QUALITY ANALYSIS OF PROBABILISTIC HYDROLOGICAL FORECASTS

A. C. Negrão[1&2] and E. A. Leite[1]
1. *Technological Institute SIMEPAR*
2. *University of São Paulo - USP*

ABSTRACT: Aiming at optimizing power generation and the increased reliability of flood warning systems, decision makers are searching for hydrological models which are more accurate and representative. Probabilistic models consider the uncertainty of forecasts, therefore they are more faithful to the randomness of hydrological phenomena and more transparent with respect to the uncertainties of the forecast. However, there are still difficulties in the direct quantification of the quality of these type of model. Katz and Murphy (1997) propose the analysis of forecasts' sharpness and discrimination. Both are based on the concepts of climatological probability. The first is the forecast variability and the second relates the predictions overall average given a certain observation. In order to assess the prediction reliability, Jolliffe and Stephenson (2003) use a frequency histogram of the observation occurrences in each ensemble of forecast probability distribution. This work presents a method of accounting for these three quality indices. Moreover, they are simultaneously assessed following the concept of multi criteria to quantify the quality of probabilistic hydrological forecasts. Two probabilistic models were developed and operated by the Technological Institute SIMEPAR, Paraná, Brazil. One is based on the Bayesian theory and SOM neural network; the other is based on historical errors of deterministic forecasts and Meta-Gaussian transformation. The indices allow the monitoring of forecasts' quality of both models and the comparison to confirm which one has the best results.

Key Words: decision making, uncertainty of forecasts, quality indices

1. INTRODUCTION

Probabilistic forecasts assign a probabilistic distribution of occurrences to a future event that allow an analysis more detailed about forecasts uncertainty. The wide randomness of natural phenomena suggests that their forecasts events should be expressed in terms of probabilities. Although most operational hydrological forecasting systems produce deterministic forecasts, hydrologists have been interested in recent years by probabilistic forecast because they enable decision making more rational and risk-based analysis. However a learning period was needed for both forecasters and users (Krzysztofowicz, 2001).

After made the forecast is need to verify how much the forecasts corresponds to observations. Katz and Murphy (1997) defined forecast verification as the process of assessing the quality of forecasts which can be defined as the set of statistical characteristics associated with the probability distribution density that relates the forecast with the observed value of the variable. One of the main difficult of the probabilistic forecast is the verification. It is not obvious how compare, in a simple way, a forecasts probability distribution with a single observed value. For validate a probabilistic forecast system is need an analysis of observed values also in probabilistic distribution terms. If both distributions are near, the forecast is valid, if are not the forecast is inconsistent (Toth *et al.*, 2003).

This work proposes a methodology to quantify quality of probabilistic forecasts of discharge and presents two probabilistic hydrological models that are in real time operation by the Technological Institute SIMEPAR.

2. PROBABILISTIC MODELS

The Technological Institute SIMEPAR keeps in operation the Hydrological Forecast System of Iguaçu (SISPSHI) whose main structure of the hydrological model is based on a rainfall-runoff-routing (3R) model. The model is deterministic with an hourly scale and it does water stage and discharge forecasts for a set of basins of the River Iguaçu, Paraná, Brazil (Breda, 2008).

The River Iguaçu Basin has about 62000 km^2 of area drainage and 1320 km of length. It is one of the main rivers for power generation of Brazil. The SISPSHI has more focus on portion of the Upper Iguaçu to União da Vitória, city severely affected by urban flooding.

Two probabilistic models were developed to seek adding stochastic information to SISPSHI deterministic results: the SISPSHI-Bayes (Feire, 2009 and Leite, 2011) is based on Bayesian theory and takes SOM neural networks to pattern recognition; and the SISPSHI-NQT (Negrão, 2013) is based on the meta-Gaussian approach of the historical observed errors. They do stage and discharge forecasts for União da Vitória every 6 hours for a horizon of 72 hours.

Both models were developed from a technique known as ensemble forecasting. The probabilistic forecasts are based on the historical deterministic forecasts. The statistics information of the historical occurrences is combined with the information about the deterministic forecast at the moment (Toth *et al*., 2003).

3. QUALITY INDICES

Two main quality aspects of probabilistic forecasts are the reliability and resolution. The reliability verify the consistence between *a priori* predicted and *a posteriori* observed of a event and the resolution verify the frequency of events occurrence for different forecast scenarios. One alone is not sufficient for a probabilistic forecast system to be useful (Toth et al, 2003). Then, we need consider more that one quality indices for verification of forecasts.

