

# Analysis and predictive validity of Kelantan River flow using RQA and time series analysis

Mohammed Abdul Majid<sup>1,\*</sup>, M. Hafidz Omar<sup>2</sup>, M. S. M. Noorani<sup>3</sup>

<sup>1</sup>*School of Mathematical Sciences, FST, UKM, Malaysia*

<sup>2</sup>*Dept. of Mathematics & Statistics, KFUPM, Dhahran, Saudi Arabia*

<sup>3</sup>*School of Mathematical Sciences, FST, UKM, Malaysia*

*\*Corresponding author: abdulmajid\_2000@yahoo.com*

## Abstract

An analysis of the Kelantan River flow for the period 2000 to 2014 and its predictive validity have been undertaken by utilizing recurrence plot (RP), recurrence quantification analysis (RQA), and auto regressive integrated moving average (ARIMA). The study presents the analysis results for flood detection and prediction of future values on a holdout sample of the 2014 data. The ARIMA model provides a better forecast when compared with the RQA predictive model. The analysis of the daily river flow dynamics reveals an abrupt change in the system and the detection of an attractor with outliers indicating higher series values; an indication of an early warning. Models' performance was based on the RMSE criterion and ARIMA model predicted a better result. The chaotic nature of the time series has also been investigated. This study could be used in understanding disastrous consequences of a river flooding.

**Keywords:** ARIMA; flood forecasting; Kelantan River-flow data; recurrence plots; recurrence quantification analysis (RQA).

## 1. Introduction

Every year floods cause massive human and non-human losses in several parts of the world. There is a strong belief that their rate of occurrence is consistently increasing many-fold. An assessment and forecasting of flooding risk requires the use of some ensembles of sophisticated and advanced techniques in an attempt to have a thorough understanding of extreme meteorological and hydrological events. According to Amiri & Mesgari (2018), studying the variability and changes in climate extremes is of great importance. The information and awareness about this natural calamity is more prominent now than ever, due to the revolution in advanced computing power. The multiple factors contributing to flooding are the changes in climatic conditions of certain hotter regions resulting in increased incessant and unrelenting rainfall. Thus, an early detection of causes and trends, which could lead to flooding, is crucial, and this could be possible with a long time series, (Kundzewicz, 2004).

The study presents the result of the application of nonlinear time series analysis methods on river streamflow

data based on the theory of dynamical systems for flood detection and prediction of future values. Extreme values represent high flows such as peak discharges and sudden changes in the dynamics of the river.

### 1.1 Motivation for the research

The studies on the mechanism through which floods occur are critical in developing flood protection and management systems. Some reasonable prediction in developing flood early warning systems is important for expeditious preparations and readiness, thus reducing the disastrous consequences prior to their onset using nonlinear predictive methods.

Several methods have been used in the literature such as ARIMA (Ahlert & Mehta (1981), and Kurunç *et al.* (2005)), SARIMA (Otok & Shabri A. *et al.* (2012)), Internet of Things network (Purkovic *et al.* (2019)), and support vector regression (Yu *et al.* (2006)), application of stochastic flood forecasting model for Kelantan River (Sazali *et al.* (2018)), and artificial neural network (Huang

*et al.* (2004)). Mosavi *et al.* (2018) conducted a survey report on the use of machine learning models for the river flow forecasting. Adamowski *et al.* (2008) developed a short-term river flood forecasting method based on wavelet and cross-wavelet analysis. Zhi-Qiang *et al.* (2016) have used recurrence concept for forecasting extreme volatility in financial risk management. Adenan & Noorani (2013) have used a nonlinear prediction method for a monthly river flow. One of the criteria of evaluating the accuracy of the model simulation of the Kelantan River rainfall-runoff inundation by TAM *et al.* (2019) was RMSE. The study showed that Global Satellite Mapping of Precipitation-Near Real Time (GSMP-NRT) can be reliable.

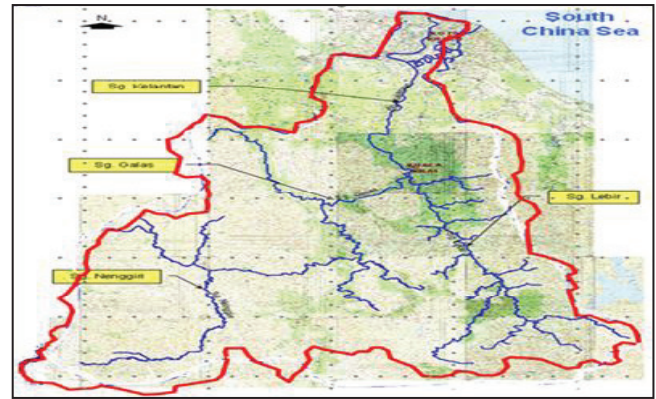
In this paper, approaches based on a nonlinear method combined with ARIMA modeling have been used for Kelantan River flow analysis and its forecasting. Specifically, the analysis and a model for daily river flow forecasting based on the RP, RQA, and ARIMA are presented. Poincaré (1890) described recurrence as a fundamental characteristic of many dynamical systems. The notion of recurrence is one fundamental notion in the theory of dynamical systems. Marwan *et al.* (2007) stated that many dynamical systems in weather and basic sciences are modeled by RQA.

To the best of the authors' knowledge, the methods based on RQA and ARIMA have not been used in the literature for analysis and prediction of the river dynamics in general, and the Kelantan River flow in particular. Our choice of using both lies in the fact that RQA can be used to identify subtle patterns of recurrence in a data series (Webber & Zibult, 1994; Zibult & Webber 1992), and it provides a more objective depiction of the dynamics of a system. It is independent of limiting constraints like the data set size, does not require stationarity, linearity, and any other usual assumptions on the probability distribution of data in a time series. ARIMA models, on the other hand, provide an acceptable approach to time series forecasting, and the aim is to describe the data autocorrelation structure. A comparison between these approaches is expected to enrich the time series analysis.

