

The m -delay Autoregressive Model with Application

Manlika Ratchagit¹, Benchawan Wiwatanapataphee¹ and Nikolai Dokuchaev^{1,*}

Abstract: The classical autoregressive (AR) model has been widely applied to predict future data using m past observations over five decades. As the classical AR model required m unknown parameters, this paper implements the AR model by reducing m parameters to two parameters to obtain a new model with an optimal delay called as the m -delay AR model. We derive the m -delay AR formula for approximating two unknown parameters based on the least squares method and develop an algorithm to determine optimal delay based on a brute-force technique. The performance of the m -delay AR model was tested by comparing with the classical AR model. The results, obtained from Monte Carlo simulation using the monthly mean minimum temperature in Perth Western Australia from the Bureau of Meteorology, are no significant difference compared to those obtained from the classical AR model. This confirms that the m -delay AR model is an effective model for time series analysis.

Keywords: Delay autoregressive model, least squares method, brute-force technique.

1 Introduction

Time series analysis has been widely applied in many branches such as agriculture, environment and so on. For agriculture application, the rainfall forecast is one of a popular topic in time series analysis for the irrigation management propose. Based on using historical rainfall data, a time series model has been used to predict the amount of rainfall for agriculture target planning. This would probably help the farmer to deal with their resources and gain more agricultural products [Moeletsi, Mellaart, Mpandeli et al. (2012)]. For Environment, the weather forecast is important as the temperature is the key factor for government planning and management in many aspects such as tourism, vegetation and public health [Ustaoglu, Cigizoglu and Karaca (2008)].

Various approaches based on time series methods have been proposed to fit the data set. These approaches include Box-Jenkins method, neural networks, hybrid technique and fuzzy time series method. Box-Jenkins method is one of the famous methods consisting of

¹School of Electrical Engineering, Computing and Mathematical Sciences, Curtin University, Perth, Australia.

*Corresponding Author: Nikolai Dokuchaev. Email: N.Dokuchaev@curtin.edu.au.

Received: 20 October 2019; Accepted: 17 December 2019.

autoregressive (AR) model, moving average (MA) model, autoregressive moving average (ARMA) model and autoregressive integrated moving average (ARIMA) model [Box, Jenkins and Reinsel (2008)]. This method has been applied to a wide range of applications including economic especially a stock process sequence [Ip, Zhang, Sowinski et al. (2015)], engineering [Amo-Salas, López-Fidalgo and Pedregal (2015); Bacher, Madsen and Nielsom (2009); de Waele and Broersen (2003)], industry [Acedański (2013)] and scientific [Hurvich and Tsai (1989) and Sawa (1978)]. The artificial neural networks (ANN) approach is the main tool for nonlinear data as it is a brain-inspired system [Tealab (2018)]. The hybrid technique has been established since the individual model may not be adequate to analyze all the actual observations [Taskaya-Temizel and Casey (2005)]. This method combines at least two individual forecast methods to improve forecasting accuracy [Pan, Zhang and Xia (2009)]. ANN approach integrated with either AR or ARIMA method are called AR-ANN model [Qi and Zhang (2001)] or ARIMA-ANN model [Zhang (2003)]. Fuzzy time series is a forecasting method using fuzzy principle as the basis. It is suitable for numerical and linguistic data [Saxena and Easo (2012)]. Apart from the development of forecasting methods, various estimating model-parameter approaches used in the time series model have been proposed such as the maximum likelihood method, the method of moment and the Yule-Walker procedure. For the maximum likelihood method, studies aim to find the parameter values giving the distribution that maximize the probability of observing the data. The method of moment is a simple procedure. By equating the sample moments to the corresponding population moments, the estimating model parameter is solved [Cryer and Chan (2008)]. The Yule-Walker method (or autocorrelation method) fits an AR model to the window input data by maximizing the error in the least squares sense [Broersen (2008)].

Time series data are commonly classified into two groups; stationary and non-stationary data. For stationary data, its mean and variance are consistent. Its sequence does not reveal an upward (or downward) trend and seasonal pattern. In contrast to the stationary data, mean and variance of non-stationary data vary over time. Two common ways to check the stationary of the data are time series plot and roots of the characteristic equation [Maddala and Wu (1999)]. As possible roots of the characteristic equation may be real and/or complex numbers, the data is stable if all real roots are greater than one or the modulus of each complex root is greater than one. The data is considered to be non-stationary data if at least one root falls between minus and plus one or fall inside the unit circle [Box, Jenkins and Reinsel (2008)]. As the AR model required stationary data to predict future data, single/double differencing method and mathematical transformation are two common procedures to convert non-stationary data to stationary data [Hassler (1994)]. The differencing technique is used to remove the trend component of the data while mathematical transformations including the square/cube root transformation and log transformation are applied when the variance of data is unstable [Pourahmadi (2001)].

As the classical AR model uses more parameters in the formula to predict future observation, statistical techniques have been applied to obtain these parameters. In this study, we propose the m -delay AR model in which only two parameters are determined.

The m -delay AR formula based on the least squares method is derived and an optimal delay algorithm based on a brute-force technique is developed. The remainder of this paper is structured as follows. Section 2 concerns the derivation of the m -delay AR model to obtain two parameters of the m -delay AR model. The performance of parameter estimation is presented in Section 3 and the empirical study is reported in Section 4. Discussion and conclusion are given in Section 5.

