Journal of Hydroinformatics | 7.3 | 2005

Fuzzy set based error measure for hydrologic model evaluation

Ramesh Teegavarapu and Amin Elshorbagy

ABSTRACT

Traditional error measures (e.g. mean squared error, mean relative error) are often used in the field of water resources to evaluate the performance of models developed for modeling various hydrological processes. However, these measures may not always provide a comprehensive assessment of the performance of the model intended for a specific application. A new error measure is proposed and developed in this paper to fill the gap left by existing traditional error measures for performance evaluation. The measure quantifies the error that corresponds to the hydrologic condition and model application under consideration and also facilitates selection of the best model whenever multiple models are available for that application. Fuzzy set theory is used to model the modeler's perceptions of predictive accuracy in specific applications. The development of the error measure is primarily intended for use with models that provide hydrologic time series predictions. Hypothetical and real-life examples are used to illustrate and evaluate this measure. Results indicate that use of this measure is rational and meaningful in the selection process of an appropriate model from a set of competing models.

Key words | error measures, fuzzy sets, hydrologic modeling

Ramesh Teegavarapu
Department of Civil Engineering,
University of Kentucky,
Lexington,
KT, 40506-0281,
USA

Amin Elshorbagy (corresponding author)
Centre for Advanced Numerical Simulation (CANSIM), Department of Civil & Geological Engineering,
University of Saskatchewan,
Saskatoon, SK,
Canada S7N 5A9
Tel: +1 306 966 5414
Fax: +1 306 966 5427
E-mail: *elshorbagy@engr.usask.ca*

INTRODUCTION

Design of water resource management systems primarily depends on the accurate prediction and assessment of hydrologic inputs (e.g. streamflows and rainfall). Temporal variations of streamflow and rainfall values define a spectrum of hydrologic conditions (e.g. drought, normal and flood). Drought and flood conditions form two extreme ends of this spectrum and normal or average conditions reflect all the flows in between. Forecasting of hydrologic extremes over time is essential for a number of reasons. For example, low flows are important from a water quality management or drought assessment perspective whereas high flows assume importance from a flood protection point of view. Watershed models (Singh 1995) of varying complexity and time series prediction models (Salas *et al.* 1980) are generally used for hydrologic forecasting applications such as flow forecasting, synthetic flow generation and design-flood estimation. The severity of flood depends on the magnitude of flows and the watershed region in which it occurs. Similarly, critical flow values designated for water quality improvement are debatable and subjective. Hydrologists, water resource management practitioners and planners are often faced with the difficulty of defining these extreme conditions, their importance for a particular application and the reliability of the outputs obtained from models (Melching *et al.* 1991; Watts 1997; Grayson & Bloschl 2000). Selection of the best model from a set of competing models available for a particular application and hydrologic process is often a difficult task. Grayson & Bloschl (2000) indicate the need to establish a link between the purpose of the model and the measures used to quantify the performance of the model. Their comments provide strong motivating factors for undertaking the current study.

In general, traditional error measures (e.g. mean squared error and mean relative error) based on the predicted and actual observations are used to evaluate the performance of "conceptual" hydrologic and time series forecasting models. Information gathered from error measures has been used in the past to refine and re-calibrate conceptual models in watershed hydrology (Sorooshian & Gupta 1995). As additional data-driven forecasting techniques (e.g. neural network based methods) (Govindaraju & Rao 2000) and conceptual models become available for hydrologic predictions, researchers are being forced to evaluate the performance of these new techniques in light of the existing techniques in a number of ways. The development of conceptually refined error measures is also important from this perspective.

Conventional error measures used to evaluate the performance of models developed for a number of hydrologic processes (e.g. Nash & Sutcliffe 1970; Maidment 1993; Karunanithi *et al.* 1994; Bastarache *et al.* 1997) have limited use and may not always provide a comprehensive assessment of the performance of the model developed for a specific application. Some of the existing error measures provide an objective assessment of the models and are absolute measures. For example, the use of a single error measure (e.g. MSE or MRE) of model performance fails to provide a clear assessment of some of the weaknesses and strengths of models developed for flood forecasting (Watts 1997; Elshorbagy *et al.* 2000). Also, the significance attached to the prediction accuracy of a model depends on the type of the application (e.g. flood, drought) for which the model is intended. Both the areas of error significance and hydrologic conditions (e.g. flood, drought or normal flows) are imprecise and are not clearly defined for a particular application. The use of fuzzy set theory (Zadeh 1965) to address these issues is explored in this study.