## 2. Data and study area

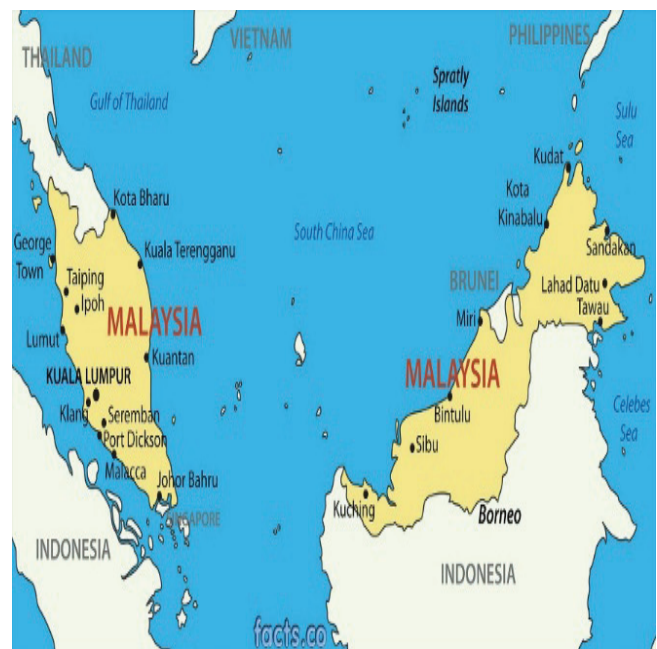
Aminah *et al.* (2016) reported on flood events for Malaysian Kelantan River Watershed for the last Decade (2001–2010). The report concludes that a forecast for

more severe floods may occur in the future. The flood occurred in Kelantan in 2014, caused major losses and according to Baharuddin *et al.* (2015), it was the most significant and the largest recorded flood in the history of Kelantan. Kelantan, located on the north-eastern region of Peninsular Malaysia, (Figure 1) covers an area of 15,099 km<sup>2</sup> in Malaysia (Figure 2). It comprises of ten administrative districts.



**Fig. 1.** Map of Malaysia and location of Kelantan River.

<https://www.expatgo.com/my/2015/09/02/8-maps-for-malaysia/>



**Fig. 2.** Map of Malaysia.

<https://www.expatgo.com/my/2015/09/02/8-maps-for-malaysia/>



Fig. 3. The map of rivers in Malaysia.

<https://www.mapsofworld.com/malaysia/river-map.html>

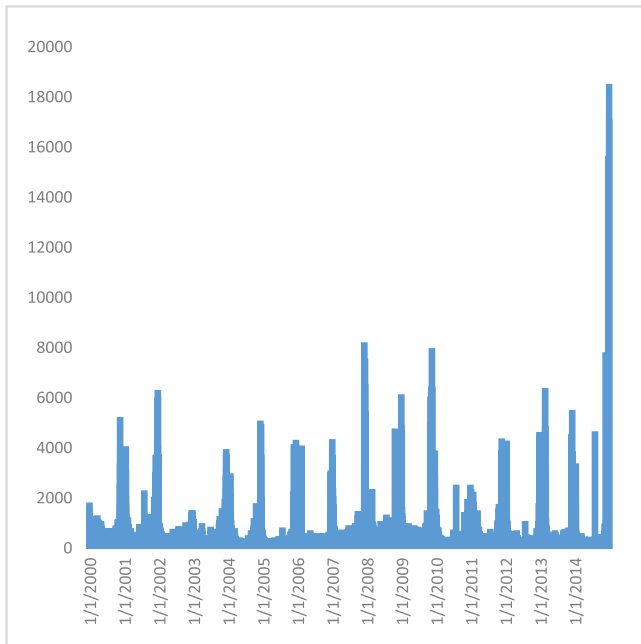


Fig. 4. Kelantan River daily streamflow (in m<sup>3</sup>/s).

We analyzed the daily average values of streamflow collected at different points along the river, based on the location of meteorological stations. The time series data for this study consist of daily Kelantan River flow (in m<sup>3</sup>/s) from January 1, 2000, to December 31, 2014, as shown in Figure 4. From the Figure 4, it appears that the series is pretty much hovering around the same mean river flow level except at certain times such as the end of 2014. Elshorbagy *et al.* (2002) believe that the data

collected from observation stations must be used without any imputations. Therefore, no data imputation techniques were used in this study.

For the purpose of validation analysis, holdout data for the whole of 2014 year (365 daily data values) was used. That is, models to be built in this paper are blinded from the 2014 data, and their respective model building algorithm has been based on data from January 1, 2000, to one day prior to 2014. The holdout 2014 daily data will then be used to test the selected model in terms of their predictive capabilities. The advantage of doing so is to see how well the models fair in predicting or forecasting future values. This is crucial in helping us define an early warning mechanism to detect aberrant river flow patterns that may cause flooding. Some important descriptive statistics of the daily Kelantan River flow data from 2000 to 2014 are given in Table 1.

Table 1. Descriptive statistics of daily river flow data.

Data points	5478.00
Median	267.95
Mode	196.90
Mean	454.03
Standard deviation	795.08
Maximum value	18339.40
Minimum value	49.90
Variance	632146.83
Coefficient of variation	1.75

### 3. Methods

#### 3.1 Phase space reconstruction:

The nonlinear analysis of the dynamical systems, using recurrence concept, can be studied based on phase space trajectories (Marwan *et al.*, 2007). The study of a dynamical system such as river flow presents several challenges primarily due to a lack of knowledge of many variables involved in the system; however, the solution lies in the reconstruction of the system in phase space using the time series (Sivakumar, 2002). The phase space reconstruction of a one-dimensional time series is performed using Takens embedding theorem (1981) of estimating time delay given by the equation (1) using an *m*-dimensional embedding phase space with a carefully chosen time delay  $\tau$ .

$$X^m(t) = [x(t), x(t + \tau), \dots, x(t + (m - 1)\tau)] \quad (1)$$

A suitable embedding dimension will yield global properties of a dynamical system and will preserve its topological geometry in phase space reconstruction. It has a direct relation with the trajectory of an attractor in phase space and neighborhood of points. Too small or too big embedding of data will lead to false recurrences and an incorrect geometric representation (Marwan *et al.*, 2007).

### 3.2 Recurrence plot

Introduced by Eckmann *et al.* (1987), recurrence plots use an advanced technique of nonlinear data analysis to visualize the recurrence of states in a square matrix in which the matrix elements correspond to those times at which a state of a dynamical system recurs (columns and rows correspond then to a certain pair of times). They contain information about the dynamics of the system and dynamic characteristics defined as

$$R_{i,j}(\epsilon) = \Theta(\epsilon - \|\bar{x}_i - \bar{x}_j\|), i, j = 1, \dots, N \quad (2)$$

where  $N$  is the number of measured points  $\bar{x}_i$ ,  $\epsilon$  is the threshold distance, and  $\Theta$  is the Heaviside function in which  $\Theta(x) = 0$  if  $x < 0$  and  $\Theta(x) = 1$ , otherwise.