2 Derivation of the m -delay autoregressive (MAR) model

Generally, an explicit formula of the present value x_t is determined by the standard AR model [Box, Jenkins and Reinsel (2008)] , i.e.,

$$x_t = \epsilon_t + \sum_{i=1}^m \phi_i x_{t-i}, \quad (1)$$

where x_t represents the present value at instant time t , $\{x_{t-1}, x_{t-2}, \dots, x_{t-m}\}$ is the list of the past observations, $\phi_i (i = 1, \dots, m)$ are i^{th} coefficients of AR model and ϵ_t is a gaussian white noise process which is assumed to be the normal distribution, $N(0, \sigma^2)$.

In this study, we propose a m -delay AR model to approximate the present value x_t by using only the first term at $t - 1$ and the last term at $t - m$,

$$x_t = \epsilon_t + \phi_1 x_{t-1} + \phi_m x_{t-m}. \quad (2)$$

The first and the last coefficients of the AR model are estimated by Eq. (3) based on the least squares method [Khalil and Moraes (1995)]. The principle concept of the least squares procedure is to minimize the sum of square error functions

$$S_c(\hat{\phi}_1, \hat{\phi}_m) = \sum_{t=m+1}^n \left[x_t - \hat{\phi}_1 x_{t-1} - \hat{\phi}_m x_{t-m} \right]^2. \quad (3)$$

To find the optimal values of $\hat{\phi}_1$ and $\hat{\phi}_m$ in Eq. (3). We differentiate Eq. (3) with respect to $\hat{\phi}_1$ and $\hat{\phi}_m$ and set them to zero. This step is shown in Eq. (4).

$$\begin{cases} \frac{\partial S_c}{\partial \hat{\phi}_1} = -2 \sum_{t=m+1}^n \left[x_t - \hat{\phi}_1 x_{t-1} - \hat{\phi}_m x_{t-m} \right] x_{t-1} = 0, \\ \frac{\partial S_c}{\partial \hat{\phi}_m} = -2 \sum_{t=m+1}^n \left[x_t - \hat{\phi}_1 x_{t-1} - \hat{\phi}_m x_{t-m} \right] x_{t-m} = 0. \end{cases} \quad (4)$$

We finally obtain the m -delay formula for approximating $\hat{\phi}_1$ and $\hat{\phi}_m$, i.e.,

$$\left\{ \begin{array}{l} \hat{\phi}_1 = \frac{\sum_{t=m+1}^n x_t x_{t-1} \sum_{t=m+1}^n x_{t-m}^2 - \sum_{t=m+1}^n x_{t-1} x_{t-m} \sum_{t=m+1}^n x_t x_{t-m}}{\sum_{t=m+1}^n x_{t-1}^2 \sum_{t=m+1}^n x_{t-m}^2 - \left[\sum_{t=m+1}^n x_{t-1} x_{t-m} \right]^2}, \\ \hat{\phi}_m = \frac{\sum_{t=m+1}^n x_{t-1}^2 \sum_{t=m+1}^n x_t x_{t-m} - \sum_{t=m+1}^n x_t x_{t-1} \sum_{t=m+1}^n x_{t-1} x_{t-m}}{\sum_{t=m+1}^n x_{t-1}^2 \sum_{t=m+1}^n x_{t-m}^2 - \left[\sum_{t=m+1}^n x_{t-1} x_{t-m} \right]^2}. \end{array} \right. \quad (5)$$

We call $\hat{\phi}_1$ and $\hat{\phi}_m$ as the new approximation of the m -delay AR coefficients.

As the standard AR model is a stationary time series process, we now determine the stationarity condition of the m -delay AR model. In this study, we investigate the stationarity condition of our proposed model (2) by computing the roots of AR characteristic equation.

To discuss the stationarity, we define the AR characteristic polynomial as

$$\phi(t) = 1 - \phi_1 t - \phi_m t^m, \quad (6)$$

where the characteristic equation is

$$1 - \phi_1 t - \phi_m t^m = 0. \quad (7)$$

To solve Eq. (7), a sufficient stationarity condition of our model is

$$|\phi_1| + |\phi_m| < 1, \quad (8)$$

where the delay (m) is between 3 and 120.

We illustrate the sufficient stationarity condition in Fig. 1.

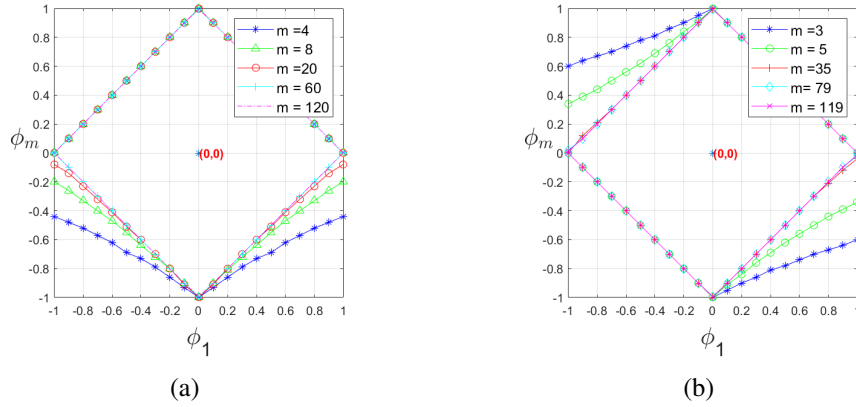


Figure 1: Linear relationship between ϕ_1 and ϕ_m for different delays (m); (a) m is even number, (b) m is odd number

To illustrate the performance of the MAR model and the AR model, two different errors based on the root mean square error are computed by

$$RMSE = \sqrt{\frac{\sum_{l=1}^L \left[\left(\phi_1 - \hat{\phi}_1^{(l)} \right)^2 + \left(\phi_m - \hat{\phi}_m^{(l)} \right)^2 \right]}{L}}, \quad (9)$$

$$E_m = \sqrt{\frac{\sum_{t=m+1}^n (\hat{x}_t - x_t)^2}{n - m}}, \quad (10)$$

where ϕ_1 and ϕ_m are parameters of the MAR model, $\hat{\phi}_1$ and $\hat{\phi}_m$ are approximated parameters, \hat{x}_t are the forecast observations, n is the sample size, m is the delay and L is the number of simulations.

