Short-term Water Level Prediction Using Fuzzy Adaptive System and Artificial Neural Networks Approaches

[1]R. Erinawati and [2]J. Fenton

[1]Civil and Environmental Engineering Department, The University of Melbourne, [2]School of Anthropology, Geography and Environmental Studies, The University of Melbourne,
E-Mail: [1]r.erinawati@civenv.unimelb.edu.au , [2]fenton@unimelb.edu.au

Keywords: *Neural Networks, Fuzzy Adaptive System, River Flow Forecasting*

EXTENDED ABSTRACT

This paper presents a comparative study on short-term water level prediction using fuzzy adaptive systems (FAS) and artificial neural networks (ANNs). The short-term water prediction (three days or less prediction) is crucial for flood warning system. This prediction is usually calculated using conceptual or physical-based models.

Although conceptual or physical-based models can lead to better understanding of hydraulic and hydrological processes, these types of models are often constrained by data availability, funding limitation and human resources. In this paper, the ANNs and FAS methods are applied to handle such situations where data are limited, such as in developing and under-developed countries. The water level predictions are only based on the information from upstream. Once the water level is known, the discharge can be computed using a rating-curve.

In this paper, four river reaches from four different catchments on Java are used as case studies. Three rivers (Bogowonto, Bengawan Solo, and Telomoyo) use daily data, while the Ciliwung River employs three hourly water level data.

In order to assess the performance of the model, three performance indicators were used: root mean square error (RMSE), mean absolute error (MAE), and coefficient determination (R^2). To allow direct comparisons for four rivers, unit free formulas from RMSE and MAE are also employed. These are known as root mean square percentage error (RMSPE) and mean absolute percentage error (MAPE).

ANNs and FAS produced similar pattern results. However, ANNs gave slightly better results for almost all rivers. Out of four rivers, only the Bengawan Solo performed very well up to three days ahead predictions (RMSPE < 25%). The model of the Ciliwung River could perform relatively well. For ANN models, both the Bogowonto and Telomoyo River could actually gave reasonable results with RMSPE < 25% for one day ahead. Their performance decreased significantly for a longer period. However, for FAS models, the results from both rivers were unacceptable for all time lags, with RMSPE > 50%.

It is concluded from this study, that the application of system identification techniques in data limited areas is not always possible. Careful data selection, location determination and travelling time should always be carried out. Both ANNs and FAS do not perform very well in extrapolating. The nature of ANNs and FAS is trying to identify the relationship of inputs and outputs. If the real physical relationship between inputs and outputs does not exist, it is likely that these techniques will fail.

1. INTRODUCTION

In the last decade floods have occurred in many countries around the world, resulting in significant economic loss and human life. Accurate flood predictions need to be improved in order to minimize the impact of flooding. It is important that flood predictions and warnings should be made as accurately and as far ahead as possible, so that impacts can be mitigated.

Many methods have been proposed to predict incoming floods. Choice of these methods should be considered based on many aspects, such as data, funding, human resources, degree of accuracy, and level of importance. Conceptual or physical based models may lead to a better understanding of hydraulic and hydrological processes; however, most of these types of models are relatively expensive, require a high level of expertise and various kind of data which may not always be available.

If accuracy of the model is important and physical data are available, then using physically-based or hybrid models may be the best choice. In many cases, especially in developing and under-developed countries, the data available are limited. However, a high degree of accuracy may be required. In this case system identification models will be the best option.

The amount of data is highly correlated with funding. In developing countries, hydrological data collection might be not one of funding priority, due to limited funds. So, it is difficult to have a relatively complete hydrological data set.

Building a physical-based model requires a high level of expertise. Highly educated human resources are limited in some developing countries. However, both developed and developing countries normally require models to have a high degree of accuracy. So, even though system identification techniques may not be able to lead to better understanding of underlying physical phenomena, this type of model is suitable for regions which have several limitations.

The idea of using system identification techniques in hydrology is not new. People have been using them since 1970, after Box and Jenkins published their famous book *Time Series Analysis* in 1969 (Hall, 1997). However, the growth in popularity of using data-based techniques only started in the early 80s. With the development of database computer technology, following the fast growth of computer technology and information systems, many of the hydrology data sets have been reconstructed using the database system. This system allowed hydrologists to trace and extract any data for identification, which are then used for system simulation and stream flow prediction.

