MATHEMATICAL MODELING AND COMPUTING, Vol. 12, No. 2, pp. 490–497 (2025)

A Kernel Grey Model with Genetic Algorithm Optimizer and its applications in forecasting the palm oil price in Malaysia

Shabri A.

Mathematics Department, Faculty of Science, Universiti Teknologi Malaysia, 81310, Johor, Malaysia

(Received 15 December 2024; Revised 21 May 2025; Accepted 22 May 2025)

Accurate forecasting is difficult since palm oil prices are consistently highly nonlinear. It is important to choose the right forecasting models since there are several available. The grey model has proven to be a good forecasting model. Nevertheless, the majority of extant grey models are fundamentally linear models, which limits their ability to capture nonlinear trends. This paper introduces a nonlinear extended parametric grey model known as the kernel grey model (KGM). However, the prediction of the KGM model is dependent on the kernel function and the KGM parameters. This study presents a genetic algorithm-based enhanced KGM model and verified it with data on Malaysia's palm oil prices from 2000 to 2019. The multivariable linear regression (MLR) and support vector machine (SVM) forecasting models were chosen for comparison based on mean absolute percent error and root mean square percent error. The results reveal that KGM outperforms the other two models in training and testing data performance, and it can significantly enhance forecast accuracy.

Keywords: *multivariable linear regression, kernel grey model, support vector machine, palm oil.*

2010 MSC: 62M10, 62G08 **DOI:** 10.23939/mmc2025.02.490

1. Introduction

The palm oil industry has become one of the most important sectors in Malaysia due to the volume of production and the number of exports to other nations. The price of palm oil is a significant factor in the palm oil industry. This price has been employed to determine the nation's revenue and for numerous decision-making processes [1]. Palm oil, like many other economic commodities, has a fluctuating and changing price due to its complicated interaction with other economic elements such as demand and supply rates, substitute product prices, and crude oil prices. Policymakers, dealers, and farmers are just a few of the industry participants who can benefit from an accurate palm oil price forecasting model. However, forecasting palm oil prices is quite challenging because of the complex and dynamic nature of the underlying elements that influence the pricing. In the last three decades, numerous studies have been conducted to forecast palm oil prices using statistical and machine learning methodologies. Statistics methods such as the autoregressive moving average (ARIMA) model [2–4], the generalised autoregressive conditional heteroscedasticity (GARCH) model [5,6], the linear regression (econometric) model [7], the autoregressive distributed lag (ARDL) model [8], and the Holt–Winters method [9]. Although statistical models are highly versatile, their main limitation is the presumed linear form of the model. The approximation of linear models to complicated real-world problems is not always successful. For machine learning methods, artificial neural networks (ANN) [10,11], fuzzy-ruled based [12], random forest method [13] and support vector machine (SVM) [14] have been widely explored in palm oil forecasting, which has considerable advantages in dealing with complex nonlinear situations. Even

The author sincerely appreciates the support for this research project provided by the Ministry of Higher Education Malaysia (MOHE) and Universiti Teknologi Malaysia (UTM) under the Fundamental Research Grant Scheme (FRGS/1/2019/STG06/UTM/02/3).

though these models are capable of producing more accurate results, they have certain limitations in certain situations. For example, selecting a large number of parameters experimentally through trial and error is difficult and require complex calculations in order to achieve good forecasting accuracy [15]. Recently, the grey system theory has been utilized in various disciplines [16–20]. The classical grey prediction model, GM(1,1), has been utilized for decades, and numerous updated variants of the model have been proposed [19–23]. Kernel learning is another major method in the machine learning model. The grey model combined with the kernel approach to create a kernel grey model (KGM), demonstrating that the KGM model can improve grey model performance [24–26]. Although the KGM model is an acceptable forecasting model, selecting the kernel function and parameters is a critical issue in KGMs. Only the grid search approach [24, 26], the try-and-error method [27] and the grey wolf optimizer (GWO) [28] have been investigated as methods for optimizing KGM parameters. To improve the forecasting performance of the KGM model, this paper proposes the optimal parameter estimation approach based on the optimal theory of genetic algorithms (GA). The GA considered in this article is one of the most used algorithms for solving numerous optimization problems [29].