Katz and Murphy (1997) propose the sharpness (refinement) and discrimination as options to verification the resolution of probabilistic forecasts. Both are supported on the concept of climatological frequency of occurrence that in this work is the probability distribution of historical observed discharge from 1998 to the moment of forecast. Figure 1 shows the distribution of the complement of cumulative probability (*1-F(q))* of a given discharge *Q* be greater than a historical observed discharge *q*.

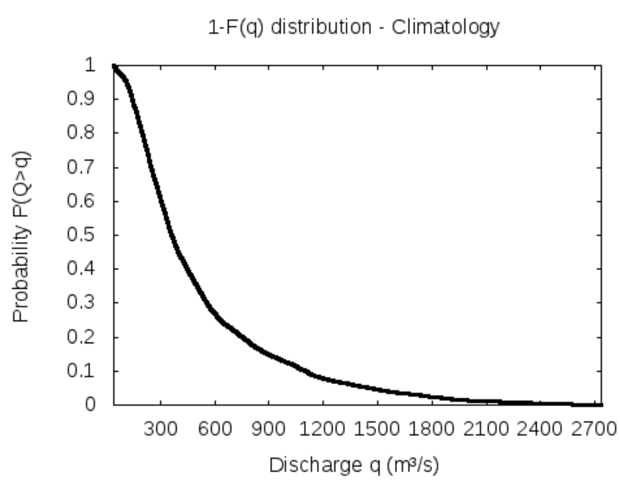

Figure 1: Climatology curve from 1998 to 2012.

The sharpness is an indicator of the variability of forecasts and it is not related an observed value. A forecast is considered perfectly sharp when it has only probabilities zero and one as in a deterministic forecast. On the other hand, a forecast without sharp is one that is constant of climatology probability (Katz and Murphy, 1997).

In this work, to account for the sharpness, we related the sharp with the area under the distribution of the complement of cumulative probability (*1-F(q)*, for *q* as the observed discharge) around the forecast with 0.5 occurrence probability value. When this area is zero the forecast is perfectly sharp and the parameter is one (Figure 2(a)). When the area is the area of climatology curve then the forecast don't have sharp and the parameter is zero (Figure 2(b)). Thus, the Equation [1] was using to account the indice sharpness (Freire, 2009). When the indice is closer by one, better the sharpness of forecast.

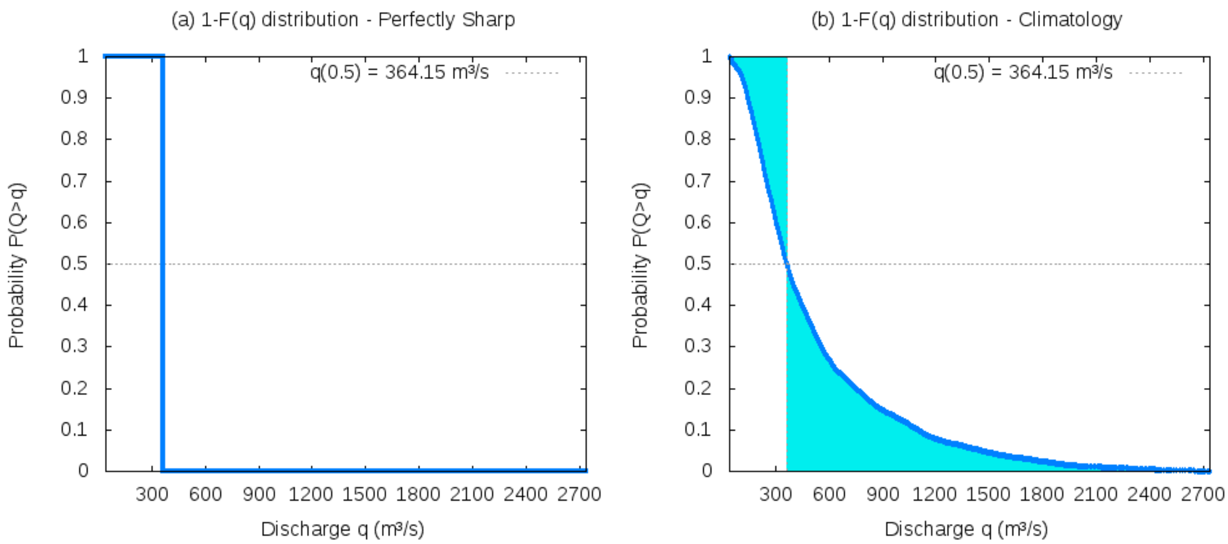

Figure 2: Equivalent area of probabilistic forecast (a) for the Sharpness of climatology curve and (b) for a perfectly sharp, considering the discharge with probability of 0.5.

