Flooding probability of urban area estimated by decision tree and artificial neural networks

Jeng-Chung Chen, Shu-Kuang Ning, Ho-Wen Chen and Ching-Sung Shu

ABSTRACT

Remote sensing, such as from satellite, has been recognized as useful for monitoring the changes in hydrology. In this study, we propose a way that is able to estimate flooding probability based on satellite data from the observation network of the World Meteorological Organization. Through a two-stage probability analysis, we can depict the area with high flooding potential in near-real time. In the first stage, decision trees offered a prompt and rough estimation of the flooding probability; in the second stage, artificial neural networks handle the rainfall forecast in a small-scale area. Case studies, simulating two rainfall events on 20 May 2004 and 11 July 2001, proved that our proposed method is promising for mitigating the flooding damage along urban drainage within the downtown area of Kaohsiung city.

Key words | artificial neural networks, decision trees, flooding probability, satellite images

Jeng-Chung Chen (corresponding author)
Ching-Sung Shu
Department of Environmental Engineering and Science, Fooyin University,
151 Chin-Hsueh Rd. Ta-Liao, Kaohsiung 831,
Chinese Taiwan
E-mail: *jcchen@mail.fy.edu.tw*

Shu-Kuang Ning
Department of Civil and Environmental Engineering,
National University of Kaohsiung,
Chinese Taiwan

Ho-Wen Chen
Department of Environmental Engineering and Management,
Chaoyang University of Technology,
Chinese Taiwan

INTRODUCTION

More and more cities in the world are suffering from flooding risks. In Taiwan, an unusual storm on 11 July 2001 flooded more than 1500 buildings in the city of Kaohsiung and caused five fatalities; another storm on 17 September 2001 severely damaged the city of Taipei and killed 55 people. These flooding events pushed us to fulfill a long-term observation of floods in Kaohsiung city from 2001 to 2004 (Chen *et al.* 2004). In our preliminary study, we found that simulation of floods based on measurements from rain gauges is too slow to meet the demand of flooding control. That is to say, the use of observed precipitation does not allow sufficient lead-time to predict floods and issue warnings (Abebe & Price 2005). To reduce flooding damage and improve urban drainage management, there is a need to increase the forecasting accuracy of rainfall models for use in small-scale urban areas (Kawamura *et al.* 1996).

Since the 1970s, remote sensing has been recognized as a promising technique for monitoring the changes in hydrology (Pietroniro & Leconte 2000). Remote sensing of rainfall can be made using two approaches: ground-based weather radar and satellite imagery. But geographic conditions, e.g. mountains or ocean, will limit the site of the radar. For those cities located at the coast, like Kaohsiung city in Taiwan, satellite imagery seems to be more available than ground-based weather radar for monitoring the movement of clouds over the far ocean because that satellite has good resolution and coverage over space and time (New *et al.* 2001). Nowadays the World Meteorological Organization (WMO) has an observation network composed of geostationary meteorological satellites – METEOSAT (ESA), INSAT (India), GMS (Japan), GOES-E (USA) and GOES-W (USA) – and polar-orbiting meteorological satellites – NOAA (USA) – for monitoring severe weather around the earth. To receive weather data from satellite is not difficult any more. About 10 years ago, the Taiwan Central Weather Bureau (TCWB) started using the imagery of GMS-5 to monitor the movement of typhoons

doi: 10.2166/hydro.2008.009

around Taiwan. Up to June 2003, the job of satellite GMS-5 was temporarily replaced by GOES-9 because the launch vehicle for MTSAT failed. Right now MTSAT-1R is serving the duty of observing atmospheric phenomena around Taiwan. It can provide imagery for the Northern hemisphere every thirty minutes. In this study, we used the satellite data from the IR1 and IR3 channels of GMS-5, GOES-9 and MTSAT-1R. Satellite data must be calibrated for differences in viewing angle of each satellite because the coverage area of GOES-9 is 15° further east than that of GMS-5 and MTSAT-1R.

For the purpose of rainfall forecasting, some statistical characteristics of infrared (IR) data from satellites were applied to distinguish between deep and shallow precipitation (Kurino 1997). Although sensors on weather satellites do not measure hydrological data directly, they are capable of interpreting the changes in hydrology (Pietroniro & Leconte 2000). For example, they can identify the structural properties of convective systems, tropical rainfall distributions and rain cloud locations (Carvalho & Jones 2001; Ba & Gruber 2001). In the interpretation of satellite data, clouds with cold tops in the satellite IR image may produce more rainfall than those with warmer tops (Tarruella & Jorge 2003). Based on the information from satellite infrared and water vapor channels, Feidas (2003) successfully developed an automatic software tool for monitoring convective cloud systems. In the rainfall forecasting study of Wei *et al.* (2006), they proposed a multi-spectral spatial convolution approach. His study incorporated cloud-top temperatures from three infrared channels from weather satellites to be the inputs of the rainfall model. The above studies caused us to adopt satellite data to do rainfall forecasting. We therefore used the data from the IR1 channels of weather satellites to infer cloud-top temperature. Then, using the data from the IR3 channels of weather satellites we monitored the change in water vapor.

Overall there is a need to turn an enormous amount of satellite data into meaningful information. Colquhoun (1987) proposed a decision tree approach to examine which one of the meteorological parameters should be considered for developing thunderstorms, severe thunderstorms and tornadoes. In discerning different types of precipitating systems, Miller & Emery (1997) applied an automated neural network cloud classifier to analyzing Advanced Very High Resolution Radiometer (AVHRR) type imagery. Later Hong *et al.* (2004) applied an artificial neural network cloud classifier to estimating fine scale rainfall distribution. In addition, Baldwin *et al.* (2005) chose hierarchical cluster analysis to classify convective and non-convective features. Using artificial neural networks to catch the relationship between rainfall and satellite data has been proved to be workable by Grimes *et al.* (2003) and Bellerby (2004). Remarkably decision tree, classification, cluster and artificial neural networks mentioned above are common techniques in data mining. Significantly data mining is a promising process to assist us in understanding the complex nature in hydrology (Bankert *et al.* 2004; Babovic 2005). The above studies inspired us to integrate data mining techniques and satellite data for the estimation of flooding probability within the downtown area of Koahsiung city. A two-stage probability analysis, composed of decision trees and artificial neural networks, was proposed to fulfill this job. In this study, processing satellite information for starting artificial neural networks is very time consuming. Satellite information is also very expensive here. So we need decision trees to get a prompt and rough estimation of the flooding probability. It can reduce unnecessary jobs in the subsequent analysis of artificial neural networks. When the flooding probability estimated by DT is high, we start the rainfall forecast within the downtown area of Kaohsiung city via artificial neural networks. Overall our proposed method is promising in mitigating the flooding damage along urban drainage within the downtown area of Kaohsiung city.

