Multiple data fusion for rainfall estimation using a NARX-based recurrent neural network – the development of the REIINN model

M R C O Ang[1], R M Gonzalez[1], P P M Castro[2]
[1] Department of Geodetic Engineering, University of the Philippines, Diliman, Quezon City 1101, Philippines
[2] Institute of Civil Engineering, University of the Philippines, Diliman, Quezon City 1101, Philippines

E-mail: concon.ang@gmail.com

Abstract. Rainfall, one of the important elements of the hydrologic cycle, is also the most difficult to model. Thus, accurate rainfall estimation is necessary especially in localized catchment areas where variability of rainfall is extremely high. Moreover, early warning of severe rainfall through timely and accurate estimation and forecasting could help prevent disasters from flooding. This paper presents the development of two rainfall estimation models that utilize a NARX-based neural network architecture namely: REIINN 1 and REIINN 2. These REIINN models, or Rainfall Estimation by Information Integration using Neural Networks, were trained using MTSAT cloud-top temperature (CTT) images and rainfall rates from the combined rain gauge and TMPA 3B40RT datasets. Model performance was assessed using two metrics – root mean square error (RMSE) and correlation coefficient (R). REIINN 1 yielded an RMSE of 8.1423 mm/3h and an overall R of 0.74652 while REIINN 2 yielded an RMSE of 5.2303 and an overall R of 0.90373. The results, especially that of REIINN 2, are very promising for satellite-based rainfall estimation in a catchment scale. It is believed that model performance and accuracy will greatly improve with a denser and more spatially distributed in-situ rainfall measurements to calibrate the model with. The models proved the viability of using remote sensing images, with their good spatial coverage, near real time availability, and relatively inexpensive to acquire, as an alternative source for rainfall estimation to complement existing ground-based measurements.

1. Introduction

Among the elements of the hydrologic cycle, rainfall is the most difficult to model due to the dynamics of the atmospheric processes that generate it and its variation over wide range of scales both spatially and temporally [1]. With the palpable warming of the climate system [2] now evident from observations of increases in global average air and ocean temperatures, the widespread melting of snow and ice, and rising global average sea level, more aberrations in the hydrological cycle such as increased area-average mean rainfall are anticipated particularly in tropical Asia [3]. A more accurate rainfall estimation is necessary especially in localized catchment areas where variability of rainfall is extremely high [4] in order to be forewarned of severe rainfall and help prevent disasters due to flooding.

Several studies have pointed out the vulnerability of the Philippines to disasters that could come with climate change [5], [6]. Recent flood events that accompanied tropical cyclone Megi in October 2010, tropical cyclone Nesat in September 2011, and tropical cyclone Nalgae which came immediately after Nesat are ominous. All these instances were preceded by extreme rainfall, and underscored the need for more accurate rainfall estimates.

Rainfall is measured from different sources such as rain gauge networks, ground-based radar systems, and remotely-sensed satellite images. Although conventional rain gauge networks and radar systems provide precise amounts of rainfall, they are sparsely located and provide only point-scale data that become inaccurate when extrapolated over a wide area [7]. Moreover, radar infrastructures are costly and their coverage is limited by topography. These limitations of in-situ rainfall observations over wide remote regions in the Philippine countryside make satellite images a promising source of rainfall information.

Several rainfall estimation algorithms that utilize satellite images have been developed to provide estimates over a wide area, such as the Tropical Rainfall Measuring Mission (TRMM) [8] and the Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN) [7]. However, these algorithms give global estimates that do not characterize the variability of Philippine rainfall. Locally dependent algorithms are necessary for locally focused applications. This paper presents a framework to estimate rainfall in a large watershed by using remote sensing, geospatial analysis tools in a geographic information system, and artificial neural networks.

2. The case study area

The Cagayan River Basin, with a huge catchment area of about 27,300 km^2, is located in the north-eastern part of the island of Luzon in the Philippines. Its main drainage channel is the Cagayan River, the longest and largest river in the country. Being at the foot of two mountain ranges, the Sierra Madre Mountain and the Cordillera Mountain Range, the Cagayan river basin is replete with resources. It drains a fertile valley that produces a variety of crops such as rice, corn, banana, coconut, citrus and tobacco. It is regarded as the Philippines' "Food Basket and Gateway to the North".