$$S = 1 - \frac{1}{n}\left(\frac{\sum_{i=1}^{n} A_i}{Ac}\right) \qquad [1]$$

where *S* is the sharpness; *n* is the number of forecasts; *Ai* is the area of forecast *i*; and *Ac is the* area of climatology curve.

The discrimination is related with the difference between the overall mean of forecasts and the mean for a specific observation (Katz and Murphy, 1997). In this work, we used the same sharpness area method to account the discrimination. Nevertheless, now the reference is the last observed discharge. The Figure 3 shows an example for a hypothetical observed discharge of 750 $m^3 s^{-1}$ and the account of indice is obtained by Equation [2]. When the indice is closer by one, better the discrimination of forecast.

$$D = 1 - \frac{1}{n}\left(\sum_{i=1}^{n} \frac{A_i}{Ac_i}\right) \qquad [2]$$

where *D* is the discrimination and Ac_i is the area of climatology curve to observed discharge at the moment.

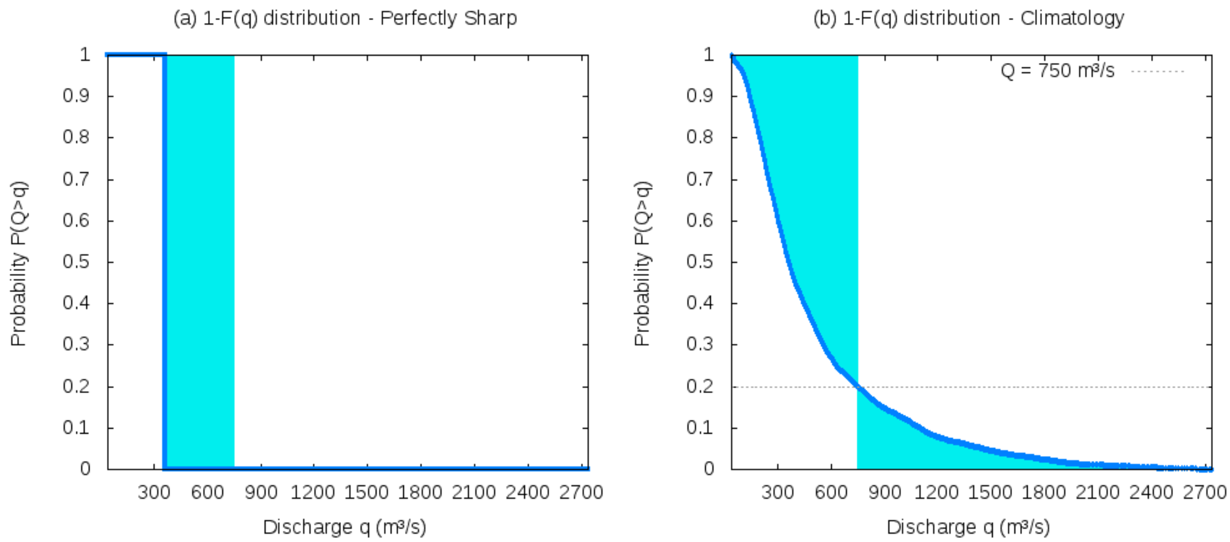

Figure 3: Equivalent area of probabilistic forecast considering an observed discharge (a) for the discrimination to climatology curve and (b) for the discrimination to perfectly sharp.

Toth *et al.* (2003) presents the analysis rank histogram as an alternative to account reliability. For this, the occurrence probabilities were divided by intervals to define ensembles. In this work we consider ten ensembles where the probability interval varies each 0.1. The forecast is more reliable when the discharge occurs closest to the percentage indicated by the probabilistic forecast, in other words, when the rectangle of the histogram is the closest of the line defined by ensembles interval. In our case the line is 0.1. Freire (2009) represented the distance between the rectangles and line by areas too (Figure 4). The maximum reliability is when this area is zero (Figure 5(a)) and we don't have reliability when all occurrences are concentrated in a single ensemble (Figure 5(b)).