3 Parameter estimation

This section concerns the effectiveness of the m -delay formula (see Eq. (5)) via Monte Carlo simulation, and the optimal delay using brute-force technique.

3.1 Effectiveness of the m -delay formula ($\hat{\phi}_1$ and $\hat{\phi}_m$)

To examine the coefficients ϕ_1 and ϕ_m obtained from the m -delay formulation (see Eq. (5)), we set the initial value of m data to zero ($m \ll n$). The observations x_t determined by Eq. (2) are generated using Monte Carlo technique. Fig. 2 presents the examination process using computation scheme as shown in Tab. 1. We perform the same process for the different sample sizes with different delays.

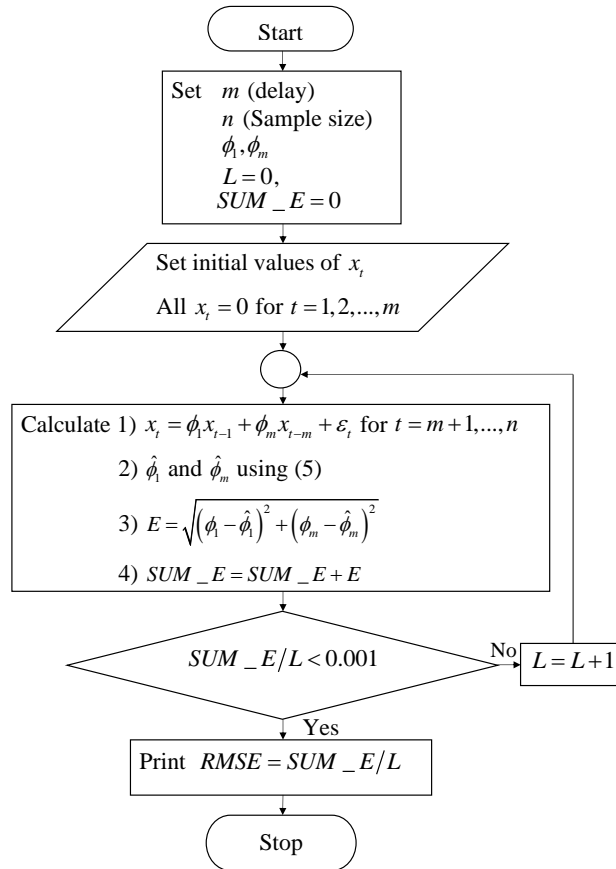


Figure 2: The flowchart for examining the effectiveness of the m -delay formulation

Table 1: Computation scheme used in generating Fig. 3

Sample size	50	100	300	500	1000	2000	5000	10000
	5	5	5	5	5	5	5	5
Delay	20	20	20	20	20	20	20	20
			120	120	120	120	120	120

The simulation results indicate that the average $\hat{\phi}_1$ and $\hat{\phi}_m$ approach the actual ϕ_1 and ϕ_m .

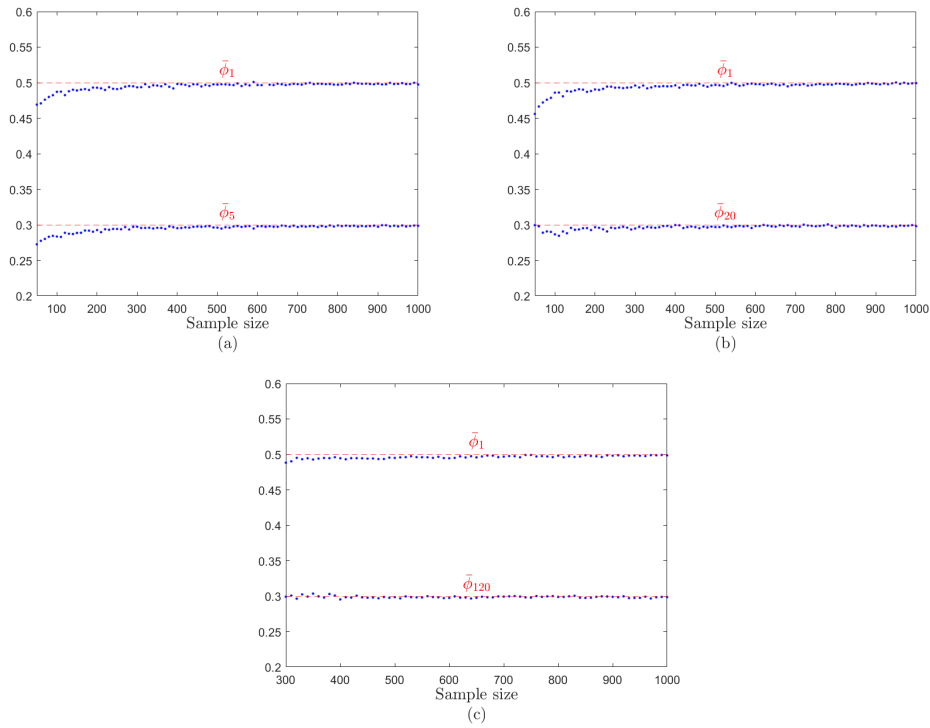


Figure 3: Scatter plots of the average m -delay coefficients $\hat{\phi}_1$ and $\hat{\phi}_m$: (a) $m = 5$; (b) $m = 20$; (c) $m = 120$

