7th International Conference on Hydroinformatics
HIC 2006, Nice, FRANCE

A SWOT ANALYSIS OF ANN VERSUS CONCEPTUAL, PHYSICALLY ORIENTED MODELLING BASED ON LYSIMETER DATA

JITENDRA KUMAR
M.Tech. Student, Department of Civil Engineering,
IIT Kanpur, India

MANFRED OSTROWSKI
Professor, Fachgebiet für Ingenieurhydrologie und Wasserbewirtschaftung,
TU Darmstadt, Germany

ASHU JAIN
Associate Professor, Department of Civil Engineering,
IIT Kanpur, India

Abstract: This paper presents a comparative study of Artificial Neural Networks (ANN) and conceptual, physically oriented modelling techniques for soil moisture modelling. Data from three lysimeters in the Federal Republic of Germany was employed in the present study. A three layered feed forward ANN using error back-propagation training was used for neural network modelling of the lysimeters. A real coded Genetic Algorithm was used for automatic parameter estimation of conceptual, physically based model. The results show that the ANN models performed best in terms of various performance criteria for the modelling of soil moisture for the lysimeters. Acceptable results were obtained from the conceptual physically based model provided with the estimated parameters which can be helpful in practice in situations when insufficient data is available for the purpose of ANN modelling.

1. INTRODUCTION

Hydrologic soil moisture models are of high importance for understanding the soil-plant-water relationships and for the management of irrigation and water resources systems. Hydrologic simulation models utilize a set of parameters and inputs that are specific to the location being simulated. Model results are highly sensitive to soil hydraulic parameters. Therefore, accurate quantification of soil hydraulic properties is essential to model hydrological processes (Jhorar et al., 2004). Soil characteristics often show very heterogeneous spatial distribution and anisotropy. Modelling errors can be mainly caused by scarce information on soil characteristics rather than by insufficient model accuracy (Ostrowski, 1991).

Soil hydraulic function can be determined in a classic way by laboratory experiments on the soil core samples. In principle, this is a direct way but practically difficult, time consuming and costly. Specialized apparatus are required for collection of the soil samples and for subsequent experiments. Besides that, it is hard to assume the parameters of the soil core obtained under laboratory conditions to be representative of field conditions. Another approach for the estimation of the soil hydraulic parameters are pedotransfer functions. Pedotransfer functions relate soil hydraulic properties to more easily available soil properties through regression equations (Nemse et al, 1999).

Recently there has been intensive research into the development of automatic calibration methods for hydrologic models (Yapo et al., 1996; Madsen, 2000). Researchers have studied the use of various optimization algorithms and also various objective functions for the automatic calibration of the hydrologic models (Gupta and Sorooshian,1983; Sorooshian and Gupta 1985; Yapo et al.,1996; Yu and Yang, 2000; Madsen, 2000; Khu and Madsen 2005). Also Artificial Neural Networks (ANNs) have been proposed as efficient tools for modelling and prediction in hydrology as black box models. Recent studies have demonstrated the efficiency of ANNs in modelling hydrologic processes (Jain et al., 2004,2006).

The objective of the study presented in this paper are (a) estimation of soil hydraulic parameters using genetic algorithm with different objective functions (b) modelling of soil-moisture hydrologic processes using neural networks (c) analysis and comparison of the two modelling techniques.

2. DESCRIPTION OF THE DATA USED FOR THE STUDY

The data used for the analysis stem from the Senne lysimeter station in Germany, located in the north-western part of the country in the state of North-Rhine-Westfalia. Data from three lysimeter troughs were used in the present study. All lysimeters have grass cover on the surface with an area of 1m^2. The troughs can be weighed and surface runoff can be measured. Table 1 gives an overview of the lysimeters modelled.

Table 1: Characteristics of the lysimeters modelled

Station	Senne		
Trough	1	2	3
Soil column	2.0 meters		
Soils 65-78	Sand	Loess	Loess
Soils 79-85	Sand	Sand	Loess
Surface Area	1m^2		
Vegetation	Grass		

The daily values of the data for lysimeters were available for a period of 20 years from 1965-1984. Hydro-meteorologic data used were rainfall at ground and 1 m level, temperature, potential evapotransiration, change of weight, percolation and surface runoff.