METHODOLGY

Decision trees

Decision Trees (DT) are able to form a set of simple rules on classifying attributes among huge amounts of data. It has the advantage of making no assumptions regarding the distribution of the predictor variables. In this study, we used the function *treefit* presented in the Matlab® statistical toolbox to build DT for the rainfall conditions happening in Kaohsiung city. We grew the structure of DT by the CART (Classification and Regression Trees) algorithm that had been proposed by Breiman *et al.* (1984). The CART algorithm is an exhaustive recursive partitioning routine, and it can divide each parent node into two child nodes by posing a series of yes–no

questions. According to the comments of Burrows *et al.* (1994), we know that DT developed by the CART algorithm have a good ability in predicting the maximum surface ozone concentration. His comments encourage us to apply DT on splitting our rainfall event into a suitable scale of rain. During the subsequent analyses, we divided the rain into three scales: fine rain (i.e. total accumulated rainfall is less than 30 mm and the maximum rainfall intensity is less than 10 mm/h), heavy rain (i.e. total accumulated rainfall is in between 30 mm and 100 mm, and the maximum rainfall intensity is greater than 10 mm/h) and torrential rain (i.e. total accumulated rainfall is greater than 100 mm). In our experience, the flooding in Kaohsiung city usually takes place when heavy or torrential rain continues for more than 2 h.

Next we need a database for fitting decision trees. There are 1077 rainfall events happening around Kaohsiung city during the period 1995–2004. Among them several rainfall events have problems with ambiguous rain scale or without complete monitoring data in practice. So we have to remove some unavailable data. In addition, the role of decision trees analysis in this study just provides a rough estimation of the flooding probability. Finally we chose only 100 rainfall events, except for the rainfall events on 20 May 2004 and 11 July 2001, remaining in the subsequent study.

After a procedure pruning the decision trees, six rules survived. They can give us the prediction of rainfall intensity during the later part of the first hour. Among them, CT_{01} (cloud-top temperature of pixel scale over the studied area in the earlier part of the first hour before precipitation) is the most critical feature in judging the probability of heavy rain happening in the later part of the first hour. RM_3 (relative moisture in the earlier three hours before precipitation) is the other critical feature. Under the following conditions, heavy rain will happen in the later first hour:

Rule 1: If $CT_{01} < -60$ then $RI_1 = 22–48$.

Rule 2: If $CT_{01} > -60$ and $RM_3 > 97$ then $RI_1 = 13–45$.

Rule 3: If $CT_{01} > -60$ and $RM_3 < 97$ and $CCR_{11} < 0.51$ and $AVECT_{23} > 29$ then $RI_1 = 23$.

Rule 4: If $CT_{01} > -60$ and $RM_3 < 97$ and $CCR_{11} < 0.51$ and $AVECT_{23} < 29$ and $RM_3 > 89$ then $RI_1 = 14–26$.

Rule 5: If $CT_{01} > -60$ and $RM_3 < 97$ and $CCR_{11} > 0.9$ and $AVECT_{23} > 17$ then $RI_1 = 41$.

Rule 6: If $CT_{01} > -60$ and $RM_3 < 97$ and $CCR_{11} > 0.9$ and $AVECT_{23} < 17$ and $CCR_{21} < 0.9$ then $RI_1 = 14$.

Next we moved on to researching what kind of weather conditions can keep raining heavily for two hours. After a procedure pruning decision trees, we got three rules. Among these three rules, WS_1 (wind speed at the earlier one hour before precipitation) becomes the most critical feature in sustaining rainfall till the next hour. Heavy rain will happen in the later second hour under the following conditions:

Rule 7: If $WS_1 > 5.5$ then $RI_2 = 22–65$.

Rule 8: If $WS_1 < 3.2$ and $CT_{01} < -53$ and $WS_1 < 3.4$ then $RI_2 = 14–43$.

Rule 9: If $WS_1 < 3.2$ and $-53 < CT_{01} < 29$ and $AVECT_{41} > 17$ then $RI_2 = 7–30$.

Then we continued to get the weather conditions sustaining rainfall till three hours. Only four rainfall features remained in the induced rules by DT. RM_1 (relative moisture at the earlier one hour before precipitation) is the most critical feature which causes the heavy rain happening during the later third hour. The DT rules are shown as follows:

Rule 10: If $RM_1 > 97$ then $RI_3 = 52$.

Rule 11: If $82 < RM_1 < 97$ and $AVECT_{11} > 9.5$ and $CT_{01} < -18$ then $RI_3 = 20$.

Rule 12: If $82 < RM_1 < 97$ and $AVECT_{11} < 9.5$ and $WS_1 > 5.9$ then $RI_3 = 22$.

Overall there are 8 critical rainfall features composed of CT_{01}, RM_3, CCR_{11}, $AVECT_{23}$, WS_1, $AVECT_{41}$, RM and $AVECT_{11}$ existing in the above twelve rules of DT. Once the estimations of heavy rain lasting for more than two hours were true through DT analyses, the flooding would come soon. At that time, we need a more precise rainfall forecast model, like ANNs, to know which area could be facing flooding.