According to Adenan & Noorani (2013), two parameters have to be determined, a value of  $\tau$ , as the most optimal time delay, which can provide a separation of neighboring projections in any dimension embedded in the phase space, and the other parameter is embedding dimension  $m$  which is varied. The optimal embedding dimension for phase space reconstruction can be determined by choosing the increasing values of embedding dimension and selecting the value that yields a reasonably acceptable prediction (Sivakumar, 2002).

The visual inspection of the structures in recurrence plots (RP) is usually not considered the only objective tool for making conclusions about the nature of the system. The more objective approach is the application of RQA and their indices calculated for quantitative confirmation.

### 3.3 RQA (quantitative measures of complexity)

RQA can be used to identify subtle patterns of recurrence in a data series (Webber & Zbilut, 1994), and it provides a more objective depiction of the dynamics of a system than the visuals represented through recurrence plots. These measures are based on the recurrence point density and the diagonal and vertical line structures of the RP

(Marwan *et al.*, 2007).

RQA can project the input signal into higher-dimensional space (e.g., embedding capabilities), and it functions like a microscope (snooping out higher-dimensional subtleties in the dynamics that are not obvious in its first-dimensional representation). For a detailed description of RQA measures, readers can refer to Marwan *et al.* (2007).

RQA incorporates two essential modelling features. The first is that quantification does not depend upon mathematical transformations like the Fourier transform that is linear, wavelets, and so on. The second is that quantification can be carried out in multiple dimensions through the method of time delays that is important when treating chaotic-like systems that are multidimensional and nonlinear (Webber & Zbilut, 2002).

### 3.4 Dynamical invariants from RP: Entropy

The Shannon information entropy-ENT (Shannon, 1948) is the length of the diagonal lines (equation 3), is a measure of complexity of the system and can be related to Shannon entropy (Marwan *et al.*, 2007).

$$ENT = -\sum(P_{bin})\log_2(P_{bin}) \quad (3)$$

In RQA, an estimate of the correlation entropy  $K_2$  can be calculated, and its measure is primarily used as an estimate of predictability of a time series (Thiel *et al.* (2003). Theoretically, for a perfect deterministic system,  $K_2=0$  can be interpreted as only one possibility for the trajectory to evolve.  $K_2 \rightarrow \infty$  for a pure stochastic system means that the number of possible future trajectories increases to infinity, so quickly. If the value of  $K_2$  is finite, then the system is considered chaotic. For chaotic systems dynamics, the inverse of  $K_2$  has a unit of time and can be interpreted as the mean prediction horizon/time of the system (Bloh *et al.* (2005); Li *et al.* (2008)).

The embedding dimension ' $m$ ' of the daily data and the time delay are calculated using Visual Recurrence Analysis (VRA) software (Eugene, 2002). The VRA software has been used for nonparametric time series prediction. The software uses a number of predictor models and kernel functions. The model that provides the smallest RMSE and the NMSE that provides a value less than 1 were chosen, and it was found that the Kernel Regression as a predictor model and the Tricube as the Kernel function are the best combination as they provided better forecasting accuracy.

#### 4. Phases of the study

There are three phases in this study: (1) model building phase, (2) predictive phase, and (3) recurrence quantification analysis phase. The model building phase will use Kelantan daily river flow data prior to 2014, while the predictive phase will feature the performance of the chosen models on the holdout 2014 daily river flow data. The models considered in this paper can be categorized as (1) ARIMA models and (2) RQA-based nonlinear nonparametric time series models. In the model-building phase, each considered model will be subjected to certain criteria to select the best performing models.

##### 4.1 Model building phase

###### 4.1.1. ARIMA models

First, we check the most prominent characteristics of the time series data. We check the correlation between the original data and the  $k^{\text{th}}$  lagged data by the use of the autocorrelation function. This tells us how much the current data relies on its own past values and usually helps us in identifying the order of the moving average component of the time series.

In addition to this, we also examined the correlation between the original data and the  $k^{\text{th}}$  lagged data while correcting for intervening lag correlation. This is done through the partial autocorrelation function. This partial autocorrelation function is useful to detect the autoregressive nature of the data, if any.

The criteria for choosing the best ARIMA model are (1) root mean square error (RMSE) and (2) correlation between the fitted values and the original data. The RMSE criterion measures the overall discrepancy between the original data and the predicted data. The model with the best fidelity to the data has the smallest RMSE value. The correlation, which typically ranges from -1 to 1, between the fitted and original data values shows how strong is the relationship between the model and the actual data.

##### 4.2 Predictive phase

In this phase, the models chosen in each category are tested on the 2014 holdout daily river flow data. That is, the target of this analysis is to compare the capabilities of the selected models in predicting the 2014 daily data values. To provide a common evaluation yardstick between models, the prediction root mean square error (RMSE) is used. The criterion measures the overall discrepancy between original data and the predicted data. The model

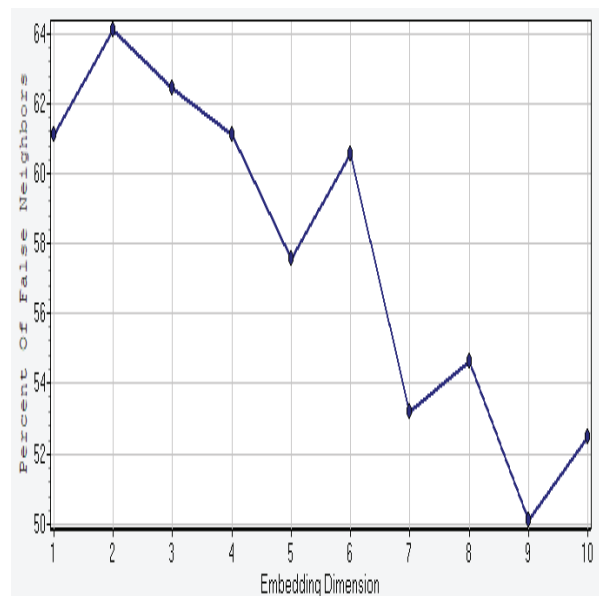
with the best prediction is expected to have the smallest prediction RMSE value.

#### 5. Results and discussion

The dynamic complexity of the river flow of the time series consists of 5478 data points computed and analyzed through constructing the RP and extracting the RQA measures. The RQA measures that were calculated in this study are %RR, %DET,  $L_{\text{MAX}}$ ,  $V_{\text{MAX}}$ ,  $L_{\text{MEAN}}$ , LAM, and RATIO (Table 2) including the dynamical invariant ENTR. The analysis was carried out using the CRP toolbox for MATLAB developed by Marwan: Cross Recurrence Plot Toolbox for MATLAB®, Ver. 5.22 (R32.4), <http://tocsy.pik-potsdam.de/CRPtoolbox/>, and accessed 25-05-2020.

