How to cite this paper:

Siti Nor Fathihah Azahari, Mahmod Othman, & Rizauddin Saian (2017). An enhancement of sliding window algorithm for rainfall forecasting in Zulikha, J. & N. H. Zakaria (Eds.), Proceedings of the 6th International Conference on Computing & Informatics (pp 23-28). Sintok: School of Computing.

AN ENHANCEMENT OF SLIDING WINDOW ALGORITHM FOR RAINFALL FORECASTING

Siti Nor Fathihah Azahari[1], Mahmod Othman[2], and Rizauddin Saian[3]

[1]*Universiti Teknologi MARA, Malaysia, fathihah08@gmail.com*
[2]*Universiti Teknologi Petronas, Malaysia, mahmod.othman@utp.edu.my*
[3]*Universiti Teknologi MARA, Malaysia, rizauddin@perlis.uitm.edu.my*

ABSTRACT. Various rainfall forecasting models or techniques are presented by researchers to obtain the best result of forecasting. Despite of various techniques and methods, not all of them produce the satisfy result of rainfall forecasting. Therefore, this study proposed a forecasting rainfall method based on sliding window algorithm (SWA), in order to obtain the best rainfall prediction value. The problem statements in this study are related to unsatisfactory accuracy rainfall output on previous study of SWA. Hence, SWA is enhanced in order to produce highly accurate prediction rainfall. The proposed method is tested by using three different rainfall gauge station data that are taken from the Department of Drainage Irrigation (DID), Perlis, Malaysia. Then, the rainfall forecasting result is validated by using mean square error (MSE), and relative geometric root mean squared error (relative GRMSE). The validation analysis shows that the proposed method has a higher forecasting accuracy than the previous method of sliding window algorithm.

Keywords: sliding window, fuzzy time series, rainfall forecasting

INTRODUCTION

Rainfall is a nature phenomenon that always happened around us. The forecasting of the amount of rainfall distribution is very important since it could affect everyday life (Asklany, Elhelow, Youssef, & Abd El-wahab, 2011). Previously, many researchers have proposed rainfall forecasting models. For example, Dani and Sharma (2013) utilized a fuzzy time series model to forecast rainfall in Ambikapur. The study is based on fuzzy time series high order of time-invariant model where the concept is originally introduced by Song and Chissom (1993) in order to solve difficulties of forecasting problem when using linguistic value of historical data. Besides, Artificial Neural Network (ANN) is also widely used by researcher in rainfall forecasting (French, Krajewski, & Cuykendall, 1992; Hung, Babel, Weesakul, & Tripathi, 2008; Lee, Cho, & Wong, 1998; Luk, Ball, & Sharma, 2001). Meanwhile, Zhang (2003) had deployed hybrid methodology of Autoregressive integrated moving average (ARIMA) and artificial neural networks (ANNs) to improve forecasting accuracy.

In the research of Datar, Gionis, Indyk, and Motwani (2002), a model of forecasting is introduced by them namely sliding window that is setting for data streams, which aggregates and statistics are computed over a "sliding window" of the N most of recent items in the data stream. A number of interesting results presented on estimating functions over a sliding window for a single stream are obtained. A basic problem that they considered is determines the number of 1's in a sliding window which known as the Basic Counting problem. Sliding window algorithm is widely used in various temporal applications like medical (d'Arcy, Collins, Rowland, Padhani, & Leach, 2002) weather forecasting (BenYahmed, Abu Bakar, Hamdan, Ahmed, & Syed Abdullah, 2015; Kapoor & Bedi, 2013), and database system (Arasu & Manku, 2004).