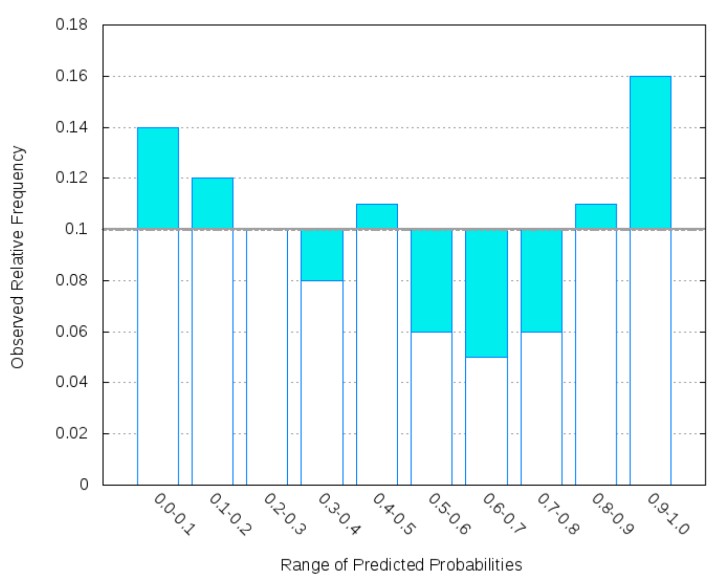

Figure 4: Histogram to assess the equivalent area to the reliability degree of a probabilistic forecast.

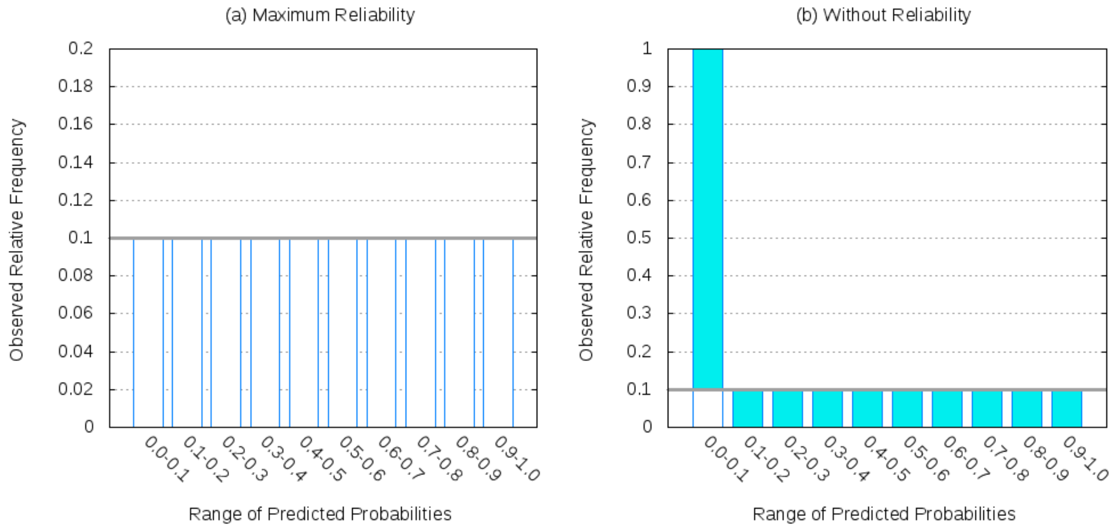

Figure 5: Equivalent area to reliability degree for a probabilistic forecast (a) with maximum reliability and (b) without reliability.

The quality indice of reliability (Equation [3]) was developed for a range between zero and one. It is zero when the forecast is without reliability and one when the reliability is maximum. Thus, when the indice is closer by one, better the reliability of forecast.

$$R = \frac{0.18 - A}{0{,}18} \qquad [3]$$

where *R* is the reliability and *A* is the area outside the plan of the observed discharge histogram.

The three quality indices were accounted to each forecast horizon totaling 36 values. The quality function based on the multi criteria concept proposed by Tkach (1997) was developed for a simultaneous evaluation of the indices and for a more direct comparison of models' quality. How the reliability and discrimination are more important than sharpness for these models, we adopted weight two to the first and second and one to the last. Weights were also attributed to each forecast horizon. This was done because the higher the horizon the higher the uncertainty of the forecast. Equation [4] was developed for a range between zero and one.

$$f = 1 - \sqrt[3]{\frac{2\sum_{h=1}^{12}(13-h)R_h^3 + 1\sum_{h=1}^{12}(13-h)S_h^3 + 2\sum_{h=1}^{12}(13-h)D_h^3}{5*78}} \qquad [4]$$

where *f* is the quality function that combined all indices; h is the order of the horizon, where h=1 for the horizon of 6 hours, h=2 for the 12 hours, etc.