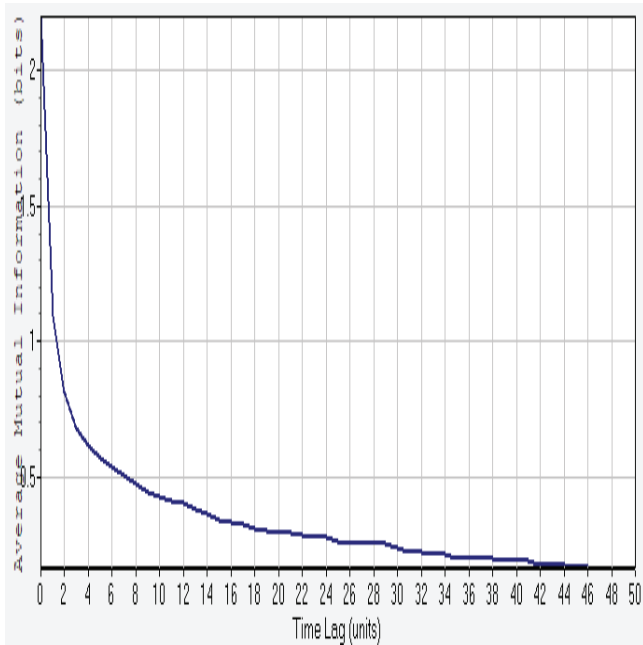
##### 5.1 Optimal parameters calculation

The choice of parameters will affect the overall performance of the framework. Therefore, we are using the optimal time delay and embedding dimension proposed by the methods such as average mutual information and false nearest neighbors respectively. The time delay, where the average mutual information reaches its first minimum, can be considered as optimal time delay. The embedding dimension  $m = 9$  (Figure 5) and time delay  $\tau = 27$  (Figure 6) were obtained using methods of false nearest neighbor analysis and average mutual information respectively. The first minimum of the mutual information function (Fraser & Swinney, 1986) found is considered as the “optimal” value of the time delay for the state space reconstruction.



**Fig. 5.** Optimum global embedding dimension  $m=9$  calculated using the method of false nearest neighbors.

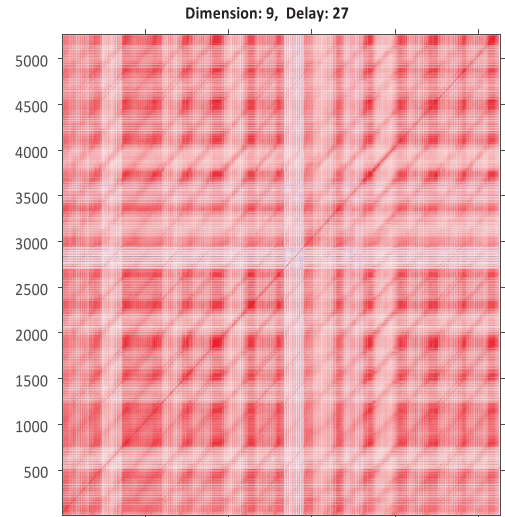
The method of false nearest neighbor was used for choosing the minimum embedding dimension (Kennel *et al.*, 1992) to find the nearest neighbor of every point in  $d$ -dimension and to see the vicinity of neighborhood points in a higher dimension.



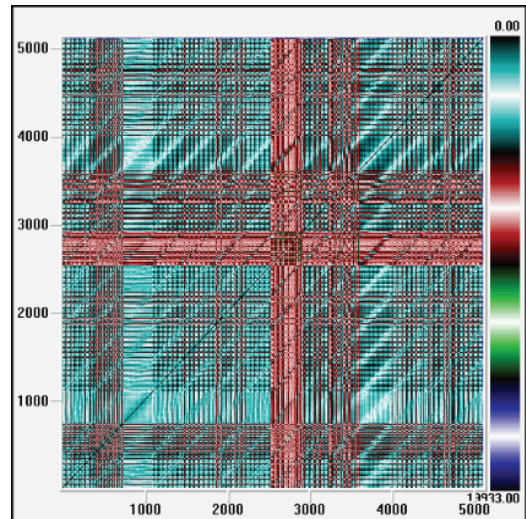
**Fig. 6.** The first average mutual information minimum was found at time lag 27.

5.2. Recurrence plot

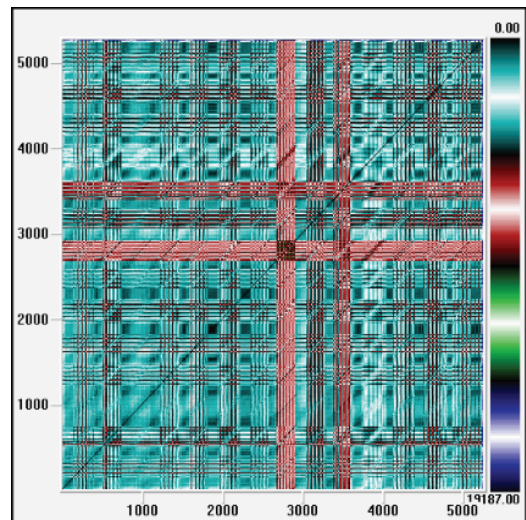
From the visual inspection analysis of the line structures in the recurrence plots (Figures 7a, 7b, and 7c), the existence of the horizontal, vertical, and diagonal line structures is revealed. There are no apparent single isolated points that indicate uncorrelated randomness in the system. There are more visible diagonal lines parallel to the line of incidence (the main diagonal line), which is an indication of the similar evolution of states at different times, which imply that the process is more deterministic in nature, suggesting higher predictability. The vertical and horizontal lines in the recurrence plot exhibit that some states do not change or may change slowly for some time, which is an indication for laminar states.



**Fig. 7a.** Dimension = 9, delay =27



**Fig. 7b.** Dimension = 38, delay = 10



**Fig. 7c.** Dimension = 23, delay = 10

**Fig. 7 (a–c):** Recurrence plots of the Kelantan River flow time series (using CRP and VRA software).

Several thick vertical lines in the recurrence plots observed in figure 7 (a-c) can be considered peak streamflow of the river, sudden changes in river dynamics (Fragkou, 2018). This was further confirmed by the recurrence rate values.