2. Some existing palm oil price forecasting models

In this section, a few Palm Oil Price forecasting models that will be employed in the application section are briefly reviewed.

2.1. Multivariable linear regression

Let $y_1^{(0)}(k) = \big(y_1^{(0)}(1), y_1^{(0)}(2), \ldots, y_1^{(0)}(n)\big)$ be the actual dataset for the dependent variable (primary factor), and let $X_i^{(0)} = \big(x_i^{(0)}(1), x_i^{(0)}(2), \ldots, x_i^{(0)}(n)\big)$ for $i = 2, 3, \ldots, N$ be the related factors. Then, the definition of the multivariable linear regression (MLR) model is

$$y_1^{(0)}(k) = b_1 + \sum_{i=2}^{N} b_i x_i^{(0)}(k), \quad k = 1, 2, 3, \ldots, n \tag{1}$$

where $b_1, b_2, \ldots, b_N$ are computed using the ordinary least squares (OLS) technique,

$$[b_i]^T = \left(B^T B\right)^{-1} B^T Y. \tag{2}$$

The MLR model's estimated values can be represented as

$$\hat{y}_1^{(0)}(k) = b_1 + \sum_{i=2}^{N} b_i x_i^{(0)}(k), \quad k = 1, 2, 3, \ldots, n. \tag{3}$$

2.2. Support vector machines

Support vector machine (SVM) can be divided into two parts known as support vector classification (SVC) and support vector regression (SVR) [30]. The main principle of SVM is to map the original data into a high-dimensional feature space via nonlinear mapping. Supposing the training data are $\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$, the SVR formula can be expressed as

$$y_i = \sum_{i=1}^{n} w \cdot \phi(x_i) + b, \tag{4}$$

where w_i are the weights, b is the threshold value, and ϕ_i are the nonlinear mapping functions. The weights w_i can be obtained by minimizing the following quadratic programming problem:

$$\min \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*)$$

subject to

$$|y_i - w \cdot \Phi(x_i) - b| \leqslant \varepsilon + \xi_i, \quad \xi_i, \xi_i^* \geqslant 0, \quad i = 1, 2, \ldots, n,$$

where $C > 0$ is the penalty factor, ξ_i and ξ_i^* are the slack variables, and ε is the intensive loss function. By solving the optimization problem, the estimation function of SVR can be obtained as follows:

$$y=\sum_{i=1}^{n}(\alpha_i-\alpha_i^*)\,K(x_i,x_j)+b, \tag{5}$$

where α_i, α_i^* are positive and solve the quadratic programming problem

$$\min_{\alpha_i,\alpha_i^*}\varepsilon\sum_{i=1}^{n}(\alpha_i+\alpha_i^*)-\sum_{i=1}^{n}y_i(\alpha_i^*-\alpha_i)+\frac{1}{2}\sum_{i,i'=1}^{n}y_i\,(\alpha_i^*-\alpha_i)\,(\alpha_i'^*-\alpha_i')\,K(x_i,x_j) \tag{6}$$

subject to

$$\sum_{i=1}^{n}(\alpha_i-\alpha_i^*)=0,\quad \alpha_i\geqslant 0,\quad \alpha_i^*\leqslant C,$$

where $K(x_i,x_j)=\sum_{i=1}^{D}\phi(x_i)\cdot\phi(x_j)$ is a kernel function. Examples of kernel functions include polynomials, Radial Basis Functions (RBF), sigmoid, and linear functions. The RBF kernel function is employed in this paper because it is computationally simple and reliable when dealing with nonlinear data. The RBF is defined mathematically as

$$K(x_i,x_j)=\exp\left(-\frac{\|x_i-x_j\|^2}{2\sigma^2}\right),$$

where σ^2 is the width of the RBF, and $\|x_i-x_j\|$ is the Euclidean distance between two points x_i and x_j.