Using the Modified Coronas Classification scheme, the western portion of the Cagayan River Basin is characterized by short dry season during the periods from December to February or from March to May and not having a pronounced maximum rain period. Small portions on the eastern side have no dry season and a very pronounced maximum rain period from December to February. Minimum rainfall occurs during the period from March to May.

3. Materials and Methodology

3.1. Conceptual framework

Figure 2 shows the conceptual framework followed in the development of the REINN model--Rainfall Estimation by Information Integration using Neural Networks. It combines indirect methods of rainfall estimation, i.e. using infrared satellite imageries, and the more direct techniques, i.e. using passive microwave estimates of rainfall rates, calibrated with ground-based measurements using machine learning. Using current datasets, the model can estimate the amount of current rainfall.

3.2. Satellite images and in-situ datasets

Freely available satellite images and datasets of different spatial and temporal resolutions as well as in-situ rainfall data were utilized in the early stage of REINN model development. The MTSAT images are hourly gridded datasets of four infrared (IR) channels. Gridded datasets are latitude and longitude oriented meaning they were already geometrically corrected based on the satellite's position information [9].

The 3B40RT data product of TMPA was also used. 3B40RT is a 3-hourly global 0.25° x 0.25° area-averaged combined microwave precipitation estimate. The rainfall rate value for each grid box is a pixel-weighted average from a combination of all available estimates from the Microwave Imager on

IOP Conf. Series: Earth and Environmental Science **17** (2014) 012019 doi:10.1088/1755-1315/17/1/012019

TRMM, Special Sensor Microwave Imager on DMSP, Advanced Microwave Scanning Radiometer (AMSR-E) on Earth Observing System (EOS) Aqua and Advanced Microwave Sounding Unit-B (AMSU-B) on the NOAA satellite series over the 3-hour period centered at synoptic times 00Z, 03Z, …, 21Z [10], [11].

Hourly rainfall measurements from only eleven rainfall stations are available for the whole 2.7 million hectares of Cagayan River Basin. They were obtained from the Philippine Atmospheric, Geophysical and Astronomical Services Administration (PAGASA) and the National Irrigation Administration (NIA).

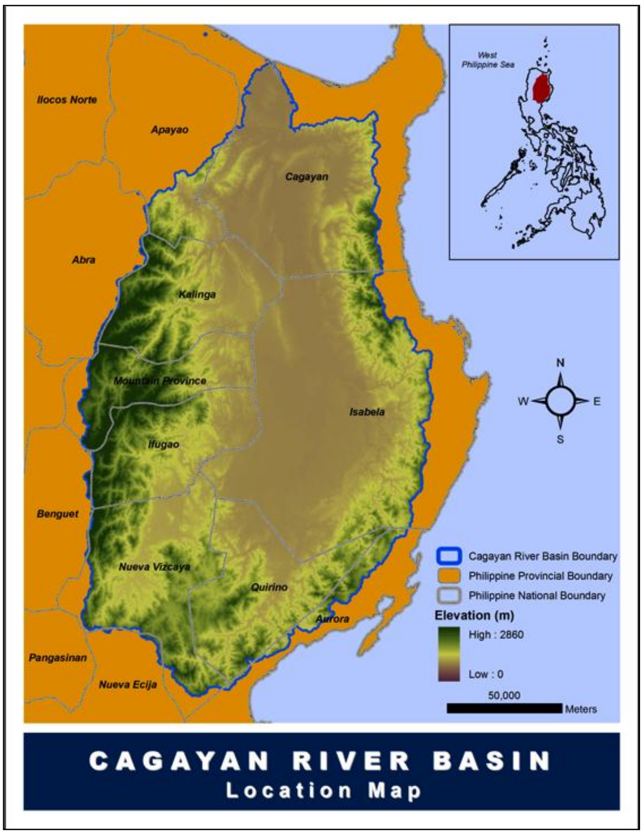

Figure 1. The Cagayan river basin.

