Evolutionary computation role in improving an accuracy of forecasting mortality data

Article *in* International Journal of Advances in Soft Computing and its Applications · January 2018

2 authors:

Pasura Aungkulanon
Srinakharinwirot University

Pongchanun Luangpaiboon
Thammasat University

Int. J. Advance Soft Compu. Appl, Vol. 10, No. 2, July 2018
ISSN 2074-8523

Evolutionary Computation Role in Improving an Accuracy of Forecasting Mortality Data

Pasura Aungkulanon[1] and Pongchanun Luangpaiboon[2]

[1]Faculty of Industrial Technology, Phranakhon Rajabhat University
Bangkok, 10220, THAILAND
e-mail: pasurachacha@hotmail.com

[2]Industrial Statistics and Operational Research Unit (ISO-RU), Faculty of Engineering, Thammasat University
Pathumthani, 12120, THAILAND
e-mail: lpongch@engr.tu.ac.th

Abstract

It may be difficult to forecast mortality with conventional methods such as linear regression due to characteristics of data and appropriate tools for the prediction. In this article, two metaheuristic approaches namely sequential evolutionary procedures based on virus optimisation and elephant swarm water search are proposed to forecast mortality in Thailand. Related factors such as the number of populations, births, healthcare resources, and patients are included in the model. The results show that mean absolute errors of the virus optimisation algorithm are better, but statistically significant only on the quadratic model. Metaheuristic approaches are the powerful tool for predictions mortality and aid decision making.

Keywords: *Health forecasting, Mortality, Health resources, Elephant Swarm Water Search Algorithm, Virus Optimisation Algorithm.*

1 Introduction

Forecasting is the process of making prediction about future based on past and present data and analysis of the trend [1]. Forecasting is essential for decision making in various field such as security, finance, business and government budgeting, operation planning, agriculture, and transportation. Due to the increasing complexity and variety of real-world problems, there are many methods and techniques developed for forecasting in recent years. Time-series forecasting, the naïve approach, moving averages, weighted moving averages, exponential smoothing, exponential smoothing with trend adjustment, seasonal variations in data, artificial neural networks (ANNs)[2], autoregressive integrated moving average (ARIMA) [3], linear regression and quadratic regression are commonly used [4]. In general, however, the accuracy of these forecasting methods relies on long past data and has a limitation of handling non-linear data. Due to these reasons, metaheuristic approaches to forecasting trend turn out to be a successful tool. Metaheuristic algorithms have been developed to offer optimal or near-optimal solutions to complex problems and decision-making. Due to their powerful features, metaheuristic approaches have been commonly used in many fields of science and engineering, but the use of these potent algorithms for offering solutions to the complex health problems is still limited.

Mortality or death is affected by a variety of factors. They may be biological, physiological, environmental, etc. Genetic algorithm (GA) is one of the first and original proposed metaheuristics to solve complex health problems in a variety of medical disciplines [5]. Tan et al [6], conducted a study to investigate the relationship between soil trace elements and cervical cancer mortality in China based on genetic algorithm-partial least squares and support vector machines. Bozcuk et al [7], employed genetic algorithm technique to predict the risk of in-hospital mortality in a general population of cancer patients with non-terminal disease based on a combination of six blood tests. Virus optimisation (VOA) and elephant swarm water search (ESWSA) algorithms are novel and based on artificial life of animals and virus characteristics. Extensive comparisons of both algorithms were conducted with over well-known metaheuristic algorithms and showed that the VOA and ESWSA are viable solutions for complex real-life problems [8, 9].

Mortality statistics are a demographic indicator that is frequently used as a health outcome indicator. Empirical evidences show mortality is related with numbers of medical doctors and nurses, prevalence of disease, population size, and fertility. The aforementioned information is stored as disclosable data in social and quality of life database systems under the Office National Economic and Social Development Board. The present research was aimed at comparing the performances of two metaheuristic methods, VOA and ESWSA algorithms in searching for appropriate coefficients in linear and quadratic regressions. Appropriate performance is measured based on mean absolute error (MAE), mean square error (MSE), mean absolute percentage error (MAPE) and the correlation coefficient [10].

