefully about what information is added. Conversely, forecasts that require this extra interpretation are in fact incomplete, leaving too much interpretation to the user of the forecasts. Ideally, seasonal forecasts should provide pieces of information that are a good summary of the probability estimate. For smooth distributions, a mean and variance are more appropriate than probability triplets. See also Weijs et al. (2010a) for more discussion on how forecasts should be presented to be most informative.

The information contained in the mean-variance forecast and in a normal forecast distribution is the same. The MRE-update results in a weighted ensemble that exactly incorporates this information. The pdf-ratio method diverges from the forecast information by not exactly matching the given moments. From an information-theoretical perspective, it is unclear how to justify this divergence. A multi-objective optimization was performed to find a Pareto front that represents the tradeoff between lost information from the forecast and lost initial uncertainty. An analysis of the methods in this objective-space revealed that in some cases, the pdf-ratio method reduces uncertainty more than is justified by the partial information taken from the forecast. This results in a solution that is not Pareto-optimal. Also the Croley parametric method lies above the Pareto front. It uses the full information from the forecast, but reduces uncertainty more than information permits. One of the symptoms of this false certainty are the ensemble members which get weight zero, although nothing in the forecast rules them out. By definition of the chosen objectives, the MRE-update results in a Pareto-optimal solution in which the complete information from the forecast is used and no other information is added to the weighted ensemble.

The pdf-ratio method has the advantage that it is fast and does not require an optimization search for all individual weights. An adaptation of the pdf-ratio method that includes a search for parameters that result in an exact match of the target moments is possible. This is equivalent to finding the values of the Lagrange multipliers for the analytical solution of the MRE-update (see Appendix A) and results in an optimization problem that is much easier to solve. In this paper, the equivalence for the univariate and bivariate normal distributions was demonstrated, but it is anticipated that similar results can be found for other distributions for which sufficient statistics exist.

The MRE-update can incorporate information that can be formulated in terms of constraints. This paper showed also how skew can be included. In addition, a multivariate example was given in which information in both means and variances was matched while also preserving initial cross-correlation. Generation weighted ensembles using information from non-parametric forecast pdfs, such for example obtained from kernel density estimation, remains an open issue.

6.1. Recommendations for future work

While this paper introduced information theory as a tool in the problem of generating weighted ensemble forecasts and discussed the theoretical benefits of this approach, an application to a real-world test case would give more insight in the practical value. This value could be established by investigating verification statistics for the weighted ensembles obtained from the different methods. These verification statistics would need to compare the weighted ensemble forecast, consisting of weights

and time series (or time series of spatial fields) with observations. Because the design of such an experiment is complex and involves a number of choices that need to be discussed, it is beyond the scope of the present paper.

For future experiments to test the practical difference between the approaches, the following issues should be considered. Firstly, when the weighted ensembles are compared with observations, the quality of the original forecast has an important impact on the verification score. It may be difficult to separate the contribution to the skill of the weighted ensemble of the forecast that goes in on the one hand, and the weighting method on the other. Errors in the weighting method might compensate errors in the forecast. It could for example happen that the forecasts are over-dispersive, and a weighting method that is theoretically adding too much information by being under-dispersive compensates this and would score best in a verification. Secondly, the weighted ensembles themselves are difficult to evaluate, and need to be translated back to probability distributions to evaluate them with verification scores. This translation involves some subjective choices on how to interpret the ensemble. Moreover, the forecast only contains information on the coarse scale, while the fine scale structure of the ensemble members is only intended to provide a realistic stochastic structure, but not to provide a correct realization that should be directly compared with measurements. Thirdly, Because ensemble weighting methods are mainly applicable to situations of relatively weak forecast skill, where the full range of historical time series is still important, the skill scores will have a relatively large variation between different instances. To really demonstrate the advantage in practice, large time series of forecasts and observations are therefore needed to conclusively show differences in performance between the weighting methods.

The Matlab source codes for the MRE-update, as well as implementations of the other methods compared in this paper, are available from the website `www.hydroinfotheory.net/mre`.

Acknowledgements

Appendix A. Equivalence between MRE-update and pdf-ratio solutions for the normal case

We try to solve

$$\min_{q_1 \ldots q_n} \left\{ \sum_{i=1}^{n} q_i \log\left(\frac{q_i}{p_i}\right) \right\}$$

subject to the constraints

$$\sum_{i=1}^{n} q_i = 1$$
$$q_i \geq 0$$
$$\sum_{i=1}^{n} q_i x_i = \mu_1$$
$$\sum_{i=1}^{n} q_i (x_i - \mu_1)^2 = \sigma_1^2$$

this leads to the Lagrangeans $\forall i$

$$\frac{\partial}{\partial q_i} \left\{ \sum_{i=1}^{n} q_i \log\left(\frac{q_i}{p_i}\right) + \lambda_1 \left(\sum_{i=1}^{n} q_i - 1 \right) + \lambda_2 \left(\sum_{i=1}^{n} q_i (x_i - \mu_1)^2 - \sigma_1^2 \right) + \lambda_3 \left(\sum_{i=1}^{n} q_i x_i - \mu_1 \right) \right\} = 0$$

