ニューラルネットワークによるGPVデータを用いた筑後川流域の降水量予測
RAINFALL ESTIMATION IN CHIKUGO RIVER BASIN BY ATMOSPHERIC DOWNSCALING USING NEURAL NETWORKS

Izumi Ishikawa Student Member, Department of Civil Engineering, Kyushu University
Jonas Olsson Department of Water Resources Engineering, Lund University
Akira Kawamura Member, Institute of Environmental Systems, Kyushu University
Kenji Jinno Member, Institute of Environmental Systems, Kyushu University

1. Introduction

A relatively simple and efficient way to estimate local and regional rainfall, as well as other hydrometeorological variables, is by establishing statistical relationships with large-scale atmospheric variables. This is known as statistical atmospheric downscaling. In the present study, such downscaling is performed for the Chikugo River basin (~3000 km^2) located in the northern part of Kyushu Island. This basin is important for the water supply of Fukuoka City, which sometimes experience severe droughts. For proper management of the water resources during droughts, accurate rainfall prediction is important. In the downscaling procedure, three large-scale meteorological variables are used as input: precipitable water, and zonal and meridional wind speeds. Output is the mean rainfall intensity in Chikugo River basin during a 12-hour period. To find the optimum input-output relationships, artificial neural networks are employed. The outputs are compared with the mean observed Chikugo River basin rainfall data.

2. Experimental setup

2.1 Data.

The downscaling output, i.e., mean 12-hour rainfall intensity in Chikugo River basin (CRb), was obtained as the average of the intensity measured in 11 stations (so-called AMeDAS network) located within the basin. The large-scale atmospheric state was specified by grid point meteorological data (GPV) at 00Z and 12Z (9am. and 9pm. JST) in a region spanning approximately 105-160°E and 20-55°N (Fig.1), provided by Japan Meteorological Agency. Three variables were considered for the downscaling, as these have been previously found most efficient for rainfall prediction (Uvo et al.[1]): (a) precipitable water (vertically integrated humidity), and (b) zonal and (c) meridional wind speed at 850 hPa. These variables, specified at a 100×100 km resolution over the area (43x51 points), were correlated with the mean rainfall in CRb. The mean values over areas correlated with 95% statistical significance (Fig. 1) were used as input in the downscaling procedure. Data from four summer seasons (Jun-Aug 1996-1999; 717 values) were used.

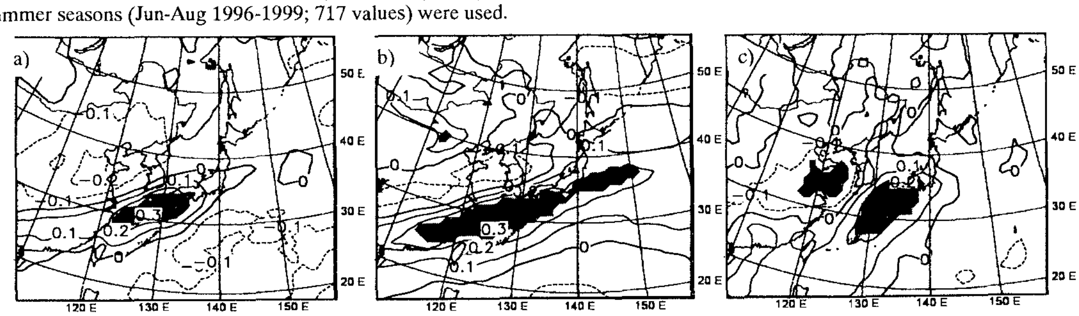

Fig.1 Correlation fields representing the covariance between mean rainfall in Chikugo River basin and GPV variables: a) precipitable water, b)zonal, and c) meridional wind speed (at 850 hPa). Shading denotes 95% statistical significance.

2.2 Methodology -Artificial neural networks-

An artificial neural network (ANN) is a flexible mathematical tool for identifying and utilizing complex and nonlinear relationships between data sets. An ANN model is usually made up of a number of layers of processing elements (neurons) with multiple connections between the elements of each layer. Information enters through the input layer of theANN, is passed through hidden layers which have weighted connections, and is transformed by means of particular transfer functions. The response of the ANN for a particular input is produced in the output layer.

In the present experiments, feedforward ANNs were trained by a backpropagation algorithm for the determination of ANN parameters (weights and biases). The ANNs consisted of an input layer with three neurons representing the three meteorological variables described in Section 2.1, an output layer with one neuron representing the mean rainfall in CRb, and in between hidden layers whose number and sizes were optimized (see below). To every neuron was allocated a log-sigmoid transfer function having a continuous output confined between zero and one.

We used the first 75% of the data (1996-1998) for ANN calibration and the last 25% (1999) for independent validation. A common problem when calibrating ANNs is so-called overfitting, which means that the ANN adjusts to noise in the data and becomes poor at generalizing to other data sets. To avoid overfitting, the calibration set was divided into one training set (80%) and one testing set (20%). During calibration, (1) the training set was used to repeatedly adjust the values of weights and biases so the output of the ANN becomes as close as possible to the target values, i.e., the actual mean CRb rainfall during the training period, and (2) ANN performance for the testing set was continuously checked, and when this performance started to decrease training was stopped. Performance was mainly assessed by comparing ANN output and actual target in terms of the correlation coefficient (cc) and root mean square error (rmse).

