九州大学学術情報リポジトリ
Kyushu University Institutional Repository

Rainfall Estimation in the Chikugo River Basin by Atmospheric Downscaling Using Artificial Neural Networks

Ishikawa, Izumi
Department of Urban and Environmental Engireering : Graduate Student

Olsson, Jonas
Swedish Meteorological and Hydrological Institute : Researcher

Jinno, Kenji
Institute of Environmental Systems : Professor

Kawamura, Akira
Institute of Environmental Systems : Associate Professor

他

https://hdl.handle.net/2324/1113

出版情報：九州大学工学紀要. 62 (2), pp.85-95, 2002-06-26. 九州大学大学院工学研究院
バージョン：
権利関係：

Memoirs of the Faculty of Engineering, Kyushu University, Vol. 62, No. 2, June 2002

Rainfall Estimation in the Chikugo River Basin by Atmospheric Downscaling Using Artificial Neural Networks

by

Izumi ISHIKAWA*, Jonas OLSSON**, Kenji JINNO***, Akira KAWAMURA****
Koji NISHIYAMA***** and Ronny BERNDTSSON******

(Received March 25, 2002)

Abstract

For the proper water resources management of the Chikugo River basin, the prediction of both drought and heavy rainfall needs to be carried out by the conventional and engineering method which can be useful to for the practitioners who work on the water resources management and flood control. A relatively simple and efficient way to estimate local and regional rainfall, as well as other hydrometeorological variables, is now intensively discussed. This method utilizes the grid data point value (GPV) to predict the regional rainfall based on the so called atmospheric downscaling. In this paper, artificial neural networks (ANNs) are employed. As the input variables, three large-scale meteorological variables, precipitable water, and zonal and meridional wind speeds, are used. Output is the mean rainfall intensity in the Chikugo River basin during a 12-hour period. In the model, the serially combined ANNs were employed to predict the rainfall amount exactly. The result from the serially combined ANNs is slightly better than the result from the neumerical weather prediction model of the Japan Meteorological Agency by comparing the values of CC and RMSE.

Keywords: Atmospheric downscaling, GPV data, precipitable water, wind speeds, Artificial neural network, correlation analysis

1. Introduction

In pace with increasing demands on water resources at the basin scale demands are also increasing on the predictive techniques for runoff estimation. To increase the forecast lead time for runoff, modelers are more and more forced to use precipitation estimates from large-scale atmospheric data bases. The problem remains to best downscale these data to

*Graduate Student, Department of Urban and Environmental Engineering
**Researcher, Swedish Meteorological and Hydrological Institute
***Professor, Institute of Environmental Systems
****Associate Professor, Institute of Environmental Systems
*****Research Associate, Institute of Environmental Systems
******Professor, Department of Water Resources Engineering, Lund University

actual rainfall at basin scale to be used as input to runoff models.

A way to estimate local and regional rainfall is to use grids of meteorological data point value (GPV) as input to a downscaling procedure. These GPVs cover the entire atmosphere with 20 km grid meshes in the horizontal and 36 layers in the vertical. Among the variables that can be used are wind speeds, temperature, dew point depression, and pressure.

At present, however, there exists no optimal method to downscale large-scale meteorological data to basin-scale rainfall. Several studies have been made using artificial neural networks (ANN) to arrive at actual rainfall. The advantage of using ANN is that it is simple and can be updated repeatedly to quantify even complicated nonlinear relationships.

In the present paper we use GPV data from the Japan Meteorological Agency to downscale mean rainfall during 12-hour periods for the Chikugo River basin in south Japan by use of ANN. We compare the results with a physical meteorological model for rainfall prediction. Finally, we close with a discussion on the practical implications of the results.

2. Methods and experimental area

2.1 Artificial neural network

An artificial neural network (ANN) is a flexible mathematical tool for identifying and utilizing complex and nonlinear relationships between input and output 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 entering through the input layer of the ANN, is passed through hidden layers which have weighted connections, and is transformed by means of particular transfer functions.

In this study, the ANNs used consised of three layers, an input layer with three neurons representing meteorological variables described below, a hidden layer, and an output layer with one neuron representing actual rainfall of the Chikugo River basin. An example the ANN model is shown in **Figure 1**. In all experiments, feedforward ANNs were trained by a backpropagation algorithm for the determination of weights and biases[1]. In the ANNs, every neuron was allocated a log-sigmoid transfer function defined as

$$a=\frac{1}{1+e^{-n}} \tag{1}$$

where n is the input value to the neuron and a is the output value from the neuron.

