Use of geostationary meteorological satellite images in convective rain estimation for flash-flood forecasting

T. Wardah [a,*], S.H. Abu Bakar [a], A. Bardossy [b], M. Maznorizan [c]

[a] Faculty of Civil Engineering, Universiti Teknologi MARA, Science and Technology Building, 40450 Shah Alam, Malaysia
[b] Hydrology and Geohydrology Institute, Stuttgart University, Germany
[c] Satellite Division, Malaysian Meteorological Department, Malaysia

Summary   Frequent flash-floods causing immense devastation in the Klang River Basin of Malaysia necessitate an improvement in the real-time forecasting systems being used. The use of meteorological satellite images in estimating rainfall has become an attractive option for improving the performance of flood forecasting-and-warning systems. In this study, a rainfall estimation algorithm using the infrared (IR) information from the Geostationary Meteorological Satellite-5 (GMS-5) is developed for potential input in a flood forecasting system. Data from the records of GMS-5 IR images have been retrieved for selected convective cells to be trained with the radar rain rate in a back-propagation neural network. The selected data as inputs to the neural network, are five parameters having a significant correlation with the radar rain rate: namely, the cloud-top brightness-temperature of the pixel of interest, the mean and the standard deviation of the temperatures of the surrounding five by five pixels, the rate of temperature change, and the sobel operator that indicates the temperature gradient. In addition, three numerical weather prediction (NWP) products, namely the precipitable water content, relative humidity, and vertical wind, are also included as inputs. The algorithm is applied for the areal rainfall estimation in the upper Klang River Basin and compared with another technique that uses power-law regression between the cloud-top brightness-temperature and radar rain rate. Results from both techniques are validated against previously recorded Thiessen areal-averaged rainfall values with coefficient correlation values of 0.77 and 0.91 for the power-law regression and the artificial neural network (ANN) technique, respectively. An extra lead time of around 2 h is gained when

---

* Corresponding author. Tel.: +60 3 5543 5260; fax: +60 3 5543 5275.
  E-mail addresses: wardah_tahir@yahoo.com, warda053@salam.uitm.edu.my (T. Wardah).

the satellite-based ANN rainfall estimation is coupled with a rainfall-runoff model to forecast a flash-flood event in the upper Klang River Basin.

Introduction

Flash-floods have caused countless grievances to the urban folks of Malaysia: being trapped in long hours of traffic jam congestion, damage to parked cars, and the destruction of homes and business properties. In February 2006, a severe flash-flood occurred at the Shah Alam municipality in the lower Klang River Basin, which caused damages to properties worth approximately 25-million US dollars. The 104 mm rain recorded within 3 h had caused the water level to rise to a depth of about 1.6 m in some residential and commercial areas, thereby destroying thousands of homes and properties. Despite extensive structural flood-mitigating measures, the capital city of Kuala Lumpur, which is located at the upper Klang River Basin, is still frequently hit by severe flash-floods, as indicated by the various events that occurred in 2007.

The current flood forecasting system for the Klang River Basin is operated on the basis of river levels forecasted from the output of an integrated rainfall-runoff and flood-routing model, which receives inputs from a telemetric rainfall network. Similar systems with different flood routing models are applied in other parts of the country, and they would be able to provide several hours of lead time. However, due to the short response time available for small and medium urban catchments during the flash-floods, the available forecasts are not sufficient for most rescue operations. This study is an attempt to develop an algorithm for rainfall estimation using the GMS-5 IR images, as a potential input for an improved flood forecasting-and-warning system.

Rainfall is one of the most difficult variables to forecast; due to its large variability, both in space and in time. Recent advances in radar meteorology have improved the rainfall forecasting process remarkably. Nevertheless, radars can measure the rainfall intensity only after the formation of raindrops. So as to forecast rainfall as early as possible, prediction of the development of convective clouds and the associated precipitation system assumes great significance.

Numerical weather prediction rainfall forecasts are considered too coarse in their spatial and temporal resolution; hence they are rarely used as direct inputs in rainfall-runoff models for flood forecasting. Nakakita (2002) observes that the present status of the quality of rainfall forecasts can be improved in accuracy by extrapolation techniques using satellite and radar data.

A meteorological satellite with fine temporal and spatial resolution should be considered as complementary to radar and rain gage measurements because it covers a broader area that may be inaccessible or that may cause difficulties with the traditional rainfall measurement such as the oceans or rigid mountains. Nevertheless, unlike gage measurement, rain estimation using satellite is indirect.

Many studies have been carried out on satellite-based rainfall estimation, as published by others including Arkin (1979), who has developed the geostationary operational environmental satellite (GOES) precipitation index (GPI) technique, and Griffith et al. (1978), who have developed the Griffith—Woodley technique. The latter technique defines the cloud as anything colder than 253 K and estimates the associated radar echo for each cloud. Negri et al. (1984) have simplified the Griffith—Woodley technique by eliminating cloud tracking and have produced a precipitation scheme that treats each cloud as if existing only in one image, resulting in the Negri—Adler—Wetzel (NAW) algorithm. Porcù et al. (1999) have combined the NAW technique with radar data for nowcasting. Todd et al. (2001) have reported on cloud-indexing algorithms for the polar-orbiting National Oceanic and Atmospheric Administration (NOAA) satellites and for geostationary satellite imagery whereas Adler and Negri (1988) and Anagnostou et al. (1999) have developed a one-dimensional cloud model, which relates the cloud-top temperature to the rain rate and rain area; that is called the convective stratiform technique (CST). Ba and Nicholson (1998) have used an index of convection based on the METEOSAT IR data and rainfall measurements to analyze the convective activity and its relationship to rainfall. The RAINSAT technique (Lovejoy and Austin, 1979; Bellon et al., 1980) is claimed to reduce the number of false alarms in the pure IR techniques by screening out cold but not highly reflective clouds and those that are highly reflective clouds with a relatively warm top. The algorithm has been developed on the basis of a supervised classification trained by radar to recognize precipitation from both the visible (VIS) brightness and the IR cloud-top brightness-temperature.

Studies on satellite-based rainfall estimation techniques are constantly evolving and improving with the aim of getting a better accuracy in the rainfall estimates, with a finer spatial and temporal resolution. However, due to the indirect nature of the relationship between the satellite-measured radiance in both the visible and infrared regions and the corresponding rainfall, it is observed that the above techniques are not universally applicable; techniques developed for the extratropic regions might not perform well in the tropics. In addition, techniques developed to estimate the monthly rainfall might not be applicable for the hourly rainfall estimation.

Intense convective rain

Physical process of convective rain

Fox and Collier (2000) discussed the physical process leading to convective rain formation, as it is this type of rainfall that most often leads to flash-flooding. The parcel theory of convection for rainfall estimates relates the rainfall to two simple parameters: the vertical velocity of the ascending air parcel of the precipitation system and the water available to form precipitation (Wiesner, 1970; Fox and Collier, 2000). The peak rain-rate equation is given as

$$R_p = \int_{Z}^{Z_T} \rho(z)w(z)\frac{\delta\chi_s}{\delta z}\,dz \qquad (1)$$

where R_p is the peak rain-rate generated by the convective system in mm/h. The parameters contributing to the calculation of R_p are the density of air (ρ), the saturated mixing ratio (χ_s), and the vertical velocity of the ascending air parcel (w). The integration takes place between the condensation level and the tropopause. This formula implies that precipitable water and vertical velocity are two significant quantities in estimating the rate of rainfall generated by a convective system. It is assumed herein that heating increases through the morning, leading to convection being initiated during the afternoon; which then this process leads to cloud formation and subsequent rain in the late afternoon and early evening, as is observed in tropical regions (Minnis and Harrison, 1984; Reed and Jaffe, 1981; Fox and Collier, 2000).

Deep convection

Deep convection is a mechanism to the transport of momentum, heat, and moisture from the earth's surface to the free troposphere, thus becoming the main source of energy for large-scale circulations and weather phenomena. The exchange of air between the tropical boundary layer and the free troposphere by convective transport leads to the redistribution of chemical compounds in the atmosphere. Rains and thunderstorm are the most frequent representations of deep convection. The presence of thick cumulonimbus clouds that reach to high, cold altitudes is an indication of deep convective activity. Fig. 1 illustrates the Byers—Braham model of a thunderstorm cell, indicating air movements and precipitation. The temperatures of the cloud-tops causing severe storms are very low, sometimes below the tropopause temperature.

Convective overshoots are visible in the high resolution IR images as very cold regions of about 10 km diameter, and they represent an evidence for active deep convection. The cirrus anvils associated with well-developed con-

vection can reach up to 10^5 km^2, and they last long after the actual convection has ceased, whereas the actual deep convection takes place in smaller regions, where the rising air parcels penetrate the entire troposphere in under an hour (van Hees et al., 1999). Griffith et al. (1978) have found that a convective system is more active and produces the greatest rainfall rates when the tops become colder and continue to expand.
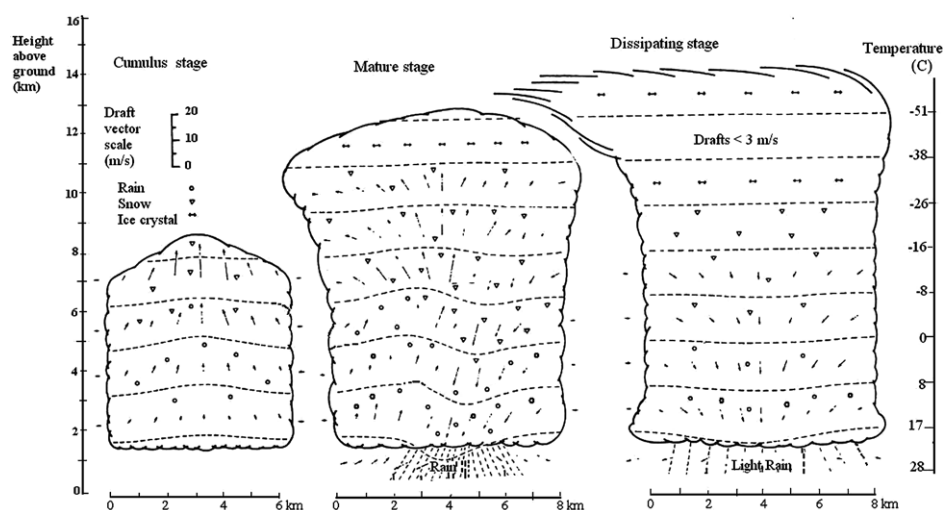
Intense rainfall events at the upper Klang River Basin

An analysis of heavy rainfall events (intensity > 30 mm/h) recorded at station 3116003 in the upper Klang River Basin from 1980 to 2004 shows that most of the events occur in the late afternoon, and early evening, as illustrated in Fig. 2. The analysis indicates that 81.7% of the 175 heavy rain events that have been recorded are of the convective type which occur between 15:00 and 21:00 LT (local time).