2.3. Kernel grey model

The kernel grey model (KGM) is represented by

$$y_1^{(0)}(k)+a\,z_1^{(1)}(k)=w^T\phi(k)+u, \tag{7}$$

where w is a weight vector, u is a bias, and $z_1^{(1)}(k)=0.5x_1^{(1)}(k)+0.5x_1^{(1)}(k-1)$, $k\geqslant 2$. For estimating the parameters of the KGM model, the following regularized problem is considered:

$$\min J(a,w,u,e)=\frac{a^2}{2}+\frac{w^2}{2}+\frac{C}{2}\sum_{k=2}^{n}e_k^2 \tag{8}$$

subject to

$$e_k=y_1^{(0)}(k)+a\,z_1^{(1)}(k)-w^T\phi(k)-u,$$

where C is the regularization parameter. The solution of the KGM can be written as

$$\hat{y}_1^{(1)}(k)=\alpha^{(k-1)}x_1^{(0)}(1)+\sum_{j=2}^{k}(\phi(j)+\mu)\,\alpha^{(k-j)}, \tag{9}$$

where

$$\alpha=\frac{1-0.5a}{1+0.5a},\quad a=\sum_{k=2}^{n}\lambda_k z_1^{(1)}(k),\quad \phi(k)=\frac{\sum_{j=2}^{n}\lambda_j K(x_j,x_k)}{1+0.5a},\quad \mu=\frac{u}{1+0.5a},$$

$$[u,\lambda]^T=(B^TB)^{-1}B^TY,$$

$$B=\begin{pmatrix}0 & 1_{n-1}^T\\ 1_{n-1} & \Omega+\frac{I_{n-1}}{C}\end{pmatrix},\quad \lambda=\begin{pmatrix}\lambda_2\\ \vdots\\ \lambda_n\end{pmatrix}^T,\quad 1_{n-1}=\begin{pmatrix}1\\1\\ \vdots\\1\end{pmatrix}_{n-1},$$

$$\Omega=\left(K(x_i,x_j)-z_1^{(1)}(i)z_1^{(1)}(j)\right)_{(n-1)\times(n-1)}.$$

The forecast value of KGM is computed using the formula:

$$\hat{y}_1^{(0)}(k)=\hat{y}_1^{(1)}(k)-\hat{y}_1^{(1)}(k-1). \tag{10}$$

Two important parameters, C and σ^2, which directly affect KGM's ability, are provided by Eq. (7). To identify the optimum options, the Genetic Algorithm (GA) is employed as a parameter search

Table 1. GRA between the price of crude palm oil and all variables.

GRA	PPO	SPO	PSO	PCO
$r\big(X_1^{(0)}, X_i^{(0)}\big)$	0.671	0.683	0.724	0.623

technique. To improve modeling accuracy, the objective function with the lowest Mean Absolute Percentage Error (MAPE) is chosen. The basic procedures for applying the GA to an optimization problem are as follows.

Step 1: Construct the fitness function and generate a random population of chromosomes with both C and σ^2 within acceptable limits.

Step 2: The MAPE of reference points and check points evaluates an individual's fitness based on a set of parameters C and σ^2.

Step 3: During the selection process, chromosomes with a high degree of fitness are more likely to be chosen.

Step 4: After the genetic procedure, the fitness of the new offspring would be determined once more. Then, a new generation would be formed by the best N outcomes of parents or offspring.

Step 5: Steps 1–4 should be repeated until the end condition is satisfied or the predefined maximum number of iterations has been met. If the requirements for termination are satisfied, the procedure is completed. The KGM parameters with the highest fitness (C and σ^2) are the outcome of the process.