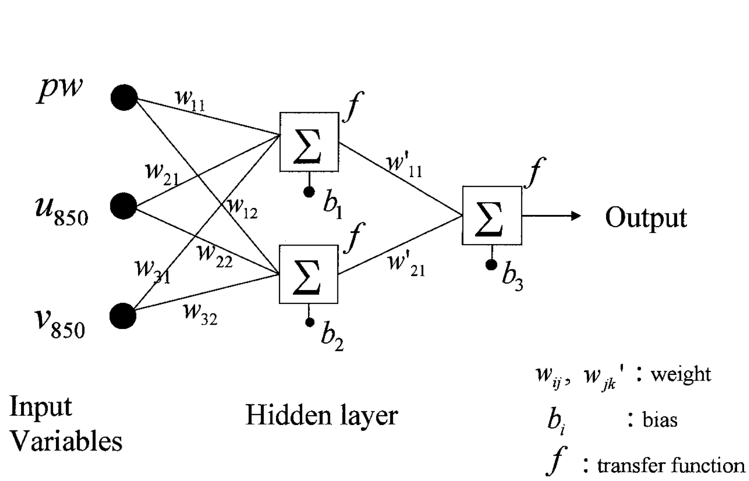


Figure 1 An example ANN model consisting of three meteorological variables

Output from the log-sigmoid function is continuous and confined between zero and one.

A time series of totally 717 30-min rainfall values was divided into three parts for training, testing, and validation of the ANN. Data corresponding to 25% of the entire series, were reserved for validation (validation set) and were not used in any step of the training process. The remaining data were divided in two parts: 80% were used for training the ANN (training set) and the remaining 20% for testing (test set). This arrangement is often used in ANN applications to avoid overfitting the training data, that is, to avoid the ANN reproducing noise present in the data.

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. Tests were made using both one and two hidden layers with up to eight neurons in each layer.

To compare the target data (observed rainfall) with the ANN output (predicted rainfall) different indexes were used. The root mean square error ($RMSE$) was calculated as

$$RMSE=\sqrt{\frac{1}{m}\sum_{j=1}^{m}(T_j-O_j)} \tag{2}$$

where m is the number of data, T_j is the target value, and O_j is the output of ANN. The ability of the model to forecast events with zero rainfall was calculated using $P(0)$ according to

$$P(0)=\frac{N_O(0)}{N_T} \tag{3}$$

where N_T is the number of observed (targeted) zero rainfall (0 mm) events in the target data set and $N_O(0)$ is the number of predicted zero rainfall (0 mm) by the ANN. The total rainfall rate index Trr was used to characterize the sum of predicted vs. targeted rainfall. It is defined as

$$Trr=\frac{\sum O_j}{\sum T_j} \tag{4}$$

Finally, the hit rate (HR) index was used to characterize the ANN's ability to partition between zero and non-zero rainfall. It is defined as

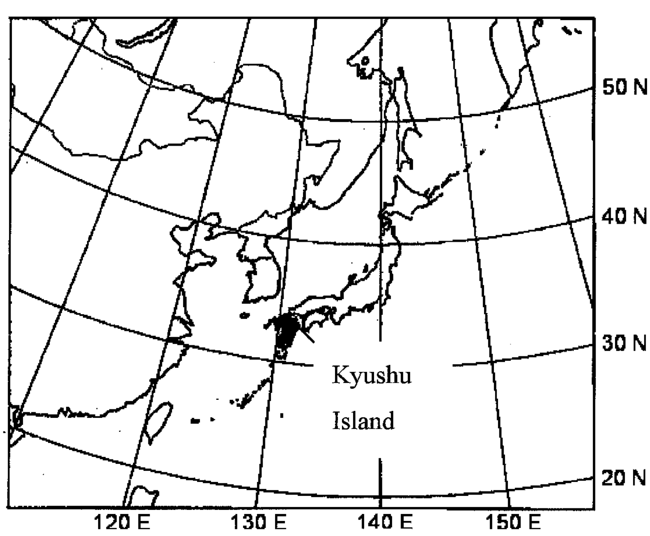


Figure 2 The area in which GPV meteorological databases were available (Kyushu Island is marked in black)

