Constructing with River Flood Prediction Models

Thinn Htet Htet San, Mie Mie Khin
University of Computer Studies, Mandalay
thinnhtetsan2000@gmail.com, miemie.khin9@gmail.com

Abstract

With the incidence of severe weather and flooding on the increase around the world, there is a need to improve flood forecasting and warning. Floods cause physical damage, loss of basic sanitation that leads to disease, economic hardship due to rebuilding costs and food shortages. By improving flood forecasts it becomes possible to take mitigating actions in advance of the flood and hence avoid millions of pounds worth of damage and even human fatalities.

In this paper, a time series and Markov models for river flood prediction are constructed. These models focus on the prediction of events and can capture the fact that time flows forward. The output will be approximate and show that there is a close agreement between the predicted and actual river flooding amount. The system compares the results of time series model and Markov model with the actual weather station results and also shows the best model for river flood prediction over Ayeyarwady River in Myanmar.

Keywords *– flood, prediction, time series, Markov Chain*

1. Introduction

A flood is a natural event that can have far reaching effects on people and the environment. A flood is caused by a combination of heavy rainfall causing river / oceans to over flow their banks, and can happen at any time of the year. Floods in Myanmar generally occurs during the South West Monsoon season of June to October when the monsoon troughs or low pressure waves superimposed on the general monsoon pattern resulting in intense rainfall over strategic areas of the river catchments. According to recent statistics, temperature in Myanmar have been rising gradually during the past 30 years by about 0.6 Celsius since 1977, resulting in irregular weather patterns, including a shorter but more intense monsoon which had increased the risk of flooding. Floods have affected loss of crops and valuable property and even loss of people. Loss of human life and property etc. can be reduced to a considerable extent by giving reliable advance information about the coming floods.

Flood prediction and control is one of the greatest challenges facing the world today, which have become more frequent and severe due to the effects of global climate change and human alterations of the natural environment.

In this work, daily flood level prediction for each weather station is implemented by applying time series model and Markov model. It is characterized by real time of data flow, which is from the recording equipment in testing of complex systems, and the rapid processing data, at processing of satellite information. Weather stations are selected by along with Ayeyarwaddy river which often face local flooding in rainy seasons as our region of interest. Flood level can be modeled by using various time scales such as yearly, monthly, weekly and daily.

This paper is organized in the following sequence. A briefly survey of many researchers is presented in section 2. The application area is dealt with in section 3. This is followed by the time series and Markov chain in section 4 and 5. Overview of the proposed system and conclusion is described in section 6 and 7.

2. Related Works

M. Kannan et al. [3] predicted the rainfall by using multiple linear regression (MLR) models and computed Pearson coefficient for five years data and then compared with predicted data using regression approach. The predicted values lie below computed values. According to the results, it does not show accuracy but show an approximate value. Predicting flood, cyclone, forest fire detection, and global warming was avoided.

M. A. Kulkarni et al. [4] proposed wind speed prediction using four different statistical techniques; Curve fitting, Auto Regressive Integrated Moving Average Model (ARIMA), Extrapolation using periodic curve fitting and Artificial Neural Networks (ANN). They computed the Root Mean Square Error (RMSE) in prediction of zonal component of wind speed for all the months.

In [10] Song et al. presented key findings and methods for modeling and forecasting from 2000-2008, including time-series, econometric, and combination techniques. This survey identified some new research directions, which included improving the forecasting accuracy through integrating both qualitative and quantitative forecasting approaches, tourism cycles, seasonality analysis, impact assessment, and risk forecasting.

N. Sen presented a long-range summer monsoon rainfall forecast model based on power regression technique with the use of Ei Nino, Eurasian snow cover, north west Europe temperature, Europe pressure gradient, 50 hPa Wind pattern, Arabian sea SST, east Asia pressure and south Indian ocean temperature in previous year. The experimental results showed that the model error was 4% in [7]

S. Nkrintra, et al. [6] described the development of a statistical forecasting method for SMR over Thailand using multiple linear regression and local polynomial-based nonparametric approaches. SST, sea level pressure (SLP), wind speed, EiNino Southern Oscillation Index (ENSO), and IOD were chosen as predictors. The experiments indicated that the correlation between observed and forecast rainfall was 0.6.

T. Sohn, et al. [9] has developed a prediction model for the occurrence of heavy rain in South Korea using multiple linear and logistics regression, decision tree and artificial neural network. They used 45 synoptic factors generated by the numerical model as potential predictors.

In [11] W. T. Zaw developed a prediction model for determining Rainfall over Myanmar using multiple linear regressions where 15 predictors has been used. As a result of several experiments, the predicted rainfall amount is close to actual value.

