한국환경과학회 봄 학술발표회지 제16권(제1호), 199~202, 2007
Proceedings of the Korean Environmental Sciences Society Conference, May, 18~19, 2007

OE4)

Uncertainty Reduction of the Nonlinear Time Series Model

Sung-Won Kim
Department of Civil and Railroad Engineering, Dongyang University

1. Introduction

The uncertainty, based on the neural networks theory, can be ascribed to not only the modeling process but also the limited data used for the training performance of neural networks model. Kim and Cho (2003) performed the uncertainty analysis for predicting the flood stage by using the neural networks model with time-delayed patterns in small river basin, South Korea. And, Kim and Kim (2006) carried out the reliability analysis of flood stage forecasting by the uncertainty analysis of input data information using Elman Discrete Recurrent Neural Networks Model (EDRNNM). In this study, we apply the neural networks model, which has the advantage in the substantial approach without including arbitrary physical backgrounds, but includes some uncertainty for the architectural aspect and input variables of neural networks model. Also, if we choose the appropriate architecture of neural networks model, including the number of hidden layer and hidden layer nodes, and apply such architecture with paying attention to the performance of training and testing, the uncertainty related to the architecture arrangement of neural networks model can be treated as a trivial problem and only the uncertainty of input variables can be emphasized.

The objective of this study is to execute the uncertainty analysis for the input layer nodes of the original COMBINE-GRNNM-GA (Type-1) and eliminate the least influencing input layer node using simulation-based method, and then set up the optimal COMBINE-GRNNM-GA (Type-1) for each 14 major meteorological station respectively. Also, we present the statistical model to simply induce the optimal correlation relationship between the measured pan evaporation data and the estimated alfalfa reference evapotranspiration data for each 14 major meteorological station and construct the two-dimensional map for the pan evaporation and the alfalfa reference evapotranspiration and then provide the reference data for drought analysis and the construction of irrigation and drainage networks system in South Korea.

2. The optimal COMBINE-GRNNM-GA

2.1. Construction of the optimal COMBINE-GRNNM-GA

The COMBINE-GRNNM-GA (Type-1), which was developed from Kim and Kim (2007), is composed of nine input layer nodes and two output layer nodes. And, because the genetic algorithm (GA) is embedded during the training performance, the optimal smoothing factors and the overall smoothing factors for each input node are decided during the training performance. From the training performance, the closer the smoothing factor of input layer node is to zero, the less the underlying input layer node influences model results, that is, the greater the uncertainty of input layer node is (Neuroshell 2, 1993). In this study, we examined the uncertainty analysis for input layer nodes using simulation-based method. First, we eliminates the underlying input layer node which represents the lowest smoothing factor calculated from the training performance of the original COMBINE-GRNNM-GA (Type-1). And, we reconstructs the original COMBINE-GRNNM-GA (Type-1) into the modified COMBINE-GRNNM-GA (Type-1) with 8 input layer nodes and then retrains the modified COMBINE-GRNNM-GA (Type-1) with 8 input layer nodes carefully. After comparing the results of statistical analysis calculated from this stage with the those of statistical analysis calculated from Kim and Kim (2007), if it is within the range of error tolerance, we eliminates the input layer node again which represents the lowest smoothing factor calculated from the training performance of the modified COMBINE-GRNNM-GA (Type-1) with 8 input layer nodes.

2.2. Analysis of the optimal COMBINE-GRNNM-GA

According to the analysis results of the selected optimal COMBINE-GRNNM-GA (Type-1) by the uncertainty analysis of input layer nodes in this study, there are many meteorological stations which show the slightly better results compared with those of the original COMBINE-GRNNM-GA (Type-1). There are, however, only a few meteorological stations which don't show the better results compared with those of the original COMBINE-GRNNM-GA (Type-1). According to the overall analysis results of 14 major meteorological stations, the maximum temperature and the sunshine duration were regarded as the necessary climatic variables to estimate the pan evaporation and the alfalfa reference evapotranspiration for all of the 14 major meteorological stations. Also, the mean relative humidity and the mean wind speed were regarded as the necessary climatic variables for many meteorological stations except for two and three meteorological stations. But, the mean dew point temperature and the minimum relative humidity were regarded as the unnecessary climatic variables for many meteorological stations in this study. So, the optimal COMBINE-GRNNM-GA (Type-1) could be constructed for each 14 major meteorological station using the uncertainty analysis of input layer nodes, and the necessary climatic variables showed the slight difference for each 14 major meteorological station respectively.