TRADITIONAL ERROR MEASURES

Different error measures carry different information and may provide contradictory preferences for models from a set of competing models used for a particular application. Two widely recognized and commonly used error measures for hydrologic applications are the MSE and the MRE, given by equations (1) and (2), respectively:

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(\hat{\phi}_i - \phi_i)^2 \tag{1}$$

$$\mathrm{MRE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{(\hat{\phi}_i - \phi_i)}{\phi_i}\right| \tag{2}$$

where n is the total number of observations, $\hat{\phi}_i$ is the predicted value and ϕ_i is the actual value of the observation. In general, MSE, or its root RMSE, provides information about the performance of a model in high flow situations whereas MRE is a good indicator of a model's performance at low or moderate streamflows (Carpenter & Barthelemy 1994; Karunanithi *et al.* 1994). One disadvantage of MSE and MRE is that they attach equal importance to residuals or error values of equal magnitude, irrespective of their relevance to specific aspects of particular model application. MSE is one of the most commonly used error measures in hydrologic modeling. Many researchers (e.g. Kite 1978; Karunanithi *et al.* 1994; Bastarache *et al.* 1997; Shamseldin 1997) in the past have used the MSE for performance evaluation of models developed for hydrologic applications. Additional measures such as percent error in volume, percent error in matching the maximum flow and the correlation between observed and simulated flows were used by Hsu *et al.* (1995) and Degagne & Simonovic (1994). Sugawara (1995) used an error criterion that combines MSE and its logarithmic form, MSEL, to evaluate the performance of a watershed model.

Carpenter & Barthelemy (1994) have used function approximation examples to assess error measures based on the quality of fit at specific points and overall fit in a particular region of interest. Model efficiency criterion, R^2 (Nash & Sutcliffe 1970), was used by Shamseldin (1997) to assess the performance of different rainfall–runoff models. The criterion combines two MSE measures that are based on the mean value of actual and individual observations. Simonovic & Todini (1994) used 7 statistical measures (water balance error, relative absolute error, explained variance, coefficient of determination, correlation coefficient, mean value and standard deviation) for automatic calibration of a rainfall–runoff model. Elshorbagy *et al.* (2000) developed pooled mean squared error (PMSE) to evaluate flood forecasting models based on artificial neural networks (ANN) and regression techniques. This measure combines two standard error measures, MSE and MRE, by ranking the pool of all relative errors obtained from the predictions of all available models. The squared error (SE) is then multiplied by the rank and normalized to provide one single error measure. A review by ASCE (1993) recommends three measures – deviation of run-off volumes, Nash–Sutcliffe coefficient and coefficient of gain from daily mean – to evaluate model predictions and also cautions that many published papers provide inadequate information about the quality of model predictions.

In many scientific disciplines where time series forecasting is considered valuable (e.g. finance, business), MSE has been a preferred error measure to draw conclusions about different forecasting methods (Armstrong & Collopy 1992a). Other measures that are commonly used in the business and finance industry include: (i) mean absolute percentage error (MAPE), (ii) geometric mean of relative absolute error (GMRAE) and median of absolute percentage error (MdAPE). These measures have been used in different forms in hydrology in the past (e.g. Hsu *et al.* 1995). Armstrong & Collopy (1992b) provide a detailed study of these measures for making comparisons of 11 forecasting methods for a time series. They indicate that MSE is less reliable for evaluating forecasting methods when compared to unit-free MRE measure. Several studies (Armstrong & Collopy 1992a, b; ASCE 1993; Armstrong & Fildes 1995; Watts 1997) indicate the limitations of traditional error measures and their biases in evaluating forecasting methods. Watts (1997) indicates that none of the traditional error measures is perfect. In light of the earlier discussion regarding MSE and MRE and their inability to clearly identify the significance of the calculated error values or their relevance to a specific model application, it is apparent that an error measure capable of overcoming the previously discussed difficulties is required.