S. Banik et al. (2009) [1] have developed rainfall forecasting model using ANN, ANFIS and GA processes and the results obtained by these models are also compared to the statistical. The ANFIS forecasting model and the GA forecasting model can be used to forecast monthly monsoon rainfall more accurately than the ANN model and the statistical model, forecasting method, namely linear multiple regression model.

D. Nayak et al. (2013) have done survey on rainfall predictions using different neural network architectures over twenty-five years. From the survey it has been found that most of the researchers used back propagation network for rainfall prediction and got significant results. The survey also gives a conclusion that the forecasting techniques that use MLP, BPN, RBFN, SOM and SVM are suitable to predict rainfall than other forecasting techniques such as statistical and numerical methods in [5].

El-Shafie et al. (2011) [8] have developed an adaptive neuro-fuzzy inference system (ANFIS) and artificial neural network (ANN) model to forecast the rainfall for Klang River in Malaysia on monthly basis. The result showed performance of ANFIS method is better than ANN method and concluded that ANFIS method is superior to the ANN method in forecasting monthly rainfall.

3. Application Area

Myanmar is located between latitudes 09º 32'N and 28º 31'N and longitudes 92º 10'E and 101º 11'E. The location and topography of the country generate a diversity of climate conditions and seasonal changes in the monsoon wind directions crate summer, rainy and winter seasons. According to historical statistics, heavy moonsoon rains in Myanmar especially from mid-July to mid-September have caused flooding along the Ayeyarwady, Chindwin, Thanlwin, Sittaung and Yangon rivers and their tributaries [2].

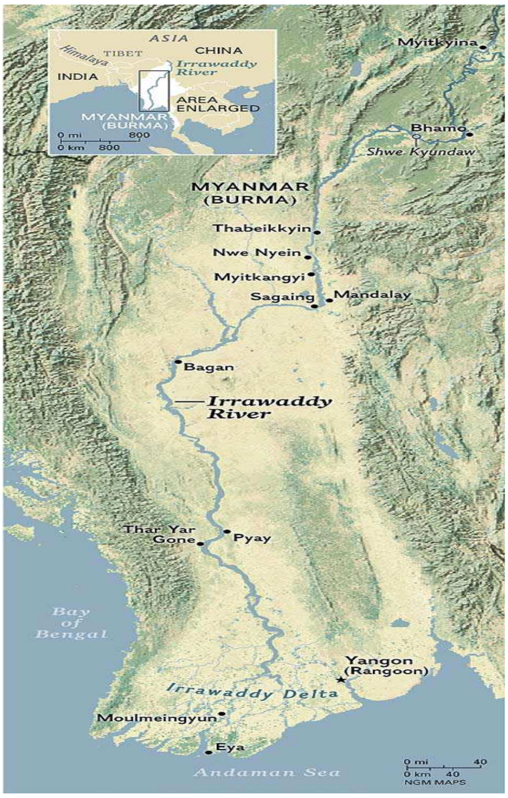

Figure 1. Geographical Area of Study

Ayeyarwady is the country's largest river and most important commercial waterway. It starts in the northern part of the country, Kachin State, and flows south to the Andaman Sea on the Ayeyawady coast. Its drainage area of about 413,000 km^2 covers a larger part of Myanmar. Due to monsoonal rains, which occur between mid-May and mid-October, the volume of the Ayeyarwady and its tributaries varies greatly throughout the year.

The river flows through the cities: Putao, Myitkyina, Bhamaw, Katha, Tagaung, Kyaukmyaung, Mandalay, Sagaing, Chauk, Bagan, Nyaung-U, Magway, Pyay, Hinthada, Pantanaw. Details of some weather stations and geographical area of study in this work are shown in Table 1 and Figure 1, respectively.

In this work, daily flood level prediction for each weather station is implemented by applying time series model and Markov model. Using this model we can use evidence of a past days observed values to update the statistical beliefs we have of what might occur in future days. With these updated statistical beliefs we can then make forecasts.

No.	Station Name	Region	Location
1.	Myitkyinar	Upper	25.3'N 97.4'E
2.	Sagaing	Central	21.5'N 95.6'E
3.	Mandalay	Central	21.9'N 96.1'E
4.	Magway	Central	18.83'N 93.78'E
5.	Aunglan	Lower	19.38'N 95.23'E
6.	Pyay	Lower	18.8'N 95.21'E

Table 1. Details of Met/Hydro Stations

The main factors that cause flooding are heavy rainfall, sudden or heavy snow melt, and dam failure, the possibility of levee failure. All of these factors can suddenly increase discharge of water into streams, within streams, and out of streams. Furthermore, when the discharge causes the river to rise above flood stage water runs onto the floodplain.