Study area and data

Study area

The Klang River Basin is located toward the west coast of peninsular Malaysia, in the central part of Selangor state. The basin consists of the most developed and rapidly urbanizing regions including the Federal Capital of Kuala Lumpur. Most of the important areas in the basin have been frequently affected by recurrent flooding of the Klang River and its tributaries. The Klang River has a total stream length of about 120 km and has a total catchment area of about 1200 km^2 at the mouth of Port Klang. The main river has headwaters in the mountainous areas to the north-east of Kuala Lumpur, where the maximum elevation rises to above 1400 m. The lower reaches flow through relatively flat land with some swampy areas. Most of the catchments are hilly, and the rivers are generally fast-flowing. The annual mean rainfall in the basin is about 2300 mm.

Figure 1   The Byers—Braham model of a thunderstorm cell, indicating air movements and the precipitation forms (Rogers and Yau, 1988).
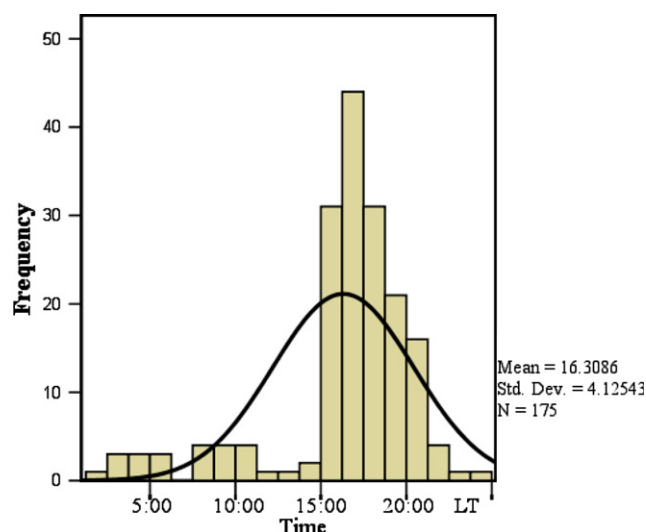
Figure 2   Timing of intense rain (>30 mm/h).

This study focuses on the upper Klang River Basin, as shown in Fig. 3, which comprises an area of 468 km². Ten rain-gaging stations and one flow-gaging station are available in the upper Klang River Basin for the benefit of research into the field.

Data

Geostationary meteorological satellite images
Geostationary meteorological satellites have high altitudes around 36,000 km and they maintain a fixed geographic location while orbiting at the same speed that the earth rotates, so that they stay above the same spot on the earth. The satellites make observations at 30 min of intervals throughout each day, enabling meteorologists to observe clouds at all levels of the atmosphere over both land and the oceans.

Radiation can be measured at specific wavelengths by using a filter placed in front of the satellite sensor. In satellite meteorology, the larger the radiance from an element, the darker is the pixel of an IR image. Lower clouds are war-mer and thus emit more radiation than higher clouds, which are much colder. Low clouds appear gray on an IR satellite picture whereas high clouds appear white. By closely analyzing the satellite images, meteorologists estimate cloud heights and thicknesses.

Meteorological satellite images have been acquired from the Malaysian Meteorological Department (MMD), which receives these images from the Japanese National Space Development Agency. The GMS-5 is positioned at 140° east and has a spin-scan radiometer that returns VIS and IR images. The images used for this study are the hourly GMS-5 IR images (channel 10.3—11.3 μm) with a spatial resolution of about 7×5 km. Images from the year 2003 and from the wet months (March, April, October, and November) of 2004 and 2005 have been collected. The validation data that have been used are the latest GMS generation, named Multifunctional Transport Satellite (MTSAT), IR images for the rainy days period in the year 2006. Fig. 4 shows an example of the GMS-5 IR image (enhanced color) with the look-up table provided by the MMD.

Radar reflectivity and display images
Peninsular Malaysia has a network of six weather surveillance radars (WSR), and one Doppler type radar. The radar data for this research have been taken from the Subang WSR81S Band S radar, located at Lat 3.120N and Long 101.55E and at an altitude of 32 m. The acquired data are the hourly displayed radar images (2 km × 2 km resolution) and the 15-min-interval values of radar reflectivity. Fig. 5 shows an example of radar display with the look-up table.

Because the radar technique is an indirect estimation of the rainfall rate, a verification procedure for the technique is carried out by comparing the radar rain estimation with the point gage measurement. Radar displays have been processed using a computer program available at the MMD to zoom in to the upper Klang River Basin, and the stations are located individually. Selected rainfall events (categories moderate to very heavy) for the year 2006 measured by a point rain gage are compared with the radar rain estimation. A total of 59 data pairs have been produced for the comparison. Fig. 6 is the plot of radar rain estimation versus the gage measurement, with a 95% line-prediction interval.
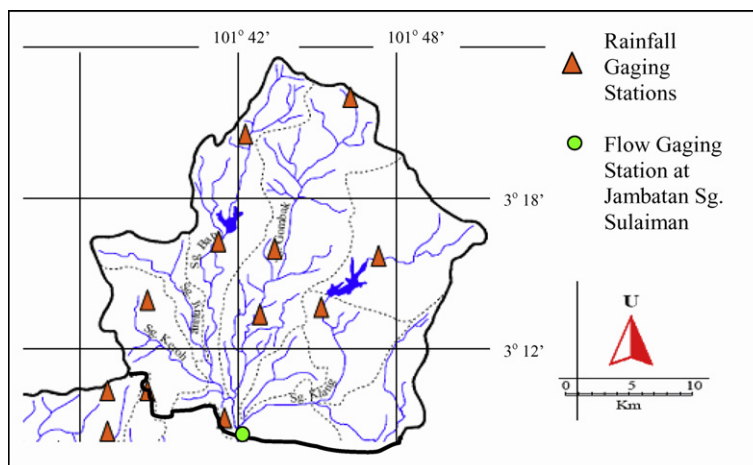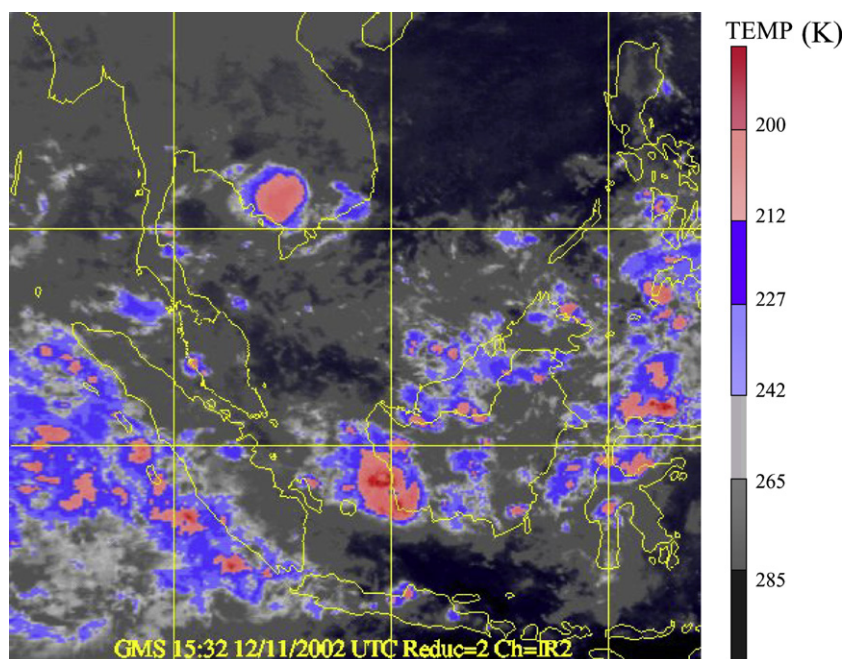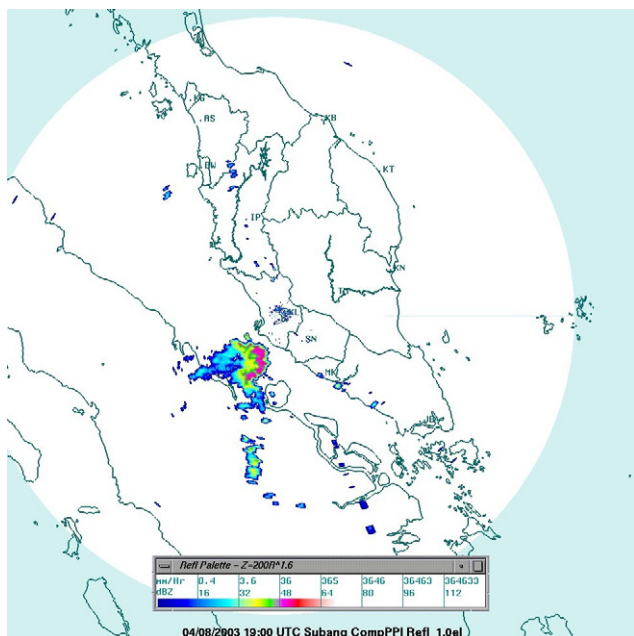

Figure 3   The upper Klang River Basin with ten rainfall-gaging stations and a flow-gaging station.