This paper discusses the performance of artificial neural networks (ANNs) and fuzzy logic (FL) when they are applied in areas where data are limited, such as in developing countries.

2. ARTIFICIAL NEURAL NETWORKS

ANNs are defined as massively parallel-distributed information-processing systems that resemble biological neural networks of the human cognition (ASCE and Govindaraju, 2000). Although McCulloch and Pitts firstly introduced the idea of artificial neural networks over fifty years ago (McCulloch and Pitts, 1943), the large-scale development started only in 1982, when Hopfield introduced iterative procedures for neural networks (Hopfield, 1982).

There are several types of ANNs. Multi-layer perceptrons (MLP) are considered the most widely used in water resources applications (Gupta and Sorooshian, 2000). The MLP with three layers are employed in this study, consisting one input layer, one hidden layer, and one output layer (Figure 1). The common back-propagation algorithm is used as a learning rule. There are two phases involved in the back-propagation algorithm, a feed forward phase where the information propagates forward to calculate the output signal and a backward phase where the connection weights are updated to minimize the difference between computed output and the given output.

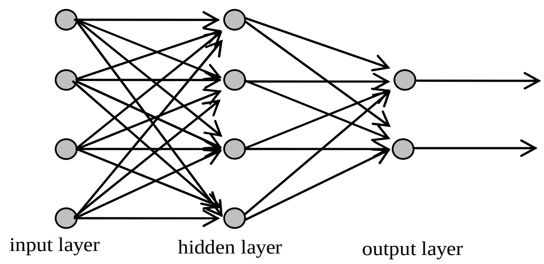

Figure 1. Multi layer perceptron

3. FUZZY RULE BASED-SYSTEM

FL was introduced by Lotfi Zadeh in 1965 (Zadeh, 1965). FL extends the general form of Boolean logic, true and false to handle the concept of vagueness and uncertainty. This approach takes a value between 1 (full belongingness) and 0 (non belongingness), rather than a crisp value. The degree of belongingness is called the membership function. Fuzzy rules are collections of linguistic IF and THEN arguments. A general form of the fuzzy rule can be expressed as IF “X” THEN “Y”.

X is the premise and Y is the consequence of the rule. Since it is based on verbal arguments, this rule allows imprecision and uncertainty in the variables.

There are five steps involved in fuzzy rule based-system: fuzzify inputs, apply fuzzy operators, apply implification method, aggregate outputs, and defuzzify outputs (Figure 2).

In the reality, with the complexity of real world, it is usually not easy to construct rules due to limitations of manipulation and verbalization by an expert. Several methods have been proposed to extract rules directly from numerical data.(Abe, 1997; Abe and Ming-Shong, 1995). This method is normally called a fuzzy adaptive system (FAS).

4. PERFORMANCE INDICATOR

Three performance indicators are used: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Coefficient of Determination (R^2).

RMSE is defined as,

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(O_i - P_i\right)^2} \tag{1}$$

MAE can be calculated as,

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}\left|O_i - P_i\right| \tag{2}$$

where N is the number of data points, O_i is the observed value, and P_i is the predicted value.

Usually RMSE ≥ MAE, and the degree by which RMSE exceeds MAE is an indicator of the extent to which outliers exist in the evaluation set.

In some cases where the comparisons for several rivers are required, the unit free formula of RMSE (RMSPE) and MAE (MAPE) are also employed.

R^2 assesses the goodness of fit by indicating the deviation of the estimates values from the line of the best fit or the regression lineThe value of R^2 is between zero and unity. A value close to unity indicates a satisfactory result, while a low value implies an inadequate result.

5. CASE STUDIES

Java is the most densely populated island in Indonesia. It covers an area of 134,045 km^2. In fact, it is actually only 7% of the total area of Indonesia. But, based on the 1995 census, the total population in Java exceeded 140,000,000, more than half of the total population of Indonesia.

Java stretches from 7°12' to 8°48' south latitude and from 107°00' to 114°42' east longitude. It has a tropical climate with two monsoon seasons: a wet season from November to March and a dry season from June to October. The average temperature in Java is 21^o to 33^{o}C and varies little during wet to dry seasons. Average rainfall in the lowlands varies from 1,780 to 3,175 mm per year, while in some mountainous areas rainfall may reach 6,100 mm per year. The humidity is very high, with the average humidity 80% yearly.