The main objective of this study is to develop an error measure that will help evaluate and assess the relative performance of competing models for a specific application. An error measure combining two error measures (mean squared error and mean relative error) derived from actual and predicted values is developed using concepts from fuzzy set theory. This error measure facilitates the involvement of the modeler in the assessment of the model performance and its suitability to a particular situation.

DEVELOPMENT OF FUZZY SET-BASED ERROR MEASURE

A new way of looking at the performance evaluation of models is to address two issues in the process of error measure development: (i) information specific to hydrologic conditions and model application (Watts 1997) and (ii) the modeler's preferences or perceptions attached to the accuracy of the prediction (Loague & Freeze 1985). The former issue can be addressed by incorporating information about hydrologic conditions within the error measure, whereas the latter can be handled using specific tolerance levels for accuracy in defining the model performance. If information is not vague and preferences are clearly defined as crisp numbers, then numerical weights can be used in the error measures to refine and improve the performance evaluation process. If not, concepts such as fuzzy set theory (Zadeh 1965) that help quantify imprecision can be used.

ERROR MEASURES AND FUZZY SETS

Error measure calculations are based on the individual observed and model-generated outputs relevant to a specific hydrologic process. Forecasting models should therefore be evaluated using model predictions of individual or a set of outputs (e.g. streamflows) that define different hydrologic conditions (drought, normal and flood). Also, the intended use of the model for a specific application defines the importance of these conditions. A set of these predicted values comprises a region of interest, in which the model performance needs to be assessed (e.g. a flood forecast model should place more emphasis on the prediction accuracy of high flows). However, this region of importance or interest is often imprecise and is not clearly defined for a particular application. For example, the range of low flow values for developing water quality management measures is not clearly defined. Hence, the evaluation of model performance for a particular application is subjective, especially when the importance of the model-simulated outputs for a specific application is not clearly defined. Fuzzy set theory (Zadeh 1965; Zimmermann 1984) can be used to handle the imprecision and vagueness associated with the definition of these boundaries of importance.

Another important aspect that needs to be included in the error measure is the modeler's perception of the accuracy of the model predictions. This perception is subjective and will depend on the application for which the model is intended. A qualifier that attaches a significance or tolerance level to each individual error value is needed. Fuzzy membership function (Zadeh 1965) is perceived to be an appropriate tool that can signify error values calculated at a specific flow range according to the application under consideration. Error values may be qualified by a membership function according to the modeler's perception of the relative importance of different error values.

Membership functions

Membership functions in fuzzy set theory (Zadeh 1965; Zimmermann 1984) are generally used to express degrees of relevance attached to any element belonging to a set. These functions can be used to model the connection between the preferences or perceptions of the modeler and the accuracy of the prediction as well as the importance attached to specific hydrologic conditions of the application under consideration. Loague & Freeze (1985) strongly advocate the importance of the modeler's complete involvement in analyzing the results of the model and his or her perceptions attached to the performance of the model. Membership functions can be defined to attach the level of importance or penalty to a particular parameter (flow or relative error). Figure 1 shows how membership functions can help address the imprecise areas of hydrologic conditions (e.g. drought, normal and flood). For example, in Figure 1(a) it is evident that the modeler intends to emphasize the importance of low flows for a particular application (drought assessment). The region of interest (Figure 1(a)) is defined by the interval [β_o, β_r]. A higher value of membership ($\sim$1) is attached to this region of interest (flows between 0 and 100 cfs). Similar regions can be identified for flood and normal flow conditions as [β_r, β_R], [β_o, β_R], respectively. A variety of membership functions with different shapes are possible if these regions are considered imprecise. Normal condition (Figure 1(c)) reflects all flow values between the two extremes: drought and flood. The range for relative error, [$\theta_{\min}$, $\theta_{\max}$], is problem-specific. Practical guidance for the derivation of membership functions can be obtained from widely used criteria (e.g. 7Q10 or 4Q3 criterion in water quality management) for low flows. Any flow above the low flow range specified by these criteria can be used for defining the range for high flows. The debatable demarcation between high and low flow ranges (Watts 1997)