$$1 + \log q_i - \log p_i + \lambda_1 + \lambda_2 (x_i - \mu_1)^2 + \lambda_3 x_i = 0$$
$$\log p_i - 1 - \lambda_1 - \lambda_2 (x_i - \mu_1)^2 - \lambda_3 x_i = \log q_i$$

where λ_1, λ_2 and λ_3 are the Lagrange multipliers corresponding to the constraints for the sum, the mean and the variance, respectively. The constraint for nonnegativity is never binding and is left out of the Lagrangean. This leads to a solution of the form

$$q_i = p_i e^{-1 - \lambda_1 - \lambda_2 (x_i - \mu_1)^2 - \lambda_3 x_i} \tag{A.1}$$

The Langrange multiplier can subsequently be solved numerically to match the constraints.

The result of the pdf-ratio method for normal initial and target distributions has the same form as equation A.1, but not necessarily the same values for λ_1, λ_2 and λ_3, because the solution is not forced to satisfy the constraints. When the parameters of the target normal distribution that is used as an input for the pdf-ratio method are modified so that the moments of the resultant weighted ensembles match μ_1 and σ_1 exactly, the result of the pdf-ratio method matches exactly the result of the MRE-update. The search for the two parameters of the normal distribution and the normalization constant are exactly the three degrees of freedom that are required to find λ_1, λ_2 and λ_3 numerically. An analytical solution is not possible because the constants depend on all individual x_i.

Similarly, for the Croley parametric method with the same constraints, and objective function

$$\min_{q_1 \ldots q_n} \left\{ \sum_{i=1}^{n} (q_i - p_i)^2 \right\}$$

it can be found that the solution is of the form

$$\frac{\partial}{\partial q_i} \left\{ \sum_{i=1}^{n} (q_i - p_i)^2 + \lambda_1 \left(\sum_{i=1}^{n} q_i - 1 \right) + \lambda_2 \left(\sum_{i=1}^{n} q_i (x_i - \mu_1)^2 - \sigma_1^2 \right) + \lambda_3 \left(\sum_{i=1}^{n} q_i x_i - \mu_1 \right) + \lambda_i q_i \right\} = 0$$

$$2q_i - 2p_i + \lambda_1 + \lambda_2 (x_i - \mu_1)^2 + \lambda_3 x_i + \lambda_i = 0$$
$$2p_i - \lambda_1 - \lambda_2 (x_i - \mu_1)^2 - \lambda_3 x_i + \lambda_i = 2q_i$$

where λ_i are the extra Lagrange multipliers which ensure that all q_i are nonnegative. The solution to the Croley method are weights that are a parabolic function of the values x_i. The quadratic objective function thus leads to a solution that is a second order (quadratic polynomial) approximation of the solution found by the MRE-update. '

15

Appendix B. Supplementary material

Supplementary data associated with this article can be found, in the online version, at `http://dx.doi.org/10.1016/j.jhydrol.2013.06.033`.

References

Ahrens, B., Walser, A., 2008. Information-based skill scores for probabilistic forecasts. Monthly Weather Review 136, 352–363.

Avellaneda, M., Bu, R., Friedman, C., Grandchamp, N., Kruk, L., Newman, J., 2001. Weighted monte carlo: A new technique for calibrating asset-pricing models. intern. J. of Theor. and Appl. Finance 4, 91–119.

Cloke, H.L., Pappenberger, F., 2009. Ensemble flood forecasting: A review. Journal of Hydrology 375, 613–626.

Cover, T.M., Thomas, J.A., 2006. Elements of information theory. Wiley-Interscience, New York.

Croley, T.E., 1996. Using noaa's new climate outlooks in operational hydrology. Journal of Hydrologic Engineering 1, 93–102.

Croley, T.E., 1997. Mixing probabilistic meteorology outlooks in operational hydrology. Journal of Hydrologic Engineering 2, 161–168.

Croley, T.E., 2003. Weighted-climate parametric hydrologic forecasting. Journal of Hydrologic Engineering 8, 171–180.

Day, G.N., 1985. Extended streamflow forecasting using NWSRFS. Journal of Water Resources Planning and Management 111, 157–170.

Faber, B.A., Stedinger, J.R., 2001. Reservoir optimization using sampling sdp with ensemble streamflow prediction (esp) forecasts. Journal of Hydrology 249, 113–133.

Georgakakos, K.P., Krzysztofowicz, R., 2001. Special issue: Probabilistic and ensemble forecasting. Journal of Hydrology 249.

Hamlet, A.F., Huppert, D., Lettenmaier, D.P., 2002. Economic value of long-lead streamflow forecasts for columbia river hydropower. Journal of Water Resources Planning and Management 128, 91.

Hamlet, A.F., Lettenmaier, D.P., 1999. Columbia River streamflow forecasting based on ENSO and PDO climate signals. Journal of Water Resources Planning and Management 125, 333–341.

Jaynes, E.T., 1957. Information theory and statistical mechanics. Physical Review 106, 620–630.

