KKU ENGINEERING JOURNAL 2016;43(S1):111-113 Research Article

KKU Engineering Journal

https://www.tci-thaijo.org/index.php/kkuenj/index

Using GM(1,1) with sample standard deviation to forecast downtrend rainfall for small sample in Khon Kaen, Thailand

Mathee Pongkitwitoon and Watcharin Klongdee*

Department of Statistics, Faculty of Science, Khon Kaen University, Khon Kaen 40002, Thailand.

Received April 2016
Accepted June 2016

Abstract

The objective of this research is to forecast the rainfall, Khon Kaen cumulative annual rainfall data, using GM(1,1) and Markov chain for small sample size based on sample standard deviation. We compare the accuracy of prediction models between this approach–sdGM(1,1) and standard GM(1,1) in terms of MSE and PARE. We find that sdGM(1,1) is better than GM(1,1).

Keywords: GM(1,1), Sample standard deviation, Small sample, Rainfall forecasting, Downtrend forecasting

1. Introduction

The climate studies in Thailand aim to find out current distributions of precipitation in particular rainfall and how these might change in the future. Almost current methodologies, unfortunately, require large data and complicated calculations to forecast the rainfall but these cannot predict the result accurately, especially in case of fluctuated and downtrend data–for example, Ono and Kazama developed an extreme daily rainfall distribution in Thailand's mountainous areas using 20-year APHRODITE data set from 150 rain gauges [1]. The theory of grey systems was proposed by Deng whose partial information is known and studied the uncertainty of such systems. Grey prediction model use the first order differential equation of one variable to simulate an unknown information–denoted by GM(1,1)–which is suitable for forecast in this situation, limited historical data [2]. There are many methods to improve accuracy of forecasting GM(1,1). For example, Huo and Zhan improved GM(1,1) by using Fourier series to correct the residuals based on genetic algorithm. Zhao et al. optimized GM(1,1) by using MFO algorithm with rolling mechanism. Yao et al. used segmental correction to improve accuracy [3-5]. On the other hand, Markov chain model using the transition probabilities –stochastic matrix– which can be applied to forecast the random varying time series system. Therefore, there are many studies using the Markov chain to improve the forecasting accuracy. For example, Sous et al. improved GM(1,1) using Markov chain and middle points matrix–denoted MCGM(1,1). Wang et al. forecasted gold price based on GM(1,1) and Markov chain [6-7].

In this research, firstly, we will consider the cumulative annual rainfall in Khon Kaen from 2000 to 2015 which is a downtrend data set with small sample size as shows in the Figure 1, Northeastern Meteorological Center (Upper Part) data set, and approach a new GM(1,1) model, GM(1,1) and Markov chain with downtrend data set, based on sample standard deviation–denoted sdGM(1,1). Secondly, the residual of prediction model, cumulative annual rainfall from 2000 to 2012, will be used to construct the stochastic matrix. Finally, we will predict cumulative annual rainfall from 2013 to 2015 using GM(1,1) and sdGM(1,1) then we will compare the forecasting models by using percentage absolute relative error (PARE) and mean square error (MSE).

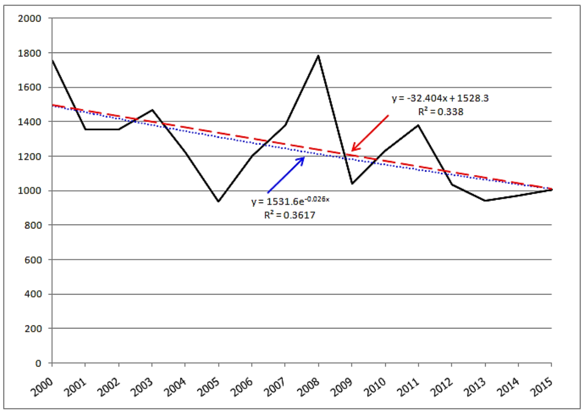

Figure 1 Cumulative annual rainfall in Khon Kaen between 2000 to 2015

2. Materials and methods

2.1 Grey Model - GM(1,1) model

GM(1,1) is the most widely used of grey model. Let $X^{(0)}$ is the nonnegative original sequence with n sample size which can be illustrated as

*Corresponding author. Tel.: +6683 356 9335
Email address: kwatch@kku.ac.th
doi: 10.14456/kkuenj.2016.34

$$X^{(0)} = \{x^0(1), x^0(2), \dots, x^0(n)\} \quad (1)$$

The accumulated generating operator (AGO) is defined by

$$X^{(1)} = \{x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n)\} \quad (2)$$

where

$$x^{(1)}(k) = \sum_{i=1}^{k} x^0(i), \quad k = 1,2,\dots,n$$

The GM(1,1) is given by

$$x^{(0)}(k) + a z^{(1)}(k) = b \quad (3)$$

where

$$z^{(1)}(k) = \frac{1}{2}\left(x^{(1)}(k) + x^{(0)}(k-1)\right), \quad k = 2,3,\dots,n$$

