QUANTITATIVE PRECIPITATION FORECAST USING NUMERICAL WEATHER PREDICTION AND METEOROLOGICAL SATELLITE FOR KELANTAN AND KLANG RIVER BASINS

Wardah Tahir*, Ahmad Kamil Aminuddin, Suzana Ramli, Jurina Jaafar
Faculty of Civil Engineering, Universiti Teknologi MARA, Shah Alam, Selangor, Malaysia

Abstract

The unusual heavy rainfall episodes over Kelantan River Basin in 2014 had caused massive destruction and several deaths. The unprecedented storm events at the north-eastern Peninsular Malaysia and many other places indicate the need for enhanced storm forecasting to improve disaster preparedness among the civilian. Quantitative precipitation forecast (QPF) from atmospheric model combined with geostationary meteorological satellite information as input to hydrodynamic model for flood forecasting system can potentially provide improved lead time for warning. In this study, a QPF model is developed using the multilayer neural network with data inputs from the numerical weather prediction (NWP) model products combined with the geostationary meteorological satellite infrared and visible image features to forecast precipitation for a flood-prone area in a tropical region. The results indicate that the model can satisfactorily produce 1-hour rainfall forecast with improved accuracy for larger forecast area. The R² for areal average rainfall for Kelantan river basin is 0.674 and for Klang river basin is 0.893 whereas the R² for point rainfall is 0.392 for Kelantan river basin and 0.495 for Klang river basin.

Keywords: Numerical weather prediction (NWP), quantitative precipitation forecast (QPF), geostationary meteorological satellite (METSAT), artificial neural network (ANN)

Accepted 10 December 2016

*Corresponding author
warda053@salam.uitm.edu.my
1.0 INTRODUCTION

The extreme heavy downpour over Kelantan River Basin Malaysia, in December 2014 had caused unprecedented flood disaster causing massive damages estimated to be more than RM1 billion, seven (7) reported deaths and evacuation of more than 150,000 people. The unusually heavy and continuous rainfall over the upstream of the basin had totalled up to more than half of average annual rainfall within only a 3-day period. The unexpected storm pattern had proven that rainfall is indeed one of the most difficult variables to forecast due to its inherent variability and complex atmospheric processes which lead to its development. Being one of the major factors causing flood disasters, there should be more effort done to enhance rainfall forecast accuracy.

In the recent years, there have been increasing efforts to improve the rainfall forecast ability and accuracy. One of the most significant and potential tools is using the high resolution non-hydrostatic mesoscale model of numerical weather prediction models (NWP). The dynamical meteorology model provides the equations representing the development processes of the atmosphere and uses numerical approximations to predict the future states of the atmospheric circulation from the knowledge of its present state. The initial boundary value inputs describe the current state of the atmosphere which represents many different characteristics of the atmosphere such as: humidity, temperature, wind velocity, pressure, and other aspects of the region for forecast. Several modeling systems have been implemented as global, hemispheric or limited area models (LAMs). LAMs run with a higher resolution over a smaller area and take boundary conditions from a larger hemispheric or global model.

During the last decades, several regional LAMs have been developed such as the MM4 (Fourth-Generation Penn State/NCAR Mesoscale Model) and later the MM5 (Fifth-Generation Penn State/NCAR Mesoscale Model) [1] and the new Weather Research and Forecasting (WRF) model [2]. Today, NWP is the most widely used prediction system, and can predict future states for up to 10 days. However, in spite of a generally very high quality product, NWP models occasionally fail to accurately predict intense precipitation particularly on the small scales. Therefore, use of other additional information such as Infrared (IR) and Visible (VIS) data from the geostationary meteorological satellite images may improve the rainfall forecast accuracy.