5.2 RQA (quantitative measures of complexity)

The recurrent points are measured using %RR= 0.27, indicating the variation in the flow of the river, and this measure describes the complexity of the system. The %DET is 0.89 as mentioned in Table 2, which is considered high. It means that the time series data is deterministic in nature, a measure of the predictability of the time series data. This measure is not a sufficient proof of the determinism, and it must be cross-validated by other measures (Marwan *et al.*, 2007). LAM is related with DET as their measures are similar. The laminarity (%LAM=0.94), the measurement of recurrent points related to the vertical line structures, and trapping time (TT=7.14), the average length of the vertical line structures, are presented by Marwan *et al.* (2002). Furthermore, both laminarity and trapping time are related with %DET.

The Length of the longest diagonal line  $L_{max} = 170$  is larger and is interpreted as a not very chaotic system; a possibility of prediction is high. The shorter size of this line implies that the dynamical system would be more chaotic. According to Eckmann *et al.* (1987) and Trulla *et al.* (1996), the length of the longest diagonal line inversely scales with the most positive Lyapunov exponent, an important variable related to prediction.

**Table 2.** The RQA measures of the river flow time series data.

RQA Measures	RQA value
%RR	0.27
%DET	0.89
$L_{mean}$	4.94
$L_{max}$	170.00
%LAM	0.94
TT	7.14
$V_{max}$	162.00
ENTR (diagonal length)	2.19

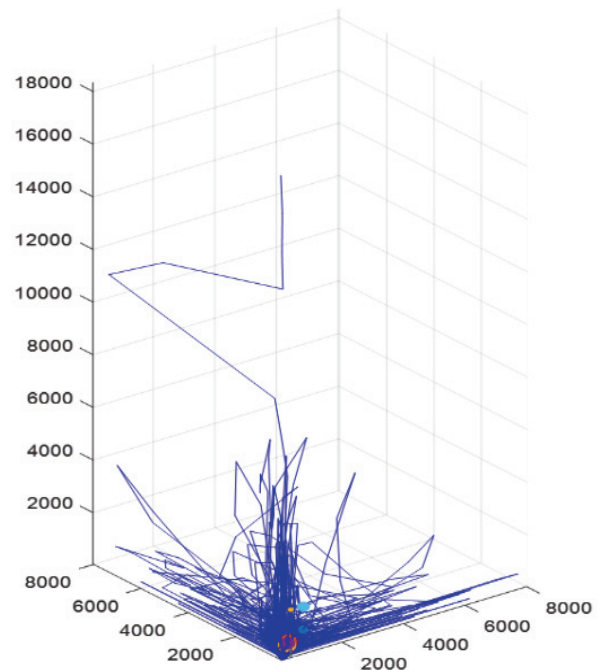
5.3 Dynamical invariants ( $K_2$ -entropy):

The value of the dynamical invariant  $K_2$  derived from the recurrence plots for the time series data is 0.203, which is calculated using <https://tocsy.pik-potsdam.de/k2.php>.  $K_2$

calculation is based on Grassberger & Procaccia method (1983a), and some related entropy measure is discussed by Hamza *et al.* (2003). It is an indication that the dynamical system is classified as chaotic. The chaotic nature of the series was further tested by the z1 test by Gottwald *et al.* (2009), and the calculated value is 0.9983. According to this test, any value for nonchaotic data approaches to 0, and it is near 1 for chaotic data; see <http://arxiv.org/pdf/0906.1418v1>.

5.5. Phase space reconstruction

Figure 8 shows 3D phase space plot reconstruction in higher dimension with time delay  $\tau=27$  and  $m=9$ , respectively. The plots show the dynamics of the daily river flow and the detection of an attractor with outliers indicating higher values of the series, an onset of a sudden change in the state of the system. SivaKumar (2002) considered the presence of the attractor in the phase space as an evidence of the low-dimensional chaotic behavior of the dynamical system. Adenan & Noorani (2013) concluded that, for a chaotic system, the attractor trajectories can be reasonably well defined, and therefore it can be predicted with confidence using the nonlinear prediction method and the chaotic theory approach. The chaotic nature of the time series data with the calculation of correlation Entropy coupled with a test of chaos has further been proved.



**Fig. 8.** 3D phase space plot of daily river flow data with time delay = 27,  $m=9$  of Kelantan River Marwan CRP MATLAB Toolbox.

5.6. ARIMA modeling

The data is divided as a training set for a model fitting, and the other part of the data as forecasting model validation. The time series river flow data of the Kelantan River comprised of the years 2000 to 2013 is used as training data to predict the values of the year 2014 using two methods, namely, RQA and ARIMA. Figure 9 shows the autocorrelation values for the Kelantan River flow data.

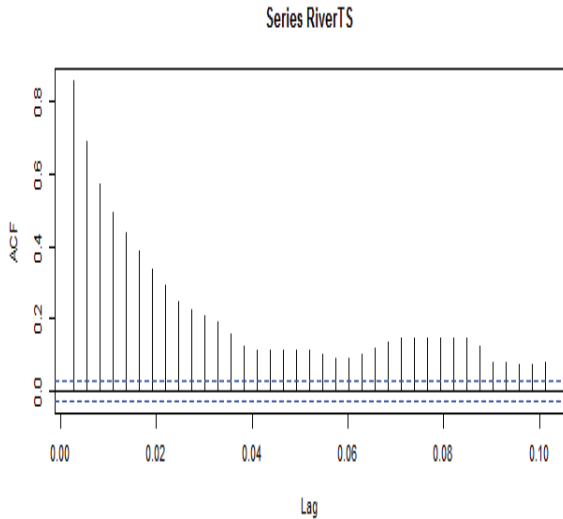


Fig. 9. Autocorrelation values for the Kelantan River flow data.

The plot shows large, but generally decaying autocorrelation values. That is, the current data relies heavily on its past values. Partial autocorrelation values for Kelantan River flow data are presented in Figure 10. The plot shows mainly large partial autocorrelations at the first to around the fifth observation lags.

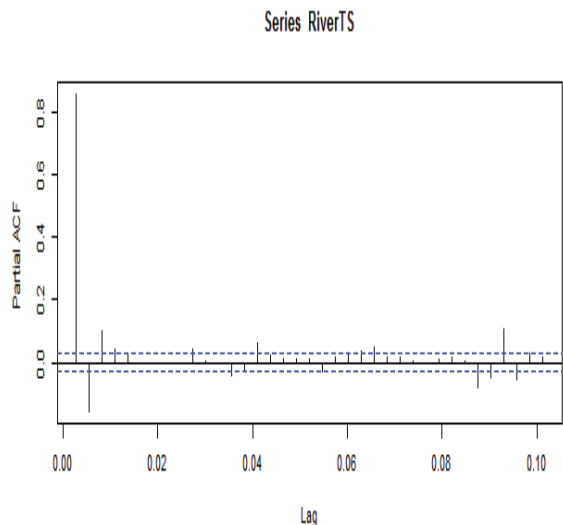


Fig. 10. Partial autocorrelation values for the Kelantan River flow data.