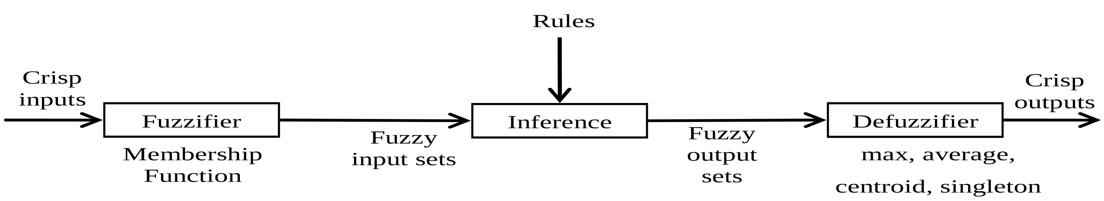

Figure 2. FL scheme work

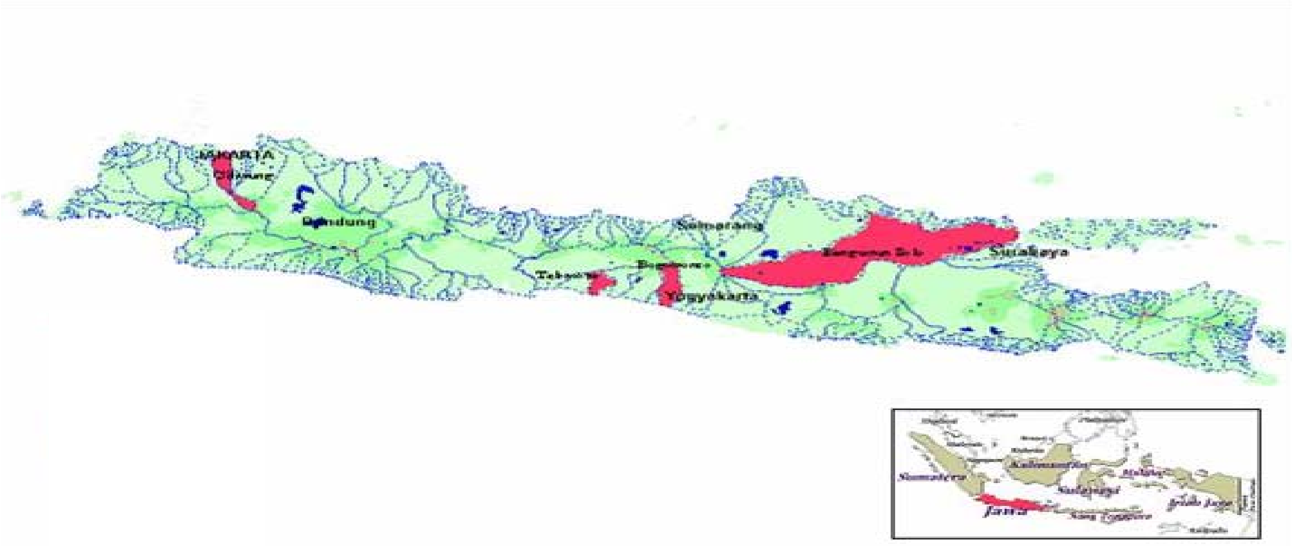

Figure 3. Java Island (Black parts denote the catchment areas used in this study)

Java has thirty river basins, which are administered into ten catchment management authorities. Figure 3 shows the catchments in Java and the areas of interest: Ciliwung (1), Telomoyo (2), Bogowonto (3), and Bengawan Solo (4).

6. MODEL DEVELOPMENT

Except for the Ciliwung River where three-hourly data are employed, the rest use daily data. In this particular study, the downstream flow at time *t*, *y(t)* is assumed to be related to the past input from upstream, *u(t-j)*,

$$y(t) = f\left(u(t-1), u(t-2), \ldots, u(t-j), \ldots, u(t-J) + e(t)\right) \quad (4)$$

where *f* is the unknown mapping function, *t* is the time index, *J* is the unknown number of the past input, *j* is the past input index, and *e* is the error to be minimized. In this study the streamflow predictions at four river reaches have been carried out for 3 hours to 72 hours ahead.