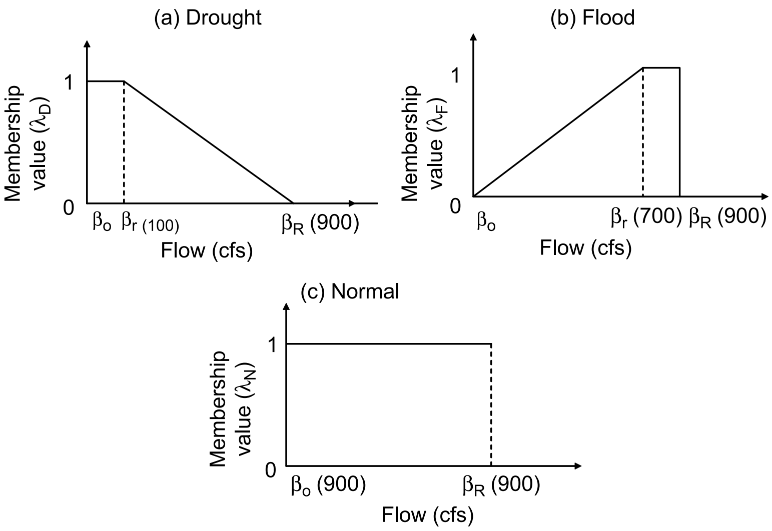

Figure 1 | Membership functions for different hydrologic conditions.

provides the modelers with an incentive as well as a challenge to develop personal preferences and finally justifies the need for techniques that can quantify these preferences.

Another membership function is used to quantify the relative importance of relative errors occurring in different flow conditions. A membership function for relative error in the present context can be referred to as a penalty or loss function because it associates a penalty value to a specific value of relative error (Figure 2).

FUZZY MEAN SQUARED ERROR (FMSE)

The error measure developed in the present study uses concepts from fuzzy set theory and therefore is referred to as fuzzy mean squared error (FMSE). The error measure combines information pertinent to problem-specific hydrologic conditions and the significance of the prediction accuracy level into one single general error measure. Two membership values that identify the preferences modeler attached to a particular flow condition and the penalty associated with the relative error are incorporated into the error. The error measure is given by

$$\mathrm{FMSE} = \frac{1}{n}\sum_{i=1}^{n}(\hat{\phi}_i - \phi_i)^2 \lambda_{RE}\lambda_{flow} \quad (3)$$

where λ_{RE} and λ_{flow} are the membership values assigned for relative error and flow values (low, normal or high) or flow conditions, respectively, and n is the number of observations. The membership values can be obtained for each individual observation and are limited to the interval, [0,1]. Linear membership functions are used in this study because they are simple to comprehend and develop. These functions for different flow conditions are shown in Figure 1. The functions suggest the importance attached to

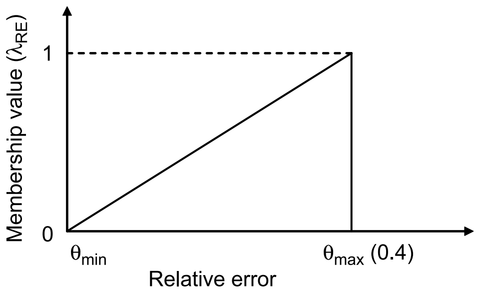

Figure 2 | Membership function for relative error.

different flow conditions by the modeler. For example, if a model is evaluated for a specific application that is intended for low flow situations, higher values of importance (close to 1) are attached to low flows and lower values to others. Similarly, a membership function for relative error (RE) can be developed based on the maximum value of RE obtained using the residuals from the models under consideration. The proposed error measure, FMSE, helps achieve the following objectives: (1) combining the effect of the MSE and MRE in one measure; (2) increasing the relative effect of the significant error residuals (using higher λ_{RE} values), which also reflects the modeler's perception of the importance attached to a specific relative error; and (3) measuring and comprehending the performance of hydrologic models in reference to the type of situations for which these models are intended to be used.