2.4. Evaluation of the model performance

Root mean square error (RMSE) and mean absolute percentage error (MAPE), two widely used model evaluation indicators, were utilized to assess the predicted accurateness of the proposed model. These model evaluation indicators can be calculated using the following formulas:

$$\mathrm{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i^{(0)} - \hat{y}_i^{(0)}}{y_i^{(0)}}\right|$$

and

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\big(y_i^{(0)} - \hat{y}_i^{(0)}\big)^2}.$$

3. Case study

This study uses five variables: the price of crude palm oil (in Malaysian Ringgit per ton) as the dependent variable, the price of soybean oil (PSO), the price of crude oil (PCO), and the production of palm oil (PPO) as the independent variables. There are 240 observations in the data, comprising from January 2000 to December 2019. This study's data comes from the Malaysian Palm Oil Council's official website and the United States Energy Information Administration. A correlation measure based on Malaysian palm oil and the factors influencing palm oil has been created using grey relational analysis (GRA). GRA is used to quantify the relationship between the price of crude palm oil and each of the independent variables. When the GRA is close to one, the dependent and associated independent variables show a strong relationship. The GRA is calculated using the following formula:

$$r\big(X_1^{(0)}, X_i^{(0)}\big) = \frac{1}{m}\sum_{k=1}^{m}\gamma\big(X_1^{(0)}, X_i^{(0)}\big), \tag{11}$$

where

$$\gamma\big(X_1^{(0)}, X_i^{(0)}\big) = \frac{\min\limits_{j\in i}\min\limits_{k}\big|X_1^{(0)} - X_j^{(0)}\big| + 0.5\max\limits_{j\in i}\max\limits_{k}\big|X_1^{(0)} - X_j^{(0)}\big|}{\big|X_1^{(0)} - X_i^{(0)}\big| + 0.5\max\limits_{j\in i}\max\limits_{k}\big|X_1^{(0)} - X_j^{(0)}\big|}.$$

Lin et al. [31] state that if the GRA is greater than 0.5, associated independent variables are considered correlated with the dependent variable.

Table 1 displays the GRA between palm oil price and related variables, with palm oil price having the greatest correlation with PSO, followed by SPO, PPO, and PCO. The training data were calculated from January 2000 to December 2019, while the testing data or forecast value were calculated from January 2020 to December 2020.

Evaluation of the model performance

The following displays the predicted results of three models that use these data. The models for MLR and KGM are developed using MATLAB. The MLR model can be written as follows:

$$y_t = 1552.823 - 0.0001x_1 - 0.0004x_2 + 2.1121x_3 - 4.1719x_4.$$

The RBF kernel is utilised for the SVM model because, in comparison to other kernels, it provides the best prediction performance [14]. The SVM model's performance is affected by the hyperparameters C, σ, and ε. To determine the most optimal combination of modelling parameters (C, σ, ε), values for C were chosen to be between $[10^0, 10^3]$, σ within $[10^{-3}, 10^4]$ in steps of 10^1, and ε set at 0.1. The optimal values are determined using the `tune_svm` function from the R package, which can be found in e1071. After standardising the sample set to $[0, 1]$, the SVR prediction results were renormalized in order to create the testing models. The ideal combination of (C, σ, ε) parameters is chosen by minimizing the mean square error (MSE) via the ten-fold cross-validation procedure. The optimal SVM values are $C = 10$, $\sigma^2 = 0.1$, and $\varepsilon = 0.1$.

The optimal KGM parameters using GA are $u = 0.9998$, $\sigma^2 = 12.1558$, $C = 94.1439$, and $\lambda = [-0.016, -0.0101, \ldots, 0.0031, -0.0030]$. The MAPE and RMSE of the MLR, SVM, and KGM models for training and testing data are shown in Figure 1.

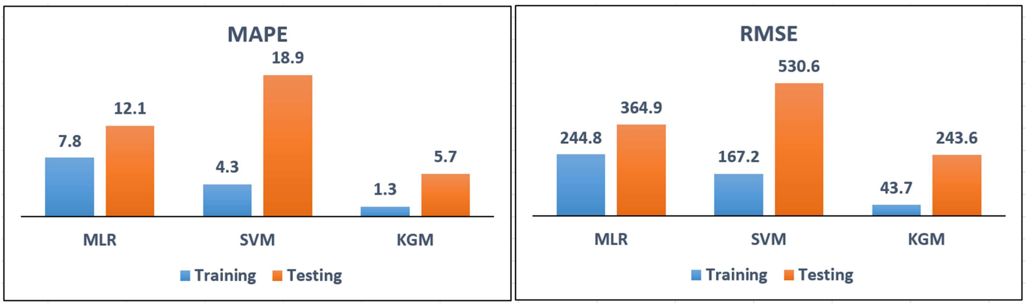

Fig. 1. The MAPE and RMSE of the MLR, SVM, and KGM models for training and testing data.