1.1 NWP Model Forecast

Evaluation of precipitation forecasts is important to monitor forecast quality over space and time, to compare the quality of different forecast systems and to enhance forecast quality [3, 4, 5, 6]. Evaluation of NWP model forecasts of precipitation is not a new topic. A number of authors have verified precipitation forecasts from a meteorological point of view as shown from the work of [7]. However several other authors have evaluated precipitation forecasts from a hydrological perspective as described by [8]. Other research [9] had highlighted the relevance of precipitation forecasts products to real-time hydrological forecasting. Golding [10] identified the critical areas where NWP products fall short, and discussed techniques being developed to improve them. Damrath et al. [11] verified 7 year of precipitation forecasts from NWP models of the German Weather Services. In other studies, [12] had compared four European and Canadian mesoscale models for precipitation forecasting to reproduce heavy precipitation events while [13] used precipitation forecasts from two French NWP models as inputs to a hydrologic model. The spatial and temporal variation in the skill of precipitation forecasts from a NWP model had been assessed by [14] who later demonstrated the benefit of using high resolution NWP model precipitation forecasts for flood and short-term streamflow forecasting. Chile and Schulze [15] had assessed the performance and accuracy of the precipitation forecasts by three NWP models over the Mgeni catchment in South Africa. Finally, a comprehensive review and the state of art in forecast verification had been discussed by [7].

1.2 Tropical Rain Type Classification

In general, tropical precipitation has two main classifications: convective and stratiform [16]. Stratiform rain results from mid latitude frontal systems, convergence into lows, up slope flow, and all situations in which the precipitation forms in a stable atmosphere. It tends to have less intense rain than convective rain and also tends to last longer. On the other hand, convective precipitation is formed in unstable atmosphere that form convective clouds e.g cumulonimbus or cumulus congestus.

1.3 Flood Forecasting

Flood forecasting systems that integrate the hydrological with the atmospheric model are now operational in many areas. The lead time between occurrence of a flood event and warning can be extended by coupling atmospheric and the hydrological models as indicated among others by [17] and [18]. Jasper et al. [17] have coupled the grid-based hydrological catchments model with forecast data from five high-resolution numerical weather prediction (NWP) models with grid cell sizes between 2 and 14 km while [18, 19] have developed a QPF using the infrared satellite images combined with NWP model products using a neural network model. Habets et al. [13] have used QPF for daily stream flow prediction over the Rhone river basin, France. The precipitation forecast is fed to a one-way atmosphere–hydrology coupled model to predict the river flow. A QPF model to forecast flood up to 24 hour
in advance, which use both NWP output, rainfall and radiosonde data has been detailed by [20]. The integrated model seems to be very comprehensive and can function as a reliable flood forecasting system. Kim and Barros [21] discussed that the complexity of handling the high thresholds and rare events is a strong limitation for operational activities using NWP, particularly for flood forecasting. In addition, the rainfall location, magnitude and time depend on how the used numerical model is able to determine the size, scale and the evolution of atmospheric systems involved. Though many studies have been done on the effectiveness of NWP models in producing QPF, several researchers comment that the use of NWP models alone do not seem to be able to provide accurate rainfall forecasts at the temporal and spatial resolution required by many hydrologic applications as described in [22].

1.4 Objective of the Study

The accuracy of quantitative precipitation forecast produced by the Malaysian Meteorological Department (MMD) is still lacking even though significant progress has been made on the technical aspects [23]. The advancement of science and technology should enable improve nowcasting capabilities for more timely, accurate and meaningful precipitation forecasting in the country, as put forth by [23]. The sparseness of ground based observations, missing records and uneven distribution of the existing rain gauge network does not provide adequate and timely information about precipitation pattern in Malaysia. Therefore the study is an effort to develop improved QPF using NWP model products and additional information from the geostationary satellite images in an artificial neural network (ANN) based model.

1.5 Study Area

Peninsular Malaysia is situated in the tropics between 1° and 7° north of the equator and at eastern longitude from 100° to 103° E. The climate of Peninsular Malaysia is characterized by uniform temperature between 21 to 32 C and very much influenced by the monsoons. The southwest monsoon (SWM) occurs from May to August while the northeast monsoon (NEM) occurs from November to February. The period of the SWM is a drier period for the whole country, while during the NEM, the eastern areas of Peninsular Malaysia receive heavier rains than the other parts of the country. Heavy rainfalls are also expected during the two inter-monsoons: March–April and September–October. The study will focus on two major river basins in Peninsular Malaysia, namely Kelantan and Klang River Basin. Figure 1 shows the location of the two river basins.