Accordingly, MAPE values obtained from predictions of annual mortality rates were compared. In doing so, information composed of several influential factors were involved such as number of population, healthcare budgets, birth rates, number of patients with non-communicable diseases, number of patients with infectious diseases, socioeconomic factors, health resource factors, etc. [11] In the following section, the forecasting model of deaths is briefly described. The basic sequential procedures of the VOA and ESWSA algorithms are given in section “The Proposed Methods”. In section “Results, Analysis and Discussions” results obtained by the death forecasting model and comparisons with VOA and ESWSA are presented and analysed. Finally, the conclusion of the study and the suggestions for future researches are given in section “Conclusion”.

2 Forecasting Model of Deaths

Forecasting plays a very significant role in aiding planning and decision-making concerning various activities in both the short- and long-terms in order to employ the data obtained in preventing and controlling potential illnesses in the future. The Ministry of Public Health gives importance to prediction of disease epidemiology and has set indicators contained within strategic plans. Thus, it is accepted that predictions play important roles in both public and private work. The present research utilised information from the social and quality of life database systems of the Office of the National Economic and Social Development Board from 1994 to 2016, and four related factors have been specified, namely number population, number of births, number of patients of any of 298 illnesses. Trends of mortality and related variables from 1994 to 2016 are shown in Fig. 1 and Table 1.

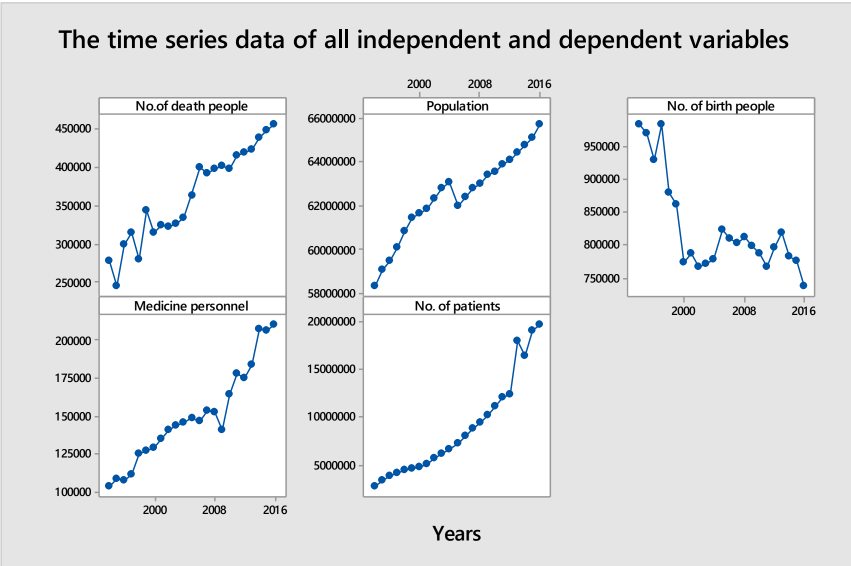

Fig. 1 The time series data of mortality and related variables from the office of the national economic and social development board

In this section, the linear and quadratic models are introduced in detail. We first describe the establishment of the model and its performance. On the establishment of the forecasting model there are four influential factors affecting the mortality in Thailand. According to Fig. 2, it can be seen that there is almost a linear relationship between four influential factors and mortality. However, only linear methods are often considered critical in order to obtain more accurate forecasts. A proper forecasting process gives the health organisation the opportunity to better understand situation dynamics and reduce uncertainty on future events, and provide the organisation's functions with useful analyses and information. After considering these issues, the mortality forecasting based on four influential factors was modeled using various forms in this research. Evolutionary elements from the virus optimisation (VOA) and elephant swarm water search (ESWSA) algorithms for estimation of Thai mortality forecasting in multiple linear form (S_{Linear}) [12] can be presented in equation (1).

$$S_{Linear} = \sum_{i=1}^{4} \alpha_i X_i + \beta \quad (1)$$

Both VOA and ESWSA for estimating Thai mortality in quadratic form ($S_{Quadratic}$) are shown in equation (2).