Data availability is limited. From four rivers, only the Ciliwung River has a relatively complete data set, including water level, cross section, bed slope, and rainfall. The Ciliwung water level data are obtained with fifteen minute intervals. However, these data were only available from the year 2002, when the Ciliwung-Cisadane flood control project was started. It was then decided to use one year for training, and one year for verification. For the other rivers, the daily data were the only information available. The bed slope could not be obtained. The Bengawan Solo has a longer period of data, three years data were selected for training, while two years were used for verification.

The Telomoyo and Bogowonto both have very short periods data (two years), one year data were used for training and one year for verification.

7. RESULTS AND DISCUSSIONS

Streamflow predictions at four rivers have been carried out up to 72 hours ahead. These predictions allow a detail assessment of the modelling performance of each model and an observation of the degradation of prediction accuracy of both models. The results for ANNs are summarized in Table 1, while for FL are summarized in Table 2.

Figure 5 displays the R^2 results from the ANN model for each river with different time lags in verification processes. For time leads up to 3 days, only the Bengawan Solo River gave a good performance with R^2 above 0.9. Its accuracy reduced slightly from its best line with increasing time leads. The Ciliwung River gave a good performance up to six hours ahead. Its accuracy

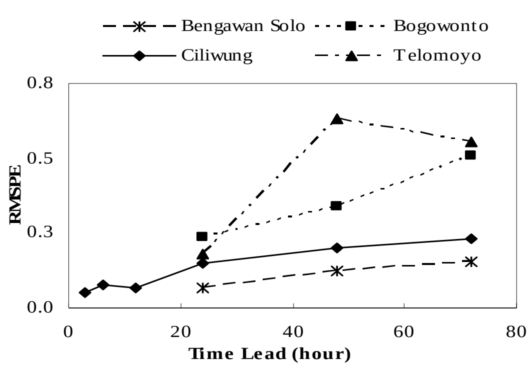

Figure 4. ANNs verification results (RMSE)

Table 1. ANNs Results

Lead Time	3 hours			6 hours			12 hours		
Indicator	RMSPE	MAPE	R^2	RMSPE	MAPE	R^2	RMSPE	MAPE	R^2
Ciliwung	0.053	0.064	0.925	0.082	0.069	0.901	0.069	0.059	0.833

Lead Time	1 day			2 days			3 days		
Indicator	RMSPE	MAPE	R^2	RMSPE	MAPE	R^2	RMSPE	MAPE	R^2
Beng.Solo	0.070	0.040	0.999	0.127	0.082	0.969	0.164	0.115	0.946
Bogowonto	0.248	0.171	0.921	0.358	0.275	0.859	0.533	0.352	0.781
Ciliwung	0.157	0.104	0.874	0.212	0.144	0.871	0.245	0.316	0.680
Telomoyo	0.188	0.072	0.878	0.666	0.408	0.508	0.584	0.324	0.660

Table 2. FL Results

Lead Time	3 hours			6 hours			12 hours		
Indicator	RMSPE	MAPE	R^2	RMSPE	MAPE	R^2	RMSPE	MAPE	R^2
Ciliwung	0.115	0.085	0.905	0.203	0.125	0.754	0.220	0.156	0.520

Lead Time	1 day			2 days			3 days		
Indicator	RMSPE	MAPE	R^2	RMSPE	MAPE	R^2	RMSPE	MAPE	R^2
Beng.Solo	0.138	0.083	0.971	0.155	0.091	0.963	0.187	0.113	0.945
Bogowonto	0.440	0.213	0.734	0.812	0.457	0.070	0.816	0.497	0.138
Ciliwung	0.322	0.201	0.446	0.398	0.248	0.249	0.488	0.348	0.050
Telomoyo	0.426	0.208	0.819	0.571	0.303	0.722	0.786	0.481	0.302

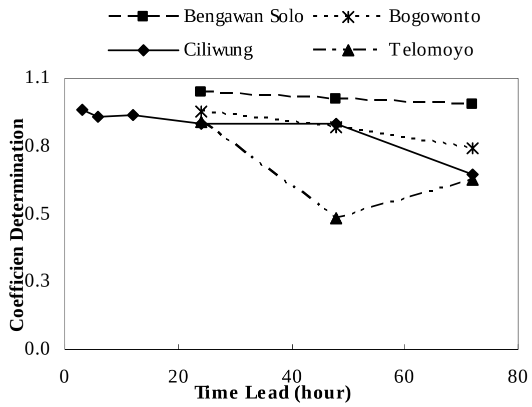

Figure 5. ANNs verification results (R^2)