Kelantan River basin which is located at the northeastern part of Peninsular Malaysia receives annual rainfall of about 2700 mm during the northeast monsoon between October and January. Kelantan River system flows northward passing through major towns such as Pasir Mas and Kota Bharu, before finally discharging into the South China Sea. Klang River Basin is located at the south-western part of Peninsular Malaysia and one of the busiest areas in Malaysia. The area receives annual mean rainfall around 1900 mm to 2600 mm. Klang river system passes through the capital city of Malaysia, Kuala Lumpur and flowing southward to the Straits of Malacca.

1.6 Backpropagation Neural Network

The backpropagation 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 known 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 called “back algorithm” and is based on the generalized delta rule.
The generalized delta rule, which determines the approximate weight adjustments necessary to minimize the errors can be explained through Figure 2 which shows a neuron (j) and its functions. The total input $P_j$, to hidden units $j$ is a linear function of inputs $x_i$ of the units that are connected to $j$ and of the weights $w_{ij}$ on these connections i.e.:

$$P_j = \sum_i W_{ij} X_i$$  \hspace{1cm} (1)$$

Neurons can be given biases ($\theta_j$) by introducing extra input to each unit which always has a value of 1. A hidden unit has a real-value output $P_{oj}$, which is a non-linear function of the total input, which can be as equation (2) if sigmoidal transfer/activation function is used.

$$P_{oj} = \frac{1}{1+e^{-(P_j+\theta_j)}}$$  \hspace{1cm} (2)$$

The aim is to find a set of weights that ensure for each input vector, the output vector produced by the network is the same or sufficiently close to the desired output vector. If there is a fixed, finite set of input-output cases, the total error in the performance of the network with a particular set of weights can be computed by comparing the actual and desired output vectors for every case. The total error $E$, is defined as:

$$E = \frac{1}{2} \sum_{c} \sum_j (O_{jj} - T_{jj})^2$$  \hspace{1cm} (3)$$

Where $c$ is an index over cases (the input-output pairs); $j$ is the index over output units; $O$ is actual state of an output unit; and $T$ is its targeted stated state. To minimize $E$ by gradient descent, it is necessary to compute the partial derivative of $E$ with respect to each weight in the network, $\partial E/\partial W_{ij}$. This can be successively computed as follows:

Differentiating equation (3) for a particular case, $c$,

$$\frac{\partial E}{\partial O_j} = (O_j - T_j)$$  \hspace{1cm} (4)$$

Next $\partial E / \partial x_i$ is computed using chain rule:

$$\frac{\partial E}{\partial x_i} = \frac{\partial E}{\partial O_j} \frac{\partial O_j}{\partial x_i}$$  \hspace{1cm} (5)$$

Differentiating equation (3) to get the value of $\partial O_j / \partial x_i$ and substituting in equation (5):

$$\frac{\partial E_j}{\partial x_j} = \frac{\partial E}{\partial O_j} O_j (1-O_j)$$  \hspace{1cm} (6)$$

shows how the change in the total input $x_i$ to an output unit, will affect the error $E$. The total input is a linear function of the states of the lower level units and also linear function of the weights on the connections. It is, therefore, easy to compute how the error will be affected by changing these states and weights. For a weight $w_{ij}$ from $i$ to $j$ the derivative is

$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial x_j} \frac{\partial x_j}{\partial O_j} \frac{\partial O_j}{\partial w_{ij}}$$  \hspace{1cm} (7)$$

and for the output of the $i_{th}$ unit the contribution to $\partial E / \partial O_i$ resulting from the effect of $i$ on $j$ is simply

$$\frac{\partial E}{\partial x_j} \frac{\partial x_j}{\partial O_j} \frac{\partial O_j}{\partial w_{ij}} = \frac{\partial E}{\partial x_j} w_{ij}$$  \hspace{1cm} (8)$$