Figure 4 An example of GMS-5 IR image (enhanced color) with the look-up table. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Figure 5 Radar display with the look-up table.

Figure 6 Radar rain estimation versus gage measurement with a 95% line-prediction interval.

The correlation coefficient for the radar rain estimation versus rain gage measurement is 0.92. This value indicates a satisfactory performance of the weather radar in the Peninsular Malaysian network for the estimation of rainfall.

Rainfall and streamflow data
Half hourly rainfall data from 1980 to 2006 have been acquired from the Drainage and Irrigation Department (DID)

of Malaysia for 10 stations in the upper Klang River Basin, as shown in Fig. 4. Flow data for the same period at the basin outlet river-gaging station are also collected for flow forecasting and validation.

NWP NCEP (National Centers for Environmental Prediction) reanalysis data
The MMD adopts the Fifth Generation Mesoscale Model (MM5) as its NWP model. For the purpose of NWP historical data retrieval, NCEP reanalysis data has been recommended by the MMD to be used for this research. The NCEP reanalysis data

are provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their website at http://www.cdc. noaa. gov/. Three relevant data namely relative humidity (RH), precipitable water content (PW) and vertical wind (VW) have been retrieved from the website. Fig. 7 is an example of PW data retrieved for February 1, 2003, averaged for 0—6 UTC with a value of around 50.5 kg/m^2 over the upper Klang River Basin.
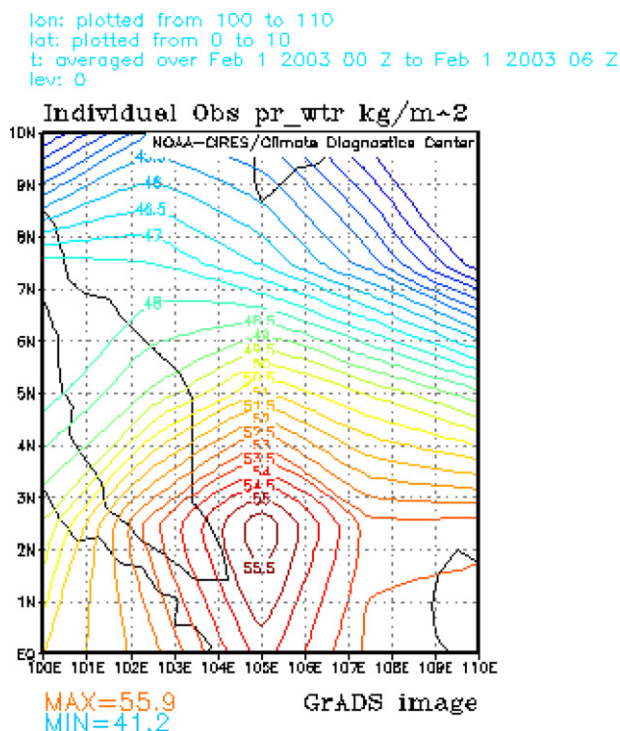

Figure 7   Sample of PW data retrieved from the NCEP web page.

Methodology

Geolocation of pixel

An important aspect of meteorological satellite image interpretation is the accurate geographic location of the image pixels. The process of geolocating pixels is called *navigation*. This process requires current orbit and altitude parameters of the satellite and also the algorithms that use these parameters to calculate the latitude and longitude of each pixel. The meteorological geostationary satellite images received by the MMD are provided with an algorithm that carries out the navigation; hence, the border lines representing countries can be placed correctly, as illustrated in Fig. 8a. To locate a specific area, the image is first cropped from 100° to 110°E and 0° to 10°N, as shown in Fig. 8b, to produce a 10° × 10° image. Since the number of pixels per image of 10° × 10° can be determined, thereby the geolocation of each pixel or the determination of pixel closest to a particular rainfall station can be calculated, knowing its latitude—longitude point on earth.

Reading the pixel cloud-top brightness-temperature

An IR image is composed of pixels with discrete digital values, which correspond directly to an equivalent blackbody radiation. The 8-bit image-display system allows a range of values from 0 to 255 for a pixel. A color or monochromatic gray shade can be assigned to each digital value to produce an enhanced image. In a grayscale IR image, cold clouds are high clouds with pixel brightness values indicated by the corresponding grayscale color tone. Using the pixel brightness value, the temperatures can be estimated. In addition to the grayscale images, some of the images acquired are in the enhanced color format, which come along with a look-up table that assigns the temperature for each color range. Several programs have been developed using
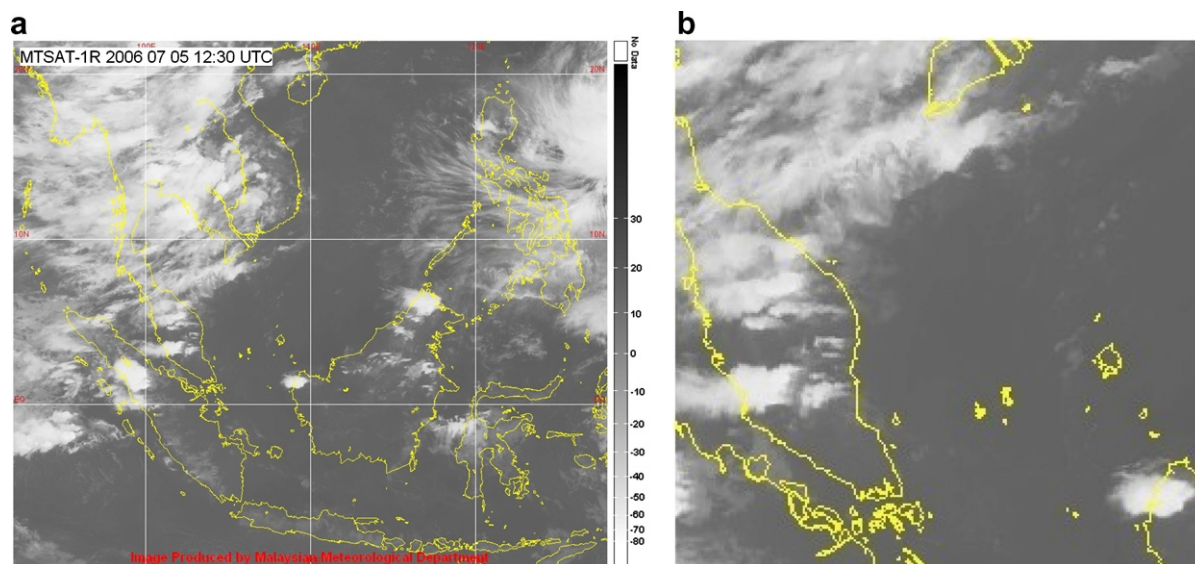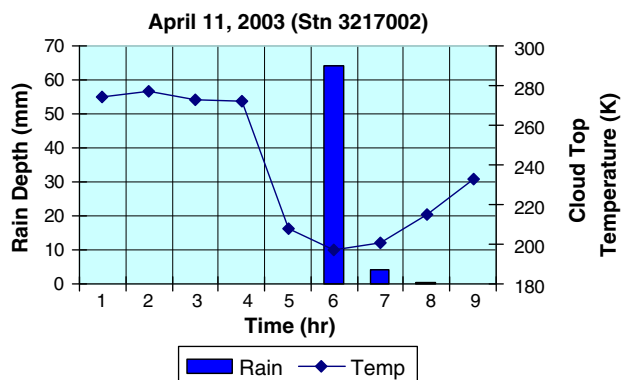

Figure 8   (a) An example of MTSAT image; (b) GMS-5 IR image cropped from 100° to 110°E and from 0° to 10°N.

Figure 9 Cloud-top brightness-temperature plotted as a function of the rain depth versus time.

the MATLAB™ image processing tool to read the temperature of the pixels of interest for each grayscale and enhanced color images. The cloud position on the earth is then estimated after some displacement error and parallax corrections as suggested by Kidder and Vonder Haar (1995).

Hourly temperature data have been produced for each image frame that is collected. As an initial observation, cloud-top brightness-temperatures of the closest pixel to a rainfall station are plotted against the rainfall events. Fig. 9 shows an example of declining cloud-top brightness-temperature during a storm event. The corresponding graph indicates that cloud-top brightness-temperature decreases and reaches its coldest value during a storm event.

Power-law regression technique

The Auto-Estimator (Vicente, 2001) of the National Environmental Satellite, Data and Information Service (NESDIS) uses the IR images obtained by the geostationary operational environmental satellite (GOES) and runs in real-time for flash-flood forecasting and operational hydrology; this pro-

cess is administered by the NOAA. This study refers to the auto-estimator methodology in the development of a modified technique named AMAE (adapted and modified auto-estimator). The procedures included in the AMAE are listed as below:

(a) Computing rainfall rates based on a non-linear, power-law regression relationship between the cloud-top temperature and radar rain rate.
(b) Applying an additional test to screen the non-raining pixels, i.e, using consecutive hourly satellite IR images to indicate vertically growing and decaying cloud systems. A pixel is considered non-raining for a negative temperature change.