The analysis of the quality function is the opposite to of indices, in other words, when *f* is closer by zero, better the quality of model. Therefore, we would minimize this function.

4. RESULTS AND DISCUSSION

The probabilistic forecasts are carried out in real time every 6 hours and available by SIMEPAR to interested users. Figure 6 shows a example of both models' forecasts for event of the biggest discharge peak of the year 2012. The graphics are hydrograph of dispersion flow which represent observed discharge, deterministic forecasts and discharge bands with their occurrence probability forecast by the probabilistic models. Besides the forecast by bands probabilities, the models also do forecasts of discharges probability distributions for each horizon. The indices are accounted by these distributions.

Analyzing Figure 6 we can see that both models presented good forecasts because the bands well wrap the observed discharge. Nevertheless, we can consider the SISPSHI-NQT obtained better results for this event because the observed discharge is closer the median and well wrap by 50% band. However, only verify some isolated events don't enough to conclude which model has presented the best results. Therefore the quality indices were developed to better compare the models and keep up their development over the forecasts.

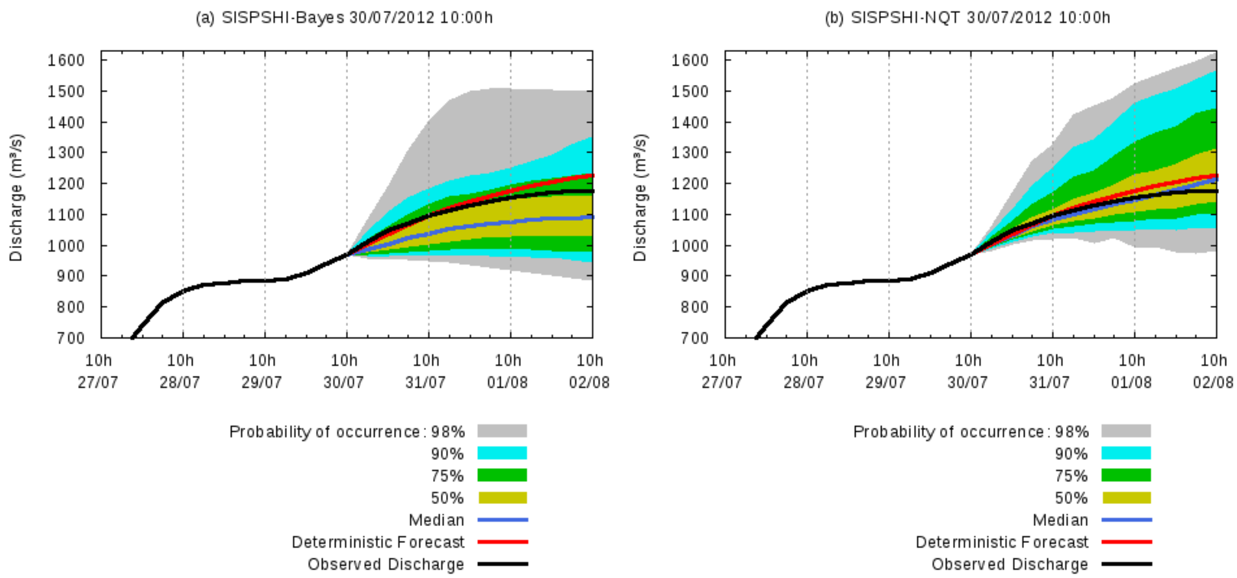

Figure 6: Discharge probabilistic forecasts held on 10/07/2012 at 10 am by (a) SISPSHI-Bayes and (b) SISPSHI-NQT.