Fig. 3 illustrates the relationship between average two m -delay coefficients, $\hat{\phi}_1$ and $\hat{\phi}_m$, and sample size for three different delays including $m = 5, 20, 120$. For the case of delay 5 and 20, the sample size starts from 50 to 1000. Figs. 3(a)-(b) indicate that sample size greater than 300 gives the reasonable results of approximated of $\bar{\phi}_1$ and $\bar{\phi}_m$. In case of the delay 120, the process starts with the sample size of 300. The estimated $\bar{\phi}_1$ and $\bar{\phi}_m$ approach the actual ones when the sample size about 500. It is noted that Eq. (5) is an effective formula for approximating the m -delay parameter ϕ_1 and ϕ_m .

3.2 Determination of optimal delay (m)

To find the optimal delay, we develop an algorithm as shown below based on brute-force technique by fixing the m_0 -delay AR parameter, ϕ_1 and ϕ_{m_0} for the initial delay m_0 and select the sample size n and the prescribed delay $\frac{n}{2} - 1$.

Algorithm 1

- Step 1 Prescribe sample size (n), initial delay (m_0), the m -delay AR parameter (ϕ_1, ϕ_{m_0}), count=0, set initial iteration ($l = 0$), the maximum number of iterations ($L = 10000$) and set $x_t = 0$ for $t=1, 2, \dots, m_0$
- Step 2 Generate $\{x_t\}_{t=m_0+1}^n$ from Eq. (2)
- Step 3 Set $m=3$
 $l = l + 1$
 Calculate $\hat{\phi}_1, \hat{\phi}_m$ from Eq. (5) and \hat{x}_t which is function of $\hat{\phi}_1, \hat{\phi}_m$
 Compute $E_3 = \sqrt{\frac{\sum_{t=m+1}^n (\hat{x}_t - x_t)^2}{n-m}}$ where $\hat{x}_t = \hat{\phi}_1 x_{t-1} + \hat{\phi}_m x_{t-m}$
 Set $\min E = E_3$
- Step 4 $m=m+1$
 Calculate $\hat{\phi}_1, \hat{\phi}_m$ from Eq. (5)
 Compute $E_m = \sqrt{\frac{\sum_{t=m+1}^n (\hat{x}_t - x_t)^2}{n-m}}$
 If $E_m < \min E$ then $\min E = E_m$ and $\min D = m$
- Step 5 If $m < \frac{n}{2} - 1$ (maximum delay) then go to Step 4
- Step 6 If $\min D = m_0$
 count=count+1 and $m_{op} = \min D$
 If $l \leq L$ then go to Step 2.
- Step 7 Compute the accuracy using Eq. (11).
-

This above algorithm begins with prescribing input parameters including the sample size (n), the initial delay (m_0) and the actual m -delay AR parameters (ϕ_1, ϕ_{m_0}) and setting the values of all m_0 observations to zero. We then generate the data set (x_t) by Eq. (2). The process starts with delay 3 at the first iteration. Two unknown parameters ($\hat{\phi}_1, \hat{\phi}_m$) are calculated by Eq. (5) and the minimum error ($\min E$) is obtained. We repeat the process by increasing the delay by one. In each iteration, the minimum error ($\min E$) and minimum delay ($\min D$) are obtained. The process stops when the delay equals the prescribed delay. The process will succeed when the minimum delay equals to the initial delay, i.e., the

optimal delay (m_{op}) is obtained for the sample size n . We then go back to Step 2 until reaching to the maximum number of iteration. Finally, the accuracy of delay estimation based on the brute-force technique is computed by Eq. (11).

To illustrate the performance of a brute-force technique. We compute the accuracy on the simulated data set for each different sample size. The accuracy of the delay estimation is computed by

$$Accuracy(\%) = \frac{\text{Number of success outcomes}}{\text{Number of possible outcomes}} \times 100. \tag{11}$$

Using a brute-force algorithm, our experiments show that sample size has a significant effect on the accuracy of the m -delay AR model as shown in Fig. 4.

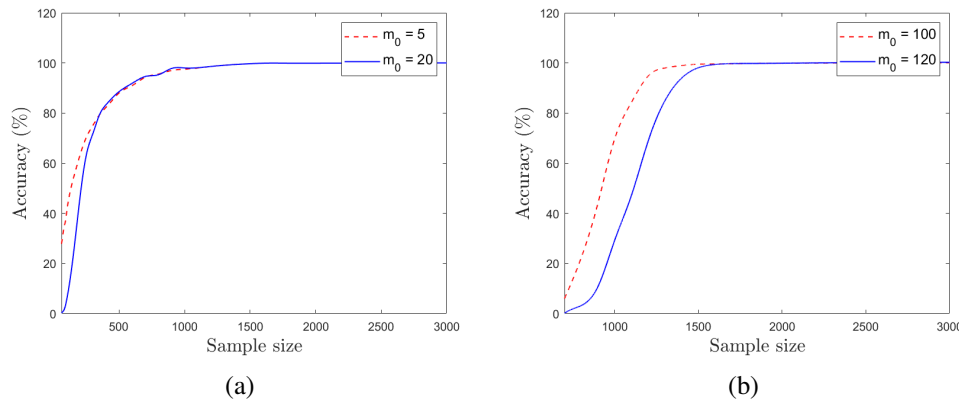


Figure 4: The relationship between the percentage of accuracy and sample sizes of the m -delay approximation