$$HR=\frac{1}{2}\left\{\frac{N_{oc}(0)}{N_t(0)}+\frac{N_{oc}(1)}{N_t(1)}\right\} \tag{5}$$

where $N_{oc}(0)$ and $N_{oc}(1)$ are the number of zero and non-zero rainfall values that the ANN correctly predicted. Similarly, $N_t(0)$ and $N_t(1)$ are the actually observed zero and non-zero values, respectively, in the target series ($N_t=N_t(0)+N_t(1)$).

Two ANN models were used in this study. The first, ANN1, was used to simulate if rainfall occurred or not. Thus, this model gave an integer output corresponding to 0 or 1, indicating zero or non-zero values of rainfall. The second, ANN2, was used to predict the rainfall intensity. Consequently, Eq. (5) was used only for ANN1.

2.2 Experimental data

The large-scale atmospheric properties are characterized by meteorological grid point values (GPV) in a region spanning approximately 105-160° E and 20-55° N, provided by the Japan Meteorological Agency (**Figure 2**). The data consist of sounding measurements (wind speeds, temperature, dew point depression, and pressure) at 00Z and 12Z (9 am and 9pm JST). These GPV data are usually used for giving initial conditions to numerical weather prediction. Using the same data, Uvo et al[2] showed that zonal and meridional wind speeds at 850 hPa (u_{850}, v_{850}) in combination with precipitable water (pw) were useful predictors for rainfall estimation. For this reason, these variables were used as input to the modeling scheme.

The atmospheric downscaling by use of ANN aimed at estimating mean rainfall in the Chikugo River basin (2840 km^2) located in Kyushu, south Japan (**Figure 3a**). The river flows

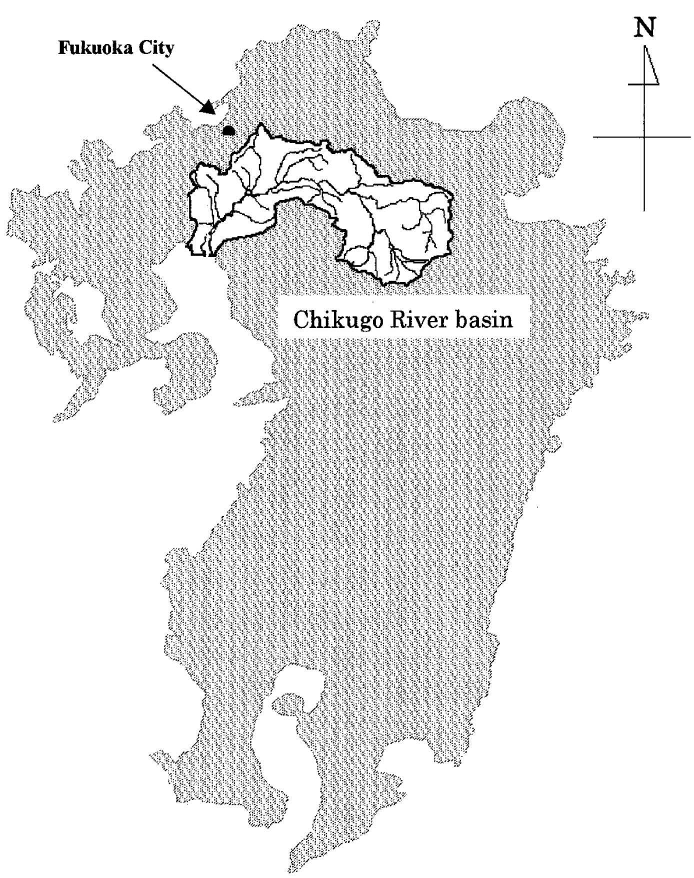


Figure 3 The location of the Chikugo River basin and 11 AMeDAS stations
a) Kyushu Island, Southern Japan, and the location of the Chikugo River basin