Conceptual analysis of rainfall model

The following conceptual analyses about the formation path of rainfall will assist us in realizing how to integrate the satellite data and ground climate data. In the tropical area, the warm clouds with a variety of different size condensation nuclei may easily form precipitation via collision–coalescence. Upon collision, the droplets coalesce into a bigger droplet and then achieve a size sufficient to induce precipitation. The ice-crystal process is the other way of forming precipitation. Theoretically ice crystals induced by the supercooled water could be found among the cold clouds in the middle and high latitudes. When

the ice crystals grow larger enough, they start falling and may collide or stick together with one another. If the ground temperature is warm enough, the ice crystals will melt before reaching the ground. Kaohsiung is located between the tropics and subtropics. This special geographic condition makes us consider that both the collision–coalescence and ice-crystal processes dominate the formation of rainfall here. For simulation of collision–coalescence, cloud accumulation in each quadrant was considered in developing our rainfall forecasting model. So we used the cloud-covering ratio inferred by satellite IR3 data to pursue the variance of cloud accumulation. On the other hand, the potential of forming ice crystals is also critical in our rainfall forecast model. For this, we used cloud-top temperatures inferred by satellite IR1 data to get the potential of forming ice crystals.

Artificial neural networks

Considering the ability of simulating nonlinear systems, this paper used artificial neural networks (ANNs) to simulate the hydraulic behavior. Usually ANNs are able to get perfect performance on simulation due to having a good database. In this study, we built a typical ANNs model that consists of three independent layers: input, hidden and output layer (see Figure 1). Input and output layers perform as micro-climatic conditions and the real rainfall measurements, respectively. There are several rainfall features including the cloud-covering ratio of quadrant i (CCR_{ij}), the minimum cloud top temperature of quadrant i ($MINCT_{ij}$), the maximum cloud top temperature of quadrant i ($MAXCT_{ij}$), the average cloud top temperature of quadrant i ($AVECT_{ij}$), the cloud top temperature of pixel-scale over the studied area (CT_{oj}), ground temperature (GT_j), ground pressure (GP_j), relative moisture (RM_j), wind speed (WS_j) and wind direction (WD_j) to be the input components of ANNs. In the above statements of rainfall features, j means the earlier jth hour before precipitation. On the other hand, rainfall intensity within the later first hour, the later second hour and the later third after precipitation (RI_k, $k = 1$, 2 and 3) were considered to be the output components of ANNs. A total of 69 items of climate data are assimilated for each rainfall event. Among them, 66 items that are potential rainfall features are the inputs.

Within ANNs, each layer is interconnected with each other via synaptic weights. W_{lm} is the weight for connecting input neuron l to the mth neuron of the hidden layer and V_{mn} represents the weight for linking the mth neuron of the hidden layer with the nth neuron of the output layer. Each processing neuron in the hidden layer transfers a weighted sum ($S_m = \sum_{l=1}^{I} W_{lm} X_l + B_m$) into H_m, where $H_m = 1/(1 + \exp(-S_m))$, X_l are the input elements (i.e. monitoring item or monitoring variable), W_{lm} are the synaptic weights between the input layer and the hidden layer and B_m is the bias. Via the same computational process, the hidden neuron value H_m may derive the value of output Y_n. Tuning the weighting value is referred to the back-propagation neural network algorithm. When we got an ANN rainfall forecast model through a recursive model training and model test, we used the R^2 value of statistical regression on the ANN predictions and the target outputs to verify the performance of the ANN model. If the R^2 value was good, the rainfall forecasting model of ANNs would be well developed.

As we know, more rainfall data is good in developing an ANN model. Under a limited budget, we chose only 131 rainfall events with significant precipitation (heavy rain or torrential rain) to be the database. Two rainfall events on 20 May 2004 and 11 July 2001 were also not included in this database. They were used to verify the ability of rainfall forecast by ANNs. Then we face a challenge. There are many parameter values which need to be estimated by the limited rainfall data during the ANNs' training. In this ANN, more than 192 synaptic weights are linked with 66 inputs and 3 outputs. We first gave a very small value to each synaptic weight at the beginning of recursive model training. During the process of back-propagation training, the values of those synaptic weights between unrelated nodes would be little revised. So the weighting value of synaptic linked with unrelated nodes remained very small after ANN training. Only the weighting values of those useful nodes could be revised a lot. That is why Garson's index integrates synaptic weighting values to mean the relative importance of each input (Garson 1991). Through a lot of test runs, the whole performance could avoid the over-fitting troubles when the ANNs' training falls in a specific situation.

100 test runs of modeling ANNs were carried out for comparing the rainfall forecasting ability of the model with satellite data and the model without satellite data, respectively. Satellite data means the cloud-covering ratio of