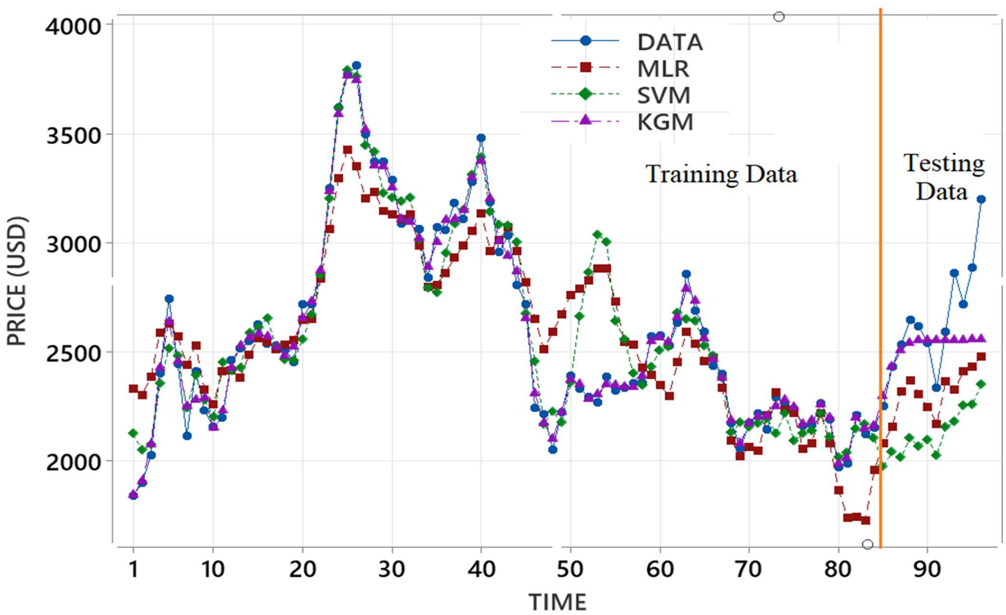

Fig. 2. The curves of real values and forecast of MLR, SVM, and KGM models.

According to Figure 1, MLR, SVM, and KGM had MAPEs of 7.8%, 4.3%, and 1.3% during training data and 12.1%, 18.9%, and 5.7%, respectively, during testing. Figure 1 illustrates that the proposed KGM model has a significantly smaller RMSE than the MLR and SVM models. As can be observed from Figure 1's results, MLR and SVM have significant RMSEs in both training and testing data, suggesting that these two models are not suitable for palm oil price prediction.

Figure 2 shows the trend of palm oil prices in the training and testing data of the three types of forecast models. It is interesting to observe that the KGM model's predicted values are quite close to the actual values when compared to the two other models.

4. Conclusions

A relatively successful advancement in grey models is represented by kernel grey models (KGM). The intelligent optimisation of the KGM was the primary focus of this paper. In order to create a KGM model based on intelligent optimisation, we proposed in this paper that the optimal parameter based on a kernel grey model be determined using the GA algorithm. When compared to MLR and SVM models, the results of our experiments showed that this model performed well and predicted the price of palm oil more accurately. In order to use the model in other situations, more research is required. In future research, researchers can investigate approaches for improving prediction stability. Furthermore, because the primary focus of this paper was medium-term prediction, our next study will concentrate on combining this model with deep learning-based algorithms for medium- and long-term prediction.

Acknowledgments

The author sincerely appreciates the support for this research project provided by the Ministry of Higher Education Malaysia (MOHE) and Universiti Teknologi Malaysia (UTM) under the Fundamental Research Grant Scheme (FRGS/1/2019/STG06/UTM/02/3).