Table 1 show the quality indices computed for both probabilistic models considering the forecasts made in the years 2011 and 2012. Remembering that when the indices are closer by one better the quality and when the quality function is closer by zero better the quality. As expected, we can show that all indices worsen with increasing the forecast horizon, especially the sharpness of SISPSHI-Bayes which go from 0.9273 in the 6 hour horizon to 0.2287 in the 72 hour horizon. However this worsening don't vary much influence the accounted of quality function in view of the fact that the adopted weights value more the reliability and discrimination and the first forecast horizons. Notwithstanding, in general, the indices of SISPSHI-NQT are better than ones of SISPSHI-Bayes what is reflect in account of quality function.

horizon (h)	SISPSHI-Bayes			SISPSHI-NQT		
	reliability	sharpness	discrimination	reliability	sharpness	discrimination
6	0.8976	0.9273	0.9604	0.9662	0.9809	0.9808
12	0.8814	0.8586	0.9209	0.9557	0.9692	0.9687
18	0.9264	0.8032	0.8913	0.9552	0.9563	0.9544
24	0.9445	0.7519	0.8640	0.9324	0.9469	0.9416
30	0.9445	0.7004	0.8364	0.9252	0.9372	0.9320
36	0.9294	0.6450	0.8051	0.9117	0.9140	0.9051
42	0.9161	0.5783	0.7661	0.9098	0.8968	0.8871
48	0.9004	0.5010	0.7201	0.9022	0.8787	0.8682
54	0.8943	0.4255	0.6741	0.8976	0.8633	0.8499
60	0.8830	0.3565	0.6330	0.8938	0.8485	0.8318
66	0.8785	0.2908	0.5928	0.8868	0.8150	0.8047
72	0.8815	0.2287	0.5558	0.8874	0.8164	0.7947
total	0.9123	0.7057	0.8359	0.9325	0.9312	0.9256
		***f* = 0.1421**			***f* = 0.0687**	

Table 1: Quality indices obtained for forecasts from 1998 to 2012 for SISPSHI-Bayes and SISPSHI-NQT.

Table 2 show the quality function evolution from one year to another. Both models had an improvement of around 5% which confirm the ability of this kind of model to improve quality of their forecasts.

	f	
year	**2011**	**2012**
SISPSHI-Bayes	0.1497	0.1421
SISPSHI-NQT	0.0721	0.0687

Table 2: Evolution in time of quality function of SISPSHI-Bayes and SISPSHI-NQT.

5. CONCLUSION

In this work, two probabilistic models were presented and a methodology of verify and compare their forecasts quality were proposed. The reliability together resolution, represented by sharpness and discrimination, proved their usefulness to verify probabilistic forecasts.

As the models are from a same theoretical basis their behavior are similar and visual comparison of the forecasts is possible but if done to isolated events it may not be enough to define which the models has presented the best results. Hence the analysis of the indices and quality function proved so important.

Through the quality function is also possible analyze an important property of the probabilistic forecasts which is their improvement over time and with increase of historical of observed and foreseen discharges. Furthermore the indices allow a more detailed monitoring of the models, with only two years is already possible verify an improvement of models quality. Despite the initial hardness to interpreting and developing indices, they proved a useful tool to verify and compare probabilistic models.

6. REFERENCES

Allan H. Murphy, 1997: Economic value of weather and climate forecasts. Chapter 2: Forecast verification. Edited by Richard W. Katz and Allan H. Murphy.

Ângelo Breda, 2008: Avaliação de melhorias para um sistema de previsão hidrológica horária. Dissertação de Mestrado em Engenharia de Recursos Hídricos e Ambientais, Universidade Federal do Paraná.

Anne C. Negrão, 2013: “Previsão hidrológica probabilística derivada da modelagem dos erros e da aproximação Meta-Gausssiana.” XX Simpósio Brasileiro de Recursos Hídricos, Bento Gonçalves, Brasil, November 17-22, 2013.

Lívia S. Freire, 2009: Uso de Rede Neural na Obtenção de Previsão Hidrológica Probabilística. Trabalho Final de Curso de Engenharia Ambiental, Universidade Federal do Paraná.

Robert R. Tkach and Slobodan P. Simonovic, 1997: “A new approach to multi-criteria decision making in water resources” Journal of Geographic Information and Decision Analysis 1, 25-44.

Roman Krzysztofowicz, 2001: “The case for probabilistic forecasting in hydrology” *Journal of Hydrology* 249, 2-9.

Zoltan Toth and Oliver Talagrand and Guillem Candille and Yuejian Zhu, 2003: Forecast Verification: A Practitioner’s Guide in Atmosferic Science. Chapter 7: Probability and Ensemble Forecasts. Edited by I. T. Jolliffe and D.B. Stephenson. John WIey & Sons, Ltd.