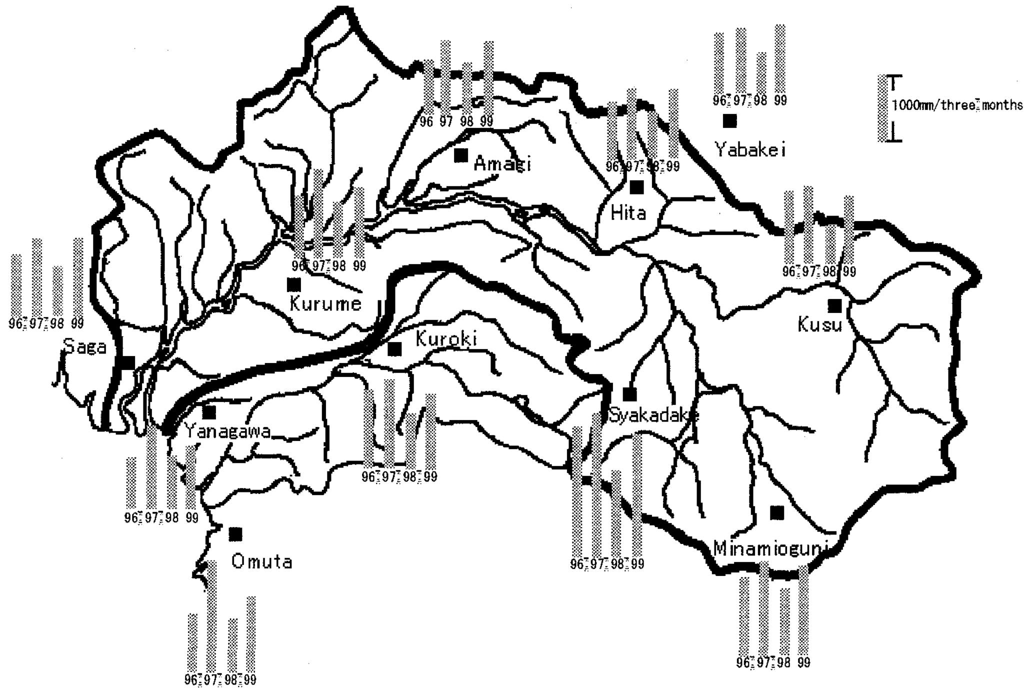


Figure 3 The location of the Chikugo River basin and 11 AMeDAS stations
b) The locations of 11 AMeDAS stations and rainfall for the period from 1996 to 1999

Table 1 The rainfall statistics of 11 AMeDAS stations.

	Amagi	Kurume	Kuroki	Yanagawa	Omuta	Yabakei	Hita	Kusu	Syaka dake	Saga	Minami oguni
1996											
annual rainfall (mm)	1650	1690	1932	1392	1508	1752	1768	1932	–	1596	2266
summer rainfall * (mm)	813	919	1158	755	873	893	854	1082	1970	928	1179
ratio	0.49	0.54	0.60	0.54	0.58	0.51	0.48	0.56	–	0.58	0.52
1997											
annual rainfall (mm)	2323	2593	2663	2351	2985	2305	2138	2435	3345	2352	2817
summer rainfall * (mm)	1117	1330	1335	1256	1676	972	1041	1140	2169	1155	1420
ratio	0.48	0.51	0.50	0.53	0.56	0.42	0.49	0.47	0.65	0.49	0.50
1998											
annual rainfall (mm)	1973	2115	2117	1778	1871	1704	1744	1695	3105	1722	2303
summer rainfall * (mm)	767	811	829	783	785	605	689	557	1298	727	981
ratio	0.39	0.38	0.39	0.44	0.42	0.36	0.40	0.33	0.00	0.42	0.43
1999											
annual rainfall (mm)	1991	1921	2161	1778	2177	1986	1980	1861	3371	2107	2515
summer rainfall * (mm)	1100	1045	1134	940	1118	995	1033	1003	1836	1173	1328
ratio	0.55	0.54	0.52	0.53	0.51	0.50	0.52	0.54	0.54	0.56	0.53
mean rainfall of four years (mm)	1984	2080	2218	1825	2135	1937	1908	1981	3274	1944	2475
mean summer rainfall of four years (mm)	949	1026	1114	934	1113	866	904	946	1818	996	1227

*three-month rainfall from June to August from a volcanic caldera in central Kyushu and plays an important role as fresh water resource in this drought-sensitive region[3]. The main climatological features of the basin, and Kyushu Island in general, are (*i*) the so-called Baiu front, which is responsible for the majority of summer rainfall, (*ii*) the strong circulation pattern associated with fall rainfall, and (*iii*) the strong influence of orographic lifting. The water resources management of the Chikugo River basin is an important subject for fishery, agriculture, industry, and drinking water supply. The water resources management is crucial both within the Chikugo River basin as well as for the Fukuoka Metropolitan area located to the north of the Chikugo River basin because of dependence on water supply.