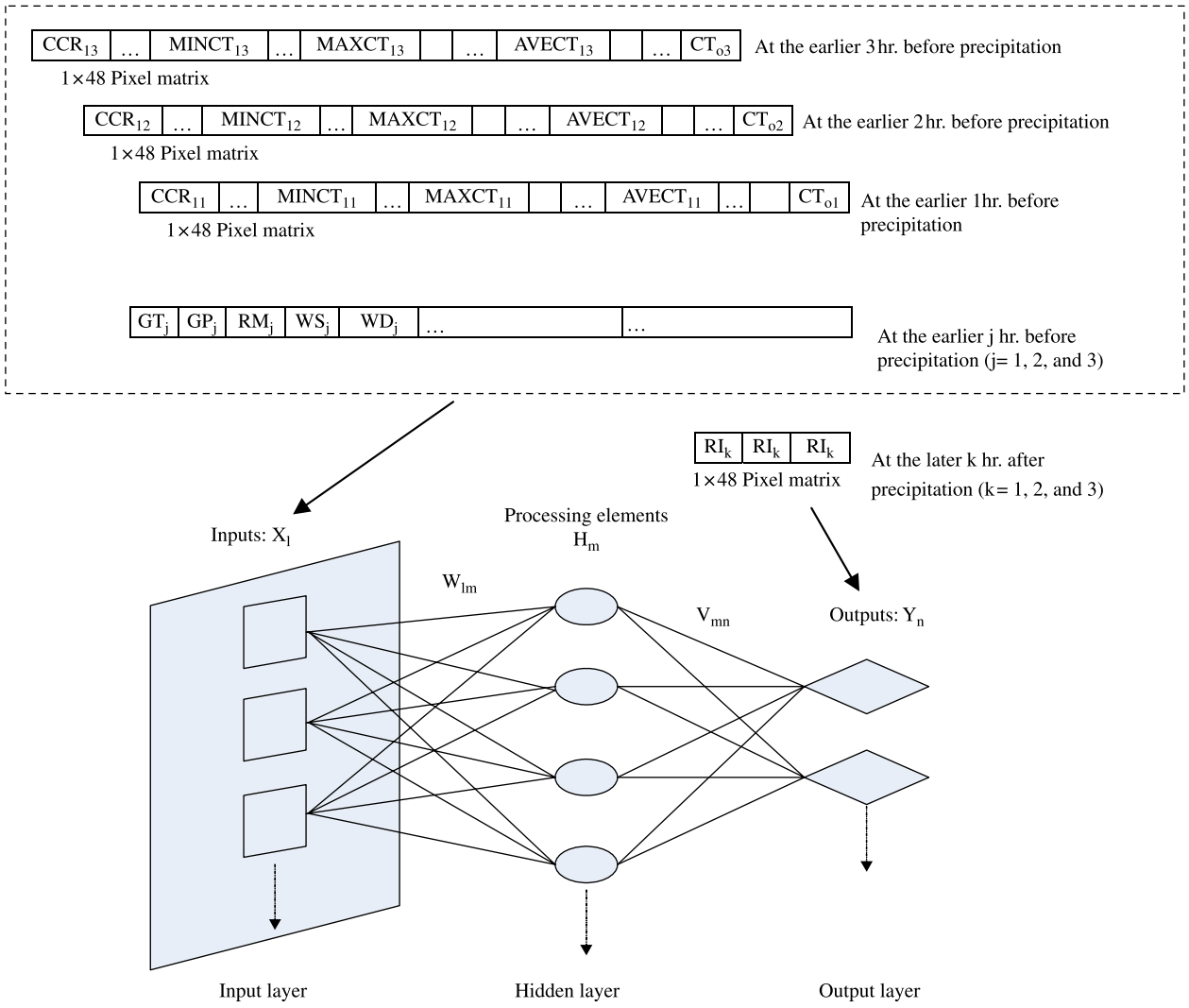

Figure 1 | Data assimilation for building a rainfall forecast model of neural networks.

quadrant i (CCR_{ij}), the minimum cloud-top temperature of quadrant i ($MINCT_{ij}$), the maximum cloud-top temperature of quadrant i ($MAXCT_{ij}$), the average cloud-top temperature of quadrant i ($AVECT_{ij}$) and the cloud-top temperature of pixel-scale over the studied area (CT_{oj}). Every test run randomly chose 16 rainfall events to be the targets for comparison. The R^2 value of statistical regression for the ANNs' predictions and the real measurements is the index of the model's performance. Figure 2 shows that satellite data is indeed useful in rainfall forecasting. All the models with satellite data can reach an R^2 value of 0.9 for the prediction of rainfall intensity within the latter first hour, second hour and third hour; in contrast, the models without satellite information only have an R^2 value of 0.4–0.7. The above finding also confirmed that ANNs with satellite data can support accurate rainfall forecasting for the estimation of flooding probability. Therefore we have to include satellite data in our ANN model despite its time-consuming data preparation.

STUDY AREA AND DATA PREPARATION

Kaohsiung City, located in Southwestern Taiwan, is a rapidly growing international harbor and business center with an area of 154 sq km and a total population of approximately 1.5 million. During 1995–2004, there were more than 1077 rainfall