To investigate the effect of sample size on the accuracy of the m -delay approximation as revealed in Fig. 4(a), we choose 24 sample sizes between 60 and 3000 giving the percentage of accuracy from 27.85% to 99.90% and 0.22% to 99.90% for delay 5 and 20, respectively. The percentage of accuracy with delay 100 and 120 is shown in Fig. 4(b). We found that using sample size from 700 to 3000 gives the percentage of accuracy 6% to 99.90% and 0.10% to 99.90%, respectively. For delay 120, we choose various sample sizes to evaluate its effect on estimation accuracy. We found that using sample size from 700 to 3000 gives the percentage of accuracy from 0.10% to 99.90%.

4 Empirical study

In this section, we first test the performance of the MAR model comparing with the AR model as shown in Algorithm 2. We then apply both models to approximate the monthly mean minimum temperature in Perth, Western Australia.

4.1 Monte Carlo simulation

Algorithm 2 below compares the performance of the MAR model (based on only 2 parameters namely ϕ_1 and ϕ_m) with the AR model (via m coefficient parameter, $\phi_1, \phi_2, \dots, \phi_m$).

Algorithm 2	
Step 1	Prescribe sample size (n), parameter of AR model (ϕ_1, ϕ_{m_0}), the initial delay (m_0) and $maxL$, set $L = 0$, $sumE_{AR} = 0$, $sumE_{MAR} = 0$ and $x_t = 0$ for $t=1, 2, \dots, m_0$
Step 2	Generate $\{x_t\}_{t=m_0+1}^n$ from Eq. (2)
Step 3	Find root mean square error from AR model ($RMSE_{AR}$) Estimate the coefficient of the AR model ($\hat{\phi}_1, \hat{\phi}_2, \dots, \hat{\phi}_{m_0}$) Compute $RMSE_{AR}$ $sumE_{AR} = sumE_{AR} + RMSE_{AR}$
Step 4	Find root mean square error from the MAR model ($RMSE_{MAR}$) Set $m=3$ Calculate $\hat{\phi}_1, \hat{\phi}_m$ from Eq. (5) Compute $E_m = \sqrt{\frac{\sum_{t=m+1}^n (\hat{x}_t - x_t)^2}{n-m}}$ where $\hat{x}_t = \hat{\phi}_1 x_{t-1} + \hat{\phi}_m x_{t-m}$ If $E_m < MinE$, Then $minE = E_m$ and $minD=m$ $m=m+1$
Step 5	If $m < \frac{n}{2} - 1$. Then go to Step 4
Step 6	Optimal $m_{op} = minD$ and Calculate $RMSE_{MAR}$ using the optimal m (m_{op})
Step 7	$sumE_{MAR} = sumE_{AR} + RMSE_{MAR}$ and $L = L + 1$
Step 8	If $L \leq maxL$. Then go to Step 2
Step 9	Compute average root mean square error of both models $\overline{RMSE}_{AR} = sumE_{AR}/L$ and $\overline{RMSE}_{MAR} = sumE_{MAR}/L$

This algorithm starts with determining input parameters including sample size (n), actual parameters of AR with delay process (ϕ_1, ϕ_{m_0}) and setting the values of all m_0 observations to zero. We then generate the data set (x_t) by Eq. (2). To find the root mean square error from the AR model ($RMSE_{AR}$), we first estimate the model parameter and we then compute the root mean square error by Eq. (10). In the MAR model, the computation process of the optimal delay starts at delay 3. The unknown parameters ($\hat{\phi}_1, \hat{\phi}_m$) are calculated by Eq. (5), and the minimum error ($minE$) and the minimum delay ($minD$) are obtained. The process is repeated until the delay reaching to the prescribed delay $\frac{n}{2} - 1$. This step returns the optimum delay with the smallest value of the error for

each iteration ($m_{op}=\min D$) and we also calculate $RMSE_{MAR}$ using the optimal delay (m_{op}). This process goes back to Step 2 until the total iterations equal the maximum number of iterations. Finally, we compute the average root mean square error of both models.

Tab. 2 compares average root mean square error obtained from the AR model (\overline{RMSE}_{AR}) and the MAR model

Table 2: Average root mean square of the *AR* model and the *MAR* model for different sample sizes

size	AR	MAR	size	AR	MAR	size	AR	MAR
30	0.09364	0.08147	650	0.09978	0.09916	4000	0.09999	0.09989
50	0.09703	0.08897	700	0.09959	0.09929	4500	0.09996	0.09978
80	0.09834	0.09320	750	0.09983	0.09930	5000	0.09997	0.09989
100	0.09871	0.09468	800	0.09989	0.09939	5500	0.09996	0.99889
150	0.09908	0.09635	850	0.09981	0.09934	6000	0.09998	0.09991
200	0.09939	0.09737	900	0.09993	0.09948	6500	0.09999	0.09993
250	0.09953	0.09793	950	0.09987	0.09945	7000	0.09999	0.09993
300	0.09955	0.09820	1000	0.09988	0.09948	7500	0.09997	0.09992
350	0.09962	0.09847	1500	0.09990	0.09963	8000	0.09996	0.09991
400	0.09978	0.09878	2000	0.09992	0.09972	8500	0.10000	0.09995
450	0.09969	0.09880	2500	0.09994	0.09978	9000	0.09996	0.09992
500	0.09982	0.09901	3000	0.09995	0.09982	9500	0.09998	0.09994
550	0.09977	0.09905	3500	0.09995	0.09983	10000	0.09999	0.09995
600	0.09975	0.09908						

The experimental results shown in Tab. 2 are generated using $\phi_1 = 0.5, \phi_m = 0.3, m_0 = 10$ and $L = 10000$. We now use the independent sample *t*-test to confirm that the results obtained from both models are not significantly different. The null hypothesis and alternative hypothesis are $\overline{RMSE}_{AR} = \overline{RMSE}_{MAR}$ and $\overline{RMSE}_{AR} \neq \overline{RMSE}_{MAR}$, respectively.