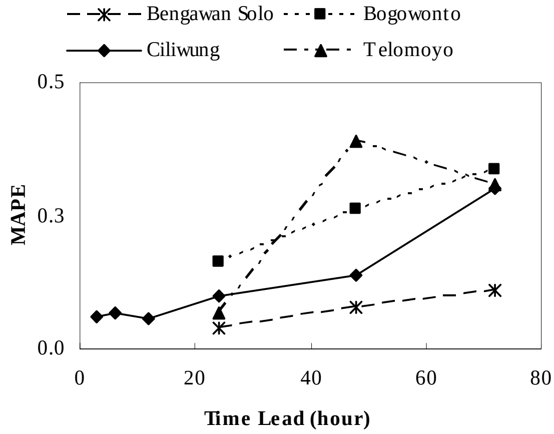

Figure 6. ANNs results (MAPE)

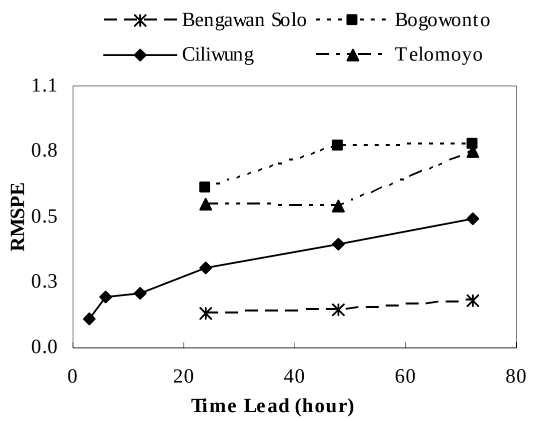

Figure 7. FAS verification results (MAPE)

reduced significantly with higher time leads (R^2=0.68 for 3 days ahead). The Bogowonto and Telomoyo Rivers produced similar patterns with the Ciliwung, but less accurate.

The results above are also indicated by the MAPE and RSMPE (Figure 4 and Figure 6). Both figures display similar error patterns. RMSPE assumes that the high degree errors are more important than small errors, while MAPE considers the absolute error on each item is equally important.

Up to 24 hours ahead, all rivers have errors less than 25%. For two days ahead, only the Telomoyo River performed unsatisfactorily (MAPE>30%).

The FAS results produced similar patterns to the ANNs. But, using FAS, only the Bengawan Solo River could produce satisfactory results (R^2>0.9) up to three days ahead prediction. The Ciliwung River could only produce good prediction for three hours ahead. After that the results departed very quickly from the observed values.

The Bengawan Solo River is the longest river in Java. It comprises more than 600 km. In this study, an approximately 135 km long section is employed. There is no information about the bed slope of the river. The Bengawan Solo has an average velocity of 0.5 m/s. It makes the average travelling time for 135 km approximately 3.1 days. In this case, up to three days ahead prediction, the inputs can represent the actual output flow of the river. The Bengawan Solo River also has relatively long historical data. Three years data were used as training. The longer training data and sufficient traveling time lead to satisfactory results.

The river reach from the Telomoyo River is the shortest section used in this study. It is only 2064m long. The data used is daily data. The slope data is not available. In general, this river produced unsatisfactory results. Based on the fact that this section is very short (about 2km long), the travel time of the water in the upstream part of this section travels is much less than one day. In addition to this problem, the data available are also very short, two years period of data are the only available information. These caused the relationship between input and output data to be not strong, and it is inadequate for ANNs and FAS to be able to recognize the bond.

Similarly with the Telomoyo River, a relatively short (4 km) section of the Bogowonto River was used. Information about flow conditions, including bed slope, velocity, and cross sections of the river were not available. For one day in advance, the calibration produced relatively good results, with R^2 = 0.878, RMSPE < 20%, and MAPE <10%. However, in verification the results were not as good as calibration with R^2 = 0.656, RMSPE > 50%, and MAPE >30%. When the time leads increased, the prediction accuracies also decreased considerably, with R^2 only 0.180 for a three day time lags. This value is considered unacceptable for modelling purposes. The data were available only for a short period (for two years), it is difficult for ANNs and FAS to actually learn the relationship between input and output.