The parameters can be calculated by

$$\begin{pmatrix} a \\ b \end{pmatrix} = (B^T B)^{-1} B^T Y \quad (4)$$

where

$$Y_{(n-1)\times 1} = \begin{pmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{pmatrix}, \qquad B_{(n-1)\times 2} = \begin{pmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -z^{(1)}(n) & 1 \end{pmatrix}$$

The differential equation

$$\frac{dx^{(1)}}{dt} + a x^{(1)} = b \quad (5)$$

is called as albinism differential equation of GM(1,1) and the time response function is

$$x^{(1)}(t) = \left(x^{(1)}(1) - \frac{b}{a}\right) e^{-at} + \frac{b}{a} \quad (6)$$

Finally, the forecasting sequence $\hat{X}^{(0)}$ can be obtained by employing the inverse accumulated generating operator (IAGO), namely

$$\hat{X}^{(0)} = \begin{cases} \hat{x}^{(0)}(1) = x^{(0)}(1) \\ \hat{x}^{(0)}(k) = \hat{x}^{(1)}(k) - \hat{x}^{(1)}(k-1) = (1-e^{a})\left(x^{(0)}(1) - \frac{b}{a}\right) e^{-a(k-1)}, \\ \quad k = 2,3,\dots,n \end{cases} \quad (7)$$

2.2 Markov Chain

Markov chain is the analysis method that analyzes the future trends and possible results. The stochastic matrix is constructed by conditional probabilities of the state transferring from state *i* to state *j*, are called transition probabilities p_{ij}, that is

$$\Pr(X_{n+1} = x_i | X_n = x_j) = p_{ij}, \; i = 1,2,\dots,n \text{ and } j = 1,2,\dots,n \quad (8)$$

For homogeneous Markov process, there is $P^{(k)}$ is its k-step stochastic matrix and the Markov forecast model is

$$P^{(n)} = P^{(n-1)} P, \; n = 2,3,\dots \quad (9)$$

where *P* is an one-step stochastic matrix or initial stochastic matrix.

2.3 The approach model – sdGM(1,1)

The stochastic matrix is constructed by residual series of original data and grey forecasting model as in 2.1, that is

$$e(i) = x^{(0)}(i) - \hat{x}^{(0)}(i), \quad i = 1,2,\dots,n \quad (10)$$

The approach model–sdGM(1,1)– has steps as followings:

Step 1 Construct the GM(1,1) as 2.1 and check the trend of data. If data is downtrend then multiply the parameter *a* by -1.

Step 2 Construct the stochastic matrix as 2.2 by dividing the residual into five states as following:

state 1: $e_1 \in (-\infty, \bar{x} - 1.5sd]$ and $\bar{x} - 1.5sd$ represents this state–e_{10},

state 2: $e_2 \in (\bar{x} - 1.5sd, \bar{x} - 0.5sd]$ and $\bar{x} - 0.5sd$ represents this state–e_{20},

state 3: $e_3 \in (\bar{x} - 0.5sd, \bar{x} + 0.5sd]$ and $\bar{x}$ represents this state–e_{30},

state 4: $e_4 \in (\bar{x} + 0.5sd, \bar{x} + 1.5sd]$ and $\bar{x} + 0.5sd$ represents this state–e_{40},

state 5: $e_5 \in (\bar{x} + 1.5sd, \infty)$ and $\bar{x} + 1.5sd$ represents this state–e_{50}

where $\bar{X}$ is a sample mean and sd is a sample standard deviation.

Step 3 Do Markov forecast for residual of the model.

Step 4 The forecast value is defined by

$$\tilde{x}^{(1)}(k) = \hat{x}^{(1)}(k) - e_{i0}, \quad k = 1,2,\dots,n \text{ and } i = 1,2,3,4,5 \quad (11)$$

3. Results

We use Anderson-Darling goodness of fit to determine the residual distribution of the GM(1,1). The A-D statistic (H_0: the data follow the specified distribution, A-D statistic=0.3743, critical value=2.5018, α=0.05) shows that the residual distribution is a normal distribution. So we are confident to use the sdGM(1,1) model in this case. The Table 1 and Figure 2 show that the sdGM(1,1) is better than GM(1,1) for one-step, two-step and three-step forecasting (2013 to 2015 prediction).