In this work, spatial dependencies not only between variables but also between weather stations are considered to model the riverine flood using collected hydrological and hydro-meteorological data which are mainly contributing to riverine flood.

4. Time Series

A time series is a sequential set of data points, measured typically over successive times. It is mathematically defined as a set of vectors $x(t), t = 0,1,2,...$ where t represents the time elapsed. The variable $x(t)$ is treated as a random variable. The measurements taken during an event in a time series are arranged in a proper chronological order. A time series containing records of a single variable is termed as univariate. But if records of more than one variable are considered, it is termed as multivariate. A time series can be continuous or discrete. In a continuous time series observations are measured at every instance of time, whereas a discrete time series contains observations measured at discrete points of time. Usually in a discrete time series the consecutive observations are recorded at equally spaced time intervals such as hourly, daily, weekly, monthly or yearly time separations.

$$Y_t = f(Y_{t-1}, Y_{t-2}, Y_{t-3}, \dots, Y_{t-n}) + e_t$$

Where Y_{t-1} is the value of Y for the previous observation, Y_{t-2} is the value two observations ago, etc., and e_t represent noise that does not follow a predictable pattern.

4.1. Time Series Model

There are many Time series models. Among these my research will be used Additive model because historical data is different from the range.

4.1.1 Additive Decomposition Method

$$\boldsymbol{Y_t = T_t + C_t + S_t + I}$$

Y_t = data at period t

T_t = trend-cycle component at period t

C_t = Cyclical component at period t

S_t = Seasonal component at period t

I_t = remainder (or irregular or error or random) component at period t

4.1.2 Constructing the Additive model

To construct the proposed additive model, the historical data at daily (period t) water level are collected. Depended on the collected historical daily data, the period trends (T_t) are calculated using Moving Average method. For getting the seasonal factors (S_t), the trends are subtracted from the collected data. And then, the environmental effects (irregular I_t) or cyclical (C_t) component can not be consider in the proposed model. Finally, the forecast (Y_t) are obtained based on seasonal factors.

5. Markov Chain

A Markov chain is a stochastic process (random process) in which the probability distribution of the current state is conditionally independent of the path of past states, a characteristic called the Markov property. A sequence of states: $X_1, X_2, X_3, \dots$(Usually over time)

Markov chain is a discrete-time stochastic process with the Markov property.

5.1 Discrete Time Markov Chain (DTMC)

A stochastic process with discrete state space and discrete time $\{X_n, n > 0\}$ is a discrete time Markov Chain (DTMC). In a DTMC, the past history impacts on the future evolution of the system via the current state of the system.

$$\mathbf{P[X_{n+1} = j \mid X_n = i_n, ..., X_0 = i_0] = P[X_{n+1} = j \mid X_n = i_n] = p_{ij}(n)}$$

The conditional distribution of any future state X_{n+1}, given the past states X_0, X_1, …., X_{n-1} and the present state X_n, is independent of the past states and depends only on the present state. The value P_{ij} represents the probability that the process will, when in state i, next make a transition into state j. $P_{ij}(n)$ is called **transition probability** from state i to state j at time n.

$$P_{ij} \geq 0, i, j \geq 0$$
$$\sum P_{ij} = 1$$

5.2 Transition Probability

The conditional probability that we will be in a future state given a current state. The transition probabilities are

$$P(X_{t+1}= j \mid X_t=i)$$

Transition probabilities are called stationary if

$$P(X_{t+1}= j \mid X_t=i) = P(X_1= j \mid X_0=i)$$

If there are only finitely many possible states of the random variables X_t then the stationary transition probabilities are conveniently stored in a *transition matrix*

$$P_{ij}= P(X_1= j \mid X_0=i)$$

5.3 Markov's State for Weather Effect

Rainfall (RF), Snow melt (SM), Tide (TD), Water level (WL), and Danger water level (DWL) are considered as Markov's state.

5.3.1 Constructing the Markov Model

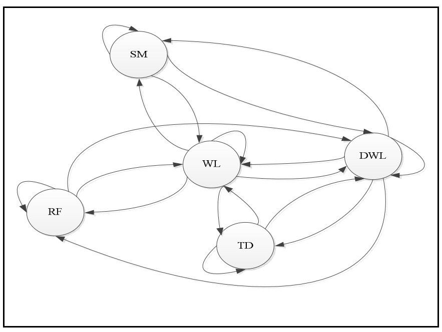

Figure 2. Markov Model of Proposed System

It is the likelihood that the system in a given current state will remain in the same state or move to another state in the next period. Table 2 shows that the weather effect for the proposed Markov model is described. In the table, depend on one state to another state is 1 and otherwise 0.