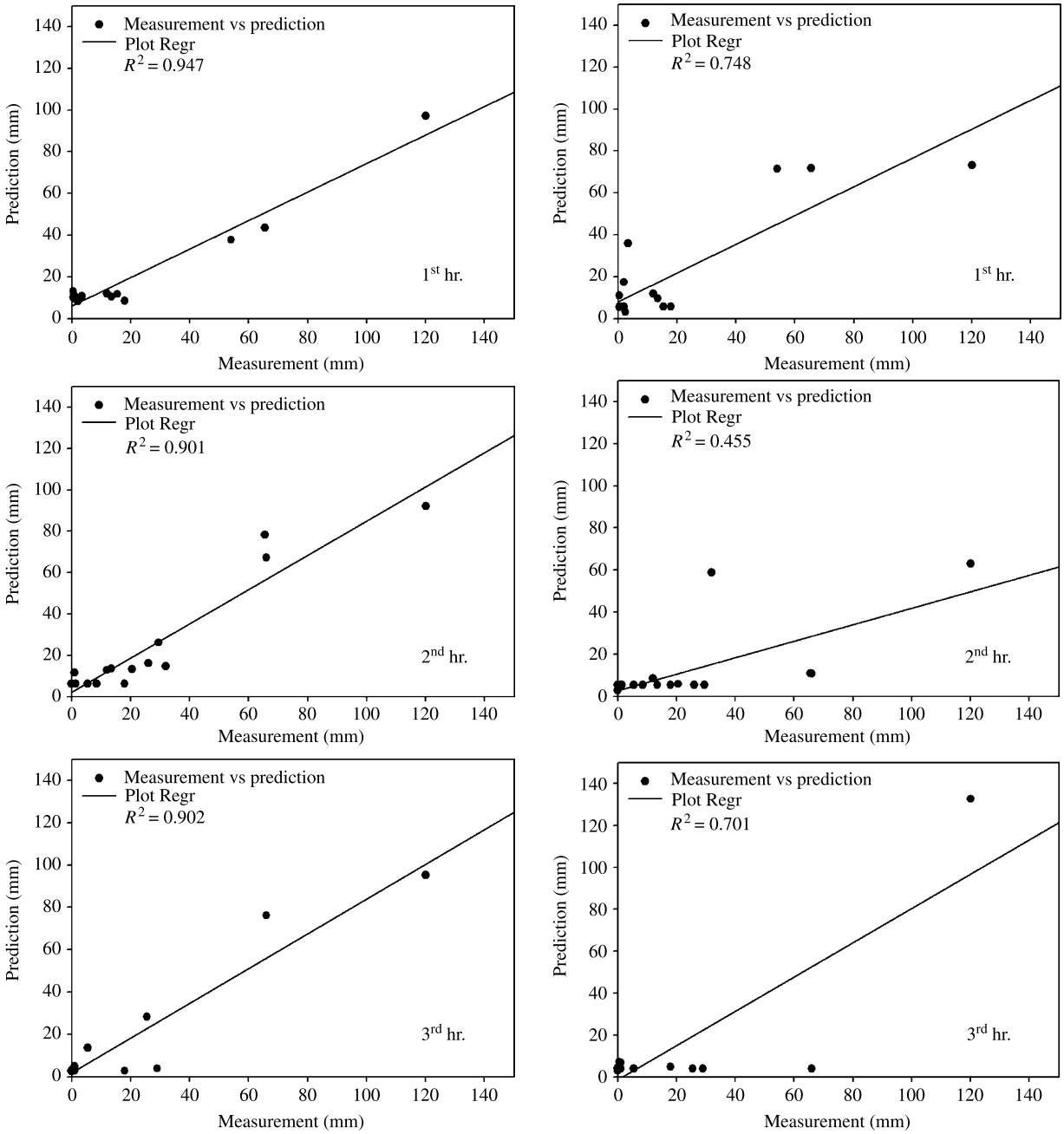

Figure 2 | The comparison between model with satellite data and model without satellite data.

events happening around Kaohsiung. Among them, 944 rainfall events (87%) belong to fine rain, 108 rainfall events (10%) are heavy rain and 25 rainfall events (3%) are torrential rain.

Love River runs through the downtown area of Kaohsiung City and then enters the harbor as shown in Figure 3. The grid area has 14 blocks where each block is one pixel in the satellite image and covers about 6.9 sq km. The total grid area studied is 96.6 sq km. These 96.6 sq km area almost include the whole Love River basin which is the most crowded area in Kaohsiung City. MS means the site of the Kaohsiung weather station that

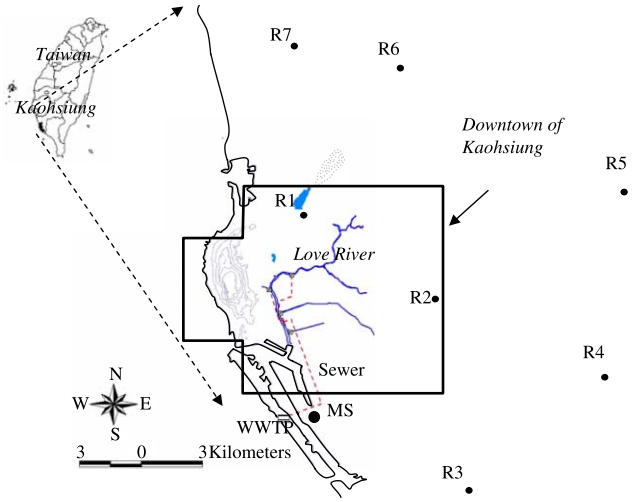

Figure 3 | The geography of Kaohsiung city.

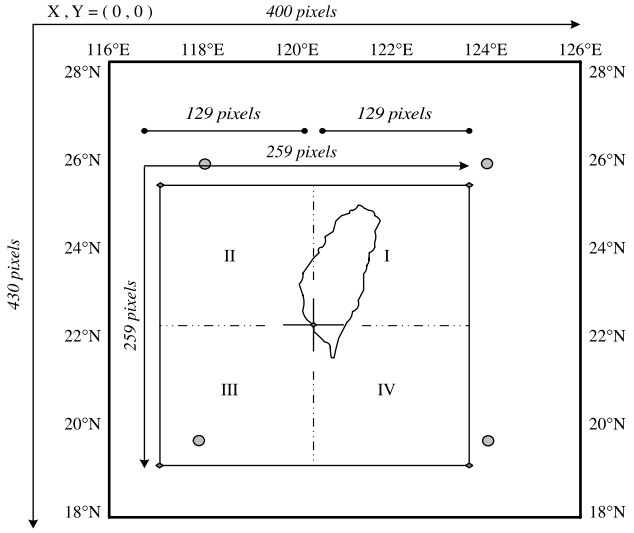

Figure 4 | The monitoring area of satellites over Taiwan provided by TCWB.

can offer real-time ground climate data. R1–R7 mean the neighboring rain gauges around the MS station.

This city has always suffered severe destruction from flooding. An unusual storm on 11 July 2001 flooded more than 1500 buildings in the city of Kaohsiung and caused five fatalities. In terms of management of the drainage system, we have to avoid flooding coming again. But to discharge combined sewage for mitigating the flooding risk will result in the water quality of the receiving river suddenly deteriorating. For the purposes of environmental protection and sightseeing, we need a more intelligent management of the drainage system (Chen *et al.* 2003). Through the analysis of high flooding by DT and ANNs, we will efficiently make the right choice for handling the potential floods.

Here we used the satellite data from the IR1 (infrared: 10.5–11.5μm) and IR3 (water vapor: 6.5–7.5μm) channels of the geostationary meteorological satellites: GMS-5, GOES-9 and MTSAT-1R. Figure 4 shows the monitoring area of the satellites over Taiwan provided by the Taiwan Central Weather Bureau (TCWB). The satellite image covering 1000 × 1000 km (longitude from 116°E to 126°E, latitude from 18°N to 28°N) was renewed each hour. Each image had a resolution of 400 × 430 pixels. In developing the rainfall forecasting model, only the data within 259 × 259 pixels scale were selected. The selected data were centered at the weather station of Kaohsiung city (120°18′29″E, 22°34′04″N).