Mean rainfall in the Chikugo River basin was determined as the arithmetic average of 11 precipitation gauges operated within the Japanese national meteorological network (AMeDAS). These gauges are evenly distributed within or near the catchment area (**Figure 3b**). Measurements were made on an hourly basis, but 12-hour totals were used to compare with the modeling using the GPV data. Each 12-h period started at the time of the GPV soundings, that is, every GPV value was used to estimate the rainfall during the next 12-h period. Data were available between April 1996 and August 1999 ; a total of 2417 12-h periods. Here, the analysis was limited to summer data (June-August), as the majority of extreme events occur during this season (of the 50 largest events, 38 occurred in summer). In total, 717 12-h values were used in this study.

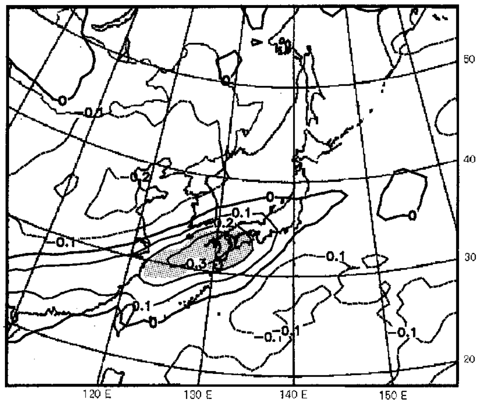
Table 1 shows rainfall properties of the 11 rainfall stations used. From this, we can see that orography is important in the area. The rainfall amount is high in the mountain area (e.g., Syakadake and Minamioguni), on the other hand, in the lower lying areas (e.g., Saga, Kurume, and Yanagawa) the rainfall amount is small. The station which has the highest rainfall is Syakadake at 1200 m altitude. The summer rainfall is almost half the annual rainfall. Consequently, the summer rainfall is very important for water resources availability in the Chikugo River basin.

In order to use efficient predictors as input to the ANN an investigation was made to clarify which meteorological variables over what specified area that were best correlated to Chikugo River basin rainfall. Variables intended as input data to ANN need to be investigated as suggested by Wilby and Wigley[4]. Uvo et al. (2000) calculated the standard correlation between the time series of rainfall in the Chikugo River basin and meteorological variables to determine the influencing variables and area. They found that precipitable water (pw), and zonal and meridional wind speed at 850 hPa, (u_{850} , v_{850}) were the most influencing variables that are specified at a 100×100 km resolution over the area (43×51 points), to the rainfall in the Chikugo River basin.

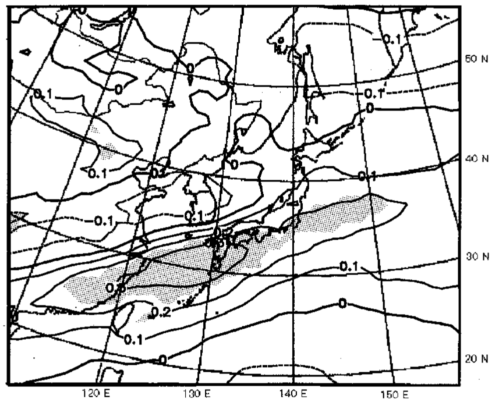
Figure 4 shows the results from the correlation analysis for summer periods from 1996 to 1999. The shaded area in **Fig. 4** indicates areas highly correlated with Chikugo River basin rainfall. These were identified by defining areas in which correlation coefficients were statistically significant at the 5%-level. The correlation coefficients correspond to values greater than 0.25. From a meteorological viewpoint the areas can be characterized as a region where atmospheric instability increases due to intrusion of large amounts of water vapor from the southern ocean forming a stationary Baiu front.