[1] Abdullah R., Lazim M. A. Production and price forecast for Malaysian palm oil. Oil Palm Industry Economic Journal. **6** (1), 39–45 (2006).

[2] Ariff N. M., Zamhawari N. H., Bakar M. A. A. Time series ARIMA models for daily price of palm oil. AIP Conference Proceedings. **1643** (1), 283–288 (2015).

[3] Sukiyono K., Arianti N. N., Sumantri B., Mustopa Romdhon M., Suryanty M. A Model Selection for Price Forecasting of Crude Palm Oil and Fresh Fruit Bunch Price Forecasting. Iraqi Journal of Agricultural Sciences. **52** (2), 479–490 (2021).

[4] Yee K. W., Samsudin H. B. Comparison between Artificial Neural Network and ARIMA Model in Forecasting Palm Oil Price in Malaysia. International Journal of Scientific Engineering and Science. **5** (11), 12–15 (2021).

[5] Khamis A., Hameed R., Nor M. E., Che Him N., Mohd Salleh R., Mohd Razali S. N. A. Comparative Study on Forecasting Crude Palm Oil Price using Time Series Models. Scientific Research Journal. **VI** (XII), 1–8 (2018).

[6] Mohamad Hanapi A. L., Othman M., Sokkalingam R., Sakidin H. Developed A Hybrid Sliding Window and GARCH Model for Forecasting of Crude Palm Oil Prices in Malaysia. Journal of Physics: Conference Series. **1123** (1), 012029 (2018).

[7] Khin A. A., Mohamed Z., Malarvizhi C. A., Thambiah S. Price Forecasting Methodology of the Malaysian Palm Oil Market. The International Journal of Applied Economics and Finance. **7** (1), 23–36 (2013).

[8] Hamid M. F. A., Shabri A. Palm oil price forecasting model: An autoregressive distributed lag (ARDL) approach. AIP Conference Proceedings. **1842** (1), 030026 (2017).

[9] Suppalakpanya K., Nikhom R., Booranawong A., Booranawong T. An evaluation of Holt–Winters methods with different initial trend values for forecasting crude palm oil production and prices in Thailand. Suranaree Journal of Science and Technology. **26** (1), 13–22 (2019).

[10] Karia A. A., Bujang I., Ahmad I. Forecasting on Crude Palm Oil Prices Using Artificial Intelligence Approaches. American Journal of Operations Research. **3** (2), 259–267 (2013).

[11] Salman N., Lawi A., Syarif S. Artificial Neural Network Backpropagation with Particle Swarm Optimization for Crude Palm Oil Price Prediction. Journal of Physics: Conference Series. **1114** (1), 012088 (2018).

[12] Wulandari R., Surarso B., Irawanto F. The Forecasting of Palm Oil Based on Fuzzy Time Series-Two Factor. Journal of Soft Computing Exploration. **2** (1), 11–16 (2021).

[13] Myat A. K., Tun M. T. Z. Predicting palm oil price direction using random forest. 2019 17th International Conference on ICT and Knowledge Engineering (ICT&KE). 1–6 (2019).

[14] Shabri A., Hamid M. F. A. Wavelet-support vector machine for forecasting palm oil price. Malaysian Journal of Fundamental and Applied Sciences. **15** (3), 398–406 (2019).

[15] Zhang G., Putuwo B. E., Hu M. Y. Forecasting with artificial neural networks: the state of the art. International Journal of Forecasting. **14** (1), 35–62 (1998).

[16] Xiao X., Duan H. A new grey model for traffic flow mechanics. Engineering Applications of Artificial Intelligence. **88**, 103350 (2020).

[17] Duan H., Xiao X., Long J., Liu Y. Tensor alternating least squares grey model and its application to short-term traffic flows. Applied Soft Computing Journal. **89**, 106145 (2020).

[18] Bezuglov A., Comert G. Short-term freeway traffic parameter prediction: application of grey system theory models. Expert Systems with Applications. **62**, 284–292 (2016).