Table 2. Weather Effect for Markov model

	RF	SM	TD	WL	DWL
RF	1	0	0	1	1
SM	0	1	0	1	1
TD	0	0	1	1	1
WL	1	1	1	1	1
DWL	1	1	1	1	1

To construct the Markov model, Markov state is needed. Therefore, rain fall, snow melt, tide, water level and danger water level are considered as Markov states. The Markov model can be constructed using above Markov states as in Figure 2. After that the transition probability are calculated. The proposed system constructs Markov model for six stations. Among these, water level state for Mandalay station is presented as sample. Transition probability matrix and its Markov model for water level at Mandalay station in 2004 is described as shown in following matrix and Figure 3.

$$P = \begin{array}{c c} & \begin{array}{ccc} F & AF & NF \end{array} \\ \begin{array}{c} F \\ AF \\ NF \end{array} & \begin{bmatrix} 0.0980 & 0.0065 & 0 \\ 0.0065 & 0.0980 & 0.0196 \\ 0 & 0.0196 & 0.7516 \end{bmatrix} \end{array}$$

Where, F = Flood
AF = Approximately Flood
NF = Non Flood

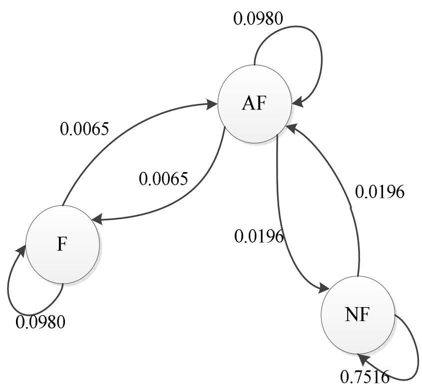

Figure 3. Markov Model of Mandalay Station in 2004

6. Proposed System Overview

In this section, the overview of the proposed system is presented as shown in Figure 4.

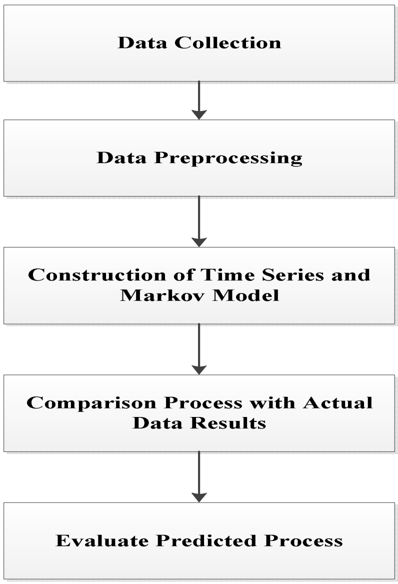

Figure 4. Overview Steps of the Proposed System

The collections of historical data are used from spatial dependencies between variables which are observable factors contributing to the flood level of Ayeyarwady river. The collected data are needed to be preprocessing. Then, Time series model and Markov model are constructed based on preprocessing data. The results of the constructed models are compared with the actual weather results. In order to show the best model for weather station, the predicted results are evaluated. The data observed for flood prediction consist of:

(1) Stage

The term stage refers to the height of a river (or any other body of water) above a locally defined elevation. This locally defined elevation is a reference level, often referred to as datum. In Ayeyarwady river, datum is varied for each station. For example, for Machanbaw station, reference level or datum, is 1219m and for Myitkyinar station, is 143.369m.

(2) bankfull stage or flood stage

The stage at which the river will overflow its banks is called bankfull stage or flood stage.

(3) rainfall

Rainfall is heavier than normal in a particular area is called Rainfall stage.

(4) runoff

When rain falls on the surface of the Earth, what remains on the surface of the Earth and eventually flows into streams is called runoff. In general, then:

Runoff = Precipitation – Infiltration – Interception - Evaporation

(5) rainfall distribution

If rainfall is heavier than normal in a particular area and infiltration, interception, and evaporation are low then runoff can be high and the likelihood of flooding will increase.

(6) lag time period

The time difference between when heavy precipitation occurs and when peak discharge occurs in the streams draining an area is called lag time.

(7) information about dams and levees

In the first step, data of each weather station which are contributing to water level prediction are collected. Then, continuous quantities of collected data are quantized into N interval ranges (DISCRETIZATION). These interval values are defined on daily amount of observation at each weather station.

In predictor analysis phase, state variables and evidence variables are defined for our models. As evidence variables, data which are mainly contributing to river flood prediction are selected from collected data. The structure is recommended by the experts and the researches in the area.