Due to the frequent presence of dry periods, i.e., 12-hour periods during which it did not rain, using the original rainfall time series as ANN target creates two main problems: (1) the ANN usually produces a small rainfall during actual dry periods, leading to an underestimated probability of dry periods, P(0), and (2) the ANN generally underestimates the most intense peaks, which are few and thus difficult to make the ANN fully reproduce. Attempting to overcome these problems, we use two ANNs in series. The first (ANN1) is trained to determine whether it will rain or not, and the second (ANN2) to determine the intensity of rainy periods. The idea behind this approach is that ANN1 will produce an accurate value of P(0), and ANN2 will have a better chance to reproduce high intensities.

Another usual difficulty when using ANNs is output variability. Particularly when using ANNs with only one or a few hidden layers and a small number of nodes in each layer, training becomes very sensitive to the initial values of the ANN weights and biases (which are essentially randomly assigned). Consequently, some trained ANNs will perform better than others, and a screening procedure is required to select the best ones. For this purpose we calibrated the ANN 25 times, each time with different initial parameters. For each calibrated ANN, performance for the whole calibration period was assessed, and the five ANNs with the highest performance were selected. The final ANN output for the validation period was the mean values of the validation period output from the selected ANNs. A critical issue in ANN applications is to optimize the number of hidden layers and the number of neurons in each layer. If too many layers and neurons, the ANN may easily overfit; if too few, the ANN may not be able to reproduce the full variability in the data. We tested using both one and two hidden layers with up to eight neurons in each layer. The best result was obtained using one hidden layer with two neurons in ANN1 and one hidden layer with four neurons in ANN2.

2.3 Results

Table 1 indicates a typical example of ANN1 output for the validation periods. The outputs of ANN1 are 1 or 0. These indicate that it will rain or not. The probability of the hit number of the validation periods was 0.77. Table 2 indicates the performance of the ANN2(only rain periods). Fig.2 shows the time series of output from the two ANN combined model (ANN1+ANN2 model) for validation period, compared with the observed mean CRb rainfall. From these results, ANN1+ANN2 can well identify when rainfall occurs or not, but is not able to fully reproduce the variability of rainfall amounts. The model tends to often generate a similar rainfall amount, less than 20 mm, which means that large observed amounts become underestimated and small amounts overestimated. This is probably related to the small number of layers and neurons that had to be used, more data would be required to use a larger ANN that could better reproduce the rainfall variability. Table 3 indicates the performance of the ANN1+ANN2.

Table 1 Output from ANN1, compared with the observed mean CRb, rainfall (rain or not rain). Hit number indicates the count of CRb= 0 and ANN1= 0 (70), as well as, the count of CRb= 1 and ANN1= 1 (57). Hit rate indicates the ratio of the hit number between CRb and ANN1. (0.77 indicates the probability of the hit number of the whole validation periods.)

Validation	CRb	ANN1	hit number	hit rate
Count of 0	94	85	70	0.75
Count of 1	71	80	57	0.80
Sum of data points	165	165	127	0.77

Table 2 The typical performance for ANN2, respectively, in terms of correlation coefficient (cc), root mean square error (rmse) of rain periods.

Validation	cc	rmse
ANN2	0.55	18.1

Table 3 Average performance for the ANN1+ANN2, respectively, in terms of correlation coefficient (cc), root mean square error (rmse)and probability of dry periods (P(0); the observed P(0) is 0.57) of whole periods.

Validation	cc	rmse	P(0)
ANN1+ANN2	0.57	12.9	0.51

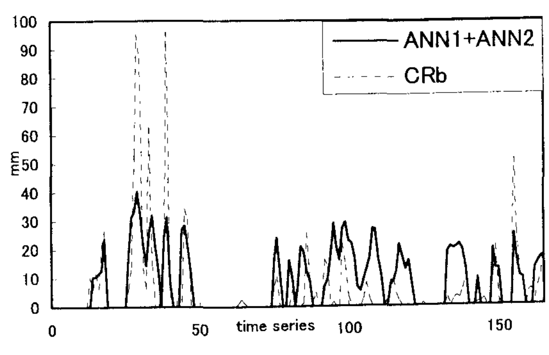

Fig.2 Time series of output from ANN1+ANN2 model, compared with the observed mean Chikugo River basin rainfall.

3. Conclusions

On the basis of the present experiments, ANN-based statistical atmospheric downscaling appears to be a useful to rainfall prediction by numerical weather prediction. The ANN-based statistical method is relatively simple, but has the output accuracy comparable to the more advanced physically based model. However, it is clear that the design, training, and application of the ANNs are very important for the method to be successful. For further studies, need to establish the applicability of the approach, for example, for other seasons and other geographical locations.

4.Reference

[1]Uvo, C.B., Olsson, J., Morita, O., Jinno, K., Kawamura A., Nishiyama, K., Koreeda, N., and Nakashima, T., Statistical Atmospheric Downscaling Estimation In Kyushu Island, Japan, Hydrol. Earth System Sci., in press, 2001.