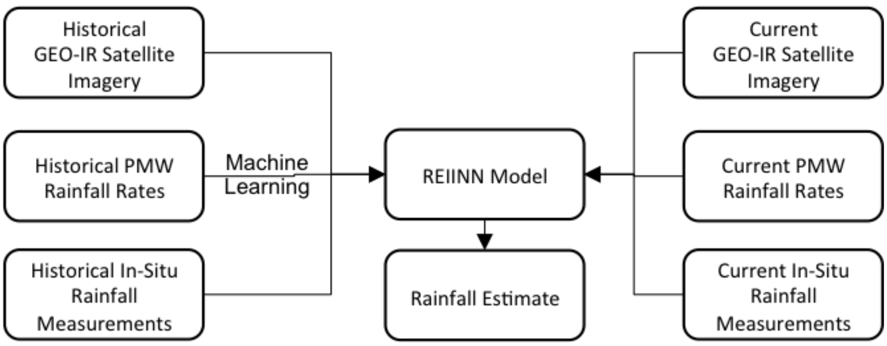

Figure 2. The REIINN conceptual framework.

3.3. REIINN framework implementation

Figure 3 shows the overall process flow diagram for the implementation of the REIINN rainfall estimation framework. Remotely-sensed satellite images of MTSAT were obtained as well as the corresponding TMPA 3B40RT rainfall rates and in-situ rainfall measurements for October 2009. The raw hourly MTSAT infrared images were pre-processed to obtain hourly brightness temperature (TB) images. The hourly brightness temperature images were then temporally aggregated to 3-hourly intervals to correspond to the TMPA resolution. A developed rain/no rain classification scheme was used to derive the 3-hourly cloud-top temperature (CTT) images. The TMPA 3B40RT microwave rainfall rates were pre-processed and calibrated with in-situ rain gauge measurements to obtain surface rainfall rates.

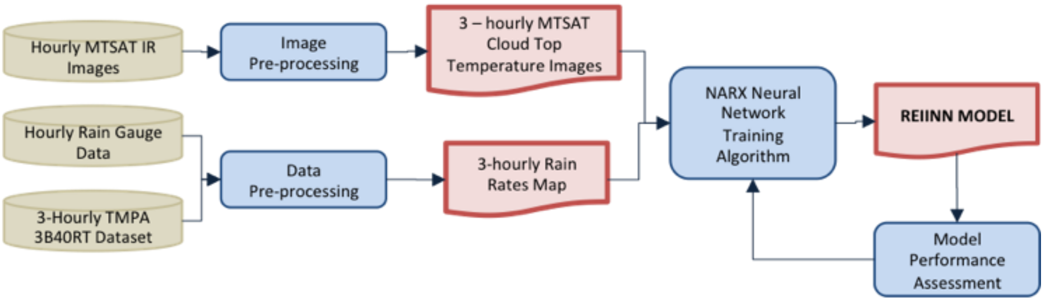

Figure 3. REIINN framework implementation.

Results of the image and data preprocessing described above were used to train the designed Non-linear Autoregressive Exogenous Model (NARX) based neural network architecture. The resulting NARX based neural network model (we call REIINN Model) defines the relationship between cloud-top temperatures and surface rainfall rates.

3.4. Designing and training the REIINN model – NARX-based neural network architecture

From the REIINN conceptual framework presented in Section 3.1, this research is trying to determine the nonlinear, dynamic relationship that exists between MTSAT IR CTT image values and the combined rain gauge–TMPA rainfall rates. Once determined, the derived function can then be used to estimate rainfall rates from current IR CTT images. Neural networks are known to be sophisticated modeling techniques capable of modeling extremely complex functions where statistical models are not anymore valid, especially in areas of function fitting and time series analysis [12]. It provides a computationally efficient way of determining an empirical, nonlinear relationship between a number of ‘‘inputs’’ and one or more ‘‘outputs’’ [13], [14].

The conceptualized REIINN model, written in mathematical form, is shown in equation (1):

$$RR_{estimated}(t) = f(MTSAT_{CTT}(t), RR_{estimated}(t-1), RR_{actual}(t-1)) \tag{1}$$

where, $RR_{estimated}\ (t)$ is the rainfall rate of interest at the current time instant t which is a function of: $MTSAT_{CTT}\ (t)$ – the MTSAT CTT image at time instant t, $RR_{estimated}\ (t-1)$ – the model output rainfall rates at the previous time instant and $RR_{actual}\ (t-1)$ – the actual rainfall rate at the previous time instant. With this kind of model, the NARX recurrent neural network is the most appropriate type of neural network to approximate the function f.