$$S_{Quadratic} = \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 + \alpha_4 X_4 + \alpha_5 X_1 X_2 + \alpha_6 X_1 X_3 + \alpha_7 X_1 X_4 + \alpha_8 X_2 X_3 + \alpha_9 X_2 X_4 + \alpha_{10} X_3 X_4 + \alpha_{11} X_1^2 + \alpha_{12} X_2^2 + \alpha_{13} X_3^2 + \alpha_{14} X_4^2 + \alpha_{15} \quad (2)$$

Table 1: Mortality and its influential factors

	Number of				
Years	Death People	Population	Birth People	Medicine Personnel	Patients of 298 Illnesses
1994	277,499	58,336,072	983,964	103,515	2,758,075
1995	244,061	59,095,419	970,760	108,510	3,413,050
1996	298,468	59,460,382	928,956	108,079	3,877,598
1997	315,467	60,116,182	983,395	112,155	4,227,410
1998	279,090	60,816,227	880,028	125,355	4,513,087
1999	344,210	61,466,178	862,260	127,779	4,663,946
2000	315,550	61,661,701	774,349	128,993	4,910,052
2001	323,846	61,878,746	786,018	135,639	5,156,157
2002	323,108	62,308,887	766,107	140,777	5,845,198
2003	326,583	62,799,872	771,787	144,454	6,220,845
2004	334,725	63,079,765	778,445	146,562	6,772,814
2005	363,647	61,973,621	822,575	149,019	7,260,862
2006	399,331	62,418,054	809,774	147,203	8,092,741
2007	392,044	62,828,706	802,924	154,413	8,911,696
2008	398,438	63,038,247	811,384	153,106	9,497,993
2009	401,981	63,389,730	797,356	141,086	10,307,684
2010	398,130	63,525,062	787,739	165,054	11,223,834
2011	414,888	63,878,267	766,370	178,388	12,078,096

2012	419,265	64,076,033	796,104	175,122	12,445,264
2013	423,213	64,456,695	818,901	184,487	17,999,153
2014	438,648	64,785,909	782,129	207,867	16,425,775
2015	448,601	65,124,716	776,370	207,019	19,126,383
2016	456,391	65,729,098	736,352	210,825	19,740,850

