# Flood Prediction Using Seasonal Autoregressive Integrated Moving Average (SARIMA) Model

**Abdulrazak Yahya Saleh, Roselind Tei**

***Abstract:*** ***This paper aims to evaluate the performance of the Seasonal Autoregressive Integrated Moving Average (SARIMA) model for the purpose of flood forecasting. Seven datasets are provided by the Drainage and Irrigation Department (DID) for Sungai Bedup, Serian, Sarawak, Malaysia; and these loads of valuable information are used to evaluate the performance of the SARIMA algorithm. A distinctive network was trained and tested using the daily data obtained from the DID from the years 2014 to 2017. The performance of the algorithm was evaluated based on the technique of Root Mean Square Error (RMSE) by comparing with the Long Short Term Memory Network (LSTM) and Backpropagation Network (BP). Among the seven datasets, the Sungai Bedup set shows a small testing error rate, which is (0.008), followed by Sungai Meringgu (0.011), Semuja Nonok (0.023), Bukit Matuh and Sungai Busit with the same value (0.025); and lastly the value of Sungai Merang is (0.029). The results prove that the SARIMA model can be employed reliably to forecast the water level of Sungai Bedup with the lowest RMSE value, which is 0.008. Meanwhile, LSTM has a RMSE value of 0.08 and Backpropagation has an RMSE value of 0.711. More discussions will be provided to demonstrate the effectiveness of the model in flood prediction.***

***Index Terms:*** ***Artificial Neural Seasonal Autoregressive Integrated Moving Average (SARIMA), Long Short Term Memory (LSTM), Backpropagation (BP)***

## I. INTRODUCTION

Flood is a natural calamity, and Malaysia experiences it almost every year in varying degrees of magnitude. Throughout Malaysia, an estimated 9% of the total land area of Malaysia, including Sabah and Sarawak, is vulnerable to flood, and approximately 4.82 million people are affected by this disaster annually [1]. Over the past decade, different kinds of modelling and data types had emerged in an attempt to forecast the flood events [2]. Excessive rainfall can cause flooding in rural regions and urban areas alike, which may undergo demographic changes from time to time. [3], 50 years of data reveal that 41% of all the natural disasters are related to severe weather conditions or water-related events such as flood. The historical records of the catchments are important information that can facilitate investigating the time series of flash floods occurring hourly [4]. An early accurate prediction of the occurrences will help to over-come logistic problems of evacuation and mitigate the impacts of flood events. Different principles have been used to forecast floods, such as computer simulations based on the watershed demographic model, principle of hydrological, hydraulic components and groundwater flow model [5]. However, these methods only can predict the occurrence of flood events for certain catchment or basin based on specific water-level values. The target in this research is employing a big pool of historical flood data to predict an accurate output; the results will then be utilised by the policy makers in implementing measures to reduce the impacts of floods, not only on the society but also on the environment. The SARIMA model is used in this research due to its ability to learn from the past data in solving complex problems, and this model has been widely used in the field of forecasting [6]. By applying the SARIMA technique, the computational models that contain numerous processing layers can learn the data given with multiple levels of abstraction [7]. The catchment of this study is part of the Sadong Basin, and it is located 80 km away from Kuching. [8], the area of the whole Sadong Basin is about 3550 km2 while the total length of the main river is 150 km, as shown in Figure 1. The datasets are obtained from the DID for the years of 2014 to 2017. Forecasting is conducted on seven different gauging stations in Serian Division; which are Sungai Bedup, Bukit Matuh, Semuja Nonok, Sungai Busit, Sungai Merang, Sungai Meringgu and Sungai Tep.

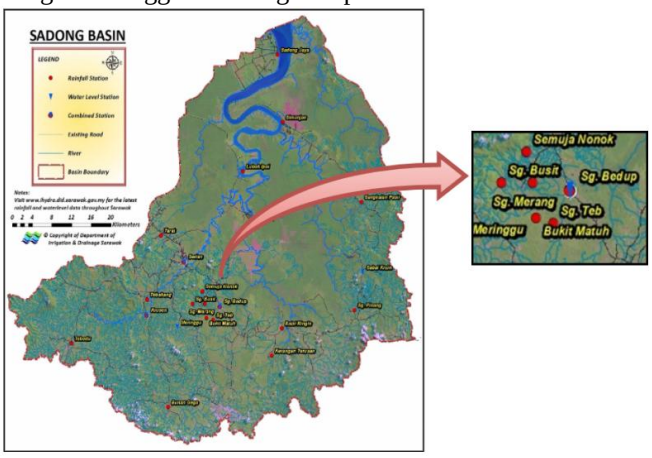


**Fig. 1: Location of the Bedup River, Sarawak, Malaysia.**

## II. METHODOLOGY

To evaluate the performance of the SARIMA model, several experiments are conducted on seven real datasets given by DID. The characteristics of the seven datasets are shown in Table 1, These are the real-world datasets given by the DID, which are similar in terms of the number of available samples, datasets characteristics (Multi-variate), and Features (2). Each station name in the dataset is presented as an input pattern; meanwhile, the features show the water level

**Revised Manuscript Received on June 05, 2019**

**Abdulrazak Yahya Saleh**, FSKPM Faculty, University Malaysia Sarawak (UNIMAS), Kota Samarahan, 94300 Sarawak, Malaysia.

**Roselind Tei**, FSKPM Faculty, University Malaysia Sarawak (UNIMAS), Kota Samarahan, 94300 Sarawak, Malaysia.