3. MODEL DEVELOPMENT

For the modelling of the soil moisture hydrologic process in the lysimeter troughs the main concern is to estimate the change in soil moisture and percolation. The input (i.e. rainfall) to the system (lysimeter troughs) flows through the different soil layers. The phenomena of infiltration, percolation, evapotranspiration etc. are affected by the climatic conditions, soil hydraulic properties and vegetation.

Conceptual physically oriented model

A conceptual physically oriented model for the lysimeter troughs was used. The model needs meteorological data along with a set of parameters in order to model the change in soil moisture and percolation of the lysimeter. Interception and snow storage processes are neglected. The parameters of the model are dependent on the soil types of the lysimeter being modelled (Ostrowski et al, 2006).

A genetic algorithm (GA) is a directed search technique based on the concept of natural selection inherent in the genetics, which combines an artificial survival of the fittest with genetic operators abstracted from nature. A genetic algorithm with real variables was used in the present study to optimize the parameters of the soil moisture model. Two approaches for the optimization of the soil hydraulic parameters were applied in the study which are termed as GA1 and GA2 model based on different objective functions. In the GA1 model the objective function to be minimized is the Sum Squared Error (SSE) for change in soil moisture.

$$SSE1 = \sum_{i=1}^{N} (\partial SMo - \partial SMc)^2$$

In GA2 model the objective function to be minimized is SSE of percolation.

$$SSE2 = \sum_{i=1}^{N} (PERCo - PERCc)^2$$

where N is the sample size in days, δSMo is the observed change in soil moisture, δSMc is the calculated change in soil moisture, PERCo is the observed percolation and PERCc is the calculated percolation. The objective functions to be minimized were chosen as Sum Squared Error which is same as in the error back propagation training algorithm of ANN to make a comparison.

Both models were applied to the three lysimeters to estimate the soil hydraulic parameters of the soil-moisture model.

An algorithm developed by Kanpur Genetic Algorithm Laboratory [12] was used in this study. A cross over probability of 0.90 and mutation probability of 0.10 was used in the present study. Simulations were carried out with an initial population size of 200. A number of trial runs were carried out with different initial population in order to obtain global optimum solution.

Artificial Neural Network model

Artificial Neural Networks rely on a highly sophisticated paradigm that borrows features from human and animal brains that enables the recognition of pattern within the data. The ANNs learn to solve a problem by developing a memory capable of associating a large number of input patterns with a resulting set of output as effects. The ANNs develops a solution system by training on examples given to it. The ANNs functions as universal approximators and are non linear in nature.
In this study a three layered feed forward ANN was employed. The ANN was trained with a stepwise training error back propagation algorithm. Separate models were developed for modelling of change in soil moisture and percolation and were employed for modelling of the three lysimeters. Cross correlation analysis of the available data was carried out to decide on the input neurons for the neural networks. Two ANN models were developed, one each for change in soil moisture and percolation and being termed as ANN1, ANN2.

The ANN1 model consists of five input neurons as potential evaotranspiration, precipitation recorded at ground level, precipitation recorded at 1m above the ground, percolation and runoff with change in soil moisture as the output. The ANN2 model consists of potential evapotranspiration, precipitation at ground level, precipitation at 1m above the ground, change in soil moisture and the runoff as inputs and has percolation as the output to the network.

The next step of the development of the ANN model is the determination of optimal number of neurons in the hidden layer. The number of neurons in the hidden layer is responsible for mapping the dynamic and complex relationship among the input and the output variables considered. The unipolar sigmoid activation function (Zurada, 1997) was used as the transfer function at both the hidden and the output layers. The study employed the stepwise learning error back-propagation training algorithm with momentum factor. The value of learning coefficient of 0.075 and momentum correction factor of 0.075 was used for the training of the network. The number of neurons in the hidden layer was varied from 1-25 to minimize the Sum Squared Error (SSE) at the output neuron. The stopping criteria adopted during the training was a maximum of 50,000 iterations or an acceptable error of SSE of 0.0005. The training data set was divided into training and validation sets to prevent over or under training of the networks.