The NARX, or the Nonlinear Autoregressive with Exogenous Inputs, neural network is a recurrent dynamic network based on the Nonlinear Autoregressive Exogenous model. This means that the model predicts the current value of a time series based on its relation to the past values of the series and current and past values of the exogenous series [15].

Two NARX-based neural network models were developed i.e. REIINN 1 and REIINN 2. REIINN 1 has a parallel architecture, wherein, at a specific instant, it takes as input the CTT of the four IR channels of MTSAT and its estimated rainfall rate that is fed back from the previous time instant as shown in Figure 4. On the other hand, REIINN 2, shown in Figure 5, has a series-parallel architecture that takes in the same CTT images, but with the inclusion of the actual rainfall rate measurements from the previous instant. The recurrent nature of the two models allows them to account for the temporal dynamics that exist in rainfall patterns.

Typically, neural networks are trained so that a particular input leads to a specific output. A network is ‘‘trained’’ to a specific task by presenting it with many examples of inputs and the corresponding desired outputs (targets). The resulting 3-hourly MTSAT CTT images (with 4 infrared channels each) and 3-hourly combined rain gauge-TMPA rainfall rates, after pre-processing, were used as input and target training data respectively. A total of 168 instances (168 3-hourly MTSAT CTT–rainfall rates pairs) corresponding to the whole month of October 2009 were used.

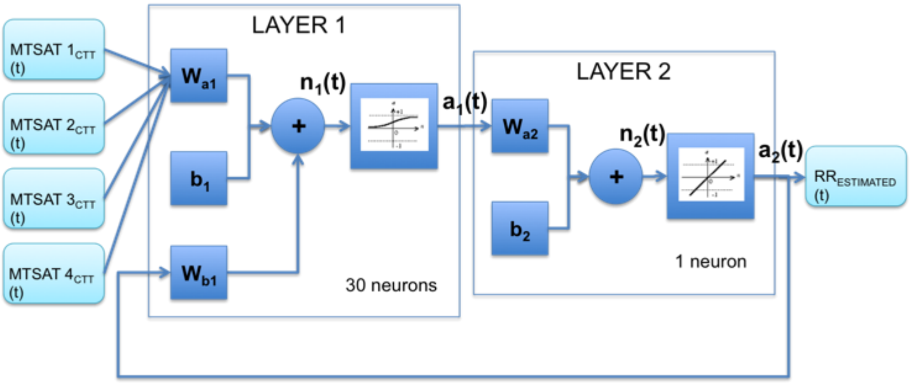

Figure 4. Schematic diagram of the REIINN 1 model.

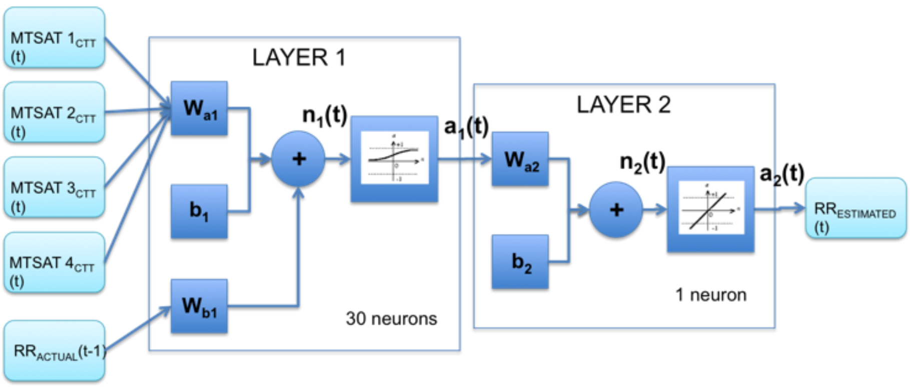

Figure 5. Schematic diagram of the REIINN 2 model.

4. Results and Discussion

Model performance of REIINN 1 and REIINN 2 were assessed using two metrics, i.e. Root-Mean-Square Error (RMSE) and correlation coefficient R.