Based on both Figures 9 and 10, we may surmise from the data that a mixed autoregressive and moving average model is probably appropriate for the data. Thus, we use the extended autocorrelation function (EACF) to find the ARIMA model that has the best autoregressive (AR) and moving average (MA) orders. Based on the Figure 11 below, we do not see a clear-cut triangle of zero with vertex pointing to the best or recommended model. However, based on the Figure 11 (as highlighted), we find the two contending models to be ARIMA (1, 0, 4) and ARIMA (5, 0, 3).

AR/MA	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	x	x	x	x	x	x	x	x	x	x	x	x	x	x
1	x	x	x	x	o	x	o	o	o	o	o	o	o	o
2	x	x	x	x	x	o	o	o	o	o	o	o	o	o
3	x	x	x	x	x	o	o	o	o	o	o	o	o	o
4	x	x	x	x	x	o	o	o	o	o	o	o	o	o
5	x	x	x	o	x	o	o	o	o	o	o	o	o	o
6	x	x	x	x	o	o	o	o	o	o	o	o	x	o
7	x	x	x	x	o	o	o	o	o	o	o	o	o	o

Fig. 11. Extended autocorrelation values for the Kelantan River flow data.

To check the comparative quality of these contending ARIMA models, we checked their estimated error variance, log-likelihood, and Akaike information criterion (AIC) as described in Table 3.

Table 3. Comparison of several quality values of competing ARIMA models.

Competing models	$\hat{\sigma}_e^2$	Log likelihood	AIC
ARIMA(5,0,3)	80727	-36115.61	72251.22
ARIMA(1,0,3)	112893	-39640.37	79290.73
ARIMA(1,0,4)	112546	-39632.02	79276.04
ARIMA(5,0,0)	112453	-39629.68	79271.36

We also added the ARIMA (5, 0, 0) and ARIMA (1, 0, 3) to aid in our comparison. It is clear from Table 3 that ARIMA (5, 0, 3) has the lowest estimated error variance, the highest log-likelihood, and the lowest AIC among the considered models, and thus it is the most desirable model considered. The Auto-Arima command in R also found that the best model chosen for the data is the ARIMA model with order (5, 0, 3).

To check the quality in the fit of the model to the data, we examined the correlation value between the values



fitted by the model and actual Kelantan River flow from January 1, 2000, to December 2013. The correlation value between the predicted values and actual values was found to be 0.877 for the ARIMA model, and this value is considered highly acceptable. The RMSE calculated value of 283.874 happened to be much smaller than that of the RQA based model. The model accuracy with optimal parameters is observed for the value  $m = 38$ ; hence, time delay=10.

5.7. Forecasting of 2014 daily river flow

Figure 12a shows the full Kelantan River flow data (red values are 2014 daily data) with confidence bands (80% and 95%) and predicted values (in blue). It is clear that most of the actual 2014 data values are within both the shorter 80% and wider 95% confidence bands except for some values at the beginning of 2014, at the latter half of 2014, and at the end of 2014. Malaysian historical records (Aminah J. *et al.*, 2016) show that, towards the end of 2014, there was record flooding around the Kelantan state. So, our ARIMA model was able to pick up this flooding information exhibiting it as aberrantly fast river flows outside of the 95% confidence bands. These aberrant levels were also very atypical of the usual river flow levels, which had a maximum less than 8000 m<sup>3</sup>/s prior to 2014 as can be seen to be outside of the chart of Figure 12b. Aberrant values outside of the 95% bands in the latter half of 2014 may be thought of as an early warning for the 2014 flooding event.

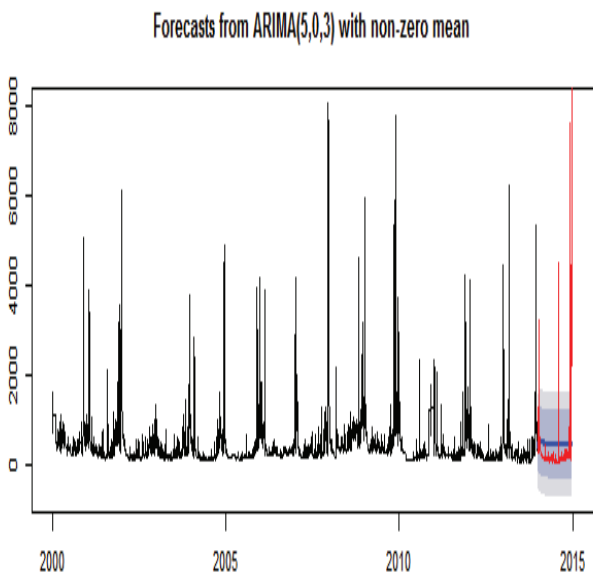
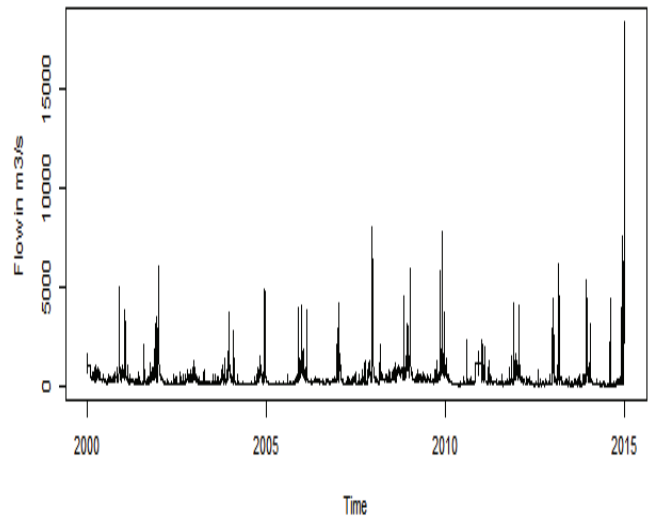


Fig. 12a): Kelantan River flow data (black), holdout (in red) 2014 data, and predicted values (in blue).

Kelantan River - Daily Stream Flow



b) Kelantan River flow data for all years including the 2014 holdout data.