4. PERFORMANCE STATISTICS

Three different standard performance statistics were employed for model development. These are Normalized root mean square error (NRMSE), Nash-Sutcliffe efficiency (E) (Nash et al., 1970), and coefficient of correlation (R).

$$NRMSE = \frac{\sqrt{\frac{1}{N}\sum(XO - XE)^2}}{\frac{1}{N}\sum XO}$$

$$E = 1 - \frac{\sum (XO - XE)^2}{\sum (XO - \overline{XO})^2}$$

$$R = \frac{\sum (XO - \overline{XO}) \times (XE - \overline{XE})}{\sqrt{\sum (XO - \overline{XO})^2 \sum (XE - \overline{XE})^2}}$$

Where XO is the observed value of the variable, XE is the estimated value of the variable from a model, $\overline{XO}$ is the average value of the variable, $\overline{XE}$ is the average estimated value of the variable, N is the total number of data points predicted, and all the summations run from 1 to N .

5. RESULTS AND DISCUSSIONS

The results from the conceptual and ANN models for the modelling of change in soil moisture and percolation in terms of various performance statistics are presented in table 3 and 4, respectively.

The data used for ANN modelling was divided into training, validation and testing data sets consisting of 3000, 750, 750 daily values respectively. The same data set of 3000 daily values as was used for the training of the ANN models was used for driving the genetic algorithm and the conceptual physically oriented soil moisture model.

Analysing the performance of the models from Table 2 for the modelling of change in soil moisture, it can be observed that all the models performed quite well in terms of various performance criteria. All models achieved an R value in access of 0.90 and E value in excess of 0.80. It can be inferred from the results that the ANN model performed ? best in terms of all error criteria for Lysimeter 2 and Lysimeter 3. While for Lysimeter 1 the GA2 model performed best. For the ANN model the optimal number of neurons in the hidden layer was found to be 25 after several trials. It was observed that the performance of the conceptual model GA1 with SSE of change in soil moisture as the objective function in the GA optimization led to better results than when the SSE of percolation was used as the objective function to be minimized in GA2. However, exceptionally it was observed that for the lysimeter 1 the performance of GA2 was better than that of GA1.

Comparing the performance of different modelling approaches from Table 3 for the modelling of the percolation, it can be noted that the performance of all methods were satisfactory although not as good as was for the modelling of soil moisture. It was observed that the ANN model performed best for all lysimeters for the prediction of percolation. For Lysimeter 1 and Lysmeter 2 an R value of 0.923 and 0.917 respectively was achieved while an E value of 0.8852 and 0.8414 was achieved. The models performed very well in terms of NRMSE also. For Lysimeter 3 a slightly lower value of performance statistics was achieved. 5-25-1 ANN network with 25 neurons in the hidden layer was found to perform best after making several trial runs varying the number of neurons in the hidden layer from 1-25. As expected the conceptual model GA2 with SSE of percolation as the objective function performed better than GA1 with SSE of change in soil moisture as objective function for lysimeter 1. However, exceptionally the performance of the GA2 model for lysimeters 2 and 3 were found to be poor.

Table 2: Performance Statistics from different models for the modelling of change in soil moisture for the three lysimeters

	Lysimeter 1			Lysimeter 2			Lysimeter 3		
Model	**NRMSE**	**E**	**R**	**NRMSE**	**E**	**R**	**NRMSE**	**E**	**R**
GA1	21.540	0.913	0.957	35.372	0.841	0.917	35.965	0.767	0.876
GA2	16.393	0.950	0.980	42.524	0.770	0.881	40.688	0.702	0.843
ANN1	-910.345	0.837	0.915	-48.050	0.922	0.964	0.657	0.901	0.968

Table 3: Performance Statistics from different models for the modelling of percolation for the three lysimeters