4.1. REIINN 1 model performance

REIINN 1 yielded an RMSE of 8.1423 mm/3h and an overall R of 0.74652. Figure 6 shows the regression plot for REINN 1, which displays the network output with respect to the target. The closer the value of R to 1, the better is the fit of the network. Here, it can be seen that the network underestimates higher rainfall rate amounts while overestimates lower rainfall rates. Also, non-rainy samples are mapped as rainy.

Figure 7 shows the REIINN 1 network's time-series response. Here, the underestimation and overestimation can be clearly seen as represented by the positive and negative vertical error lines, respectively. This shows that the network attenuates (either positively or negatively) peak rainfall values. This may be due to the nature of the NARX model, where current estimates of rainfall are dependent on previous values. This dependence on previous values makes current estimates fall not far from the values of previous estimates thus the attenuation effect.

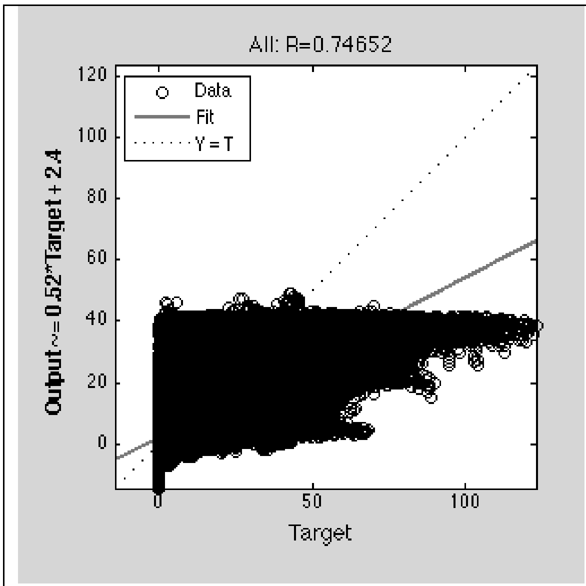

Figure 6. REIINN 1 regression plot.

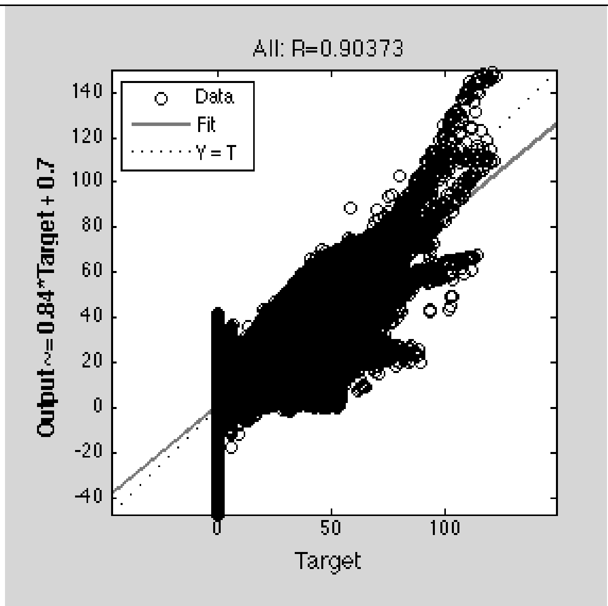

Figure 8. REIINN 2 regression plot.