Kapoor and Bedi (2013) had proposed sliding window algorithm (SWA) to predict weather condition. Four types of weather condition data from the years of 2006 to 2010 which are minimum temperature, maximum temperature, humidity, and rainfall are tested to the algorithm. The experimental results demonstrated that the average accuracy of the model applied is 92.2% which is quite efficient and precise. Similarly, Rao et al. (2015) had implemented the same SWA to study prognostication of climate in Chirala of Adhra Pradesh, India. The forecasting result shows that the accuracy average is 94.21% which also quite efficient and accurate. Meanwhile, BenYahmed et al. (2015) proposed adaptive sliding window algorithm for weather segmentation where the technique of sliding window is operate through time series sequence that detect error (change point) whenever the time series value changes, and then groups each two time points with the change point into one window. Consequently, a high number of windows and low number of time points for each window are generated. The estimation that is implemented by the change point of fast-updating algorithms obtains optimal segments of the time series in the time windows which give information to find out segmentation criteria that help in enhancing error detection and setting the window size for the research (Braverman, Ostovsky, & Zaniolo, 2009). As an addition, Ishak, Ku-Mahamud, and Norwawi (2011) applied window sliding technique to extract information from the reservoir operational database: a digital version of the reservoir operation log book. Several data sets are constructed based on different sliding window size. Artificial neural network is used as modelling tool. The findings indicate that eight days is the significant time lags between upstream rainfall and reservoir water level.

Hence in this study, the sliding window concept that was proposed by Kapoor and Bedi (2013) was enhanced in order to produce satisfactory accuracy rainfall forecasting result where in their study the algorithm is end up by identifying the predicted variation "V" that will add to the previous rainfall data to produce rainfall forecasting value. However, in this study the enhancement is done by defining the average monthly rainfall, *AM* from January 1990 to December 2013, and added to predicted variation "V" where the patterns of monthly rainfall for previous month are influencing the pattern of forecasted rainfall. Finally, the results of forecasted rainfall of the proposed algorithm is compared to the SWA by Kapoor and Bedi (2013) using mean squared error (MSE) and relative geometric root mean squared error (relative GRMSE).

PROPOSED RAINFALL MODEL

In this experiment, a concept of sliding window algorithm (SWA) based on Kapoor and Bedi (2013) study, is proposed and enhanced to produce forecasting result with better accuracy. Three different rainfall gauge station data which are Bukit Temiang, Abi Kg. Bahru and Guar Nangka are used. The Department of Irrigation and Drainage (DID), Malaysia, provided the rainfall data from January 1990 to December 2014. The objective of this study is to forecast average monthly rainfall for January to December 2014. Figure 1 shows the steps of proposed algorithm that is enhance for rainfall forecasting.

Step 1: Construct matrix Current Year, *CY* and Previous Year, *PY*, from average monthly rainfall, *AM*.
Construct a matrix size of 12x1 for current year, *CY*.
Construct a matrix size of 24x1 for previous year, *PY*.
Step 2: Identify 13 sliding window from matrix *PY*.
Step 3: Compute Euclidean distance, *Edi* of sliding window. Find mean of *Edi*.

$$Ed_{ij} = \sqrt{(x_i - y_i)^2} \quad (1)$$

$$\text{Mean of } Edi = \sum \frac{Ed_i}{n} \quad (2)$$

Step 4: Select the minimum mean value of Euclidean distance, *Edi* from sliding window.
Step 5: Compute variation for minimum Euclidean distance, *Edi* and rename as mean variation previous (*VP*).

$$\text{Variation; } V = AM_t - AM_{t-1} \quad (3)$$

Step 6: Compute variation of current year, *CY* and rename as mean variation current (*VC*).
Step 7: Determine the predicted variation, "V".

$$\text{Predicted variation, "V"} = \frac{\text{mean}\,VC + \text{mean}\,VP}{2} \quad (4)$$

Step 8: Calculate average monthly of rainfall, *AM*, 1990 till 2013.

$$AM = \frac{\sum \text{Monthly rainfall from 1990 to 2013}}{n} \quad (5)$$

Step 9: Forecast rainfall based on:
Predicted variation "V" + *AM* = forecasted rainfall, *FR*. (6)

Figure 1. An Enhancement of Sliding Window Algorithm (SWA).