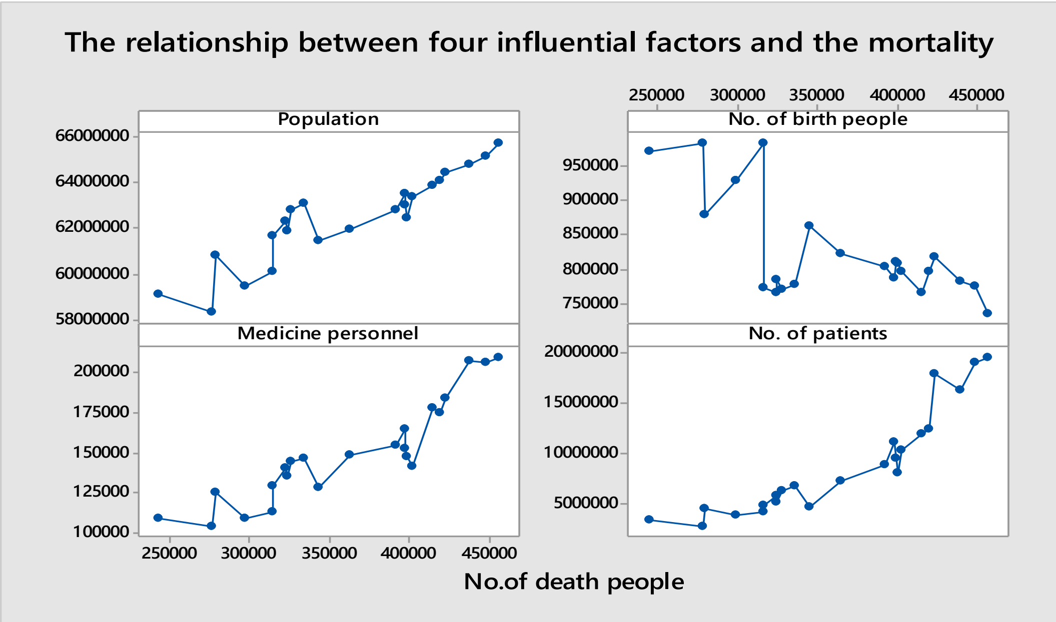

Fig. 2 The relationship between four influential factors and the mortality

3 The Proposed Methods

In order to produce optimal solutions, optimisation methods have come to play increasingly important roles in solution-finding at the industrial level. The objective is to produce several different optimal solutions. Models that indicate problems within systems attracting interest for testing include problems with and without resource limitations. Accordingly, the correlations between various factors can be presented without noise in the form of mathematical correlations containing a number of constraints. In regards to the importance of searching for the optimal solutions under the aforementioned problematic complexities, many researchers have suggested heuristic algorithms as a means for development in solving difficult and complex problems that cannot be solved with precision or that can be solved but require significant time in doing so. As a result, metaheuristics were developed to reduce calculation time, although the main purpose of metaheuristics is to produce best solutions. Good solution quality is naturally expected by developers of these methods. Therefore, in comparing metaheuristics developed to solve the same problems to determine more effective methods, the quality of solutions provided subsequent to processing or actions over the same given amount of time is compared. At present, there are several people who have invented metaheuristic methods.

3.1 Elephant Swarm Water Search Algorithm (ESWSA)

Mandal has proposed the ESWSA in the study of elephant behaviors in searching for water for survival [13]. Elephants are currently the largest land animals on the planet. With a gestation period lasting up to 22 months, elephants are considered to have the longest gestation of all animals. The average birthweight for elephants is 120 kilograms. In addition, elephant life spans range from 50 to 70 years. Due to their large size, elephants need to consume enormous quantities of food and water. Moreover, with their preference to live together in groups or herds, they have evolved and learned how to travel in search of food and water in groups. S Mandal's hypothesis about elephants consists of the following four components:

1. Elephants living in herds would scatter to forage in the forest; this style is similar to the broad or random search pattern.

2. Upon discovery of water, elephants communicate among themselves regarding the quantity and quality of whatever food and water they discover; this is akin to creating objective functions.

3. Elephants in a group will remember the quantity and quality of the food and water that they discover in order to communicate within their group. Subsequently, the group will travel toward the direction with the highest quality and quantity of food. This is similar to setting conditions or constraints in solution-finding.

4. A search for water and food is regulated by the probability (p), which is controlled by the leader of the herd on deciding whether to conduct a local or global search for food and water. Thus, the pseudo-code of the proposed ESWSA is show in Fig 3.

Procedure ESWSA Metaheuristic()

Begin;

Initialise algorithm parameters:

N:	*the number of elephant groups*
D:	*the dimensional optimization problem*
P:	*the probabilistic value*
$T_{max:}$	*the maximal number of generations*
$X_{max:}$	*the maximal number of variables*
$X_{min:}$	*the minimal number of variables*
Rand	*the random number between [0,1]*

Define the objective function of $f(\mathbf{x})$, *where* $\mathbf{x}=(x_1,........,x_D)^T$

Generate the initial population of elephant or $\mathbf{x}_i$ $(i=1, 2 ,..., n)$

Set P_{Best} *as the best position of elephant i;*

Set G_{Best} *as the global best position of elephant i;*

Set V_i *as the best position of elephant velocity i;*

While *(t< T_{max})*
For *i = 1 to N (all N elephant groups);*
If rand > P
global water search or update the elephant velocity
Else
local water search or update the elephant velocity
End if;
Update the position x_i
Evaluate objective function of f(x)
Update current best P_{Best}
Update global best G_{Best}
End for j;
End for i;
Rank the elephant groups and find the current best;
End while
Postprocess results and visualisation;
End procedure;