APPLICATION TO A HYPOTHETICAL EXAMPLE

The FMSE measure is applied to a numerical example in which two models (I and II) provide predictions for a hypothetical streamflow time series. The actual values of the flows are given in Table 1 and Table 2 along with the predicted values from Models I and II, respectively. It is evident from the tables that the performance of Model I is better in the case of low flow situations (referred to as drought in Tables 3 and 4) whereas the performance of Model II is better in the case of high flows (referred to as flood in Tables 3 and 4). This performance evaluation can be based on visual inspection of the RE values, which is possible because there are few values in the complete time series. It should be noted that there is one set of λ_{RE} values; $\lambda_{RE} \sim 1$ for the highest value of RE and $\lambda_{RE} \sim 0$ for the lowest value of RE, whereas three sets of membership values are considered to reflect the hydrologic conditions. These sets are λ_F, λ_N and λ_D for flood, normal and drought conditions, respectively (represented also in Figure 1). Each of these three sets should be used in conjunction with λ_{RE} to calculate FMSE for a specific flow condition. The membership values for relative error, λ_{RE}, are calculated using the absolute values of RE. The error measure values (MSE and MRE) are given in Table 3. In the case of MSE and MRE, the values are the same for all flow conditions

Table 1 | Actual and predicted flows from Model I, residuals, relative errors, fuzzy membership and error measure values

					λ_{flow}			FSE		
Actual flow	**Predicted flow**	**RE (%)**	**SE**	λ_{RE}	λ_F	λ_D	λ_N	FSE_F	FSE_D	FSE_N
10	9	10	1	0.25	0.00	1.00	1.00	0.00	0.25	0.25
25	24	4	1	0.10	0.03	1.00	1.00	0.003	0.10	0.10
65	64	1.5	1	0.04	0.09	1.00	1.00	0.003	0.04	0.04
98	96	2	4	0.05	0.17	1.00	1.00	0.03	0.20	0.20
200	195	2.5	25	0.06	0.28	0.78	1.00	0.43	1.22	1.56
300	291	3	81	0.08	0.31	0.67	1.00	1.89	4.05	6.08
400	384	4	256	0.10	0.33	0.56	1.00	8.53	14.22	25.6
500	450	10	2500	0.25	0.44	0.44	1.00	277.78	277.78	625.00
600	510	15	8100	0.38	0.56	0.33	1.00	1687.50	1012.50	3037.5
620	496	20	15,376	0.5	0.67	0.31	1.00	5125.33	2391.82	7688.00
650	494	24	24,336	0.6	0.78	0.28	1.00	11356.80	4056.00	14601.60
750	950	26.7	40,000	0.67	1.00	0.17	1.00	26666.67	4444.44	26666.67
820	1050	28	52,900	0.70	1.00	0.09	1.00	37094.51	3297.29	37094.51
870	600	31	72,900	0.78	1.00	0.03	1.00	56560.35	1885.35	56560.35
900	550	38.9	122,500	0.97	1.00	0.00	1.00	119097.2	0.00	119097.22

RE: Relative Error; SE: Squared error; FSE: Fuzzy squared error

because they are calculated based on the entire range of flows. It is evident from Table 3 that MSE points to one model for better performance in comparison with the other (Model II is better than Model I) whereas MRE points to the other (Model I is better than Model 2). The existence of these conflicting conclusions makes the selection process much more difficult. Karunanithi *et al.* (1994), Watts (1997) and Elshorbagy *et al.* (2000) point to similar results and conclusions in their experiments with ANN and regression models. However, the use of FMSE in this situation provides an appropriate choice of models for low, normal and high flow conditions.

It would be interesting to compare FMSE with PMSE (developed by Elshorbagy *et al.* (2000)) in the present context. The PMSE values calculated for Models I and II are 18,735 and 3524, respectively. Values of FMSE for Model I and II are 17,963 and 3223, respectively. These values are close to PMSE values in the case of normal flow conditions. PMSE and FMSE values suggest the same conclusion in the case of normal flow conditions. Also, PMSE provides the same conclusion as that of FMSE in the application where predictive accuracy at high flows is important. In the case of PMSE, all relative error (RE) values below a specific threshold (e.g. 5%) receive the same rank, suggesting that the modeler is indifferent to the value of RE as long as it is below the pre-determined threshold level. It should be noted that, unlike FMSE, the other three measures' (MSE, MRE and PMSE) values do not use information about flow conditions

Table 2 | Actual and predicted flows from Model II, residuals, relative errors, fuzzy membership and error values