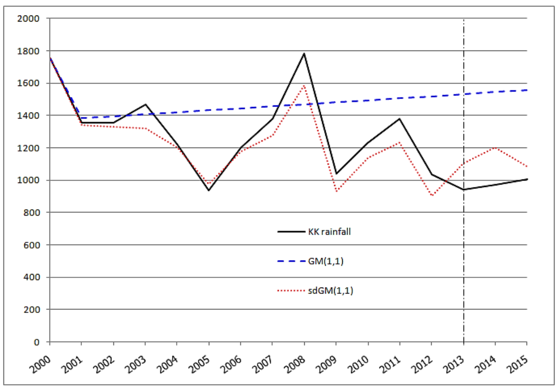

Figure 2 Comparison of Khon Kaen cumulative annual rainfall, GM(1,1) and sdGM(1,1)

Table 1 Results of prediction of GM(1,1) and sdGM(1,1)

Year	KK rainfall	GM(1,1)	Residual	sdGM(1,1)	Residual	Year	KK rainfall	GM(1,1)	Residual	sdGM(1,1)	Residual
2000	1751.0	1751.0	-	1751.00	-	2008	1780.6	1467.7	312.9	1584.46	196.1
2001	1352.6	1383.0	-30.4	1341.73	10.9	2009	1039.5	1480.2	-440.7	933.54	106.0
2002	1352.6	1394.8	-42.2	1330.39	22.2	2010	1230.1	1492.8	-262.7	1136.36	93.7
2003	1467.4	1406.7	60.7	1319.15	148.3	2011	1377.1	1505.5	-128.4	1232.55	144.6
2004	1221.8	1418.7	-196.9	1201.30	20.5	2012	1034.4	1518.4	-484.0	902.03	132.4
2005	936.5	1430.8	-494.3	976.83	-40.3	2013	943.1	1531.3	-588.2	1105.11	-162.0
2006	1201.5	1443.0	-241.5	1179.28	22.2	2014	973.1	1544.3	-571.2	1201.57	-228.5
2007	1378.9	1455.3	-76.4	1275.12	103.8	2015	1006.3	1557.5	-551.2	1084.71	-78.4

The accuracy of GM(1,1) and sdGM(1,1) determine by using MSE and PARE,

$$\text{MSE} = \frac{1}{n}\sum_{k=1}^{n}\left(x^{(0)}(k) - \hat{x}^{(0)}(k)\right)^2, \quad \text{and PARE} = \frac{1}{n}\sum_{k=1}^{n}\left|\frac{x^{(0)}(k) - \hat{x}^{(0)}(k)}{x^{(0)}(k)}\right| \times 100\%$$

The results in Table 2 show that sdGM(1,1) is better than GM(1,1) for all predicted values.

Table 2 Comparison of MSE and PARE of GM(1,1) and sdGM(1,1)

Model	MSE	PARE
GM(1,1)	325,371.6926	175.8467
sdGM(1,1)	28,197.9697	48.4489

4. Discussion

We improve GM(1,1) which is suitable for uptrend data sets for forecasting the small sample size data set based on sample standard deviation on the assumption that residual distribution is normal distribution. If residual distribution does not comply with the assumption, we can transform the residual into normal distribution and use sdGM(1,1) to do the prediction. The sdGM(1,1) is also a better forecasting model for downtrend data set.

5. Conclusions

The results of sdGM(1,1), which improving the GM(1,1) for downtrend data set with small sample size based on sample standard deviation, shows that this approach model is better than the original model. Moreover, it can forecast by using small sample size, especially downtrend data set.

6. Acknowledgements

We wish to thank Mr.Thanawut Panchapornudomlap, Director of Northeastern Meteorological Center (Upper Part) for supporting with meteorological data.

7. References

[1] Ono K, Kazama S. Analysis of Extreme Daily Rainfall in Southeast Asia with a gridded daily rainfall data set. Hydro-climatology: Variability and Change 2011;344:169-75.

[2] Deng Julong. Control Problem of Grey System. System Control Letters 1982;1(5):288-94.

[3] Huo H, Zhan S. Improved Grey Model GM(1,1) and Its Application based on Genetic Algorithm. International Journal of Digital Content Technology and its Applications (JDCTA) 2012;6:361-68

[4] Huiru Z, Haoran Z, Sen G. Using GM(1,1) optimized by MFO with Rolling Mechanism to Forecast the Electricity Consumption of Inner Mogolia. Applied Science 2016;6(20):1-18.

[5] Yao t, Liu S, Xie N. On the properties of small sample of GM(1,1) model. Applied Mathematical Modeling 2009;3:1894-903.

[6] Sous S, Thongjunthug T, Klongdee W. Gold Price Forecasting Based on the Improved GM(1,1) Model with Markov Chain by Average of Middle Points. KKU Science Journal 2014;42(3):693-9.

[7] Wang C, Chen Y, Li L. The Forecast of Gold Price Based on the GM(1,1) and Markov Chain. In: Liu S, editor. IEEE International Conference on Grey Systems and Intelligent Services; 2007 Nov 18-20; Nanjing, China. New Jersey: IEEE; 2007. p. 739-43.