3. ANN prediction of 12-hour rainfall

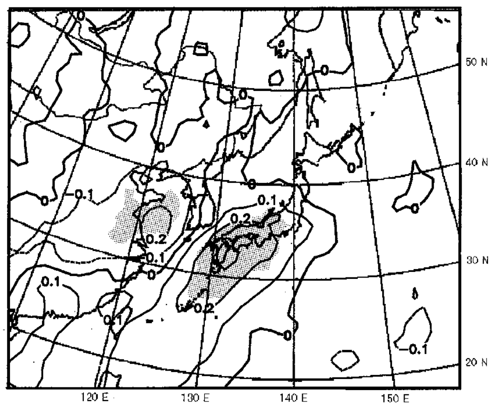
Initial tests showed that frequent dry periods, i.e., 12-hour periods with no rainfall created problems to fully reproduce the intermittency of observed rainfall. Attempting to overcome this problem, we designed a modeling system involving two ANNs coupled in a



a) The correlation between precipitable water (pw) and mean rainfall in the Chikugo River basin



b) The correlation between u_{850} and mean rainfall in the Chikugo River basin



c) The correlation between v_{850} and mean rainfall in the Chikugo River basin

Figure 4 Correlation fields representing the covariation between mean rainfall in the Chikugo River basin and GPV variables. Isolines correspond to the correlation coefficient, the shaded areas denote 95% statistical significance

Table 2 The result of ANN1 *HR* for different number of layers and neurons.

	HR of Calibration	HR of Validation
[3 2 1]	0.76	0.77
[3 3 1]	0.75	0.72
[3 4 1]	0.76	0.75
[3 5 1]	0.76	0.72
[3 8 1]	0.77	0.75
[3 2 2 1]	0.77	0.76
[3 4 2 1]	0.74	0.71
[3 3 3 1]	0.75	0.76

series. The first ANN (ANN1) was used to separate between zero (rain) and non-zero (no rain) values. The second ANN (ANN2) was then used to determine intensities for non-zero values. The ANN1 consequently feed the ANN2 with input data in form of 1 (rain) or 0 (no rain) values. 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.

All input variables were standardized to a mean value of 0 and a standard deviation of 1. This was done to ensure that every input variable would receive equal attention during the training (e.g., Maier and Dandy, 2000)[5].

3.1 Rainfall occurrence, ANN1

According to the above, ANN1 was set to predict whether it will rain or not. For this purpose, each target value (observed rainfall) of the whole period was replaced by 1 or 0, where 1 indicates rainfall more than 1 mm/12-h, and 0 indicates less than 1 mm/12-h. To represent discrete binary targets, output value from ANN1 less than 0.5 was treated as 0, and output larger than 0.5 was treated as 1.

To optimize the number of hidden layers and the number of neurons in each layer, we tested by trial and error using both one and two hidden layers with up to eight neurons in each layer. The root mean square error, *RMSE*, and hit rate, *HR*, were used as indicators to compare the different models. The best result was obtained using an ANN1 which has one hidden layer with two neurons. This model gave the highest *HR* and smallest *RMSE*.

Table 3 Output from ANN1, compared with the AMeDAS data (rain or no rain).

Validation	AMeDAS	ANN1	hit number	HR
Count of 0 (no rain)	95	86	71	0.76
Count of 1 (rain)	70	79	56	0.80
Sum of data points	165	165	127	0.77