					λ_{flow}			FSE		
Actual flow	**Predicted flow**	**RE (%)**	**SE**	λ_{RE}	λ_F	λ_D	λ_N	FSE_F	FSE_D	FSE_N
10	6	40	16.00	1.00	0.00	1.00	1.00	0.00	16.00	16.00
25	19	24	36.00	0.60	0.03	1.00	1.00	0.72	21.60	21.60
65	56	13.8	81.00	0.35	0.09	1.00	1.00	2.49	28.04	28.04
98	78	20.4	400.00	0.51	0.17	1.00	1.00	34.01	204.08	204.08
200	140	30	3600.00	0.75	0.28	0.78	1.00	750.00	2100.00	2700.00
300	195	35	11025.00	0.88	0.31	0.67	1.00	3001.25	6431.25	9646.88
400	240	40	25600.00	1.00	0.33	0.56	1.00	8533.33	14222.22	25600.00
500	400	20	10000.00	0.50	0.44	0.44	1.00	2222.22	2222.22	5000.00
600	510	15	8100.00	0.38	0.56	0.33	1.00	1687.50	1012.50	3037.50
620	558	10	3844.00	0.25	0.67	0.31	1.00	640.67	298.98	961.00
650	604	7	2116.00	0.18	0.78	0.28	1.00	288.01	102.86	370.30
750	710	5.3	1600.00	0.13	1.00	0.17	1.00	213.33	35.56	213.33
820	780	4.9	1600.00	0.12	1.00	0.09	1.00	195.12	17.34	195.12
870	830	4.6	1600.00	0.11	1.00	0.03	1.00	183.91	6.13	183.91
900	860	4.4	1600.00	0.11	1.00	0.00	1.00	177.78	0.00	177.78

[a]RE: Relative Error; SE: Squared error and FSE: Fuzzy squared error

because they are not designed to consider the importance of these conditions for the application of the model for a specific situation. In conclusion, Table 3 suggests that, based on FMSE, Model I is superior if drought (low flow) or normal flow forecast is considered, while Model II is superior in the case of flood forecast.

APPLICATION TO A CASE STUDY

Data used in the hypothetical numerical example may suggest a deliberate attempt to project the superiority of one model (i.e. Model I) over the other in some conditions (e.g. low flow situations). Similar observations can be made for Model II in the case of high flows. To eliminate this bias, and to illustrate and re-iterate the applicability of fuzzy error measure (FMSE), two models are developed for a case study region. Daily streamflow data of the Little River and Reed Creek, Virginia, USA for the period 1981–1990 are used in this study. These two streams are cross-correlated, which suggests that the flows in one stream can be estimated using the other. The cross-correlation coefficient between these two streamflow time series is found to be equal to 0.76. A length of 3000 concurrent observations of the two rivers is selected for the analysis. The Little River and the Reed Creek are used as the *reference* and *target rivers*, respectively. Seven patches, each of ten consecutive observations, are considered missing from the *target river*.

Table 3 | Error statistics for numerical example of hypothetical time series

	Model I			Model II		
Hydrologic condition	**MSE**	**MRE (%)**	**FMSE**	**MSE**	**MRE (%)**	**FMSE**
Drought	2.26×10^4	15	1159.12	4747.87	18	1781.25
Flood	2.26×10^4	15	17191.80	4747.87	18	1195.40
Normal	2.26×10^4	15	17963.60	4747.87	18	3223.70

The whole data set is divided into seven sections and one patch of data is randomly selected from each section so that the missing data may be representative of the entire data range. The complete record of the Little River is used, as the *reference river*, to estimate the missing data of the Reed Creek (*target river*).

Two widely applied and understood data-driven techniques – linear regression (LR) and artificial neural networks (ANN) – are used in the analysis. A wealth of literature on neural networks can be found elsewhere (e.g. Freeman & Skapura 1991). The concurrent data of the assumed missing segments are used as the unseen test data to verify the models. The remaining part of the data record is used to train and develop the ANN and LR models. The architecture of the network is given by the configuration, ANN (1-3-1), which suggests input and output neurons along with one hidden layer that has three neurons.

The model based on simple linear regression is given by

$$Q_R = 0.958 + 0.66\ Q_L \tag{4}$$

where Q_R is the estimated streamflow of the Reed Creek and Q_L is the measured streamflow of the Little River.