ARIMA model has reported lower values of RMSE compared to the RQA method. Hence, the prediction accuracy of the former was better than the latter. Table 4 provides the prediction accuracy of the chosen model. The RMSE calculated by the ARIMA model (1530.13) is far less than the values generated by the RQA method (1988.96) with optimal embedding dimensions and time delays.

Table 4. Comparison of RMSE values of ARIMA and RQA.

METHOD	RMSE value	NMSE value	Correlation between fitted value and original series
ARIMA	1530.136	N/A	0.877626
RQA	1988.96	0.86	N/A

Table 5 provides the RMSE and normalized error values for embedding dimensions 23 and 38 with a time delay of 10 each. Table 6 provides the RMSE and normalized error values for embedding dimensions  $m=5, 10, 15$  with a time delay of 1. For embedding dimension 23 and above with a time delay of 10, the NMSE values are less than 1. However, the RMSE for  $m=23$  and  $\tau=10$  is smaller than that of the value of  $m=38$  and  $\tau=10$ . With the time delay =1 and the embedding dimension being varied between 5 and 15, lower NMSE values are observed. It can be concluded that choosing the optimal embedding

dimension ( $m=23$ ) with time delay  $\tau=10$  provides the best prediction as the NMSE value is less than 1, and the RMSE value is 1988.96.

**Table 5.** Comparison of RMSE and NMSE result with embedding dimensions 23 and 38 with a time delay of 10 (Predictor: Kernel Regression, Kernel Function: Tricube).

Embedding Dimension (m)	Time Delay ( $\tau$ )	RMSE	NMSE
m=23	$\tau=10$	1988.96	0.86
m=38	$\tau=10$	2129.04	0.99

**Table 6.** Comparison of RMSE and NMSE values with embedding dimensions 5, 10, & 15 and the time delay 1.

Embedding Dimension (m)	Time Delay ( $\tau$ )	RMSE	NMSE
m=5, 10, 15	$\tau=1$	2129.82	0.99

## 6. Conclusion

The analysis of the daily river flow dynamics reveals an abrupt change and the detection of an attractor with outliers indicating higher series values are indicative of an early warning. The aberrant behavior of the river flow was observed, in the year 2014 that is, a holdout data. The detection of this unusual event or abrupt change is an early warning sign of the oncoming potential flooding event, and the value of the peak is observed towards the end of the year, and the critical values lie outside the confidence bands. The dynamics of the river flow time series is analyzed by using RQA method that can be summarized as follows:

- Recurrence plots of the series (Figure 7) have clearly horizontal, vertical, and diagonal lines. Its visual inspection reveals that the system is not purely stochastic due to the existence of visible lines. There are some noticeable line segments parallel to the main diagonal line, which is an indication of a nonrandom nature of the time series.
- The prediction results using the nonlinear prediction method are compared with ARIMA. The results show that the ARIMA model provides better prediction results than nonlinear nonparametric prediction method based on RQA.

Adenan and Noorani (2013) reported that the nonlinear prediction method gave better prediction than ARIMA and SVM. In our study, however, the ARIMA method gave better prediction result than the nonlinear RQA method. TAM *et al.* (2019) and Yu (2006) have used RMSE as one criterion for evaluating the performance of the predictive model. Our study does not compare either RQA or ARIMA to these methods but uses the same comparative quality criteria of RMSE.

Other methods have also been discussed in the literature such as machine learning models (Mosavi *et al.*, 2018) and wavelet and cross-wavelet analysis by Adamowski (2008).

Future studies could be done comparing these methods, but for the purpose of space conservation and focus, we have not done this in this study. In future studies, the data from other neighboring stations can also be used in addition to the investigation of other predictive factors such as rainfall precipitation, the type of soil, and other contributing factors of rapid rise in the river flow.

In summary, RQA can be used to quantify the presence of repeating patterns and behavioral transitions through statistical values. In order to distinguish normal regimes from critical regimes, the technique can be used to study the change in a dynamical system.

## ACKNOWLEDGEMENTS

The first and third authors acknowledge the academic support from UKM. No financial grant for this work was sought. The second author acknowledges the research facilities at KFUPM.

The authors wish to acknowledge usage of the CRP toolbox, developed by Norbert Marwan *et al.* (2014), the Visual Recurrence Analysis software by Eugene Kononov for RQA.

## References

- A.D. Fragkou, T.E. Karakasidis & E. Nathanail (2018).** Detection of traffic incidents using nonlinear time series analysis, *Chaos* 28, 063108; doi: 10.1063/1.5024924.
- Adamowski J. & Adamowski K. (2019).** Development of a real-time river flood forecasting transfer function-noise model with a Kalman filter for snowmelt driven floods, *Journal of Environmental Hydrology*, Volume 16 Paper 26 August 20081