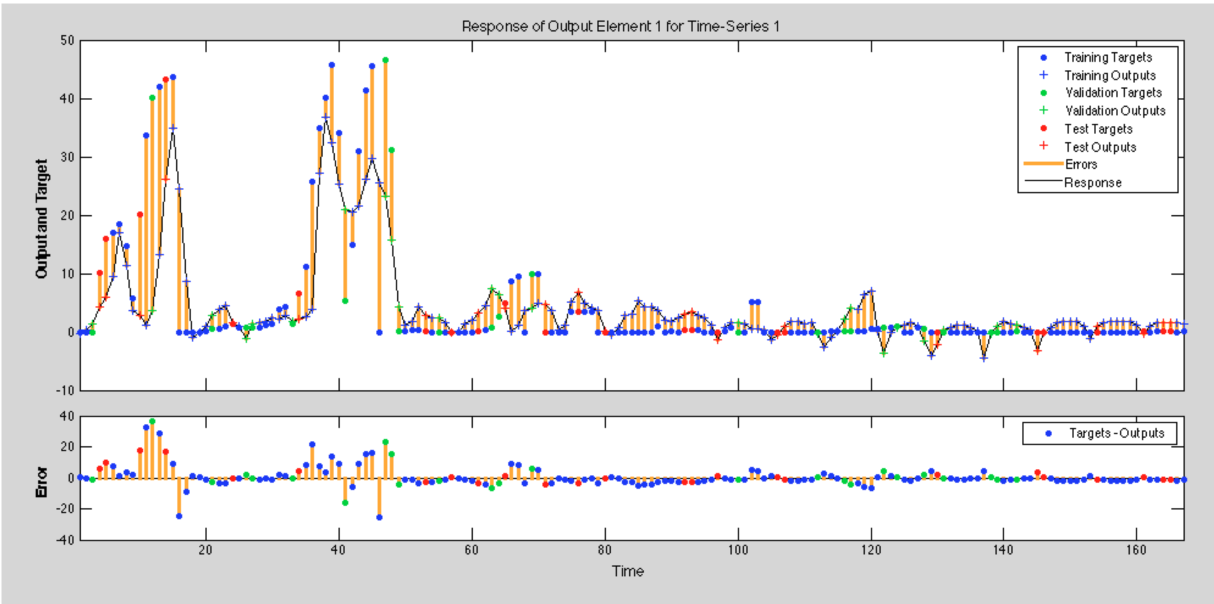

Figure 7. REIINN 1 time-series plot.

4.2. REIINN 2 model performance

REIINN 2 yielded an RMSE of 5.2303 mm/3h and an overall R of 0.90373 which are significantly better than REIINN 1. Figures 8 and 9 show the regression and time-series response plots for REIINN 2 respectively. Generally, REIINN 2 performed better than REIINN 1. This can be explained by the inclusion of actual rainfall rate measurements from the previous time instant as input to the model.

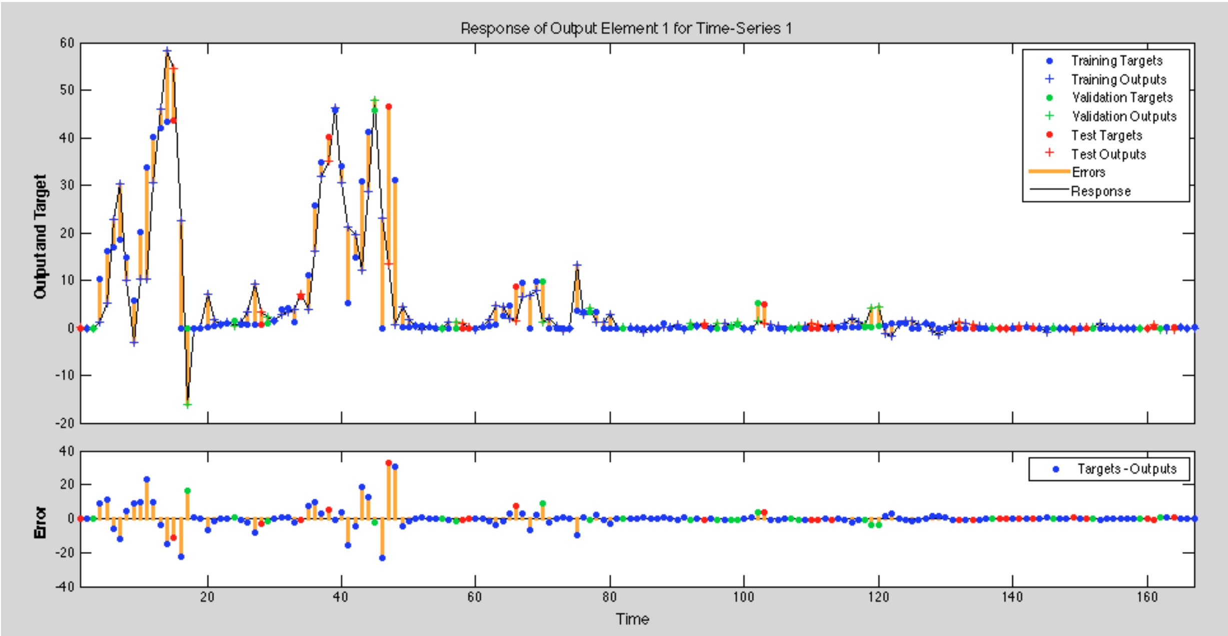

Figure 9. REIINN 2 time-series plot.