Table 4 | Error statistics based on different hydrologic conditions using ANN and regression models

	Artificial Neural Network			Regression		
Hydrologic vondition	**MSE**	**MRE**	**FMSE**	**MSE**	**MRE**	**FMSE**
Drought	26.41	0.30	3.63	29.41	0.24	2.69
Flood	26.41	0.30	9.74	29.41	0.24	12.07
Normal	26.41	0.30	9.88	29.41	0.24	12.23

Results from the neural network and regression models are summarized in Table 4. FMSE values, along with the traditional error measures, suggest the relative superiority of one model over the other in different hydrologic flow conditions of interest. It is evident that, based on the MSE criterion, the ANN model is better than the LR model, whereas the MRE supports the opposite conclusion. FMSE provides an appropriate measure to make a conclusion about the performance of the model in different hydrologic conditions: LR model for low and flows and ANN model for high and normal flows.

GENERAL REMARKS

The error measure, FMSE, developed in this paper is a conceptual improvement of the PMSE measure (Elshorbagy *et al.* 2000) in a number of ways. In the case of FMSE, the acceptable threshold level associated with relative error is flexible and can be defined by the modeler as a membership function. PMSE can be regarded as a special case of FMSE, where the membership values for all relative errors are assumed to be constant below a specific acceptable threshold value. FMSE utilizes information that helps evaluate the performance of competing models based on the preferences of the modeler and the type of application under consideration. PMSE provides a relative comparison between the performances of the models by ranking a pool of error values. Using FMSE, one can arrive at a conclusion that is absolute because the membership functions used for evaluation are constant for all the models.

Use of FMSE need not be limited to hydrologic time series prediction models. For example, in the case of any conceptual watershed model, the squared error measure can be modified to consider a specific individual parameter of interest to refine the performance evaluation process. In such a case, an additional membership function should be developed for the parameter of interest and corresponding membership values can be part of the FMSE.

Membership functions used in this study are linear and assumed to be known. However, any form of non-linear function can be used to represent the modeler's preferences. Membership functions can also be developed based on actual surveys for practical applications (Fontane *et al.* 1997) or by consulting experienced hydrologists. In general, sigmoidal or "s"-shaped functions are more appropriate to define smooth transitions in the degree of relevance or importance (Zimmermann 1984). Practical procedures for developing several varieties of membership functions are given by Cox (1999). Membership functions in this study are designed in such a way that a low value of FMSE (the lowest and ideal value being equal to zero) suggests a better performance of the model for a specific application.

CONCLUSIONS

A new error measure, named fuzzy mean squared error (FMSE), is proposed and developed to evaluate the performance of time series prediction models in water resources. The preferences of the modeler or hydrologist attached to the level of prediction accuracy and particular hydrologic conditions are incorporated through membership functions using concepts from fuzzy set theory. Membership functions derived from a number of modeler preferences can be easily aggregated to obtain a single integrated membership function, and therefore one measure can be finally obtained. The fuzzy error measure is appropriate to situations where the model performance is debatable considering the purpose for which the model is intended and the different hydrologic conditions which the model investigates. Concepts used to design FMSE can be extended to refine calibration techniques that use a weighted least squared error measure as an objective function. The applicability of this measure to hypothetical and real-life examples indicates the practical utility of this measure and the conceptual revision of traditional error measures.

ACKNOWLEDGEMENT

The second author wishes to acknowledge the financial support provided by Natural Sciences and Engineering Research Council (NSERC), Canada, through its Discovery Grant Program.

REFERENCES

Armstrong, J. S. & Collopy, F. 1992a Error measures for generalizing about forecasting methods: empirical comparisons. *Int. J. Forecasting* **8**, 69–80.

Armstrong, J. S. & Collopy, F. 1992b A commentary on error measures. *Int. J. Forecasting* **8**, 99–111.

Armstrong, J. S. & Fildes, R. 1995 On the selection of error measures for comparison among forecasting methods. *Int. J. Forecasting* **14**, 67–71.

ASCE 1993 Criteria for evaluation of watershed models. *J. Irrig. Drainage Engng.* **119** (3), 429–442.

Bastarache, D., El-Jabi, N., Turkkam, N. & Clair, T. A. 1997 Predicting conductivity and acidity for small streams using neural networks. *Can. J. Civil Engng.* **24**, 1030–1039.