Fig. 3 Procedures of the ESWSA Metaheuristic

3.2 Virus Optimisation Algorithm (VOA)

Viruses are tiny living beings that cannot be spotted, even by regular microscopes with over 100 times magnification. Instead, they can only be viewed via electron microscopes with upward of 5,000 times magnification. Viruses are made up of proteins consisting of either DNA or RNA that forms the center of each virus, thereby providing its genetic material. Additionally, viruses have an external protein layer known as capsids. The cells of viruses are distinct from those of people and other animals; they are made up of both types of the aforementioned proteins, although some viruses have an additional layer made up of fatty substances. Viruses contain no energy reserves within themselves and they neither divide nor move when outside of human, animal, plant or even bacterial cells. Viruses can multiply and cause infection only once inside the cells of the infected person. These cells function like virus production factories. Based on this information, Liang and Cuevas Juarez (2016) constructed a model on how viruses harm human cells [14]. The VOA functions by the following stages: initialization, replication and updating/maintenance.

Initialisation: Initial solutions are sought and ordered in order to determine the strength of the virus. This is divided into two types, namely, strong and common members.

Replication: At this stage, new viruses are created based on the strong and common members.

Maintenance/Updating: In this stage, antivirus mechanisms are used to regulate the number of good and bad viruses in the entire virus population and play a part in the creation of new solutions if the solution-finding system cannot additionally improve solutions. Fig. 4 below shows the flowchart of the VOA where h represents the value of iterations, H denotes the maximum value of iterations, k is persistence counter of the current solution and K represents the maximum number of consecutive non-improving iterations.

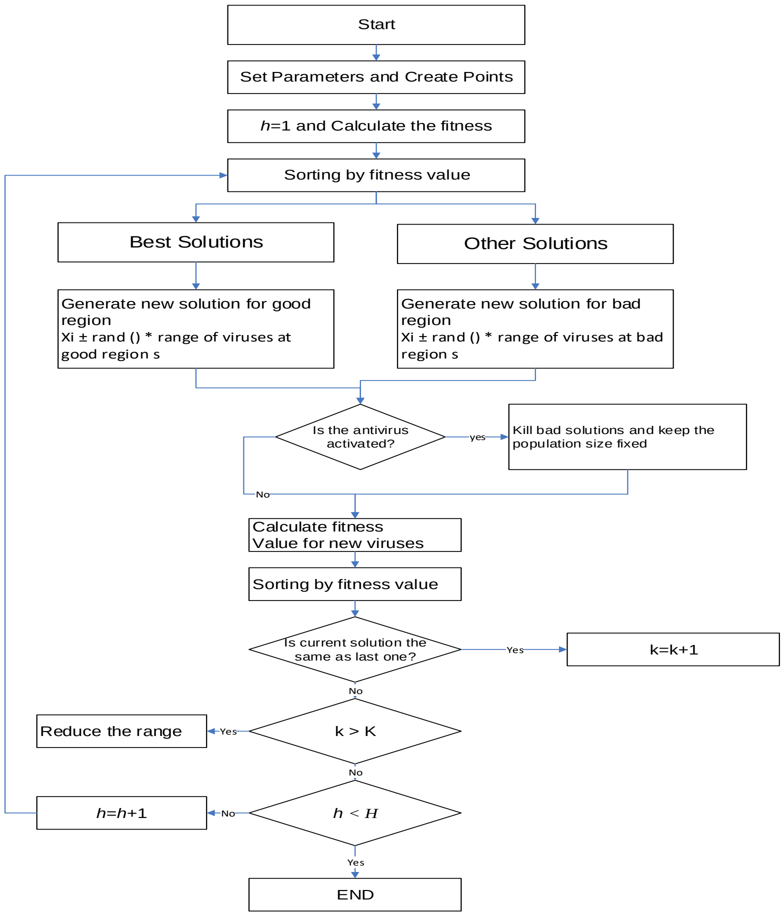

Fig. 4 Flowchart of the VOA Metaheuristic

4 Results, Analysis and Discussions

In this section, the numerical results are introduced in detail. The comparison of both algorithms of ESWSA and VOA is given. The establishment of the forecasting model with four influential factors and mortality were between 1994 and 2010. The data are collected from the Office of the National Economic and Social Development Board. The mortality forecasting based on four influential factors was modeled using linear and quadratic forms in this study. Two metaheuristic algorithms of ESWSA and VOA optimise coefficients (α_i, β) of the four influential factors $(X_i; i=1, 2, 3, 4)$. These factors consist of number of population (X_1), birth people (X_2), medicine personnel (X_3), and patients (X_4). The following mortality function is established by minimising the mean absolute error between the observed and estimated values of mortality are shown in equation (3).