Development of the power-law regression relationship

Radar images at the collocated time (about 30 min time difference between radar and satellite images) and space with the respective GMS-5 IR images for the West Peninsular Malaysia are scrutinized and analyzed for intense convective cells, as shown in Fig. 10. Cloud-top brightness-temperature values are read on the longitude and latitude grid overlay and compared with the radar rain rate at the same grid location to produce the regression equation.

Selected convective rain events with 293 sets of data were included in the regression analysis between the radar rain-rate and the IR cloud-top brightness-temperature. The results are shown in Figs. 11 and 12. Fig. 11 depicts the plot of GMS IR cloud-top brightness-temperature against the mean radar rain rate with a 95% confidence interval. The relationship between the two variables was derived through a linear regression analysis after taking the logarithm of the radar rain rate. A coefficient of determination (r^2) value of 0.94 is obtained for the equation relating the two variables as given in Eq. (2) and is shown in Fig. 12,

$$Log(R) = 7.29 - 0.028T \qquad (2)$$
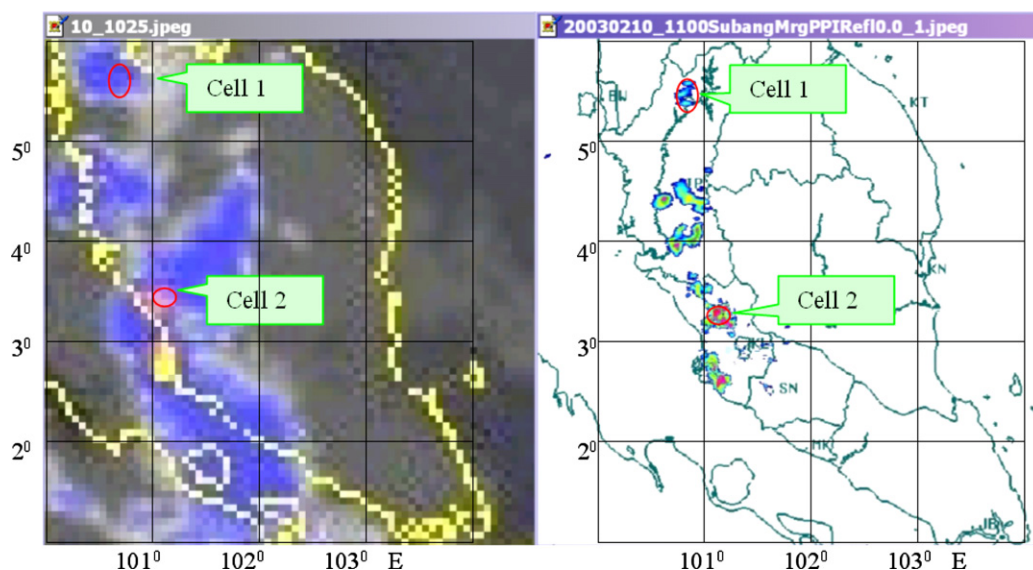
Figure 10 A pair of satellite IR images and radar displays (two images) at a collocated time, indicating two distinct convective cells.

Figure 11   GMS IR cloud-top brightness-temperature plotted against the mean radar rain rate.
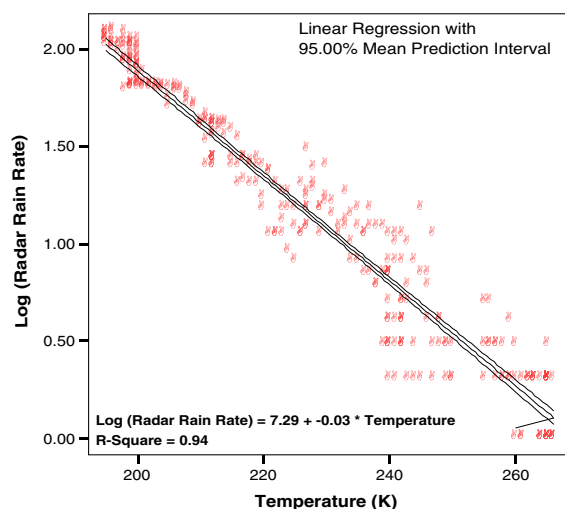
Figure 12   Regression analysis between the GMS IR cloud-top brightness-temperature and the mean radar rain rate.
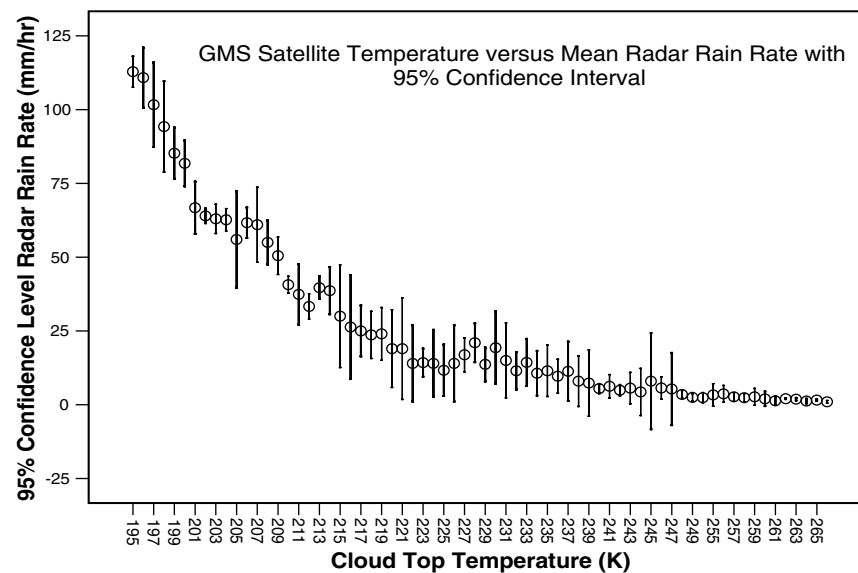
where R is the radar rain-rate and T is the IR cloud-top brightness-temperature. The equation that is derived considers both rain and non-rain pixels in the computation of the regression fit.

The ANN model for rainfall estimation

Rainfall is the outcome of a complex non-linear atmospheric process involving many parameters. Even the complicated physically based NWP model cannot precisely estimate rainfall at the required temporal and spatial resolution, especially for flash-flood simulation. An alternative to the physically based models is data driven approach which is based on machine learning. ANN is a widely used artificial intelligence technique which provides a powerful means of performing non-linear regression between a combination of prime factors to the outcome of interest.

Numerous studies have been carried out by other researchers on the use of ANN in rainfall and atmospheric process modeling, as described by the following authors: Kim and Barros (2001), on the quantitative flood forecasting

Table 1   Correlation coefficient between the parameters involved

Variables	Correlation with	Spearman's ρ	Significant (2-tailed)[*]
Pixel temperature	Radar rain rate	−0.734	0.000
Mean 5×5 pixel	Radar rain rate	−0.517	0.000
Standard deviation 5×5 pixel	Radar rain rate	−0.283	0.000
Temperature change (K/h)	Radar rain rate	0.880	0.000
Maximum sobel	Radar rain rate	0.413	0.000
Precipitable water content	Mean 5×5 pixel temperature	−0.239	0.001
Relative humidity	Mean 5×5 pixel temperature	−0.165	0.020
Relative humidity	Precipitable water content	0.728	0.000
Vertical wind	Relative humidity	0.301	0.000

[*] Correlation is significant at the 0.01 level.

Figure 13 ANN model component and back-propagation process.

using multisensor data and ANN; Sorooshian et al. (2000), on the application of ANN in satellite-based estimates of tropical rainfall; Kuligowski and Barros (1998) on short-term pr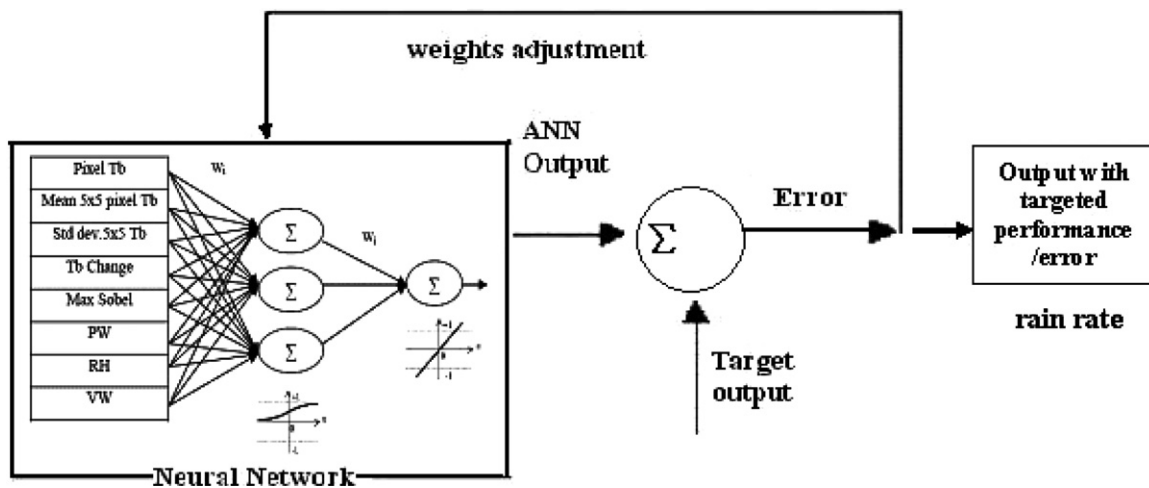ecipitation forecasting using ANN; and Bankert (1994), on cloud classification of an AVHRR (Advanced Very High Resolution Radiometer) imagery in maritime regions using a probabilistic ANN.