Next we transferred the satellite data from IR1 into the minimum, maximum and average cloud-top temperatures within four quadrants as depicted in Figure 4. Then we transferred the satellite data from IR3 into cloud-covering ratios within four quadrants. Rainfall features derived from the satellite data are defined as follows:

CCR_{ij} ($i = 1$, 2, 3 and 4), the cloud-covering ratio of quadrant i.

$MINCT_{ij}$ ($i = 1$, 2, 3 and 4), the minimum cloud-top temperature of quadrant i.

$MAXCT_{ij}$ ($i = 1$, 2, 3 and 4), the maximum cloud-top temperature of quadrant i.

$AVECT_{ij}$ ($i = 1$, 2, 3 and 4), the average cloud-top temperature of quadrant i.

CT_{oj}, the cloud-top temperature of pixel scale over the studied area

where j means the earlier jth hour before precipitation.

In addition, rainfall features like ground temperature (GT_j), ground pressure (GP_j), relative moisture (RM_j), wind speed (WS_j) and wind direction (WD_j) were observed at the weather station (MS) of Kaohsiung City and seven neighboring rain gauges (R1–R7).

RESULTS AND DISCUSSION

By decision trees (DT) and artificial neural networks (ANNs), we found several classification rules for judging the potential of heavy rain at Kaohsiung weather station

and a rainfall forecasting model to simulate rainfall accumulation around Kaohsiung weather station. In the following two case studies, we fulfilled a two-stage probability analysis for estimation of high flooding probability area. In the first stage, we used DT to infer the possibility of flooding; in the second stage, we used ANNs to predict the area with heavy rain. Overall case studies have confirmed that our proposed method is promising for mitigating the flooding damage. In addition, a time-series scenario analysis simulated by more than 10 h of weather data on 20 May 2004 and 11 July 2001 assisted us in evaluating its feasibility in handling the whole rainfall event.

Estimation of possibility for flooding by DT

Actually the rainfall event on 20 May 2004 is a typically short tropical storm (its total rainfall accumulation is less than 50 mm). Through the analysis of classification rules induced by DT, heavy rainfall could happen at 16:00–17:00. At that time, its rainfall intensity would probably be over 20 mm/h. At the rest periods, the probability of heavy rain or torrential rain would be low. The above inferences concluded that heavy rainfall would not be sustained for a long time in this case (see Table 1). If we had this inference at that time, we could decide to adopt no action for flooding control.

In contrast, the rainfall event on 11 July 2001 is torrential rain (its total rainfall was more than 500 mm). The classification rules induced by DT can infer that heavy rain could be sustained from 19:00 to 02:00 the next day (see Table 2). In particular the inference at some period can simultaneously meet several rules, that is, the possibility of heavy rainfall at that period would be very high. If we had these inferences at that time, we could make a prompt action for flooding control. These inferences seem to be very close to the facts. Then we need a more precise rainfall forecast model to know which area could face flooding.

Prediction of rainfall using ANNs

In the above section, the rain scale can be easily estimated by DT. But DT cannot tell us the whole rainfall distribution around Kaohsiung city. So we used ANNs to produce the rainfall forecast within 14 blocks shown in Figure 3. It can let us know which area could face flooding. However, its data preparation is very time consuming.

When we use ANNs to simulate the whole rainfall distribution on 20 May 2004, the results show that the whole rainfall accumulation would be not over 80 mm and could only happen within the blocks on the lower side of downtown (see Figures 5–7). This simulation reconfirmed that we might adopt no action for flooding control in this case study. If we had our proposed two-stage probability analysis at that time, we could have saved a lot of jobs and unnecessary worry.

Table 1 | Rainfall intensity predicted by DT for the case of 20 May 2004

Time (h)	Heavy rain or torrential rain	To meet the rules induced by DT
13:00–14:00	Low probability	None
14:00–15:00	Low probability	None
15:00–16:00	Low probability	None
16:00–17:00	High probability	Rule 1
17:00–18:00	Low probability	None
18:00–19:00	Low probability	None

Table 2 | Rainfall intensity predicted by DT for the case of 11 July 2001

Time (h)	Heavy rain or torrential rain	To meet the rules induced by DT
16:00–17:00	Low probability	None
17:00–18:00	Low probability	None
18:00–19:00	Low probability	None
19:00–20:00	High probability	Rule 2 ∩ Rule 7 ∩ Rule 10
20:00–21:00	High probability	Rule 2 ∩ Rule 10
21:00–22:00	High probability	Rule 2 ∩ Rule 10
22:00–23:00	High probability	Rule 2 ∩ Rule 10
23:00–00:00	High probability	Rule 1 ∩ Rule 10
00:00–01:00	High probability	Rule 7 ∩ Rule 10
01:00–02:00	High probability	Rule 10

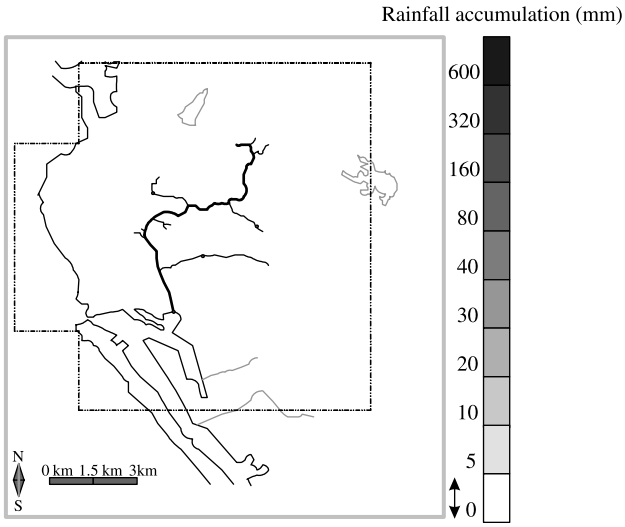

Figure 5 | Rainfall accumulation till the later one hour predicted by the ANN model on 20 May 2004.