$$f = \frac{1}{N}\sum_{i=1}^{N}|S_{Act}(i) - S_{Est}(i)| \quad (3)$$

where S_{Act} and S_{Est} are the actual and predicted number of death people (mortality), respectively; i is the number of observations; and N is the replications. The S_{Linear} and $S_{Quadratic}$ forecasting models are developed to the future mortality forecasting based on all four influential factors. Both algorithms of ESWSA and VOA are coded with visual c#2008. Setting the optimal parameters of ESWSA and VOA is also important. In all algorithms, population and maximum iteration number are set to 40 and 2000, respectively. Three parameter settings of the VOA including number of strong viruses (NV), growth rate of strong viruses (GRS) and growth rate of common viruses (GRC) were set 10, 8 and 2 respectively. The switching probability (p) of the ESWSA is set to 0.6 and the inertia weight factor decreases linearly from 0.9 to 0.4. The data from 1994 to 2010 are used to determine the coefficients of S_{Linear} and $S_{Quadratic}$ forecasting models. In the linear and quadratic forms of the VOA, coefficients obtained are given below:

S_{Linear} (VOA) $= -387929.9672 + 0.0089X_1 + 0.0849X_2 + 0.0386X_3 + 0.1405X_4$

$S_{Quadratic}$ (VOA) $= +520355893.9 - 25.162349X_1 + 120.225498X_2 + 3547.425308X_3 - 6.5921368X_4 + 1.10125\times10^7 (X_1X_2) - 4.27136\times10^5 (X_1X_3) +5.28941\times10^8 (X_1X_4) -0.000671(X_2X_3) + 7.26088\times10^5 (X_2X_4) + 2.02994\times10^5 (X_3X_4) + 0.257894(X_1^2) - 38.074136(X_2^2) - 270.605043(X_3^2) - 0.070096(X_4^2)$

The data (2011–2016) are used to validate the models. Tables 2 and 3 provide the coefficients of linear and quadratic models including the relative errors between estimated and observed data. Only the coefficient of X_4 on S_{Linear} is statistically significant at 95% confidence interval with the P-value of 0.043. The comparisons between ESWSA and VOA are shown in Figures 5 and 6 for linear and quadratic forms, respectively. It exposes that the VOA is providing better-fit estimation than ESWSA whether in linear or quadratic forms. However, an analysis of variance (ANOVA) reveals that there is sufficient evidence to show that algorithms provide statistically significant results on the quadratic form only, at the confidence interval of 95% (Table 4).

Table 2: Coefficients of linear model

Coefficients	Software	ESWSA	VOA
α1	0.0091	0.0091	0.0089
α2	0.0840	0.0840	0.0849
α3	0.0400	0.0393	0.0386
α4	0.01415	0.1405	0.1405

β	-387928	-387916.5136	-387929.9672
Mean Absolute Error (MAE)	4398825.117	4368911.1090	4351042.9429

Table 3: Coefficients of quadratic model

Coefficients	Software	ESWSA	VOA
α1	-25.514515	-25.514515	-25.162349
α2	121.054745	121.054745	120.225498
α3	3550.169225	3547.281296	3547.425308
α4	-7.5357796	-6.9289637	-6.5921368
α5	1.10125×10^{7}	2.89865×10^{6}	1.10125×10^{7}
α6	-5.03807×10^{5}	-4.3014×10^{5}	-4.27136×10^{5}
α7	5.28941×10^{8}	9.50193×10^{6}	5.28941×10^{8}
α8	-0.000610	-0.000658	-0.000671
α9	2.8263×10^{6}	-2.4057×10^{5}	7.26088×10^{5}
α10	2.04317×10^{5}	0.00011	2.02994×10^{5}
α11	0.258844	0.257988	0.257894
α12	-38.074136	-38.074136	-38.074136
α13	-271.167920	-271.945865	-270.605043
α14	-0.062015	-0.068537	-0.070096
α15	520355893.4	520355894.5	520355893.9
Mean Absolute Error (MAE)	1.032344×10^{15}	1.026953×10^{15}	1.025566×10^{15}