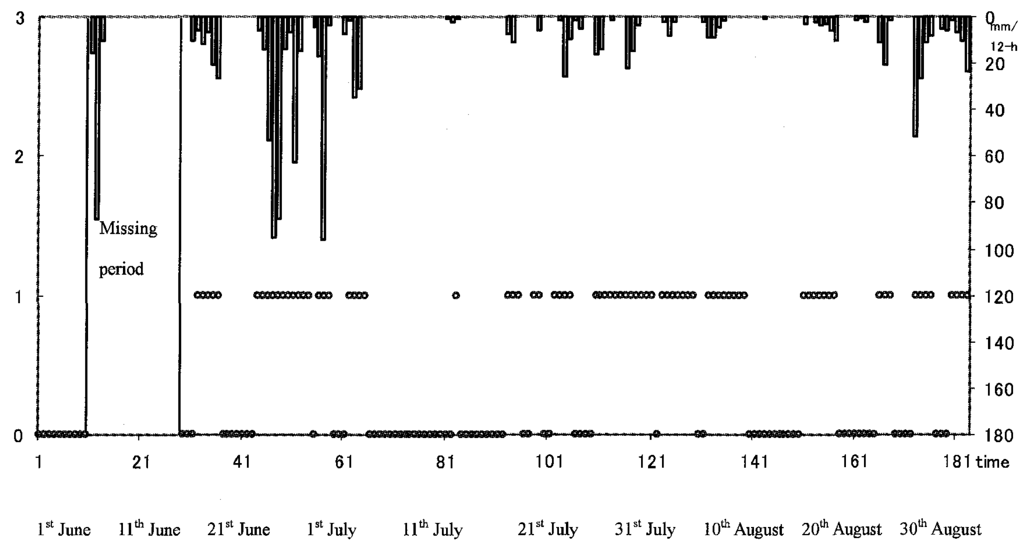


Figure 5 The result of ANN1 *HR* and rainfall time series.

Table 4 The result of 2-ANNs with different number of layers and neurons.

Validation	CC	RMSE	P(0)	Trr	peak ratio
[3 3 1]	0.51	13.2	0.52	1.38	0.34
[3 4 1]	0.55	13.1	0.51	1.40	0.40
[3 5 1]	0.56	13.3	0.52	1.47	0.45
[3 8 1]	0.60	12.4	0.52	1.30	0.43
[3 2 2 1]	0.52	13.8	0.52	1.46	0.39
[3 4 2 1]	0.56	13.2	0.52	1.46	0.40
[3 3 3 1]	0.54	13.7	0.52	1.47	0.44

Table 2 shows the result of these experiments. To summarize, we found that the structure of the ANN is not important because *HR* was between 0.71 and 0.77 for the validation period. Consequently, the most simple ANN which has one hidden layer with two neuron was chosen.

Table 3 shows the results of validation for ANN1 and **Fig. 5** the time series of observed and predicted rain or no rain. The hit number in **Table 3** indicates the count of observed rainfall equal to 0 (AMeDAS=0) and ANN1=0 (no rain), as well as, the count of observed rainfall greater than zero (AMeDAS=1) and ANN1=1 (rain). To summarize, the ANN1 had a *HR* of 0.77 and thus, could consequently rather well forecast rain or no rain for 12-h rainfall.

3.2 Rainfall intensity, ANN2

As a next step, the ANN2 was used to determine the intensity during identified rainy periods. The input variables were first standardized. The observed rainfall data (target values for periods with rain) were rescaled into the range 0.1 to 0.9. This, in order to compare directly with the output from the log-sigmoid transfer functions in ANN2. Also, for the ANN2, initial tests by trial and error were used to determine the best structure of the network model. **Table 4** shows the results for these tests. According to the table an ANN2 with one hidden layer and eight neurons was chosen. This model structure gave the highest correlation coefficient (*CC*) and smallest *RMSE* and total rainfall rate (*Trr*).

Table 5 shows a summary of prediction results for the validated rainfall time series of 1999. As seen from the table *RMSE* was 12.4 mm and a *CC* of 0.06. The predicted $P(0)$ (probability of no rain) was 0.52 while that of the observed time series was 0.57. The total rainfall rate, *Trr*,, was 1.3, about 30% larger than observed (AMeDAS). The peak rainfall was also underestimated by about 65%. **Table 5** also shows a comparison with a meteorological model run by JMA (Japan Meteorological Agency). This model was run to give output at the location of the 11 AMeDAS stations. As seen from the table *RMSE* is larger for the meteorological model. Also, several other prediction indicators show that the ANN can better predict the rainfall rate.

Figure 6 shows time series of observed vs. predicted rainfall by the ANN and the JMA meteorological model. As seen from the figure, the ANN can well identify when rainfall occurs but is not able to fully reproduce the variability of rainfall amounts. The model appears to often generate a similar rainfall rates, less than 40 mm. This means that large observed amounts (more than 40 mm) become underestimated and small amounts (less than

Table 5 Comparison of 2-ANNs with a numerical model output by JMA.