[19] Duan H., Pang X. A multivariate grey prediction model based on energy logistic equation and its application in energy prediction in China. Energy. **229**, 120716 (2021).

[20] Chen C. I., Chen H. L., Chen S. P. Forecasting of foreign exchange rates of Taiwan's major trading partners by novel nonlinear Grey Bernoulli model NGBM(1,1). Communications in Nonlinear Science and Numerical Simulation. **13** (6), 1194–1204 (2008).

[21] Ma X., Liu Z. Application of a novel time-delayed polynomial grey model to predict the natural gas consumption in China. Journal of Computational and Applied Mathematics. **324**, 17–24 (2017).

[22] Zhou W., Zeng B., Wang J., Luo X., Liu X. Forecasting Chinese carbon emissions using a novel grey rolling prediction model. Chaos, Solitons & Fractals. **147**, 110968 (2021).

[23] Zhang Y., Mao S., Kang Y., Wen J. Fractal derivative fractional grey Riccati model and its application. Chaos, Solitons & Fractals. **145**, 110778 (2021).

[24] Duan H., Wang D., Pang X., Liu Y., Zeng S. A novel forecasting approach based on multi-kernel nonlinear multivariable grey model: A case report. Journal of Cleaner Production. **260**, 120929 (2020).

[25] Ma X., Hu Y. S., Liu Z. B. A novel kernel regularized nonhomogeneous grey model and its applications. Communications in Nonlinear Science and Numerical Simulation. **48**, 51–62 (2017).

[26] Ma X., Liu Z.-b. The kernel-based nonlinear multivariate grey model. Applied Mathematical Modelling. **56**, 217–238 (2018).

[27] Wang C.-H., Hsu L.-C. Using genetic algorithms grey theory to forecast high technology industrial output. Applied Mathematics and Computation. **195** (1), 256–263 (2008).

[28] Ma X., Deng Y., Ma M. A novel kernel ridge grey system model with generalized Morlet wavelet and its application in forecasting natural gas production and consumption. Energy. **287**, 129630 (2024).

[29] Hastie T., Tibshirani R., Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer, New York (2017).

[30] Vapnik V. The Support Vector Method Of Function Estimation. Nonlinear Modeling. 55–85 (1998).

[31] Lin S. J., Lu I. J., Lewis C. Grey relation performance correlations among economics, energy use and carbon dioxide emission in Taiwan. Energy Policy. **35** (3), 1948–1955 (2007).

Ядерна сіра модель з генетичним алгоритмом оптимізації та її застосування в прогнозуванні ціни на пальмову олію в Малайзії

Шабрі А.

Кафедра математики, Факультет природничих наук,
Університет технологій Малайзії,
81310, Джохор, Малайзія

Точне прогнозування є складним, оскільки ціни на пальмову олію постійно дуже нелінійні. Важливо вибрати правильні моделі прогнозування, оскільки їх існує кілька. Сіра модель зарекомендувала себе як хороша модель прогнозування; проте більшість існуючих сірих моделей є фундаментально лінійними, що обмежує їхню здатність фіксувати нелінійні тенденції. У цій статті представлено нелінійну розширену параметричну сіру модель, відому як ядерна сіра модель (KGM). Однак прогнозування моделі KGM залежить від функції ядра та параметрів KGM. У цьому дослідженні запропоновано вдосконалену модель KGM на основі генетичного алгоритму та перевірено її на основі даних про ціни на пальмову олію в Малайзії з 2000 по 2019 рік. Для порівняння були обрані моделі прогнозування багатофакторної лінійної регресії (MLR) та машина опорних векторів (SVM) на основі середньої абсолютної відсоткової похибки та середньоквадратичної відсоткової похибки. Результати показують, що KGM перевершує дві інші моделі за продуктивністю навчальних та тестових даних, і це може значно підвищити точність прогнозування.

Ключові слова: *багатофакторна лінійна регресія; ядерна сіра модель; машина опорних векторів; пальмова олія.*