More than ten thousand fishes were killed in the receiving river, Love River, after this short tropical storm on 20 May 2004. Scientists thought that the major cause of fish mortality might be acute hypoxia due to discharging of combined sewage. If the operators had our proposed method, they would get the alarm of heavy rainfall at 16:00 as Rule 1 was true. But such an alarm would soon have been cancelled. In addition, rainfall accumulation simulated by ANNs would be below 20 mm till 16:00. So to

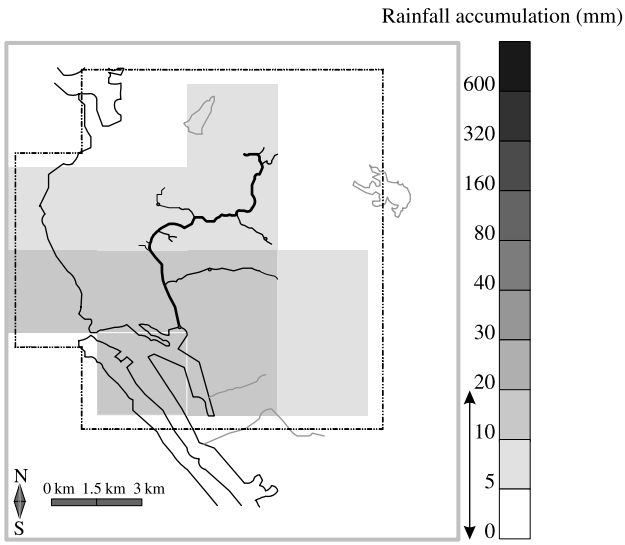

Figure 6 | Rainfall accumulation till the later three hours predicted by the ANN model on 20 May 2004.

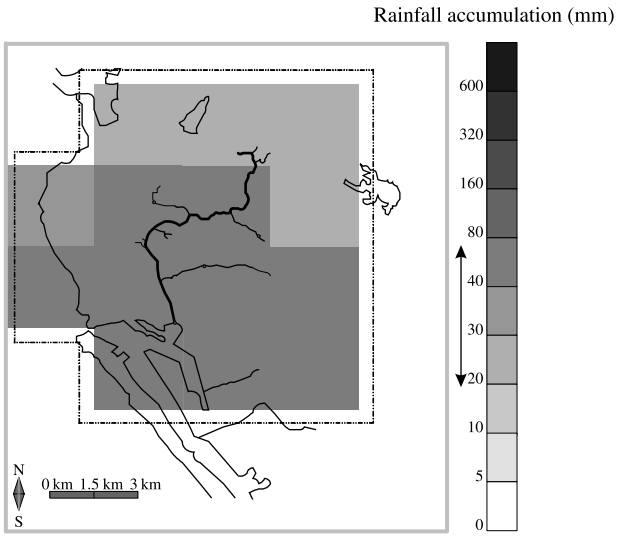

Figure 7 | Rainfall accumulation till the later five hours predicted by the ANN model on 20 May 2004.

discharge combined sewage would not be recommended at that time.

Now let's see what could have happened on 11 July 2001 if we had this two-stage probability analysis. It is torrential rain because its total rainfall was more than 500 mm. In the above section, DT has inferred that heavy rain could be sustained from 19:00 to 02:00 the next day in this case. According to the simulations of ANNs, we can see that rainfall accumulation during five hours could be over 300 mm and heavy rain would happen in the whole urban area (see Figures 8–10). These simulations highlight the flooding risk. If we knew this alarm, we could have adopted prompt action for flooding control.

Historically, an unusual storm on 11 July 2001 flooded more than 1500 buildings and caused five fatalities in Kaohsiung city. Many residents severely argued that the drainage system didn't promptly draw combined sewage out of the urban region. Via our proposed method, the residents would get the alarm of heavy rainfall at 17:00 if Rule 10 was true. Rainfall accumulation at 20:00 would reach the level of 80 mm. In addition, all the inferences reveal that this heavy rain would not cease soon. Thus to discharge combined sewage would be recommended at that time. Then authorities might promptly take emergency action to mitigate flooding damage if they had had our proposed two-stage probability analysis.

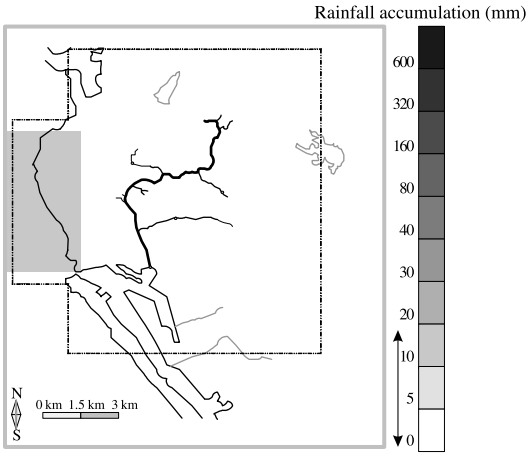

Figure 8 | Rainfall accumulation till the later one hour predicted by the ANNs model on 11 July 2001.

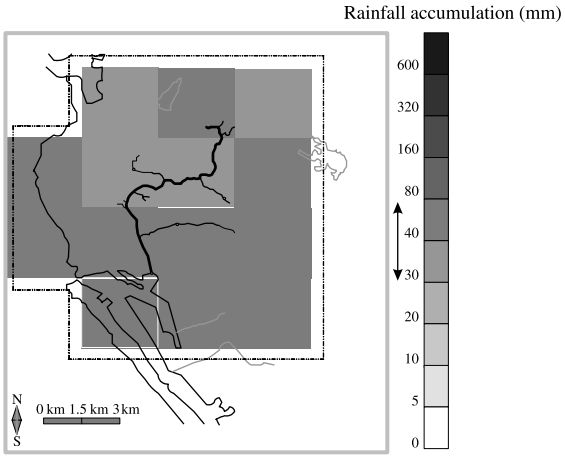

Figure 9 | Rainfall accumulation till the later three hours predicted by the ANN model on 11 July 2001.