Neural networks, which are inspired by biological nervous systems, are composed of simple elements operating in parallel. The connections between the elements determine the network function. This study uses the back-propagation neural network algorithm to estimate the rain rate from an input combination of several selected variables. The multi-layer back-propagation neural network consists of layers of neurons, with each layer being fully connected to the next layer by inter-connection strengths or weights, w. Initial estimated weight values are progressively corrected during a training process that compares predicted

outputs to target outputs, and back-propagates any error to determine the appropriate weight adjustments necessary to minimize the errors. The methodology used here for adjusting the weights is based on the *generalized delta rule*. The simplest version of gradient-descent algorithm for minimizing error is to change each weight by an amount proportional to the accumulated $\partial E/\partial w$,

$$\Delta w = -\epsilon \frac{\partial E}{\partial w} \qquad (3)$$

where ϵ is the learning rate, w is the weights, and E is the total error (Hagan et al., 1996).

Selection of input variables
Selection of input variables is crucial to ensure success of the network. Following the concept of convective rain formation described in Sections 'Physical process of convective rain' and 'Deep convection', the parameters shown in Table 1 are considered to be the important factors related to rain rate.

Rank correlation analysis
Rank correlation analysis is carried out on the variables considered to be related to the rain rate. Spearman's coefficient correlation, ρ, indicates that the radar rain rate has the highest correlation with the average negative temperature change in an hour, with ρ equal to 0.88. The positive correlation coefficient indicates that the larger the rate of temperature change, the higher the rain rate or intensity, whereas the negative correlation between the cloud-top brightness-temperature and the rain rate ($\rho = -0.734$) shows that the colder or the higher the cloud in the troposphere, the higher the rain rate. The rain rate has a significant but quite low correlation with the mean and standard deviation of the temperatures of the surrounding 5×5 pixels. The mean and standard deviation of the pixels can be assumed to grossly represent the profile and texture of the convective cloud cell. Table 1 shows the details of the correlation results.

Sobel operator
As described by van Hees et al. (1999), convective overshoots are visible in high resolution IR images (such as the
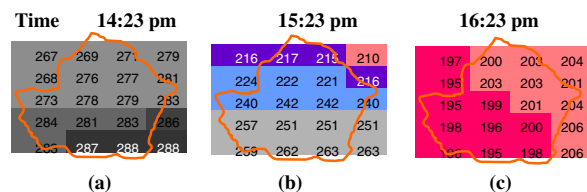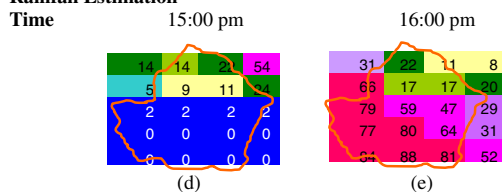
Figure 14 Areal rainfall estimation over the upper Klang River Basin.

AVHRR images) as very cold regions of about 10-km diameter and provides direct evidence for active deep convection. Though the images used in this study have spatial resolution of around 7 km × 5 km, the convective overshoot features are attempted to be represented by using sobel operator. The sobel value that indicates temperature gradient is calculated for the surrounding pixels and the maximum value is recorded. A correlation analysis between the maximum sobel value and the rain rate shows a significant positive but weak coefficient value (ρ = 0.413).

The NWP model outputs—NCEP reanalysis data

In addition to the five variables with significant correlations with the rain rate, three more variables are considered as inputs to the ANN. These variables are included to take into account the convective rain-formation process which basically involves precipitable water content and vertical wind, as described by Fox and Collier (2000). The forecast values of the PW, RH and VW from the NCEP reanalysis data are applied. Although there is no significant direct correlation between the hourly rain rate and the forecasted values of PW, RH and VW (reason could be due to the coarse spatial and temporal resolution of the NWP model output), there is a significant correlation between the RH and the mean 5 × 5 pixel temperature. A good correlation exists between the RH and PW values and a significant correlation is found between the VW and RH as displayed in Table 1. In other words, there is an inherent or indirect significant correlation between these input variables with the targeted output. In the ANN model development, the best input-parameters
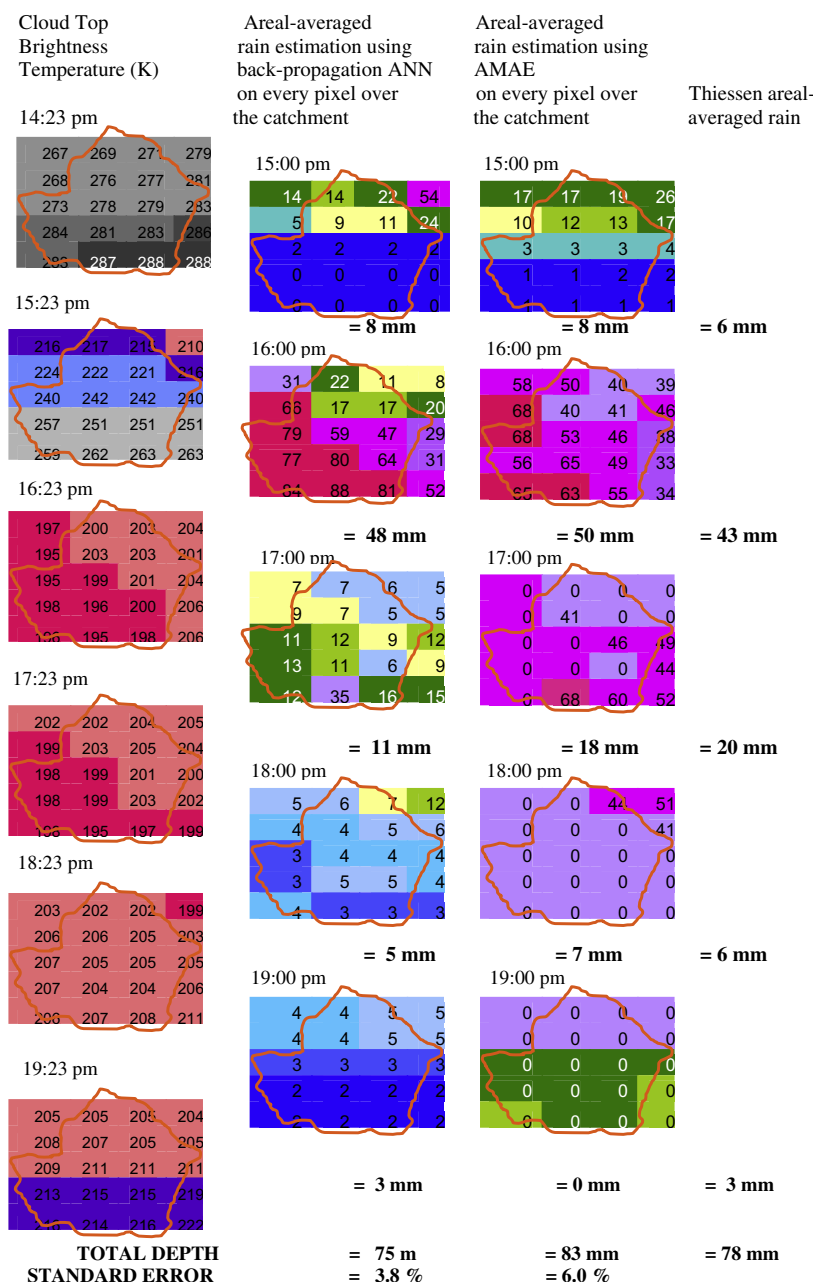
Figure 15   Example of June 10, 2003 (flood event) rain estimation using AMAE and ANN techniques.

Figure 16 Validation of the hourly areal-averaged rainfall estimation by the ANN model against the gage-measured Thiessen areal-averaged rain.

Figure 17 Validation of the total areal-averaged rainfall estimation by the ANN model against the gage-measured Thiessen areal-averaged rain.

combinations to the ANN have been tried using the lowest combination inputs, and then the combination inputs are increased step-by-step until the best performance is achieved.

The multi-layer back-propagation ANN model
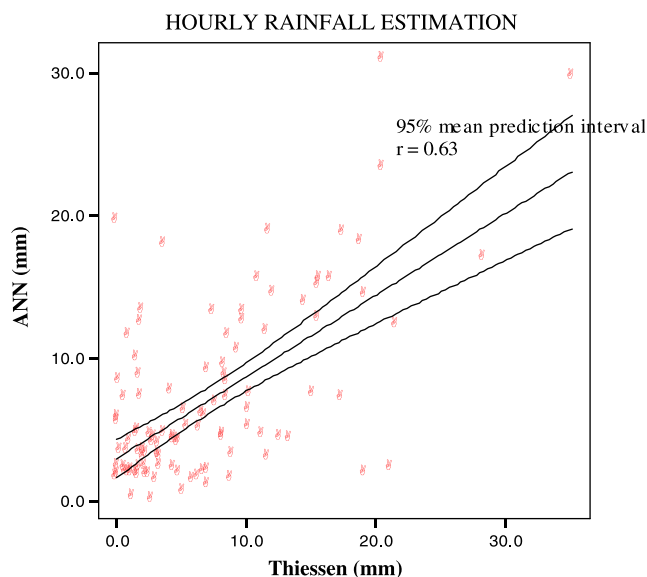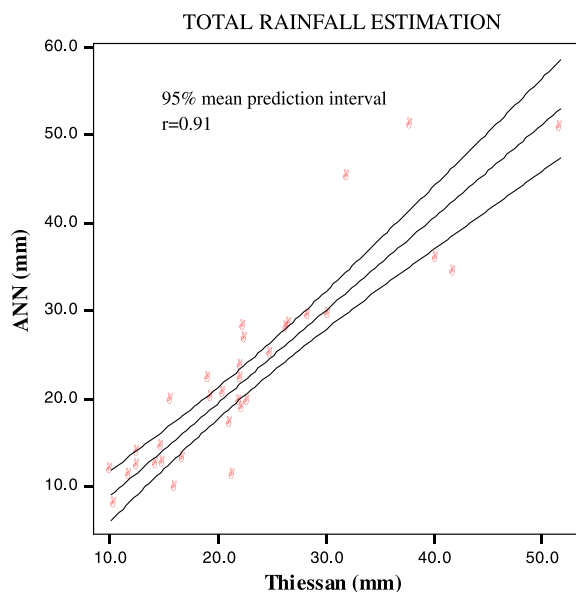
This study uses the multi-layer feed forward-type network with back-propagation learning algorithm in the development of a satellite rainfall-estimation model. The input variables are entered as an array to each neuron in the input