The main concept of sliding window algorithm that is deployed from previous study of Kapoor and Bedi is adopted in this study with several enhancement and changes in data time where in Kapoor and Bedi study, they are using a span of adjacent 2-week of previous years' data to construct the sliding window. Meanwhile, in this study, 24-month of rainfall data of previous year, *PY*, is taken to construct sliding window, and 12-month of rainfall data of current year (assuming 2013 is the current year) is selected as current year, *CY*. Another enhancement that are present in this study, are additional steps in the algorithm which are step 8 and 9 to produce prediction of rainfall distribution.

The first step of the proposed algorithm is construct a matrix size of 12x1 for current year, *CY* (rainfall distribution data of January to December 2013) and a matrix size of 24x1 for previous year, *PY* (rainfall distribution data of January 2011 to December 2012).

The second step is to identify 13 sliding window of size 12x1 from matrix *PY*. The third step is compute Euclidean distance, *Edi* of sliding window and definds the mean of *Edi*. Table 1 shows a set of sample results of sliding window that compute *Edi* and mean of *Edi*.

The fourth step, select the sliding window that shows the minimum mean value of Euclidean distance, *Edi*. The fifth step is compute the variation for selected minimum Euclidean distance, *Edi* and rename as mean variation previous (*VP*). Meanwhile, the variation of current year *CY* is also computed and rename as mean variation current (*VC*).

Hence, Table 2 shows the selected sliding window, W_i of minimum Euclidean distance, *Edi*, mean variation current, *VC* and mean variation previous, *VP*.

Table 1. Sample Results of Sliding Window that Compute *Edi* and Mean of *Edi*.

W_1		W_2		W_3			W_{13}	
R	Ed_1	R	Ed_2	R	Ed_3		R	Ed_{13}
3.4	2.79	6.0	5.37	6.2	5.58		1.9	1.29
.	.	.	.	.	.		.	.
.	.	.	.	.	.	. . .	.	.
.	.	.	.	.	.		.	.
4.0	3.47	1.9	1.37	2.9	2.31		3.3	2.76
Mean	3.27	Mean	2.78	Mean	2.34		Mean	3.50

Table 2. Mean Variation Previous, VP and Mean Variation Current, VC.

Station	Selected *Wi* for min. mean of *Edi*	Mean variation Previous (*VP*)	Mean variation current *(VC)*
Bukit Temiang	W_3	-0.30	-0.01
Abi Kg. Baru	W_{13}	0.11	0.06
Guar Nangka	W_2	-0.33	0.12

Next, the predicted variation “V” is identified based on Eq. (1). Then, the predicted variation “V” is added to the average monthly of rainfall of January 1990 to December 2013. Table 3 depicted the the result of predicted variation “V”.

Table 3. The Prediction Variation “V”.

Rainfall gauge station	Prediction variation “V”
Bukit Temiang	0.16
Abi Kg. Bahru	0.09
Guar Nangka	0.11

The predicted variation “V” is defined to be added to the average of monthly, AM, which is the final step to obtain forecasted rainfall, FR. Eq. (6) shows the calculation between predicted variation “V” and average of monthly, AM.

RESULT AND DISCUSSION

An enhancement of sliding window algorithm SWA is introduced in this study to forecast rainfall distribution in three selected rainfall gauge station in Perlis, Malaysia. Table 4 depicted the result of forecasted rainfall from the three rainfall stations. In this session, the validation of algorithm is discussed to observe the accuracy of forecasted rainfall between the enhancement SWA of this study and the previous study SWA by Kapoor and Bedi (2013).

Table 4 depicted the results for the forecasted rainfall. The algorithm is validated using mean squared error, MSE calculated using Eq. (7). According to Lazim (2007), the smallest value of MSE shows the best algorithm because less error between forecasted rainfall and actual data. For all the three stations, the proposed algorithm produced models with smaller MSE values compared to the SWA by Kapoor and Bedi (2013).