	Lysimeter 1			Lysimeter 2			Lysimeter 3		
Model	**NRMSE**	**E**	**R**	**NRMSE**	**E**	**R**	**NRMSE**	**E**	**R**
GA1	0.781	0.489	0.769	1.189	0.374	0.629	1.582	0.216	0.516
GA2	0.451	0.830	0.942	1.445	0.075	0.525	1.804	-0.020	0.468
ANN2	0.498	0.852	0.923	0.743	0.841	0.917	1.004	0.601	0.773

Table 4a: Performance Statistics from conceptual model using best estimated parameters for modelling change in soil moisture

Lysimeter 1			Lysimeter 2			Lysimeter 3		
NRMSE	**E**	**R**	**NRMSE**	**E**	**R**	**NRMSE**	**E**	**R**
139.680	0.371	0.739	104.035	0.728	0.865	102.643	0.758	0.872

Table 4b: Performance Statistics from conceptual model using best estimated parameters for modelling percolation

Lysimeter 1			Lysimeter 2			Lysimeter 3		
NRMSE	**E**	**R**	**NRMSE**	**E**	**R**	**NRMSE**	**E**	**R**
2.652	-3.940	0.129	1.789	-0.172	0.226	.902	0.034	0.297

All the modelling approaches have extensive data requirements for the purpose of calibration, which in practice is not always easily available. So, the conceptual model was used to model the lysimeters without optimizing the parameters based on the available data. The best possible estimates of the parameters based on the experience with the physical processes involved and lysimeters to be modelled were provided to the conceptual physically based model. Parameter values were taken from the German guidelines of soil classification (Ostrowski et al., 2006). The results are provided in Tables 4a and 4b. It was observed that performance of the conceptual physically based model was good for Lysimeters 2 and 3 for which an fairly good values of E as 0.728 and 0.758 respectively was achieved. The performance was also good in the terms of coefficient of correlation as well as a value 0.865 and 0.872 were achieved for the Lysimeters 2 and 3 respectively. The performance of the model for Lysimeter 1 was not satisfactory in terms of E but a fair value of R as 0.739 was achieved. However, the conceptual physically based model provided with the best pre-estimates of the parameters was unable to perform well for the prediction of percolation from all lysimeters.

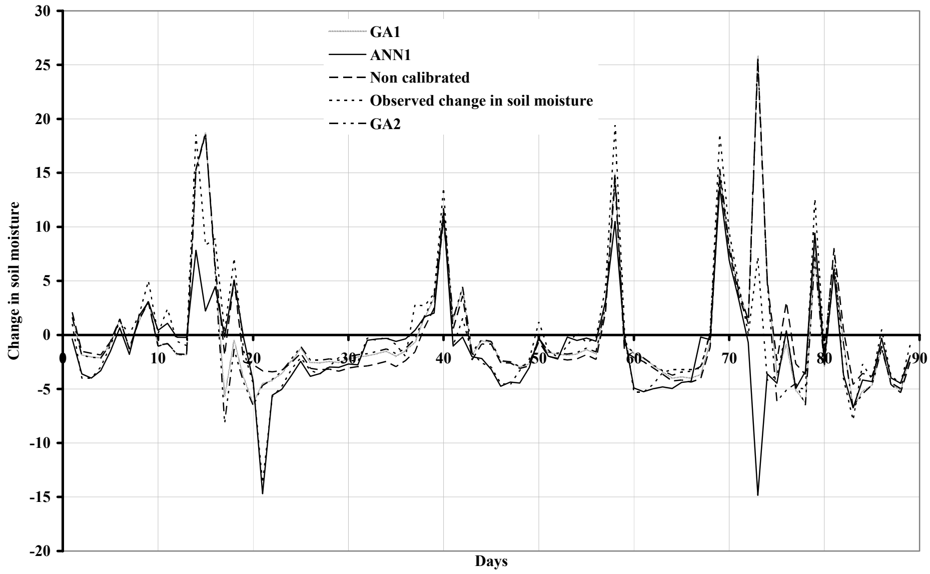

Fig.1: Observed vs calculated change in soil moisture for the various models

6. CONCLUSIONS

This paper presents a comparison of various methodologies for the purpose of soil moisture modelling. A conceptual physically based two layered model was employed for the purpose of modelling of lysimeters driven by automatic parameter estimation using real coded genetic algorithms. Alternatively, a feed forward three layered ANN was employed for modelling of soil moisture and percolation from the lysimeters.