Table 2 Accuracy measures

	ANN versus Thiessen		AMAE versus Thiessen	
	Hourly	Total	Hourly	Total
Correlation coefficient, r	0.63	0.91	0.53	0.77
Mean absolute error (MAE)	3.8	3.3	4.6	5.9
Mean squared error (MSE)	30.9	22.3	42.5	51.2
Root mean squared error (RMSE)	5.6	4.7	6.5	7.2

layer. The values are then transmitted to the next layer (hidden layer) through links after being multiplied with the weights, w, associated with each link. At each neuron in the hidden layer, the weighted transmitted values are summed up together with a weighted bias, b. A suitable transfer function is next used by the results to generate activity levels for the neuron. Subsequently, the output layer receives these activation levels through various links, and the steps are repeated as before. This training process requires a set of examples of proper network behaviour — network inputs, p, and target outputs, t. During training, the weights and biases of the network are iteratively adjusted to minimize the average squared error between the network outputs, a, and the target outputs, t.

The input variables are trained against the radar rain rate. About 204 sets of data have been retrieved for convective rain events and are used in the training process. The optimum ANN structure was determined by varying the number of neurons in each layer to determine the structure that gives the best model performance. The final network consists of an input layer with eight input variables, one hidden layer with three neurons, and one output layer. The transfer functions are log-sigmoid and linear functions for the hidden and the output layers respectively, as indicated in Fig. 13. The output of the model is the rain rate (mm/h) at the collocated time of the satellite images.

Model application

The ANN model for satellite rainfall estimation is applied to estimate the areal rainfall in the upper Klang River Basin by repeated runs on every pixel covering the area. An example is shown in Fig. 14 for the rainfall estimation during a flash-flood event, dated June 10, 2003 over the catchment areas. The first row consists of three pixel representations over the catchment, which have been arranged in time sequence with the cloud-top brightness-temperature assigned using a program developed using MATLAB™ image processing tool. The second row is the estimated rainfall rate for each pixel for the following durations: (d) between 14:23 and 15:23 (e) between 15:23 and 16:23. Note that one of the input variables to the ANN model is the temperature change per hour, thus two images in time sequence are required to calculate the variable. Since the output from the model is the rain rate for each pixel in mm/h, the areal-averaged rain depth
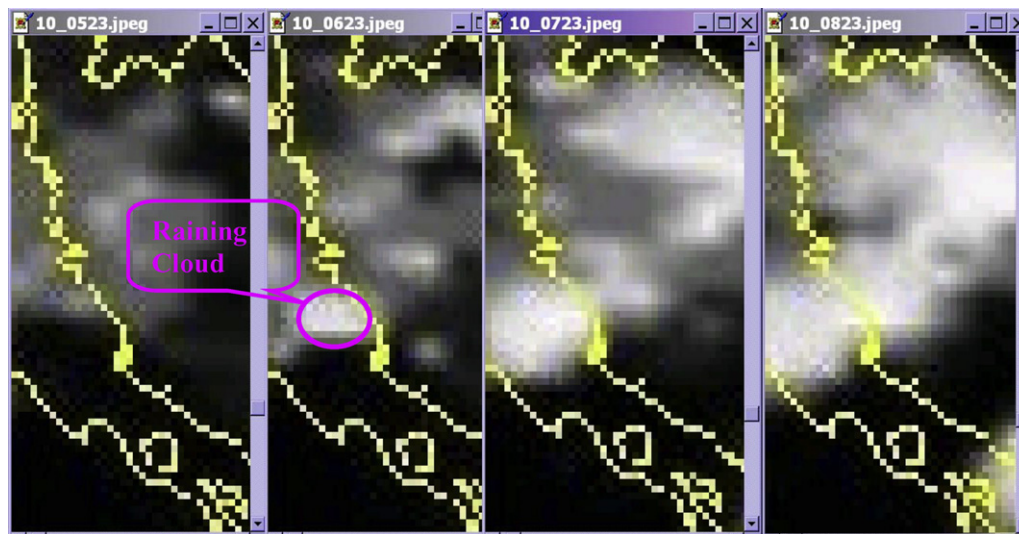
Figure 18   Sequential IR images at 1-h intervals obtained on June 10, 2003.
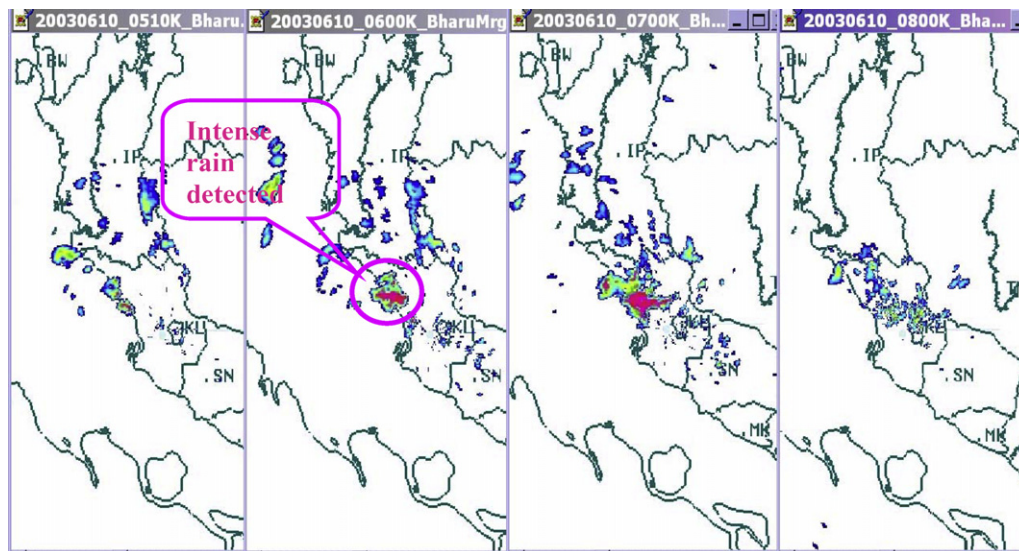
Figure 19   Sequential radar displays at 1-h intervals taken on June 10, 2003.

for the corresponding hour can then be estimated by arithmetic average of all the pixels involved. The results are shown below the diagram.

Fig. 15 illustrates the areal rainfall depth estimation over the upper Klang River Basin using the two techniques, which are subsequently compared with the recorded Thiessen areal-averaged rainfall. Apparently, both of the techniques encouragingly estimate the flash-flood event on June 10, 2003 with total standard errors of 6.0% and 3.8% for the AMAE and the ANN techniques, respectively.

Model validation

Coefficient correlation

This study advocates the use of gaged rainfall measurements as the 'truth'; hence, it uses them for validation pur-

poses. Hourly rainfall data have been collected from the DID of Malaysia for ten stations in the upper Klang River Basin, as shown in Fig. 3. Convective rainfall events of the heavy category (>30 mm/h) for at least one station are selected from the year 2006 and from the wet months of 2005 for validation purposes. A total of 107 hourly rainfall data sets from 33 storm events have been selected. To measure the closeness of the satellite rainfall estimation values obtained by the ANN model to the gaged rainfall values, the coefficient correlation, r, has been determined. Fig. 16 is the plot of the hourly areal-averaged rain estimation using the ANN model versus the gage-measured Thiessen areal-averaged rain. An r value of 0.63 is obtained, which indicates that there is no strong correlation between these two sets of values. The results are very much improved for the total areal-averaged rain depth estimation. Total, in this instance, implies the cumulative rainfall within the storm duration. The storm events included here are 2-h, 3-h, and 4-h rainfall.

a

Raining cloud cell

218	216	230	230	210	239	251	251	251	258	265	267	269	271	279	280	281
220	204	205	215	226	218	234	252	255	258	267	268	276	277	281	284	286
210	201	205	215	223	221	213	248	251	264	269	273	278	279	282	285	287
202	205	201	209	206	213	227	210	259	259	275	284	281	283	286	288	289
202	200	206	204	200	212	208	249	207	269	274	283	287	288	288	289	289
244	246	242	245	247	248	249	254	222	275	283	292	284	283	282	281	280

b

0.36	0.23	0.47	0.69	0.54	0.30	0.24	0.06	-0.29	-0.57
0.48	0.48	0.45	0.62	0.62	0.38	0.05	-0.13	-0.33	-0.54
0.42	0.38	0.63	0.79	0.73	0.45	0.05	-0.11	-0.22	-0.48
0.32	0.33	0.62	0.76	0.81	0.53	0.41	0.16	0.02	-0.39
0.49	0.24	0.26	0.48	0.60	0.38	0.56	0.27	0.21	-0.24
0.08	-0.28	-0.27	-0.30	0.24	0.09	0.49	0.32	0.34	-0.22
-0.12	-0.20	-0.30	-0.32	0.01	-0.07	0.32	0.44	0.49	0.01

c

215	218	214	213	213	217	221	220	214	216	219	216	217	215	210	210	221
206	212	199	204	207	212	219	215	214	223	227	224	222	221	216	215	223
195	197	200	200	203	207	216	215	223	232	236	240	242	242	240	239	244
195	195	195	195	195	195	203	218	227	246	249	257	251	251	251	250	253
195	195	195	196	196	195	201	224	215	251	253	259	262	263	263	263	264
195	196	195	199	195	197	202	231	211	262	266	268	270	270	270	270	271

Figure 20 (a) Cloud-top brightness-temperature at 0623 UTC showing the selected cloud cell; (b) values of correlation coefficient, *r*; (c) cloud-top brightness-temperature at 0723 UTC showing the direction of the cell movement.