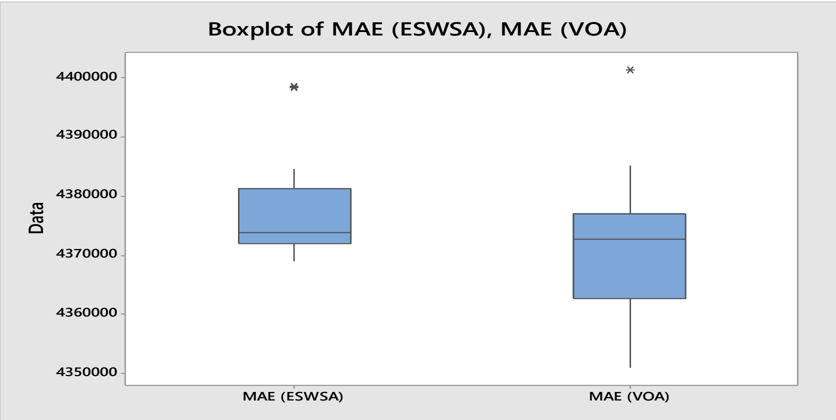

Fig. 5 Comparisons of mortality in the linear form

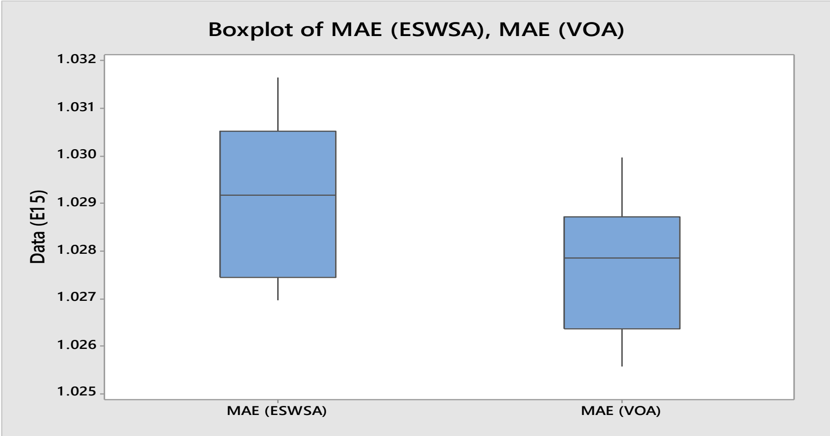

Fig. 6 Comparisons of mortality in the quadratic form

Table 4: Analysis of variance of both algorithm based on the quadratic form

Sources of Variation	Sum of Squares	Degree of Freedom	Mean Squares	F value	P-value
Algorithms	0.000014	1	0.000014	6.23	0.019
Error	0.000062	28	0.000002		
Total	0.000076	29			

5 Conclusion

This paper presents a comparative study between linear and quadratic models in mortality forecasting. Accurate forecasts of future mortality can help improve effective operations in public health issues. The evolution of intelligent metaheuristics on forecasting has been widely used to forecast better. Since mortality data present wide-range variations, we investigate the effects of different algorithms on their forecasting accuracy via the mean squared errors. Our results suggest that the virus optimisation algorithm (VOA) is the preferred approach to forecast mortality with a higher quality result and a faster convergence. Although there are only two algorithms have been provided to the mortality forecasting in this study, researching other metaheuristics also will be necessary in the future [15-17]. In addition, the number of influential factors of the model proposed is limited, so the suggestion for further works is to take more related factors into consideration to the mortality forecasting. The interval value of coefficients parameters of linear and quadratic functions can be adjusted and analysed by other statistical software and other metaheuristic algorithms. The coefficients values have impact to predict value and minimum errors factor.

ACKNOWLEDGEMENTS
The first and second authors wish to thank the Faculty of Industrial Technology, Phranakhon Rajabhat University and the Faculty of Engineering, Thammasat University, respectively, for the financial support.