5. Conclusion and Recommendations

The superior model performance presented by REIINN 2 over REIINN 1 showed that including actual in-situ rain gauge measurements taken from the previous time instant as input to the current estimation is better than feeding back the previous outputs (rainfall estimates) of the model. REIINN 2 allows a continuous re-calibration of the model with actual values of rainfall and as such, portrays the process dynamics. Moreover, it demonstrates that, although remote sensing images provide spatially distributed information on rainfall, rain gauges which are direct measurement of rainfall should not be eliminated from the model.

The results, especially that of REIINN 2, are very promising for satellite-based rainfall estimation in a catchment scale. It is believed that model performance and accuracy will greatly improve with a denser and more spatially distributed rain gauge measurements to calibrate the model with. The two models proved the viability of using remote sensing images, with their good spatial coverage, near real time availability, and relatively inexpensive to acquire, as an alternative source for rainfall estimation to complement existing ground-based measurements.

6. References

[1] Hung N Q, Babel M S, Weesakul S and Tripathi N K 2009 An artificial neural network model for rainfall forecasting in Bangkok, Thailand *Hydrology and Earth System Sciences* **13 8** 1413-25

[2] Intergovernmental Panel on Climate Change [IPCC] Climate Change 2007: Synthesis Report *Contribution of Working Groups I, II and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (Geneva, Switzerland: IPCC)

[3] Cruz R V *et al* 2007 Asia. Climate Change 2007: Impacts, Adaptation and Vulnerability *Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge, UK: Cambridge University Press)
[4] Davie T 2008 *Fundamentals of Hydrology* 2nd Edition (New York, NY: Taylor & Francis)
[5] "UNU-EHS."
[6] Germanwatch 2012 *Global climate risk index 2012: Who Suffers Most From Extreme Weather Events? Weather-Related Loss Events In 2010 and 1991 To 2010* [Online] Available: http://www.germanwatch.org/klima/cri.pdf [Accessed: 12-Jan-2012]
[7] Hsu K L and Sorooshian S 2008 Satellite-Based Precipitation Measurement Using PERSIANN System *Hydrological Modelling and the Water Cycle* 27–48
[8] Kummerow C, Barnes W, Kozu T, Shiue J and Simpson J 1998 The Tropical Rainfall Measuring Mission (TRMM) Sensor Package *Journal of Atmospheric and Oceanic Technology* **15** **3** 809-817
[9] Yamamoto M K and Higuchi A 2010 *MTSAT Gridded Dataset release note & dataset documentations*
[10] Huffman G J and Bolvin D T 2011 *Real-Time TRMM Multi-Satellite Precipitation Analysis Data Set Documentation* (Greenbelt, Maryland)
[11] Huffman G J *et al* 2007 The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-Global, Multiyear, Combined-Sensor Precipitation Estimates at Fine Scales *Journal of Hydrometeorology* **8** **1** 38-55
[12] StatSoft Inc *Neural Networks - Electronic Statistics Textbook* [Online] Available: http://www.statsoft.com/textbook/neural-networks/#intro
[13] Grimes D I F, Coppola E, Verdecchia M and Visconti G 2003"A neural network approach to real-time rainfall estimation for Africa using satellite data *Journal of Hydrometeorology* **4** 1119-33
[14] Rivolta G, Marzano F S, Coppola E and Verdecchia M 2006 Artificial neural-network technique for precipitation nowcasting from satellite imagery *Advances In Geosciences* **7** 97-103
[15] Safavieh E, Andalib S and Andalib A Forecasting the Unknown Dynamics in NN3 Database Using a Nonlinear Autoregressive Recurrent Neural Network *2007 International Joint Conference on Neural Networks* 2105-09