Carpenter, W. C. & Barthelemy, J-F. 1994 Common misconceptions about neural networks as approximators. *J. Comput. Civil Engng., ASCE* **8** (3), 345–358.

Cox, E. 1999 *The Fuzzy Systems Handbook*. Academic Press, New York.

Degagne, M. P. J. & Simonovic, S. P. 1994 A new method for mathematical modeling of complex stage-discharge relations. *Can. J. Civil Engng.* **21**, 990–998.

Elshorbagy, A., Simonovic, S. P. & Panu, U. S. 2000 Performance evaluation of artificial neural networks for runoff prediction. *J. Hydrol. Engng., ASCE* **5** (4), 424–427.

Fontane, D. G., Gates, T. K. & Moncada, E. 1997 Planning reservoir operations with imprecise objectives. *J. Wat. Res. Plann. Mngmnt., ASCE* **123** (3), 154–162.

Freeman, J. A. & Skapura, D. M. 1991 *Neural Networks. Algorithms, Applications and Programming Techniques*. Addison-Wesley, Reading, MA.

Govindaraju, R. S. & Rao, A. R. (eds.) 2000 *Neural Networks in Hydrology*. Kluwer, Dordrecht.

Grayson, R. & Bloschl, G. 2000 Spatial modeling of catchment dynamics. In *Spatial Patterns in Catchment Hydrology: Observations and Modeling*, Cambridge University Press, Cambridge, pp. 51–81.

Hsu, K. L., Gupta, H. V. & Sorooshian, S. 1995 Artificial neural network modeling of rainfall-runoff process. *Wat. Res. Res.* **31** (10), 2517–2530.

Karunanithi, N., Grenney, W. J., Whitley, D. & Bovee, K. 1994 Neural networks for river flow prediction. *J. Comput. Civil Engng., ASCE* **8** (2), 201–220.

Kite, G. W. 1978 *Frequency and Risk Analyses in Hydrology*. Water Resources Publications, Fort Collins, CO.

Loague, K. M. & Freeze, R. A. 1985 A comparison of rainfall-runoff modeling techniques on small upland catchments. *Wat. Res. Res.* **21** (2), 229–248.

Maidment, D. R. 1993 *Handbook of Hydrology*. McGraw-Hill, New York.

Melching, C. S., Yen, B. C. & Wenzel, H. G., Jr. 1991 Output reliability as guide for selection of rainfall-runoff models. *J. Wat. Res. Plann. Mngmnt., ASCE* **117** (3), 383–398.

Nash, J. E. & Sutcliffe, J. V. 1970 River flow forecasting through conceptual models. Part 1. A discussion of principles. *J. Hydrol.* **10**, 282–290.

Salas, J. D., Delleur, J. W., Yevjevich, V. & Lane, W. L. 1980 *Applied Modeling of Hydrological Time Series*. Water Resources Publications, Littleton, CO.

Shamseldin, A. Y. 1997 Application of a neural network technique to rainfall-runoff modeling. *J. Hydrol.* **199**, 272–294.

Simonovic, S. P. & Todini, E. 1994 Use of an expert system for calibration of rainfall-runoff models: ARNO model case study. In: *Advances in Water Resources Technology and Management* (eds. G. Tsakiris & M. A. Santos), Balkema, Rotterdam, pp. 59–64.

Singh, V. P. 1995 *Computer Models of Watershed Hydrology*. Water Resources Publications, Littleton, CO.

Sorooshian, S. & Gupta, V. K. 1995 Model calibration. In: *Computer Models of Watershed Hydrology* (ed. V. P. Singh), Water Resources Publications, LLC, Highlands Ranch, CO, pp. 23–68.

Sugawara, T. 1995 Tank model. In: *Computer Models of Watershed Hydrology* (ed. V. P. Singh), Water Resources Publications, LLC, Highlands Ranch, CO, pp. 165–214.

Watts, G. 1997 Hydrological modeling in practice. In: *Contemporary Hydrology* (ed. R. L. Wilby), John Wiley and Sons, New York, pp. 151–193.

Zadeh, L. A. 1965 Fuzzy sets. *Inf. Control* **8**, 338–353.

Zimmermann, H. J. 1984 *Fuzzy Set Theory and Applications*. Kluwer, Dordrecht.