The effect of snow and tide may be considered in this work since the melting of the snow and glaciers in Northern Myanmar add to the volume of Ayeyarwady river only in summer and so it cause just little fluctuation in water level of the river.

Among evidence variables of our model, rainfall distribution can be participated from other one or more stations since predicted area can be affected from rainfall amount of one or more other stations according to geographical location. Thus, spatial dependencies need to be considered not only between variables but also between weather stations.

Time series models are constructed based on the spatial dependencies between the variables defined in Predictor Analysis Process. Details of predictors for water level prediction are shown in Table 3.

Table 3. Details of Predictors for Water Level Prediction

No	Predictor	Location	Period
1	East Indian Sea Surface	90°E-110°E 10°S	Monthly
2	East Indian Sea Surface	90°E-110°E 10°S	Previous Month
3	South Oscillation	(17.5°S 149.6°W,12.4°S	Monthly

4	Southern Oscillation	(17.5°S 149.6°W,	Previous Month
5	Temperature	Predicted area	Daily
6	Dam& levee Information	Predicted area	Daily
7	Dam& levee Information	Predicted area	Previous day
8	Dam Information	Predicted area	Previous two day
9	Precipitation	Predicted area	Daily
10	Precipitation	Predicted area	Previous day
11	Precipitation	Predicted area	Previous two day
12	Tide	Predicted area	Daily
13	Snow	Predicted area	Previous two day
14	Rainfall distribution	Distribution area	Daily
15	Rainfall distribution	Distribution area	Previous day
16	Rainfall distribution	Distribution area	Previous two day
17	Water level	Predicted area	Previous day

In Markov and Time Series models, we consider water flow dynamics as a system and the water surface level is considered as the system output, the river flow, rainfall, tide, snow and *etc* are considered as the system input. The desired output will be get from the forecast water level of the two models.

7. Conclusion

River flood prediction is important for many areas of human activities such as agriculture, water resources, and hydro-electric power projects, happening of droughts or floods and others. This paper constructed Time Series model and Markov chains model. They are useful to be acceptable accuracy for flood prediction of Ayeyarwady River in Myanmar. This system is going to compare the result of models with actual weather results and also will choose the best model for river flood prediction in Myanmar. We also intend to extend and predict such time series data over time using dynamics.

References

[1] S. Banik, Mohammed Anwer, A.F.M. Khodadad Khan, Rifat Ara Rouf and Farah Habib Chanchary, “Forecasting Bangladeshi monsoon rainfall using neural network and genetic algorithm approaches”, International Technology Management Review, Volume 2 Number 1, 2009.

[2] A. N. Hlaing , “Uncertain Reasoning & Prediction for River Flooding with Bayesian Networks Model”, AICT 2010.

[3] M.Kannan, S.Prabhakaran and P.Ramachandran, “Rainfall Forecasting Using Data Mining Technique”, International Journal of Engineering and Technology, Vol.2 (6), 397-401, 2010.

[4] M. A. Kulkarni, S. Patil, G. V. Rama and P. N. Sen, “Wind speed prediction using statistical regression and neural network”, J. Earth Syst. Sci. 117, No. 4, August 2008, pp. 457-463.

[5] D. Nayak, A. Mahapatra, P. Mishra, “ A Survey on Rainfall Prediction using Artificial Neural Network”, International Journal of Computer Applications “, Volume 72, No.16, 2013.

[6] S. Nkrintra, et al., “Seasonal Forecasting of Thailand Summer Monsoon Rainfall”, in International Journal of Climatology, Vol. 25, Issue 5, American Meteorological Society, 2005, pp. 649-664.

[7] N. Sen, ” New forecast models for Indian south-west Monsoon season Rainfall”, in Current Science, vol. 84, No. 10, May 2003, pp.1290- 1291.

[8] El-Shafie, O. Jaafer and Seyed Ahmad Akrami, “Adaptive neuro-fuzzy inference system based model for rainfall forecasting in Klang River, Malaysia”, International Journal of the Physical Sciences Vol. 6(12), pp. 2875-2888.

[9] T. Sohn, J. H. Lee, S. H. Lee, C. S. Ryu, “Statistical Prediction of Heavy Rain in South Korea”, in Advances in Atmospheric Sciences, Vol. 22, No. 5, 2005, pp.703-710.

[10] Song, Haiyan, and G. Li, “Tourism demand modelling and forecasting - A review of recent research”, Tourism Management 29.2 (2008): 203-220.

[11] W. T. Zaw, T. T. Naing, “Empirical Statistical Modeling of Rainfall Prediction over Myanmar”, World Academy of Science, Engineering and Technology, Vol: 2 2008-10-27.