1999 (165 values)	2-ANNs	AMeDAS	JMA
CC	0.60	1	0.52
RMSE (mm)	12.4	0	13.6
P(0)	0.52	0.57	0.68
total rainfall (mm)	1337	1030	448
total rainfall ratio	1.30	1	0.43
peak rainfall (mm)	41.4	95.4	45.1
peak ratio	0.43	1	0.47

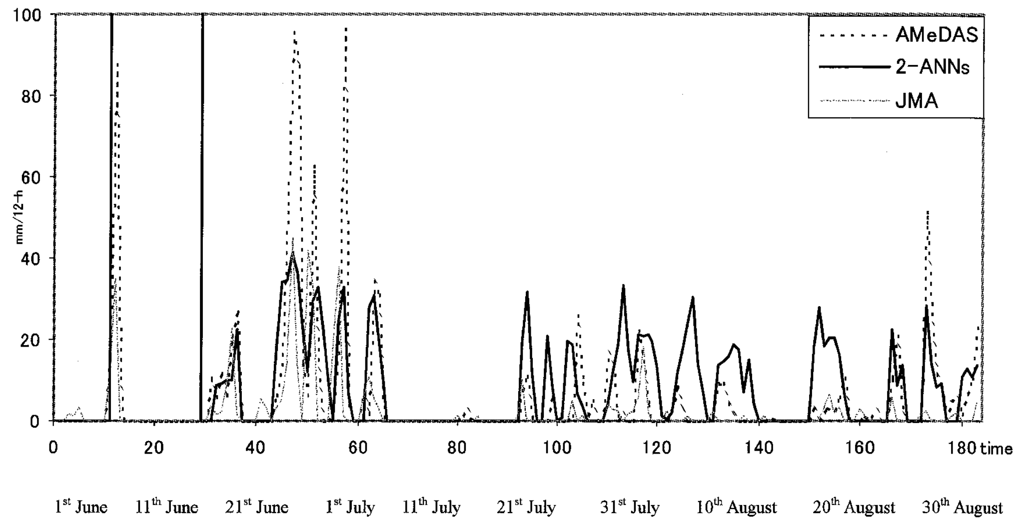


Figure 6 Time series of 2-ANN output, AMeDAS data and JMA output.

40mm) overestimated.

4. Summary and discussion

The comparison according to the above showed that the ANN model generally works slightly better than the meteorological model (JMA), even if the results are rather similar. The JMA generally produces too small rainfall rates as compared to observed values. On the other hand, the ANN, often underestimates high rainfall intensities and overestimates low rainfall intensities. This shows that it is difficult for the ANN to accurately predict the intensity. By comparing $P(0)$, we can say that the two step ANN is better than the JMA model concerning predicting whether it will rain or not. The ANN-based statistical method is relatively simple, but has output accuracy comparable to the more advanced and physically based JME model. However, it is clear that the design, training, and application of the ANN are very important for the method to be successful. Further studies are needed to establish the applicability of the approach, for example, other seasons and other geographical locations.

References

1) Hagan, M.T. and Menhaj, M., Training feedforward networks with the Marquardt algorithm, *IEEE Transactions on Neural Networks, 5*, 989-993, 1994.
Holger R. Maier, Graeme C. Dandy, Neural networks for the prediction and forecasting of water resources variables: a reviem of modeling issues and applications, Environmental Modelling & Software, 15, 101-124, 2000
2) Uvo, C.B., Olsson, J., Morita, O., Jinno, K., Kawamura, A., Nishiyama, K., Koreeda, N., and Nakashima, T., Statistical atmospheric downscaling for rainfall estimation in Kyushu Island, Japan, *Hydrol. Earth System Sci.*, submitted, 2000.
3) Merabtene, T., Jinno, K., Kawamura, A., and Olsson, J., Drought managment of water supply systems: a decision support system approach, *Memoirs Fac. Eng. Kyushu Univ.*, 58, 183-197, 1998.
4) Wilby, R.L. and Wigley, T.M.L., Downscaling general circulation model output: a review of methods and limitations, *Progress in Physical Geography*, *21,* 530-548, 1997.
5) Wilby, R.L., Wigley, T.M.L., Conway, D., Jones, P.D., Hewitson, B.C., Main, J., and Wilks, D.S., Statistical downscaling of general circulation model output: a comparison of methods, *Water Resour. Res.*, 34, 2995-3008, 1998.

Acknowledgement

We want to thank Mr. Yasuhisa Kuzuha and Ms. Michiko Hayano, National Research for Earth Science and Disaster Prevention and Mr. Mamoru Morita, Kyushu University.