The results show the superiority of ANN modelling technique over the conceptual physically based modelling in most cases. However, the application of ANN technique for modelling of soil moisture is associated with large data requirements which in many practical cases might not be given. The results of conceptual physically based model using best pre-estimated data show that the conceptual physically based model can be used with a limited but acceptable accuracy in conditions where sufficient data for modelling is not available. With best estimates based from an experienced hydrologist conceptual physically based models would prove to be more accurate and useful in scarce data situations.

The modelling results for certain cases in the study have shown unexpected results. The reasons for them need to be further explored. More different and combined objective functions to optimize the parameters of the conceptual physically based models need to be investigated. The methods have been tested on data from three lysimeters. However, the results need to be tested on more data sets. It is hoped that future research efforts will focus on some of these directions.

REFERENCES

[1] Jain A., Sudheer K.P., Srinivasulu S., "Identification of physical processes inherent in artificial neural network rainfall runoff models", Hydrological Processes, Vol.18, (2004), p571-581

[2] Jain A., Srinivasulu S.,"Integrated approach to model decomposed flow hydrograph using artificial neural network and conceptual techniques", Journal of Hydrology, Vol. 317, (2006), p291-306

[3] Jhorar R.K., van Dam J.C., Bastiaanssen W.G.M., Feddes R.A., "Calibration of effective soil hydraulic parameters of heterogeneous soil profiles", Journal of Hydrology, Vol. 285, (2004), p233-247

[4] Kanpur Genetic Algorithms Laboratory, www.iitk.ac.in/kangal

[5] Khu S.T., Madsen H., "Multiobjective calibration with Pareto preference ordering: An application to rainfall-runoff model calibration", Water Resources Research, Vol. 41,W03004, (2005)

[6] Madsen H., "Automatic Calibration of a conceptual rainfall-runoff model using multiple objectives", Journal of Hydrology, Vol.235, (2000), p276-288

[7] Nash J.E., Sutcliffe J.V., "River forecasting through conceptual models Part 1-A Discussion of Principles", Journal of Hydrology, Vol.10, (1970), p282-290

[8] Nemse A., Wösten J.H.M., Lilly A., "Development of Soil Hydraulic Pedotransfer Functions on a European scale: Their Usefulness in the Assessment of Soil Quality", 10th International Soil Conservation Organization Meeting, Purdue University, 1999

[9] Ostrowski M.W., "The effect of data accuracy on the results of soil moisture modelling", "IAHS Publication 204", Hydrological Interactions between atmosphere, Soil and Vegetation, Vienna, (1991)

[10] Ostrowski, M.W., Klawitter A., "On the estimation of soil moisture parameters in physically based hydrological models", "7th International Conference on Hydrinformatics", (2006), Acropolos-Nice, France

[11] Sorooshian S., Gupta V.K., "Automatic Calibration of Conceptual Rainfall-Runoff Models: The question of Parameter Observability and Uniqueness", Water Resources Research, Vol. 19, No.1,(1983), p260-283

[12] Yapo P.O., Gupta H.V., Sorooshian S., "Automatic calibration of conceptual rainfall-runoff models: sensitivity to calibration data", Journal of Hydrology, Vol. 181, (1996), p23-48

[13] Yu P-S., Yang T-C., "Fuzzy multi-objective function for rainfall-runoff model calibration", Journal of Hydrology, Vol.238, (2000), p1-14

[14] Zurada M.J., "An Introduction to Artificial Neural Systems", Jayco Publishing House, Mumbai, India in arrangement with West Publishing Company, St Paul, Minnesota (1997)