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{(\text{Actual rainfall}_i - \text{Forecasted rainfall}_i)^2}{\text{Number of sample}}\right| \tag{7}$$

This study also calculates relative GRMSE (Eq. (7)) to compare the proposed model and SWA by Kapoor & Bedi (2013). If relative GRMSE value is smaller than 1.0, the best model is proposed model.

$$\text{Relative GRMSE} = \left(\frac{\left(\sum \text{Actual rainfall of } M_1 - \text{Forecasted rainfall of } M_2\right)^2}{\left(\sum \text{Actual rainfall of } M_2 - \text{Forecasted rainfall of } M_1\right)^2} \right)^{\frac{1}{2n}} \quad (8)$$

Table 4. Forecasted Rainfall, *FR*.

Month / 2014	Bukit Temiang			Abi Kg. Bahru			Guar Nangka		
	Actual (2014)	M_1	M_2	Actual (2014)	M_1	M_2	Actual (2014)	M_1	M_2
Jan	0.2	1.42	0.8	0.2	1.11	0.7	0.2	1.46	1.0
Feb	0.0	2.8	8.8	0.0	2.06	7.9	0.0	2.66	3.0
March	0.4	4.94	0.7	0.3	3.75	0.4	0.3	5.11	1.4
April	6.5	4.69	3.6	6.3	5.27	7.5	6.3	5.45	4.1
May	7.4	4.92	4.5	6.6	5.23	3.9	6.6	4.57	2.8
Jun	2.9	4.39	8.0	3.5	4.46	7.5	3.5	4.09	7.2
July	3.1	6.28	5.3	3.1	6.2	5.5	3.1	5.25	3.8
August	7.6	5.54	3.8	5.6	6.35	5.3	5.6	5.92	3.8
Sept	7.3	6.63	8.8	10.3	8.79	9.0	10.3	7.48	7.9
Oct	6.3	8.71	7.6	12.5	9.01	6.5	12.5	8.03	6.6
Nov	8.2	5.67	2.1	9.1	6.62	3.0	9.1	5.62	4.7
Dis	8.5	3.32	0.7	11.9	2.96	1.4	11.9	3.34	2.3
	MSE	7.96	20.17	MSE	12.01	23.21	MSE	7.29	8.41
	Relative GRMSE	0.96		Relative GRMSE	0.97		Relative GRMSE	0.99	

M_1 is the proposed algorithm, and M_2 is SWA by Kapoor and Bedi (2013)

The forecasting results of proposed model, M_1 and SWA, M_2 by Kapoor and Bedi (2013) in Bukit Temiang, Abi Kg. Bahru and Guar Nangka are compared to the actual rainfall data in 2014. The MSE and relative GRMSE between the algorithms are calculated. As can be seen, the proposed model of this study, an enhancement of sliding window algorithm shows the smallest MSE for all rainfall station as compared to SWA of Kapoor and Bedi (2013). Meanwhile, the relative GRMSE of all rainfall station of proposed model shows that the value are smaller as compared to SWA of Kapoor and Bedi (2013). Hence, the result explained that proposed model, the enhancement of sliding window algorithm, is better than SWA by Kapoor and Bedi.

CONCLUSION

This study proposes sliding window algorithm that is improved from SWA of Kapoor and Bedi in 2013. Three type of rainfall data had been tested to the proposed model to observe the best result of rainfall forecasting. The proposed model consists of nine steps which begins with construct matrix Current Year, *CY* and Previous Year, *PY*, from average monthly rainfall, *AM*, and end up with define forecasted rainfall value by adding the predicted variation “V” and average monthly rainfall, *AM*. Hence, the result shows that the enhance sliding window algorithm of this study is highly accurate compared to previous sliding window algorithm, SWA of Kapoor and Bedi (2013) based on the result of model validation using MSE and relative GRMSE. For future work, the accuracy of SWA might be can improve if it incorporated with Artificial

Intelligent (AI) model such as fuzzy time series, or artificial neural network (ANN) technique (Kapoor & Bedi, 2013; Rao et al., 2015).