Taking into account all the connections emanating from unit $i$

$$\frac{\partial E}{\partial O_i} = \sum_j \frac{\partial E}{\partial x_j} w_{ij}$$  \hspace{1cm} (9)$$

Given $\partial E / \partial O$ for all units $j$, in the previous layer, the $\partial E / \partial O$ in the penultimate layer can be computed using equation (9). The same procedure can be repeated for the successive layers. The simplest version of gradient descent is to change each weight by an amount proportional to the accumulated $\partial E / \partial W$.

$$\Delta W = -\varepsilon \frac{\partial E}{\partial W}$$  \hspace{1cm} (10)$$

Where $\varepsilon$ is the learning rate. The convergence of equation (10) can be significantly improved, by acceleration method wherein the incremental weights at $t$ can be related to the previous incremental weights using equation (11).
\[
\Delta w(t) = e^{-\frac{\partial E}{\partial w(t)}} + \alpha \Delta w(t-1)
\]  

(11)

where \(\alpha\) is an exponential decay factor between ‘0’ and ‘1’ that determines the relative contribution of the current gradient and earlier gradients to the weight change.

## 2.0 METHODOLOGY

The methodology includes data processing from the NWP and METSAT and model development using ANN.

### 2.1 Rainfall Data

Hourly rainfall data for year 2009 have been obtained from the Drainage and Irrigation Department (DID) of Malaysia for 18 gauging stations in Kelantan and Klang River basin as shown in Figure 1 and 2. General characteristic of the rainfall stations for both catchments area such as ID numbers that are the key code indicated by DID and geographic coordinates of the stations are provided in Table 1.

### 2.2 Geostationary Meteorological Satellite Data

MTSAT/FY-2C hourly meteorological satellite images had been acquired from the Space Science and Engineering Center (SSEC), Wisconsin, USA. The SSEC received these images from the Japanese National Space Development Agency and China Meteorological Department. The images used for this study are the hourly FY-2C IR image (channel 10.8 \(\mu\)m) and MTSAT VIS image (channel 0.73 \(\mu\)m) with a spatial resolution of about 4 km.

The METSAT data processing was aided by McIDAS (Man computer Interactive Data Access System). The McIDAS is a suite of sophisticated software packages that perform a wide variety of functions with satellite imagery, observational reports, numerical forecasts, and other geophysical data. Those functions include displaying, analyzing, interpreting, acquiring and managing the data. Figure 3 shows an example of infrared image and visible image to be processed using McIDAS.

### 2.3 NWP Data

Data from MM5 and WRF models consist of atmospheric variables as listed in Table 2 were obtained from the Malaysian Meteorological Department (MMD). The data had been processed using the Grid Analysis and Display System (GrADS) software. Data output from the NWP model such as accumulated precipitable water was in units of millimeter of water and average relative humidity was in percentage. The selected output of the NWP products were from 00UTC and 12UTC for year 2009 with forecast range of hourly, 3 hourly, 6 hourly, 12 hourly and 24 hourly up to a period of 72 hours. The selected NWP model output cover area at lat 0.98 – 6.99N, lon 99.04 – 105.098 E at resolution of 4 km, and the area consists of 168 x 168 grid points. This high resolution data in temporal and spatial domains provide good opportunity to study the rainfall characteristics especially in the humid tropics. Figure 4 shows the display of 24-hr accumulated rain.