Jaynes, E.T., 2003. Probability theory: the logic of science. Cambridge University Press, Cambridge, UK.

Kelman, J., Stedinger, J.R., Cooper, L.A., Hsu, E., Yuan, S.Q., 1990. Sampling stochastic dynamic programming applied to reservoir operation. Water Resources Research 26, 447–454.

Kleeman, R., 2002. Measuring dynamical prediction utility using relative entropy. Journal of the Atmospheric Sciences 59, 2057–2072.

Kullback, S., 1997. Information theory and statistics. Dover Pubns.

Laplace, P., 1825. Essai philosophique sur les probabilités. Bachelier.

Leung, L.Y., North, G.R., 1990. Information theory and climate prediction. Journal of Climate 3, 5–14.

Pagano, T.C., Shrestha, D.L., Wang, Q., Robertson, D., Hapuarachchi, P., 2013. Ensemble dressing for hydrological applications. Hydrological Processes 27, 106–116.

Piechota, T.C., Chiew, F.H.S., Dracup, J.A., McMahon, T.A., 1998. Seasonal streamflow forecasting in Eastern Australia and the El Nino–southern oscillation. Water Resources Research 34, 3035–3044.

Reggiani, P., Renner, M., Weerts, A.H., van Gelder, P.A.H.J.M., 2009. Uncertainty assessment via bayesian revision of ensemble streamflow predictions in the operational river rhine forecasting system. Water Resources Research 45, W02428.

Roulston, M.S., Smith, L.A., 2002. Evaluating probabilistic forecasts using information theory. Monthly Weather Review 130, 1653–1660.

Shannon, C.E., 1948. A mathematical theory of communication. Bell System Technical J. 27, 379–423.

Sharma, A., 2000. Seasonal to interannual rainfall probabilistic forecasts for improved water supply management: Part 3–a nonparametric probabilistic forecast model. Journal of Hydrology 239, 249–258.

Singh, V.P., 1997. The use of entropy in hydrology and water resources. Hydrological Processes 11, 587–626.

Singh, V.P., Guo, H., 1995. Parameter estimation for 3-parameter generalized pareto distribution by the principle of maximum entropy (POME). Hydrological Sciences Journal 40, 165–181.

Singh, V.P., Rajagopal, A.K., 1987. Some recent advances in application of the principle of maximum entropy (pome) in hydrology. IAHS 194, 353–364.

Singh, V.P., Singh, K., 1985. Derivation of the pearson type (PT) III distribution by using the principle of maximum entropy (POME). Journal of hydrology 80, 197–214.

Stedinger, J., Kim, Y.O., 2002. Updating ensemble probabilities based on climate forecasts. Proc., Water Resources Planning and Management (19-22 May, Roanoke, Virginia)(CD-Rom), Environmental and Water Resources Institute, American Society of Civil Engineers, Reston, VA .

Stedinger, J.R., Kim, Y., 2007. Adjusting ensemble forecast probabilities to reflect several climate forecasts. IAHS PUBLICATION 313, 188.

Stedinger, J.R., Kim, Y.O., 2010. Probabilities for ensemble forecasts reflecting climate information. Journal of Hydrology 391, 9–23.

Van Overloop, P.J., Weijs, S., Dijkstra, S., 2008. Multiple model predictive control on a drainage canal system. Control Engineering Practice 16, 531–540.

Vrugt, J.A., Robinson, B.A., 2007. Improved evolutionary optimization from genetically adaptive multimethod search. Proceedings of the National Academy of Sciences 104, 708.

Weijs, S.V., 2011. Information Theory for Risk-based Water System Operation (PhD thesis, TU Delft). VSSD, Delft. chapter 4. pp. 57–86.

Weijs, S.V., Van de Giesen, N., 2011. Accounting for observational uncertainty in forecast verification: an information–theoretical view on forecasts, observations and truth. Monthly Weather Review 139, 2156–2162.

Weijs, S.V., Schoups, G., van de Giesen, N., 2010a. Why hydrological predictions should be evaluated using information theory. Hydrology and Earth System Sciences 14, 2545–2558.

Weijs, S.V., Van Nooijen, R., Van de Giesen, N., 2010b. Kullback–Leibler divergence as a forecast skill score with classic reliability–resolution–uncertainty decomposition. Monthly Weather Review 138, 3387–3399.

Werner, K., Brandon, D., Clark, M., Gangopadhyay, S., 2004. Climate index weighting schemes for nws esp-based seasonal volume forecasts. Journal of Hydrometeorology 5, 1076–1090.

Wilks, D.S., 2000. On interpretation of probabilistic climate forecasts. Journal of Climate 13, 1965–1971.

Wilks, D.S., 2002. Realizations of daily weather in forecast seasonal climate. Journal of Hydrometeorology 3, 195–207.

Wood, A., Lettenmaier, D., 2006. A test bed for new seasonal hydrologic forecasting approaches in the western united states. Bulletin of the American Meteorological Society 87, 1699–1712.

Wood, A.W., Lettenmaier, D.P., 2008. An ensemble approach for attribution of hydrologic prediction uncertainty. Geophysical Research Letters 35, 1–L14401.