References

[1] Zhuming, B., & Cochran, D. (2014). Big data analytics with applications, *Journal of Management Analytics*, *1*(4), 249-265.

[2] Luxhøj, J.T., Riis, J.O., & Stensballe, B. (1996). A hybrid econometric—neural network modeling approach for sales forecasting. *Journal of Production Economics*, *43*, 175–192.

[3] Ina, K., Ratnadip, A., & Ghanshyam, V. (2015). Time Series Forecasting Using Hybrid ARIMA and ANN Models Based on DWT Decomposition, *Procedia Computer Science*, *48*, 173-179.

[4] Chu, C.W., & Zhang, G.P. (2003). A comparative study of linear and nonlinear models for aggregate retail sales forecasting. *International Journal of Production Economics*, *86*, 217–231.

[5] Ghaheri, A., Shoar, S., Naderan, M., & Hoseini, S.S. (2015). The Applications of Genetic Algorithms in Medicine. *Oman Medical Journal*, *30*(6), 406–416.

[6] Tan, C., Chen, H., Wu, T., & Xia, C. (2011). Modeling the relationship between cervical cancer mortality and trace elements based on genetic algorithm-partial least squares and support vector machines. *Biological Trace Element Research*, *140*(1), 24-34.

[7] Bozcuk, H., Bilge, U., Koyuncu, E., & Gulkesen, H. (2004). An application of a genetic algorithm in conjunction with other data mining methods for estimating outcome after hospitalisation in cancer patients. *Medical Science Monitor*, *10*(6), 246-251

[8] Cuevas, J.R., Wang, H. J., Lai, Y.C., & Liang. Y.C. (2009, December). Virus optimisation algorithm: a novel meta-heuristic for solving continuous optimisation problems. In *Industrial Engineering & Management System, 2009. APIEMS2009. Tenth Asia Asia Pacific Industrial Engineering and Management Systems*, (pp. 2166–2174).

[9] Liang, Y.C., & Cuevas, J.R. (2016). Harmony search and virus optimisation algorithm for multi-objective combined economic energy dispatching problems. In *Evolutionary Computation (CEC), 2016. CEC2016. IEEE Congress on*, (pp. 2166–2174). IEEE.

[10] Danese, P., & Kalchschmidt, M. (2011). The role of the forecasting process in improving forecast accuracy and operational performance. *International Journal of Production Economics*, *131*, 204–214.

[11] Aungkulanon, S., McCarron, M., Lertiendumrong, J., Olsen, S.J., & Bundhamcharoen, K. (2012). Infectious Disease Mortality Rates, Thailand, 1958–2009. *Emerging Infectious Diseases*, 18(11), 1794-1801.

[12] Gao, J., Xie, Y., Gu, F., Xiao, W., Hum J., & Yu, W. (2017). A hybrid optimisation approach to forecast automobile sales of China. *Advances in Mechanical Engineering*, 9(8), 1–10.

[13] Mandal, S. (2018). Elephant swarm water search algorithm for global optimisation. *Sådhanå*, 43(2), 1-21, DOI: 10.1007/s12046-017-0780-z.

[14] Liang, Y.C., & Juarez, J.R.C. (2016). A novel metaheuristic for continuous optimisation problems: virus optimisation algorithm. *Engineering Optimisation*, *48*(1), 73-93, DOI: 10.1080/0305215X.2014.994868.

[15] Godfrey C. (2018). A New Plant Intelligent Behaviour Optimisation Algorithm for Solving Vehicle Routing Problem. *Mathematical Problems in Engineering*, https://doi.org/10.1155/2018/9874356

[16] Caraveo, C., Valdez, F., & Castillo, O. (2017). A New Meta-Heuristics of Optimisation with Dynamic Adaptation of Parameters Using Type-2 Fuzzy Logic for Trajectory Control of a Mobile Robot. *Algorithms*, *10*(3), 85-101.

[17] Hosseini, E. (2017) Laying Chicken Algorithm: A New Meta-Heuristic Approach to Solve Continuous Programming Problems. *Journal of Applied & Computational Mathematics*, 6, 344-352. DOI: 10.4172/2168-9679.1000344