ACKNOWLEDGEMENTS

The authors wish to thank the Ministry of Education, Malaysia for funding this study under the Long Term Research Grant Scheme (LRGS/b-u/2012/UUM/Teknologi Komunikasi dan Infomasi). The authors also like to express their deepest gratitude to the Department of Irrigation and Drainage, Malaysia that made the data available.

REFERENCES

Arasu, A., & Manku, G. S. (2004). Approximate counts and quantiles over sliding windows. Paper presented at the *Proceedings of the twenty-third ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, Paris, France.

Asklany, S. A., Elhelow, K., Youssef, I. K., & Abd El-wahab, M. (2011). Rainfall events prediction using rule-based fuzzy inference system. *Atmospheric Research*, *101*(1–2), 228–236.

BenYahmed, Y., Abu Bakar, A., Hamdan, A. R., Ahmed, A., & Syed Abdullah, S. M. (2015). Adaptive sliding window algorithm for weather data segmentation. *Journal of Theoretical and Applied Information Technology, 80*(2), 322-333.

Braverman, V., Ostrovsky, R., & Zaniolo, C. (2009). Optimal sampling from sliding windows. Paper presented at the *Proceedings of the twenty-eighth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*.

d'Arcy, J. A., Collins, D. J., Rowland, I. J., Padhani, A. R., & Leach, M. O. (2002). Applications of Sliding Window Reconstruction with Cartesian Sampling for Dynamic Contrast Enhanced MRI. *NMR in Biomedicine, 15*(2), 174-183.

Dani, S., & Sharma, S. (2013). Forecasting rainfall of a region by using fuzzy time series. *Asian Journal of Mathematics and Applications, 2013*.

Datar, M., Gionis, A., Indyk, P., & Motwani, R. (2002). Maintaining stream statistics over sliding windows. *SIAM Journal on Computing, 31*(6), 1794.

French, M. N., Krajewski, W. F., & Cuykendall, R. R. (1992). Rainfall forecasting in space and time using a neural network. *Journal of Hydrology*, *137*(1–4), 1-31.

Hung, N. Q., Babel, M. S., Weesakul, S., & Tripathi, N. K. (2008). An artificial neural network model for rainfall forecasting in Bangkok, Thailand. *Hydrology and Earth System Sciences Discussions*, *5*(1), 183–218.

Ishak, W. W., Ku-Mahamud, K.-R., & Norwawi, N. M. (2011). Mining temporal reservoir data using sliding window technique. *CiiT International Journal of Data Mining Knowledge Engineering, 3*(8), 473-478.

Kapoor, P., & Bedi, S. S. (2013). Weather forecasting using sliding window algorithm. *ISRN Signal Processing, 2013*, 5.

Lazim, M. A. (2007). *Introductory business forecasting: A practical approach*: Pusat Penerbitan Universiti, Universiti Teknologi MARA.

Lee, S., Cho, S., & Wong, P. M. (1998). Rainfall prediction using artificial neural networks. *Journal of Geographic Information and Decision Analysis, 2*(2), 233–242.

Luk, K. C., Ball, J. E., & Sharma, A. (2001). An application of artificial neural networks for rainfall forecasting. *Mathematical and Computer Modelling, 33*(6–7), 683-693.

Rao, D. V. N. K., Anusha, M., Babu, P. N., Sri, M. D., Rajesh, N., & Kumar, K. S. (2015). Prognostication of Climate Using Sliding Window Algorithm. *International Journal of u- and e-Service, Science and Technology*, *8*(4), 225-232.

Song, Q., & Chissom, B. S. (1993). Fuzzy time series and its models. *Fuzzy Sets and Systems*, *54*(3), 269-277.

Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, *50*, 159-175.