Fig. 17 shows the plot of total areal-averaged rain as estimated by the ANN model against the gage-measured rainfall. The increased convergence of the points toward the linear graph in the figure and the '*r*' value of 0.91 indicate that the model performs better for a longer duration of estimation.

Other accuracy measures

Other accuracy measures carried out in the study are the determination of mean absolute error (MAE) and root mean squared error (RMSE). Table 2 shows that the values of MAE and RMSE for the hourly estimation using the ANN model are 3.8 and 5.6, respectively, whereas the same values for the total rainfall estimation are 3.3 and 4.7, respectively. The values are considered satisfactory within a corollary that better estimation is expected for longer time durations.

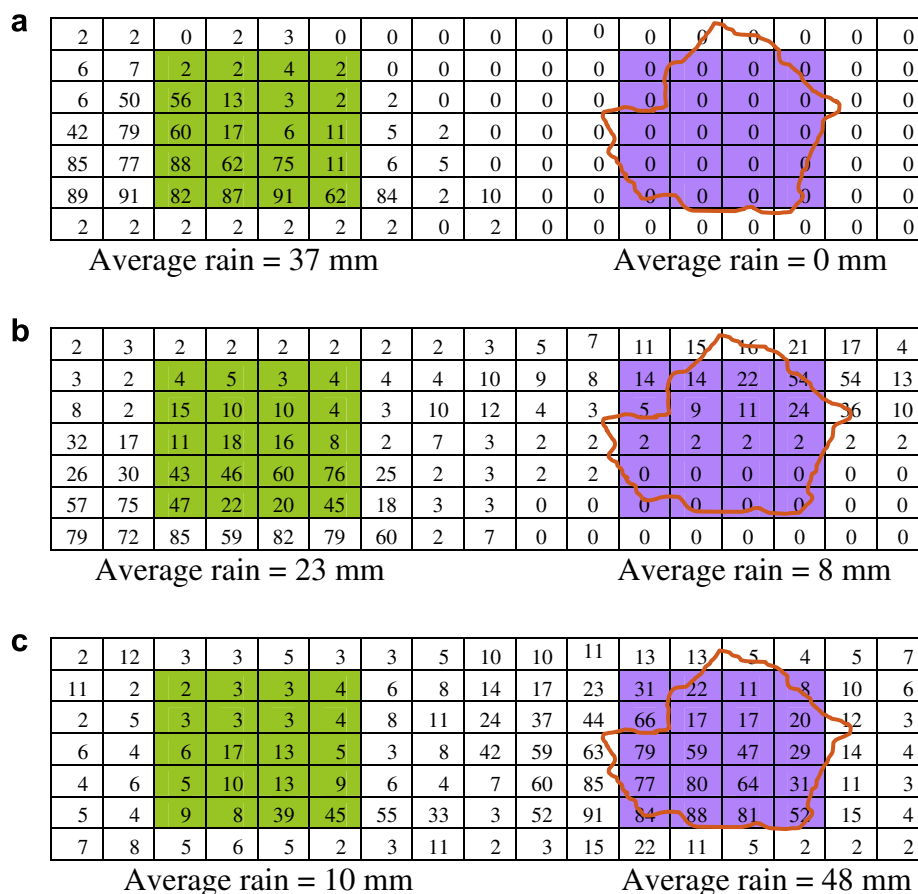## Application of the model for flash-flood forecasting

The satellite-based rainfall-estimation model has been validated against the rainfall gage measurement. The satisfactory validation results indicate that the model/technique is reasonably good. In addition to being a significant complementary tool to rain-gage measurement and radar estima-

tion, satellite-based rainfall estimation can be enhanced as a quantitative precipitation forecast (QPF). The rainfall *estimation* can become a rainfall *forecast* only when the future movement of rain can be predicted. The QPF can then be considered as an input to a flood forecasting system so as to gain an improved lead time. In this study, a cross-correlation technique is applied to track the cloud cell movement to develop a QPF.

Application of cross-correlation technique to forecast rainfall

Tracking cloud cell movement can be done using the cross-correlation method as described in Kidder and Vonder Haar (1995). The method allows for calculation of the average motion of an area of clouds in which two time sequence images are needed. A target array which contains the convective cloud cell is selected for the first image. Then, assuming that the clouds have moved, but changed little, during the time interval between the time sequence images, the cross-correlation technique is applied to locate this area in the second image.

Using a cross-correlation technique, the cloud cells are tracked during their movement, and rain can be forecasted about 2 h earlier, as described in a case study of the June 10, 2003 flash-flood event. Fig. 18 shows four sequential IR images obtained at 1-h intervals, starting from 0523

a

2	2	0	2	3	0	0	0	0	0	0	0	0	0	0	0	0	0
6	7	2	2	4	2	0	0	0	0	0	0	0	0	0	0	0	0
6	50	56	13	3	2	2	0	0	0	0	0	0	0	0	0	0	0
42	79	60	17	6	11	5	2	0	0	0	0	0	0	0	0	0	0
85	77	88	62	75	11	6	5	0	0	0	0	0	0	0	0	0	0
89	91	82	87	91	62	84	2	10	0	0	0	0	0	0	0	0	0
2	2	2	2	2	2	2	0	2	0	0	0	0	0	0	0	0	0

Average rain = 37 mm   Average rain = 0 mm

b

2	3	2	2	2	2	2	2	3	5	7	11	15	16	21	17	4
3	2	4	5	3	4	4	4	10	9	8	14	14	22	54	54	13
8	2	15	10	10	4	3	10	12	4	3	5	9	11	24	36	10
32	17	11	18	16	8	2	7	3	2	2	2	2	2	2	2	2
26	30	43	46	60	76	25	2	3	2	2	0	0	0	0	0	0
57	75	47	22	20	45	18	3	3	0	0	0	0	0	0	0	0
79	72	85	59	82	79	60	2	7	0	0	0	0	0	0	0	0

Average rain = 23 mm   Average rain = 8 mm

c

2	12	3	3	5	3	3	5	10	10	11	13	13	5	4	5	7
11	2	2	3	3	4	6	8	14	17	23	31	22	11	8	10	6
2	5	3	3	3	4	8	11	24	37	44	66	17	17	20	12	3
6	4	6	17	13	5	3	8	42	59	63	79	59	47	29	14	4
4	6	5	10	13	9	6	4	7	60	85	77	80	64	31	11	3
5	4	9	8	39	45	55	33	3	52	91	84	88	81	52	15	4
7	8	5	6	5	2	3	11	2	3	15	22	11	5	2	2	2

Average rain = 10 mm   Average rain = 48 mm

Figure 21   (a) The rain estimation using the ANN model (a) at 0623 UTC, (b) at 0723 UTC, and (c) at 0823 UTC.

UTC (or 11:23 am Malaysian time). At 0623 UTC, a raining cloud cell not far from the upper Klang River Basin is detected by its distinct features and is confirmed by the radar display in Fig. 19.

Fig. 20a shows the approximated values of the cloud-top brightness-temperature at 0623 UTC showing the selected cloud cell. Cross-correlation between the raining cloud cells for the time 0623 UTC with the surrounding pixel matrix on the next image (0723 UTC) produces the coefficient correlation, r, which is greatest in the horizontal direction indicating some movement, as shown by the arrow in Fig. 20b. Fig. 20c shows the cloud-top brightness-temperature at 0723 UTC and additionally shows the direction of movement of the cell that gradually approaches the upper Klang River Basin. The rain from this raining cloud cell is estimated using the ANN-based technique, and the estimation values are shown in Fig. 21a.

Within an hour, the cell has moved horizontally to the right. Assuming constant velocity of cell movement, it is forecasted that the raining multi-cell clouds will move to the upper Klang River Basin around 0823 UTC. Fig. 21b and c show the rainfall estimates at 0723 and 0823, respectively.

The analysis of this case study indicates that the raining cloud cell detected on the 0623 UTC image could be a part of convective multi-cell clouds, which is gradually proceeding in a certain direction and bringing along the rain. A total amount of 60 mm (37 + 23 mm) of areal-averaged rainfall depth within 2-h duration is estimated to fall from the first raining cloud over the area beneath. Through the cross-correlation technique, it is forecasted that approximately the same amount of rain is expected to fall on the upper Klang River Basin within the following 2 h. Using the forecasted rain as the input for the rainfall-runoff model, the flood flow can be forecasted ahead of time.

Derivation of the catchment-average unit hydrograph (UH)

Ponce (1989) wrote that area sizes ranging from 100 to 5000 km^2 have been used to define the limit between mid-size and large catchment. The upper Klang River Basin which comprised of an area of 468 km^2 can be categorized as a midsize catchment. Uniform spatial rainfall distribution is an assumption that characterizes a midsize catchment analysis. Following this assumption, a lumped method such as unit hydrograph (UH) is deemed suitable to be adopted. Ponce (1989) also stressed that the unit hydrograph technique is a defacto standard of hydrograph generation for midsize catchment, that has been used extensively worldwide.