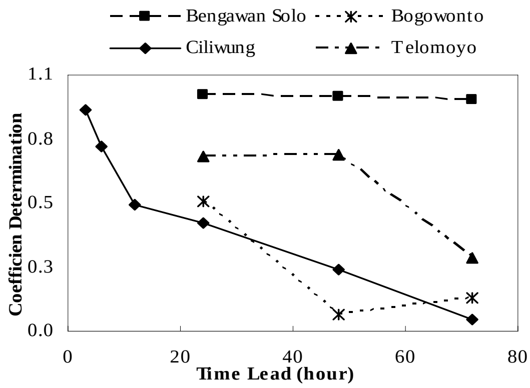

Figure 8. FAS verification results (R^2)

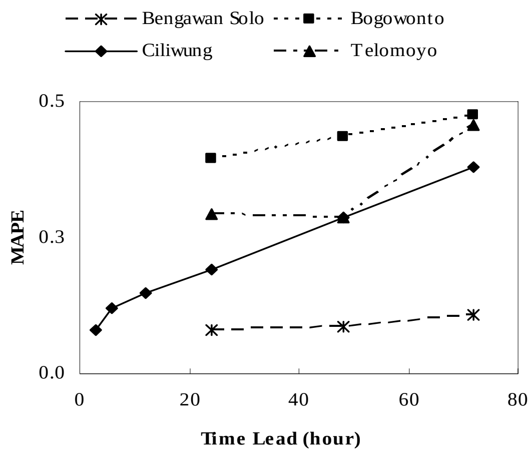

Figure 9. FAS verification results (MAPE)

8. CONCLUSIONS

This paper studies the possibility of using system identification techniques for water level predictions in the area where the data are limited. Two different techniques, ANNs and FAS were employed. Four rivers in Java, Indonesia were used as case studies.

Three rivers (Bengawan Solo, Bogowonto, and Telomoyo) were modeled using daily data, while the Ciliwung used three hourly data.

The pattern of the results from ANNs and FAS are relatively similar. Only the Bengawan Solo River could produce good results.

It can be concluded from this study that the application of system identification techniques in areas where data are limited do not always give satisfactory results. These techniques should be applied with care, especially in regard to data selection, location, and travelling time.

In some cases, using basic data (daily discharge) can actually achieve good solutions with a very high R^2 and small MAPE and RMSPE (see Bengawan Solo results). In the case of the Telomoyo and Bogowonto River, the daily data failed to estimate the downstream discharge for a short distance.

9. ACKNOWLEDGMENTS

This paper is part of a PhD study funded by AusAID. The writer would like to express the gratitude to Ir. Wisnu Subarkah, M.Eng and Ir. Jaya Sukarno, M.Eng for provided Indonesian Hydrological data. references

Abe, S. (1997), *Neural networks and fuzzy systems: theory and applications*, Kluwer Academic, Boston.

Abe, S., and L. Ming-Shong (1995), Fuzzy rules extraction directly from numerical data for function approximation, *IEEE Transactions on Systems, Man and Cybernetics*, 25(1), 119-129.

ASCE Task Committee on Application of Artificial Neural Networks in Hydrology (2000), Artificial Neural Networks in Hydrology. I: Preliminary Concepts, *Journal of Hydrologic Engineering*, 5(2), 115-123.

Gupta, H. V., and S. Sorooshian (2000). Effective and Efficient Modelling for Streamflow Forecasting. In R. S. Govindaraju and A. Ramachandra Rao (Eds.), *Artificial Neural Networks in Hydrology* (pp. 7-22), Kluwer Academic Publisher, Doordrecht.

Hall, M. J. (1997), *Statistic Hydrology*,IHE Delft, Delft.

Hopfield, J. J. (1982), Neural networks and physical systems with emergent collective computational abilities, *Nat. Acad. of Scientists*, 79, 2554-2558.

McCulloch, W. S., and W. Pitts (1943), A logic calculus of the ideas immanent in nervous activity, *Bulletin of Mathematical and Biophysics*, 5, 115-133.

Zadeh, L. (1965), Fuzzy Sets, *Information and Control*, 8, 338-353.