The result of hypothesis testing is demonstrated in Tab. 3.

Table 3: Hypothesis test using *t*-test

Variable	N	Mean	SD	df	<i>t</i> -test	<i>p</i> -value
AR	40	0.0995	0.0011	78	-0.945	0.348
MAR	40	0.1208	0.1425			

It indicates that there is not enough evidence to reject the null hypothesis at significance

level 0.05 as $p\text{-value} > 0.05$. This means that our proposed MAR model is an efficient technique for the time series prediction comparing to the AR model.

4.2 Validation of the m -delay AR model

This section concerns the prediction of the monthly mean minimum temperature in Perth, Western Australia, using the data obtained from the Bureau of Meteorology, Australia between January 1994 and June 2019 containing 306 observations. The data are divided into two sets; training set and testing set. The training set, from January 1994 to December 2017, has 288 observations. The rest, January 2018 to June 2019, is the testing set consisting of 18 observations. As the training data has a seasonal pattern as shown in Fig. 5(a), we firstly deseasonalize the original data by taking the seasonal difference ($x_t - x_{t-12}$). The result after transformation is shown in Fig. 5(b).

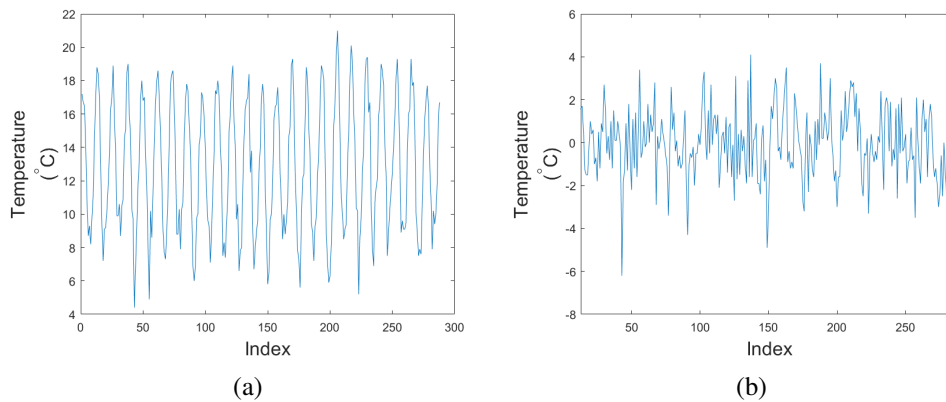


Figure 5: Monthly mean minimum temperature in Perth from January 1994 to December 2017: (a) Original data; (b) Transformed data

As the AR model requires stationary data, we then need to check the features of the transformed data: normal and independent [Razali and Wah (2011)]. To check whether normal distribution of data, we apply the Shapiro-Wilk test. The result is revealed in Tab. 4.

Table 4: Normality test for transformed data

Variable	Shapiro-Wilk Test		
	Statistics	df	$p\text{-value}$
Transformed data	0.993	276	0.234

Tab. 4 indicates that transformed data are normally distributed as $p\text{-value}=0.234$. We now check the independent of the transformed data by using one sample t -test. The result is shown in Tab. 5.

Table 5: Independent testing for one-sample group using *t*-test

Variable	n	Mean	SD	df	t	<i>p</i> -value
Transformed data	276	0.011	1.610	275	0.112	0.911

In Tab. 5, the *p*-value of 0.911 illustrates that the data are independent. We now can use these transformed data that are normally distributed and independent to find the suitable AR model.

4.2.1 Classical AR(9) model

It appears that the transformed data are suitable for this study. From our experiment, it is found that the classical AR(9) model is applicable to this data set when $\hat{\phi}_1 = 0.1871, \hat{\phi}_2 = 0.1916, \hat{\phi}_3 = -0.1097, \hat{\phi}_4 = 0.0628, \hat{\phi}_5 = -0.0229, \hat{\phi}_6 = -0.0789, \hat{\phi}_7 = 0.0938, \hat{\phi}_8 = -0.0336$ and $\hat{\phi}_9 = -0.00029934$.

Table 6: Normality test of the residuals from the classical AR order 9

Variable	Shapiro-Wilk Test		
	Statistics	df	<i>p</i> -value
Residuals	0.993	276	0.221

Table 7: One-sample *t*-test of the residuals from the classical AR order 9

Variable	n	Mean	SD	df	t	<i>p</i> -value
Residuals	276	0.009	1.534	275	0.094	0.925

In Tabs. 6 and 7, the AR(9) residuals are tested. The statistical testing results of the residuals from the classical AR(9) based on the Shapiro-Wilk test (*p*-value=0.221) and one sample *t*-test (*p*-value=0.925) demonstrate that the residuals from the classical AR (9) are normally distributed and independent.