---

**Table 1** General information of selected gauging stations

<table>
<thead>
<tr>
<th>Number</th>
<th>Station IDs</th>
<th>Latitude °</th>
<th>Longitude °</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6019004</td>
<td>6.02</td>
<td>101.98</td>
</tr>
<tr>
<td>2</td>
<td>6021010</td>
<td>6.01</td>
<td>102.10</td>
</tr>
<tr>
<td>3</td>
<td>5921009</td>
<td>5.97</td>
<td>102.14</td>
</tr>
<tr>
<td>4</td>
<td>6120014</td>
<td>6.15</td>
<td>102.10</td>
</tr>
<tr>
<td>5</td>
<td>6021013</td>
<td>6.10</td>
<td>102.11</td>
</tr>
<tr>
<td>6</td>
<td>5821007</td>
<td>5.89</td>
<td>102.16</td>
</tr>
<tr>
<td>7</td>
<td>5923012</td>
<td>5.92</td>
<td>102.04</td>
</tr>
<tr>
<td>8</td>
<td>5820006</td>
<td>5.84</td>
<td>102.07</td>
</tr>
<tr>
<td>9</td>
<td>5719001</td>
<td>5.78</td>
<td>101.97</td>
</tr>
<tr>
<td>10</td>
<td>3116003</td>
<td>3.15</td>
<td>101.68</td>
</tr>
<tr>
<td>11</td>
<td>3116007</td>
<td>3.11</td>
<td>101.70</td>
</tr>
<tr>
<td>12</td>
<td>3216001</td>
<td>3.27</td>
<td>101.69</td>
</tr>
<tr>
<td>13</td>
<td>3216005</td>
<td>3.26</td>
<td>101.68</td>
</tr>
<tr>
<td>14</td>
<td>3216007</td>
<td>3.22</td>
<td>101.66</td>
</tr>
<tr>
<td>15</td>
<td>3217001</td>
<td>3.27</td>
<td>101.73</td>
</tr>
<tr>
<td>16</td>
<td>3217003</td>
<td>3.24</td>
<td>101.71</td>
</tr>
<tr>
<td>17</td>
<td>3217004</td>
<td>3.26</td>
<td>101.77</td>
</tr>
<tr>
<td>18</td>
<td>3217008</td>
<td>3.25</td>
<td>101.72</td>
</tr>
</tbody>
</table>

---

Figure 3 (a) FY-2C Infrared Image (b) MTSAT Visible Image

---

Figure 4 Display of 24-hr accumulated rain.
Table 2 List of atmospheric variables considered as possible predictors

<table>
<thead>
<tr>
<th>Atmospheric variables</th>
<th>Level (mb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relative humidity (%)</td>
<td>Average 1000 to 500</td>
</tr>
<tr>
<td>Accumulated total cumulus precipitation (mm)</td>
<td>-</td>
</tr>
<tr>
<td>Accumulated total grid scale precipitation (mm)</td>
<td>-</td>
</tr>
</tbody>
</table>

Figure 4 Sample of 24-hour accumulated rainfall data

2.4 Artificial Neural Network (ANN) Technique for QPF Model Development

The work applied a multilayer neural network as shown in Figure 5 as the model architecture combining three inputs from NWP model products and two parameters from the METSAT images (cloud top brightness temperature and albedo) to forecast rain of one hour ahead. There are input layer, hidden layer and output layer. Each layer consists of one or more neurons. There are two types of neuron. First are passive neurons that relay data input as data output. Another is active neuron that computes data input using Activation Transfer Function (ATF) and produces an output. The most commonly use of ATF in the hidden and output neuron is sigmoid function [24]. The input into an active neuron is a summation of previous neuron’s output and its weight and the output is a computation of sigmoid function on the input. 1779 data sets for Kelantan river basin and 918 for Klang river basin had been divided into training and testing sets to come up with the optimum architecture.

Figure 5 Multi-layer ANN model of QPF

3.0 RESULTS AND DISCUSSION

Cloud top brightness temperature and albedo values of the METSAT images over the Kelantan and Klang river basins were plotted against the areal average rainfall events. Example correlation can be observed from Figure 6 (a) and (b). The graphs show that the albedo increases and the temperature decreases, with increase in rainfall depth.