- Adenan N.H. & M.S.M. Noorani (2013).** Monthly River Flow Prediction Using a Nonlinear Prediction Method, World Academy of Science, Engineering and Technology International Journal of Mathematical and Computational Sciences, 7(11):17410.
- Ahlert, R.C. & Mehta, B.M. (1981).** Stochastic analyses and transfer functions for flows of the upper Delaware River, Ecological Modelling, 14(1-2), pages 59-78.
- Aminah Shakirah J, L M Sidek1 B Hidayah Nazirul, M.Z M. Jajarmizadeh, F.C. Ros & ZA Roseli (2016).** A review on flood events for Kelantan River watershed in malaysia for last decade (2001-2010), IOP
- Conference Series: Earth and Environmental Science 32(1):012070.**
- Amiri, M.A. & Mesgari, M.S. (2018).** Analyzing the spatial variability of precipitation extremes along longitude and latitude, northwest Iran, Kuwait Journal of Science, 45(1), pp 121-127.
- B. Sivakumar (2002).** A phase-space reconstruction approach to prediction of suspended sediment concentration in rivers, Journal of Hydrology, 258:149-162.
- Bloh, W.V., Romano, M.C. & Thiel, M. (2005).** Longterm predictability of mean daily temperature data, Nonlin. Proc. Geoph. 12(471-479).
- Eckmann, J.P., Kamphorst, S.O. & Ruelle, D. (1987).** Recurrence plots of dynamical systems. Europhysics Letters, 4: 973-977.
- Elshorbagy, A., Simonovic, S.P. & Panu, U.S. (2002).** Noise reduction in chaotic hydrologic time series facts and doubts, J. Hydrol., 256(3/4), 845-848.
- F. Takens (1981).** Detecting strange attractors in turbulence. In D. A. Rand and L.-S. Young (ed.). Dynamical Systems and Turbulence, Lecture Notes in Mathematics, Springer-Verlag 898, 366-381.
- Fraser, A.M. & Swinney, H.L. (1986).** Independent coordinates for strange attractors from mutual information, Physics Review A, 33 1134-1140.
- Gottwald, Georg A. & Melbourne, Ian. (2009).** On the Implementation of the 0–1 Test for Chaos, <http://arxiv.org/pdf/0906.1418v1>. Paul Matthews, July 2009.
- H. Poincaré (1890).** On the problem of the three bodies and the equations of dynamics, Acta Mathematica, 13(270).
- Hamza, B.A. & Krim, H. (2003).** Jensen-Rhyi Divergence Measure: Theoretical and Computational Perspectives, ISIT 2003, Yokohama, Japan.
- Joseph P. Zbilut, Nitzza Thomasson & Charles L. Webber (2002).** Recurrence quantification analysis as a tool for nonlinear exploration of nonstationary cardiac signals. Medical engineering & physics, 24(1):53-60.
- Kamarul Aryffin Baharuddin, Shaik Farid Abdull Wahab, Nik Hisamuddin Nik Ab Rahman, Nik Arif Nik Mohamad, Tuan Hairulnizam Tuan Kamauzaman, Abu Yazid Md Noh & Mohd Roslani Abdul Majod (2015).** The Record-Setting Flood of 2014 in Kelantan: Challenges and Recommendations from an Emergency Medicine Perspective and Why the Medical Campus Stood Dry, Malays J Med Sci. 22(2): 1-7.
- Kennel, M.B., Brown, R. & Abarbanel, H.D.I. (1992).** Determining embedding dimension for phase-space reconstruction using a geometrical construction, Phys. Rev. A 45, 34030.
- Kononov, E. (2002).** “VRA”, v.4.2, W.Springfield, MA 01089, USA, <http://pw1.netcom.com/~eugenek>
- Kundzewicz, Z.W. (2012).** In search for a change in hydrological data. Hydrol. Sci. J. 49(1) 3-6.
- Kurunç, A., Yurekli, K. & Ozturk, F. (2005).** Effect of Discharge Fluctuation on Water Quality Variables from the Yeşilirmak River TARIM BİLİMLERİ DERGİSİ, 11(2): 189-195.
- Li, S.C., Zhao, Z.Q. & Liu, F.Y. (2008).** Identifying spatial pattern of NDVI series dynamics using recurrence quantification analysis, Eur. Phys. J. Special, 164(127-139).
- Map of Malaysia and location of Kelantan River, <https://www.expatgo.com/my/2015/09/02/8-maps-for-malaysia>.
- Map of Malaysia, <https://www.mapsofworld.com/malaysia/river-map.html>
- Map of rivers in Malaysia, <https://www.mapsofworld.com/malaysia/river-map.html>.
- Marwan, N., Romano, M.C., Thiel, M. & Kurths, J. (2007).** Recurrence plots for the analysis of complex systems, Science Direct, Physics Reports, Physics Reports, 438(237-329).
- Marwan, N., Wessel, N., Meyerfeldt, U., Schirdewan, A. & Kurths, J. (2002).** Recurrence-plot-based measures of complexity and their application to heart rate variability data. Physical Review E, 66, 026702.1-026702.8.

**Marwan Norbert, M. Carmen Romano, Marco Thiel & Jurgen Kurths (2007).** Recurrence plots for the analysis of complex systems, *Physics Reports*, **438**: 237-329.

**Mosavi Amir, Pinar Ozturk & Kwok-wing Chau (2018).** Flood Prediction Using Machine Learning Models: Literature Review, *Water* 2018, 10, 1536; DOI: 10.3390/w10111536 .

**P. Grassberger & I. Procaccia (1983a).** Estimation of the Kolmogorov entropy from a chaotic signal, *Physical Review A*, **9**(1-2): 2591-2593.

**Purkovic, D., Coates, L., Hönsch, M., Lumbeck, D. & Schmidt, F. (2019).** Smart river monitoring and early flood detection system in Japan developed with the EnOcean long range sensor technology, 2019 2nd International Colloquium on Smart Grid Metrology (SMAGRIMET), DOI: 10.23919/SMAGRIMET.2019.8720390.

**Shabri, A. & Suhartono (2012).** Streamflow forecasting using least-squares support vector machines, *Hydrological Sciences Journal*, **57**(7): 1275-1293.

**Sazali Osman, Norizan Abdul Aziz, Nurul Husaif, Lariyah Mohd Sidek, Aminah Shakirah, Faedah Hanum & Hidayah Basri (2018).** Application of Stochastic Flood Forecasting Model Using Regression Method for Kelantan Catchment, *MATEC Web of Conferences* 203, 07001, ICCEOEE 2018.

**Shannon, C.E. (1948).** A mathematical theory of communication. *Bell Systems Technical Journal*, **27**(379-423), 623-656.

system in Japan developed with the EnOcean long range sensor technology, 2nd International Colloquium on Smart Grid Metrology, Split, Croatia.

**Tam Tze Huey, Muhammad Zulkarnain Abd Rahman, Sobri Harun, Muhammad Nassir Hanapi & Ismaila Usman Kaoje.** Application of Satellite Rainfall Products for Flood Inundation Modelling in Kelantan River Basin, Malaysia, *Hydrology* 2019, 6, 95; DOI:10.3390/hydrology6040095.

**Thiel, M., Romano, M.C. & Kurths, J. (2003).** Analytical Description of Recurrence Plots of white noise and chaotic processes, *Applied Nonlinear Dynamics*, **11**(3): 20-30.

**W. Huang, B. Xu. & A. Chan-Hilton (2004).** Forecasting flows in Apalachicola River using neural networks, *Hydrological Processes*, **18**(13): 2545-2564.

**Webber, C.L., Jr. & Zbilut, J.P. (1994).** Dynamical assessment of physiological systems and states using

recurrence plot strategies. *Journal of Applied Physiology*, **76**: 965-973.

**Yu Pao-Shan, Shien-Tsung Chen & I-Fan Chang (2006).** Support vector regression for real-time flood stage forecasting, *Journal of Hydrology*, Volume 328, Issues 3-4, Pages 704-716.

**Zhi-Qiang Jiang, Askery Canabarro, Boris Podobnik, H. Eugene Stanley & Wei-Xing Zhou, (2016).** Early warning of large volatilities based on recurrence interval analysis in Chinese stock markets, 1713-1724, <https://doi.org/10.1080/14697688.2016.1175656>.

**Submitted :** 27/12/2019

**Revised :** 14/01/2020

**Accepted :** 11/06/2020

**DOI :** 10.48129/kjs.v48i1.8899