3. Homogeneity and linear model construction

3.1. Homogeneity test of the reproduced pan evaporation data

We executed the homogeneity test, which examined whether the pan evaporation data, calculated by the original and optimal COMBINE-GRNNM-GA (Type-1), were sampled from the same population compared with the measured pan evaporation data for each 14 major meteorological station respectively. The homogeneity test carried out the one-way analysis of variance (ANOVA) on the mean and variance (McCuen, 1993; Salas et al., 2001) of each pan evaporation data for 14 major meteorological stations.

3.2. Construction of bivariate linear regression analysis model

In this study, simply to estimate the alfalfa reference evapotranspiration using the total measured pan evaporation data and compare such estimated results, we adopted the bivariate linear regression analysis model (BLRAM) as the conventional and universal statistical model. BLRAM is the conventional and universal model, which can simply estimate the approximated alfalfa reference evapotranspiration by using the measured pan evaporation data for each 14 major meteorological station.

4. Map construction

We constructed the optimal COMBINE-GRNNM-GA (Type-1) in order to estimate the pan evaporation and the alfalfa reference evapotranspiration for 14 major meteorological stations in South Korea. So, by using the results of study, it is possible to construct the annual pan evaporation map and its corresponding alfalfa reference evapotranspiration map for the ungaged and unmeasured time series such as 1991 and 1992 year in South Korea.

5. Conclusion and suggestions

In this study, we constructed the optimal COMBINE-GRNNM-GA (Type-1) by using the uncertainty analysis of input layer nodes for the original COMBINE-GRNNM-GA (Type-1), the neural networks model embedding the genetic algorithm, which had been developed in order to estimate the pan evaporation and the alfalfa reference evapotranspiration at the same time. By using simulation-based method to eliminate the input layer node with the lowest smoothing factor, among the available climatic variables, the maximum temperature and the sunshine duration were regarded as the necessary climatic variables in order to construct optimal COMBINE-GRNNM-GA (Type-1) to estimate the pan evaporation and the alfalfa reference evapotranspiration for all of the 14 major meteorological station. Also, the mean relative humidity and the mean wind speed were regarded as the necessary climatic variables for many meteorological stations except for two and three meteorological stations. But, the mean dew point temperature and the minimum relative humidity were regarded as the unnecessary climatic variables for many meteorological stations in this study. The elimination of these unnecessary climatic variables can reduce the operational cost in the economic aspect.

References

Kim, S., Cho, J.S.(2003), Uncertainty Analysis of Flood Stage Forecasting using Time-Delayed Patterns in the Small Catchment. International Symposium on Disaster Mitigation and Basin-Wide Water Management Niigata 2003. IAHR/AIRH, Niigata, Japan, 465-474.

Kim, S., Kim, H.S.(2006), Uncertainty reduction of the flood stage forecasting using neural networks model. Submitted to Journal of American Water Resources Association.

Kim, S., Kim, H.S.(2007), genetic algorithm application of the neural networks model training for time series analysis. Submitted to 32nd Congress of IAHR, Venice, Italy.

McCuen, R.H.(1993), Microcomputer applications in statistical hydrology, Prentice Hall, NJ, USA.

Neuroshell 2 (1993), Ward systems group, Inc., MD, USA.

Salas, J.D., Smith, R.A., Tabios III, G.O., Heo, J.H.(2001), Statistical computing techniques in water resources and environmental engineering. Unpublished book in CE 622, Colorado State University, Fort Collins, CO, USA.