Albedo in the visible image is represented white for high reflectance and black for low reflectance. Bright clouds or high albedo in visible images are more likely to precipitate than dark cloud since the cloud thickness is related to its brightness in the visible image [25].

The taller clouds or higher tops indicating colder temperature will be brighter in infrared images. The images are represented black for high temperature and white for low temperature. Since thick and tall clouds characterize the cumulonimbus raining cloud, colder and thicker clouds (higher albedo) will give brighter images on infrared and visible images of METSAT and more likely to precipitate [25].
Figure 6 Rainfall depth versus (a) albedo (b) cloud top temperature

Figure 7 (a) and (b) and Figure 8 (a) and (b) indicate good correlation for areal average rainfall compared to point rainfall for both catchment areas for which $R^2$ for Kelantan river basin is 0.674 and for Klang river basin is 0.893. The use of QPF for point areal rainfall is found to be less accurate with $R^2$ for Kelantan river basin is 0.392 and for Klang river basin is 0.495. The results tallied with the findings in other studies using satellite based rainfall estimation [18, 19]. Satellite based rainfall estimation is more appropriate to be applied for areal average or basin-scale rainfall estimation due to the indirect estimation of high altitude geostationary meteorological satellite located around 36,000 km above the ground. The cloud position on the earth is better representing an area rather than a point although the displacement error and parallax corrections had been made as suggested by [25]. In addition, NWP rainfall forecasts are also considered too coarse in their spatial and temporal resolution, therefore more suitable to be applied for basin scale [18].

Table 3 shows that the values of MAE and RMSE for the areal rainfall for both catchments are relatively low than point rainfall with a range value of 2.589-3.436 for RMSE and 0.8651-1.872 for MAE. It was also observed that the performance of the ANN based QPF is better at Klang river basin compared to Kelantan river basin. The findings could be correlated with the geographical location of the two basins. Kelantan river basin which is located at the northeast of Peninsular Malaysia is bounded by the vast South China Sea and under the strong influence of northeast monsoon season; therefore more variation in rainfall pattern and type is expected. Whereas Klang River Basin which is located at the southwest of Peninsular Malaysia and far from the influence of the wet monsoonal season, tends to have more consistent type of convective rain.
Table 3 Comparison of performance for ANN QPF models at both catchment areas

<table>
<thead>
<tr>
<th>ANN based QPF versus Rainfall depth</th>
<th>Kelantan river basin</th>
<th>Klang river basin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Point rainfall</td>
<td>Areal rainfall</td>
<td>Point rainfall</td>
</tr>
<tr>
<td>Root mean square error (RMSE)</td>
<td>7.471</td>
<td>3.436</td>
</tr>
<tr>
<td>Mean Absolute error (MAE)</td>
<td>3.025</td>
<td>1.872</td>
</tr>
</tbody>
</table>

4.0 CONCLUSION

The research work attempted to combine the NWP model products and the geostationary meteorological satellite infrared and visible data for QPF model development. The model had been developed using the multilayer ANN and performance had been measured for two different river basins in Peninsular Malaysia. The QPF ANN based techniques had performed satisfactory in forecasting areal-average rainfall depth for convective rainfall in a given duration, and output results had been validated against the gauged rainfall. The best performance of the model is for forecasting 1-hour ahead areal rainfall event for Klang river basin with $r^2 = 0.893$. The study found that the combination of NWP model products and METSAT image features has great potential for enhanced QPF model.

Acknowledgment

The authors would like to thank the Ministry of Higher Education Malaysia for FRGS Grant, 600-RMI/FRGS 5/3 (138/2015) and ERGS Grant 600-RMI/ERGS 5/3 (21/2012) and the Research Management Institute (RMI) UTM for the support provided. We would also like to thank the Malaysian Meteorological Department (MMD) and the Drainage and Irrigation Department (DID), for providing us with the data. Special thanks also to the University of Wisconsin Madison, Space Science and Engineering Center for the great technical help and support as an alternative METSAT image provider.

References

Hydrometeorological and Landslides Hazards. Sydney, Australia.