4.2.2 The *m*-delay AR model

From Eq. (5) and using the transformed data, we obtain the *m*-delay coefficient $\hat{\phi}_1 = 0.194561$ and $\hat{\phi}_9 = 0.027676$ which satisfy the characteristic Eq. (7) with inequality condition $|\phi_1| + |\phi_m| < 1$.

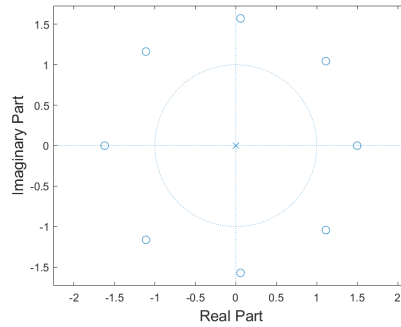


Figure 6: All roots of the characteristic Eq. (7) in the MAR model

Fig. 6 demonstrates that all roots of Eq. (7) lie outside the unit circle. This confirms the stationarity condition of the proposed the MAR model. The analyses of residuals using Shapiro-Wilk test and one sample t -test are shown in Tabs. 8 and 9.

Table 8: Normality test of the residuals from the 9-delay AR model

Variable	Shapiro-Wilk Test		
	Statistics	df	p -value
Residuals	0.995	276	0.485

Table 9: One-sample t -test of the residuals from the 9-delay AR model

Variable	n	Mean	SD	df	t	p -value
Residuals	276	0.010	1.579	275	0.106	0.916

The results from Tabs. 8 and 9 illustrate that the residuals from the MAR model are normally distributed and independent with p -values of 0.485 and 0.916, respectively.

4.2.3 Predictive modeling

Using the test set of data with 18 observations as shown in column 2 of Tab. 10 from January 2018 to June 2019, we apply the AR(9) model and the proposed MAR model to predict the monthly mean minimum temperature as shown in columns 3 and 4 of Tab. 10. Fig. 7 and Tab. 10 compare the forecasting monthly mean minimum temperature obtained from the MAR model and the AR model with observed data.

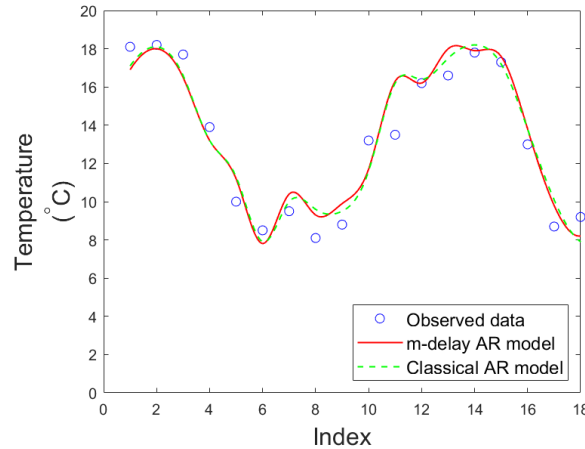


Figure 7: Scatter plot of observed data, and a dashed line (the MAR model) and a solid line (the AR model)

From the results as shown in Tabs. 2 and 3, it confirms that there is no significant difference in results obtained from the MAR model and the AR(9) model. Consequently, the m -delay AR model is an effective model for time series prediction.

5 Conclusion and discussion

In the classical AR model with a large delay, a number of coefficient parameters are needed to determine. In this paper, we consider a special case of the classical AR model, particularly when the delay is large. The optimal m -delay AR model using two parameters is proposed. This would improve the results and reduce the computation time. We outline the future development and the possible application below.

- For parameter estimation, we only modify the least square method to approximate the coefficients of the m -delay AR model. There are several techniques to approximate these AR parameters, namely the maximum likelihood method, the Yule-Walker estimation, and the method of moment.
- Based on the simulation results in Section 4, the root mean square error is utilized. For future development, adding more information criteria can be useful to choose an optimal delay such as the Akaike Information Criteria (AIC), the Bayesian Information Criteria (BIC), the Minimum Description Length (MDL) [Dickie and Nandi (1994)], the Predictive Least Square (PLS), the Predictive Densities Criteria (PDC) and the Sequentially Normalized Maximum Likelihood (SNML) criteria [Giurcăneanu and Razavi (2008)], the Final Prediction Error (FPE) [Akaike (1970)] and the criteria autoregressive transfer function (CAT). [Burshtein and Weinstein (1985)]
- In this work, we assume a random process to be a sequence of independent elements

Table 10: Forecasting monthly mean minimum temperature in Perth obtained from the MAR model and the AR(9) model

Date	Observation data	Predicted values	
		The MAR model	The AR(9) model
January, 2018	18.1	16.9	17.1
February, 2018	18.2	18.0	18.1
March, 2018	17.7	16.5	16.6
April, 2018	13.9	13.2	13.2
May, 2018	10.0	11.2	11.3
June, 2018	8.5	7.8	7.9
July, 2018	9.5	10.4	10.0
August, 2018	8.1	9.3	9.6
September, 2018	8.8	9.9	9.5
October, 2018	13.2	11.7	11.6
November, 2018	13.5	16.3	16.2
December, 2018	16.2	16.2	16.4
January, 2019	16.6	18.0	17.5
February, 2019	17.8	17.9	18.2
March, 2019	17.3	17.6	17.2
April, 2019	13.0	13.8	13.8
May, 2019	8.7	9.8	10.1
June, 2019	9.2	8.2	7.9
Root mean square error (RMSE)		1.1499	1.1282
Mean absolute deviation (MAD)		0.9667	0.9389

with normal distribution via zero mean and constant variance. It is possible for changing a random process to other distribution such as uniform distribution or exponential distribution. Replacing random process would probably affect the parameter estimation for the m -delay AR model.