Rainfall and runoff data from the year 2003 to 2006 for the stations shown in Fig. 4 have been analyzed to select good single-peak rainfall events for the derivation of a catchment-average UH. The resultant 2-h duration catch-
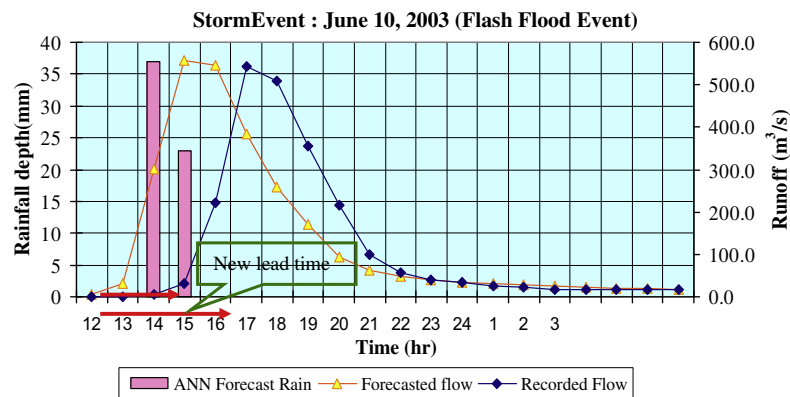
Figure 22 Forecasted flow plotted against the observed flow with the new lead time indicated.

ment-average UH has been applied as the rainfall-runoff model in this study.

Coupling the rainfall forecast with the catchment-average UH for flash-flood forecasting

The rain amount forecasted from the satellite-based ANN rainfall-estimation technique is next assessed for effective rainfall after a consideration of the infiltration processes involved. Through the analysis of rainfall-runoff data from the year 2003 to 2006, the phi-index for the upper Klang River Basin is estimated to be within the range of 0.35—2.0 cm/h. The phi-index increases with an increase in rainfall intensity. Using the maximum value of the phi-index for the June 10, 2003 flash-flood event, the forecasted rain depth (6.0 cm) is deducted by the infiltration amount to obtain the effective rainfall as shown next:

$$\text{Effective rain} = \text{Rain depth} - (\text{infiltration depth})$$
$$= 6.0\,\text{cm} - (\text{phi-index} \times \text{duration})$$
$$= 6.0\,\text{cm} - (2.0\,\text{cm/h} * 2\,\text{h})$$
$$= 2.0\,\text{cm}.$$

The ordinates of the UH are multiplied with the 2-h effective rainfall to obtain the excess runoff. The excess runoff ordinates are then added to the estimated baseflow to produce the direct runoff. This forecasted direct runoff is compared with the observed flow hydrograph of June 10, 2003. Fig. 22 shows the hydrographs of the forecasted flow versus the actual observed flow. At time 16:00 h, it can be forecasted that a peak flow of around 500 m^3/s is about to occur at the flow-gaging station 2 h later (around 18:00 h). The flow record obtained from the DID shows that an actual peak flow of around 556 m^3/s had actually occurred at about 18:00 h; therefore, an extra lead time of 2 h has been gained using the satellite-based rainfall forecast. Consequently, a total of approximately four (4) hours of lead time (from the time of rain centroid) has been attained by coupling the satellite-based rainfall forecast and the UH technique.

The satellite-based ANN rainfall estimation is applied to the rain forecast for three other flash-flood events that occurred at the upper Klang River basins and the results indicate that a lead time of around 3—4 h can be obtained

Table 3 Forecasting the average catchment-rain depth using satellite-based ANN rainfall estimation for other flash-flood events

Date (h)	Average catchment-rain depth forecasted (mm)	Lead time
May 10, 2006	31	3.5
August 24, 2006	50	3
August 25, 2006	45	3

by coupling this model and the average catchment UH as given in Table 3.

Concluding remarks

In this work, a satellite-based quantitative rainfall estimate developed using the artificial neural network has been detailed. The ANN technique has been selected due to its capability to recognize differing patterns of relationships and function approximations between combinations of factors (input) against a target value (output). The model performed satisfactorily in estimating the areal-averaged rainfall depth for convective rain events in a given duration, and the results output were validated against the gaged rainfall. A value of 4.7 for the RMSE can be considered as a typical magnitude for an error in estimation (Wilks, 1995). Hence, if the rain estimation is 30 mm, and error of 16% is expected in the estimation.

Model application and certainty

The satellite-based quantitative ANN rainfall forecast has been coupled with a UH rainfall-runoff model to forecast flood. In summary, an extra 1—2 h lead time has been gained in addition to the 2-h lead time if using the UH technique only. As a result, a total of 3—4 h of lead time for flash-flood forecasting can be achieved for the upper Klang River Basin, which is known for its very quick response to intense rainfall events.

The main contributions of this work are the exploratory and correlation analysis between the geostationary meteorological satellite IR images and intense convective rain for an equatorial country such as Malaysia and the development

of ANN satellite-based rainfall-estimation models incorporating the knowledge from actual observations. Further comprehensive work is required for the establishment of these rainfall-estimation models as reliable quantitative rainfall forecasts that can be used as inputs to a flood forecasting-and-warning system with a better lead time.

Acknowledgements

References

Adler, R.F., Negri, A.J., 1988. A satellite infrared technique to estimate tropical convective and stratiform rainfall. J. Appl. Meteorol. 27, 30—51.

Anagnostou, E.N., Negri, A.J., Adler, R.F., 1999. A satellite infrared technique for diurnal rainfall variability studies. J. Geophys. Res. 104, 31477—31488.

Arkin, P.A., 1979. The relationship between fractional coverage of high cloud and rainfall accumulations during GATE over the B-scale array. Mon. Weather Rev. 106, 1153—1171.

Ba, M.B., Nicholson, S.E., 1998. Analysis of convective activity and its relationship to the rainfall over the Rift Valley lakes of East Africa during 1983—90 using the Meteosat infrared channel. J. Appl. Meteorol. 37, 1250—1264.

Bankert, R.L., 1994. Cloud classification of a AVHHR imagery in maritime regions using a probabilistic neural network. J. Appl. Meteorol. 33, 909—918.

Bellon, A., Lovejoy, S., Austin, G.L., 1980. Combining satellite and radar data for the shortrange forecasting of precipitation. Mon. Weather Rev. 108, 1554—1556.

Fox, N.I., Collier, C.G., 2000. Estimating medium-range catchment flood potential. J. Hydrol. 237, 1—16.

Griffith, C.G., Woodley, W.L., Grube, P.G., Martin, D.W., Stout, J., Sikdar, D.N., 1978. Rain estimation from geo synchronous satellite imagery — visible and infrared studies. Mon. Weather Rev. 106, 1153—1171.

Hagan, M.T., Demuth, H.B., Beale, M., 1996. Neural Network Design. Thomson Publishing, pp. 19—21.

Kidder, S.Q., Vonder Haar, T.H., 1995. Satellite Meteorology, An Introduction. Academic Press.

Kim, G., Barros, A.P., 2001. Quantitative flood forecasting using multisensor data and neural networks. J. Hydrol. 246, 45—62.

Kuligowski, R.J., Barros, A.P., 1998. Experiments in short-term precipitation forecasting using artificial neural networks. Mon. Weather Rev. 126, 470—482.

Lovejoy, S., Austin, G.L., 1979. The delineation of rain areas from visible and IR satellite data from GATE and mid-latitudes. Atmos. Ocean 17, 77—92.

Minnis, P., Harrison, E.F., 1984. Diurnal variability of regional cloud and clear-sky radiative parameters derived from GOES data. Part II: November 1978 cloud distribution. J. Appl. Meteorol. 23, 1012—1031.

Nakakita, E., 2002. Radar hydrology precipitation and water resources. Textbook for 12th IHP Training Course, HyARC Nagoya University and UNESCO, pp. 111—129.

Negri, A.J., Adler, R.F., Wetzel, P.J., 1984. Rain estimation from satellite: an examination of the Griffith—Woodley technique. J. Clim. Appl. Meteorol. 23, 102—116.

Ponce, V.M., 1989. Engineering Hydrology: Principles and Practices. Prentice Hall, pp. 154—204.

Porcù, F., Borga, M., Prodi, F., 1999. Rainfall estimation by combining radar and infrared satellite data for nowcasting purposes. Meteorol. Appl. 6, 289—300.

Reed, R.J., Jaffe, K.D., 1981. Diurnal variation of summer convection over West Africa and the tropical eastern Atlantic during 1974 and 1978. Mon. Weather Rev. 111, 2527—2534.

Rogers, R.R., Yau, M.K., 1988. A Short Course in Cloud Physics. Butterworth-Heinemann, p. 290.

Sorooshian, S., Hsu, K.-L., Gao, X., Gupta, H.V., Imam, B., Braithwaite, D., 2000. Evaluation of PERSIANN system satellite—based estimates of tropical rainfall. Bull. Am. Meteorol. Soc. 81, 2035—2046.

Todd, M.C., Kidd, C., Kniveton, D., Bellerby, T.J., 2001. A combined satellite infrared and passive microwave technique for estimation of small-scale rainfall. J. Atmos. Ocean. Technol. 18, 742—755.

van Hees, R.M., Lelieveld, J., Collins, W.D., 1999. Detecting tropical convection using AVHRR satellite data. J. Geophys. Res., 9213—9228.

Vicente, G.A., 2001. Satellite rainfall estimation for flash flood application European basic auto estimator within the frame of the SAFNWC. Visiting Scientist's Report, NASA/GSFC Distributed Active Archive Center DAAC, USA.

Wiesner, C.J., 1970. Hydrometeorology. Chapman and Hall, London, p. 232.

Wilks, D.S., 1995. Statistical Methods in the Atmospheric Science. Academic Press, pp. 233—281.