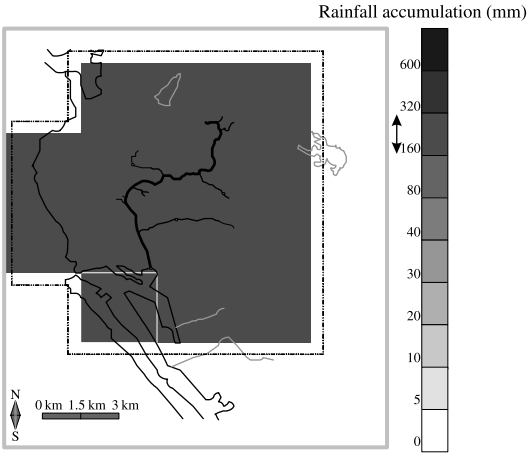

Figure 10 | Rainfall accumulation till the later five hours predicted by the ANN model on 11 July 2001.

CONCLUSION

This paper proved that satellite data is really useful for rainfall forecasting. Based on satellite data, classification rules induced by DT can infer the potential of flooding and the ANNs model can get the resolution of pixel scale (6.9 sq kmr) and predict well the hourly rainfall intensity in the downtown area in Kaohsiung city. In case study I, our proposed method is able to identify a short storm and avoid discharging combined sewage into the receiving river. In case study II, our proposed method is able to detect an unusual torrential rain storm and let fast action be taken over the forthcoming flooding damage. Overall our proposed method is promising to improve urban drainage management.

ACKNOWLEDGEMENTS

The authors express deep gratitude to the National Science Council in Taiwan for the financial support (93-2211-E-242-002).

REFERENCES

Abebe, A. J. & Price, R. K. 2005 Decision support system for urban flood management. *J. Hydroinf.* **7** (1), 3–15.

Ba, M. B. & Gruber, A. 2001 GOES multispectral rainfall algorithm (GMSRA). *J. Appl. Meteorol* **40** (8), 1500–1514.

Babovic, V. 2005 Data mining in hydrology. *Hydrol. Process.* **19**, 1511–1515.

Baldwin, M. E., Kain, J. S. & Lakshmivarahan, S. 2005 Development of an automated classification procedure for rainfall systems. *Mon. Weather Rev.* **133** (4), 844–862.

Bankert, R. L., Hadjimichael, M., Kuciauskas, A. P., Thompson, W. T. & Richardson, K. 2004 Remote cloud ceiling assessment using data-mining methods. *J. Appl. Meteorol.* **43** (12), 1929–1946.

Bellerby, T. J. 2004 A feature-based approach to satellite precipitation monitoring using geostationary IR imagery. *J. Hydrometeorol.* **5** (5), 910–921.

Breiman, L., Friedman, J., Olshen, R. & Stone, C. 1984 *Classification and Regression Trees*. Wadsworth International Group. Monterey, CA.

Burrows, W. R., Benjamin, M., Beauchamp, S., Lord, E. R., McCollor, D. & Thomson, B. 1994 CART decision-tree statistical analysis and prediction of summer season maximum surface ozone for the Vancouver, Montreal, and Atlantic regions of Canada. *J. Appl. Meteorol.* **34** (8), 1848–1862.

Carvalho, L. M. V. & Jones, C. 2001 A satellite method to identify structural properties of mesoscale convective systems based on the Maximum Spatial Correlation Tracking Technique (MASCOTTE). *J. Appl. Meteorol.* **40** (10), 1683–1701.

Chen, J. C., Chang, N. B., Chang, Y. C. & Lee, M. T. 2003 Mitigating the impacts of combined sewer overflow to an urban river system via web-based share-vision modeling. *J. Civil Engng. Environ. Syst.* **20** (4), 213–230.

Chen, J. C., Chang, N. B. & Chen, C. Y. 2004 Minimizing ecological risk of combined sewage overflow in an urban river system by a system-based approach. *J. Environ. Engng. ASCE* **130** (10), 1154–1169.

Colquhoun, J. R. 1987 A decision tree method of forecasting thunderstorms, severe thunderstorms and tornadoes. *Weather Forecasting* **2** (4), 337–345.

Feidas, H. 2003 A software tool for monitoring the features of convective cloud systems with the use of Meteosat images. *Environ. Model. Software* **18**, 1–12.

Garson, G. D. 1991 Interpreting neural-network connection strengths. *AI Expert* **6** (4), 47–51.

Grimes, D. I. F., Coppola, E., Verdecchia, M. & Visconti, G. 2003 A neural network approach to real-time rainfall estimation for Africa using satellite data. *J. Hydrometeorol.* **4** (6), 1119–1133.

Hong, Y., Hsu, K. L., Sorooshian, S. & Gao, X. 2004 Precipitation estimation from remotely sensed imagery using an artificial neural network cloud classification system. *J. Appl. Meteorol.* **43** (12), 1834–1853.

Kawamura, A., Jinno, K., Berndtsson, R. & Furukawa, T. 1996 Parameterization of rain cell properties using an advection-diffusion model and rain gage data. *Atmos. Res.* **42**, 67–73.

Kurino, T. 1997 A satellite infrared technique for estimating "deep/shallow" precipitation. *Adv. Space Res.* **19** (3), 511–514.

Miller, S. W. & Emery, W. J. 1997 An automated neural network cloud classifier for use over land and ocean surfaces. *J. Appl. Meteorol.* **36** (10), 1346–1362.

New, M., Todd, M., Hulme, M. & Jones, P. 2001 Precipitation measurements and trends in the twentieth century. *Int. J. Climatol.* **21**, 1899–1922.

Pietroniro, A. & Leconte, R. 2000 A review of Canadian remote sensing applications in hydrology, 1995-1999. *Hydrol. Process.* **14**, 1641–1666.

Tarruella, R. & Jorge, J. 2003 Comparison of three infrared satellite techniques to estimate accumulated rainfall over the Iberian Peninsula. *Int. J. Climatol.* **23**, 1757–1769.

Wei, C., Hung, W. C. & Cheng, K. S. 2006 A multi-spectral spatial convolution approach of rainfall forecasting using weather satellite imagery. *Adv. Space Res.* **37** (4), 747–753.

First received 9 June 2006; accepted in revised form 10 July 2007