- Our model assumes that the noise coefficient is finite. The challenging task is unknown variance. Consequently, we may extend model parameters and estimate these unknown parameters including two coefficient parameters, the delay and the noise coefficient, namely ϕ_1 , ϕ_m , m and σ^2 .
- Our model can be applied to historical financial data to analyze the dependence of time-scale volatility measurements.
- The experiment results show that the sample size is one of the key factors. It would be interesting to consider the connection of the error (RMSE) with the sample size.

Acknowledgement: The authors would like to thank the associate editor and the anonymous referees for their valuable comments and suggestions.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- Akaike, H.** (1970): Statistical predictor identification. *Annals of the Institute of Statistical Mathematics*, vol. 22, no. 1, pp. 203-217.
- Bacher, P.; Madsen, H.; Nielson, H.** (2009): Online short-term solar power forecasting. *Solar Energy*, vol. 83, no. 10, pp. 1772-1783.
- Box, G.; Jenkins, G.; Reinsel, G.** (2008): *Time Series Analysis: Forecasting and Control*. Wiley, Hoboken, New Jersey.
- Broersen, P.** (2008): Finite-sample bias propagation in the Yule-Walker method of autoregressive estimation. *17th World Congress: International Federation of Automatic Control (IFAC)*, vol. 41, no. 2, pp. 2744-2749.
- Burshtein, D.; Weinstein, E.** (1985): Some relations between the various criteria for autoregressive model order determination. *IEEE Transaction on Acoustics, Speech and Signal Processing*, vol. 33, no. 4, pp. 1017-1019.
- Cryer, J.; Chan, K.** (2008): *Time Series Analysis with Application in R*. Springer New York, New York.
- de Waele, S.; Broersen, P.** (2003): Order selection for vector autoregressive models. *IEEE Transactions on Signal Processing*, vol. 51, no. 2, pp. 427-433.
- Dickie, J.; Nandi, A.** (1994): A comparative study of AR order selection methods. *Signal Processing*, vol. 40, no. 2-3, pp. 239-255.
- Hassler, U.** (1994): (MIS) specification of long memory in seasonal time series. *Journal of Time Series Analysis*, vol. 15, no. 1, pp. 19-30.
- Hurvich, C.; Tsai, C.** (1989): Regression and time series model selection in small samples. *Biometrika*, vol. 76, no. 2, pp. 297-307.
- Ip, E.; Zhang, Q.; Sowinski, T.; Simpson, S.** (2015): Analysis of feedback mechanisms with unknown delay using sparse multivariate autoregressive method. *Public Library of Science*, vol. 10, no. 8, pp. 1-20.
- Khalil, M.; Moraes, F.** (1995): Linear least squares method for time series analysis with an application to a Methane time series. *Journal of the Air & Waste Management Association*, vol. 45, no. 1, pp. 62-63.
- Maddala, G.; Wu, S.** (1999): A comparative study of unit root test with panel data and a new simple test. *Oxford Bulletin of Economics and Statistics*, vol. 61, pp. 631-652.
- Moeletsi, M.; Mellaart, E.; Mpandeli, N.; Hamandawana, H.** (2012): The use of rainfall forecasts as a decision guide for small-scale farming in limpopo province, South Africa. *Journal of Agricultural Education and Extension*, vol. 19, no. 2, pp. 133-145.

- Pan, F.; Zhang, H.; Xia, M.** (2009): A hybrid time-series forecasting model using extreme learning machines. *2nd International Conference on Intelligent Computation Technology and Automation*, vol. 1, pp. 933-936.
- Pourahmadi, M.** (2001): *Foundations of Time Series Analysis and Prediction Theory*. A John Wiley & Sons, Inc, Canada.
- Qi, M.; Zhang, G.** (2001): An investigation of model selection criteria for neural network time series forecasting. *European Journal of Operational Research*, vol. 132, no. 3, pp. 666-680.
- Razali, N.; Wah, Y.** (2011): Power comparisons of Shapiro-Wilk, Kolmogorov-Sminov, Lilliefors and Anderson-Darling test. *Journal of Statistical Modeling and Analytics*, vol. 2, no. 1, pp. 21-33.
- Sawa, T.** (1978): The exact moments of the least squares estimator for the autoregressive model. *Journal of Econometrics*, vol. 8, no. 2, pp. 159-172.
- Saxena, P.; Easo, S.** (2012): A new method for forecasting enrollments based on fuzzy time series with higher forecast accuracy rate. *International Journal of Computer Technology & Applications*, vol. 3, no. 6, pp. 2033-2037.
- Taskaya-Temizel, T.; Casey, M.** (2005): A comparative study of autoregressive neural network hybrids. *Neural Networks*, vol. 18, no. 5-6, pp. 781-789.
- Tealab, A.** (2018): Time series forecasting using artificial neural networks methodologies: a systematic review. *Future Computing and Informatics Journal*, vol. 3, no. 2, pp. 334-340.
- Ustaoglu, B.; Cigizoglu, H. K.; Karaca, M.** (2008): Forecast of daily mean, maximum and minimum temperature time series by three artificial neural network methods. *Meteorological Applications*, vol. 15, no. 4, pp. 431-445.
- Zhang, G.** (2003): Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, vol. 50, pp. 159-175.