Original Research Article

A Hybrid Model Based on Grey Wolf Optimizer and Lagrangian Support Vector Regression for European Natural Gas Consumption Forecasting

Abstract

With the intensification of the Russia-Ukraine war, there is a large-scale energy shortage in Europe, and the natural gas supply in Europe has a natural gas crisis due to the cut-off of the Nord Stream No.1 pipeline. Therefore, it is necessary to accurately predict the consumption of natural gas. In order to fulfill this requirement, this paper uses the Lagrangian Support Vector Regression model with $Sorensen$ kernel based on the Nonlinear Auto-Regressive model and Grey Wolf Optimizer for 5-step forecasting of monthly natural gas consumption in all European countries. Under three time lags, comparing the 5-step predict results of $GWO\text{-}LSVR$ with SVR, RF, $LightGBM$, $XGBoost$, and MLP, those five models' hyperparameters also optimized by GWO, it found that $GWO\text{-}LSVR$ has smallest $MAPE$ in almost all cases, and the numerical results of $MAPE$ generated by $GWO\text{-}LSVR$ is from 5.844% to 11.622%, the smaller the forecasting step size, the better the effect. Moreover, compares the difference of GWO and WOA, it is found that GWO can obtained better model hyperparameters and smaller $MAPE$ results. To sum up, the proposed $GWO\text{-}LSVR$ model has strong generalization performance and robustness, and is a reliable natural gas consumption prediction model.

Keywords: Lagrangian Support Vector Regression; Grey Wolf Optimizer; Nonlinear Auto-Regressive; Kernel Learning; Natural Gas Consumption in Europe.

1. Introduction

As a clean and efficient low-carbon energy, natural gas is a very important part of the global energy structure, accounting for about 25% of European energy consumption. Coupled with the intensification of the situation in Russia and Ukraine, there is a serious shortage of natural gas supply in Europe. Therefore, it is necessary to accurately predict the consumption of natural gas [1].

Forecasting is the basis for understanding and decision-making, and natural gas consumption forecasting is based on dynamic analysis of dynamic natural gas data models and quantitative calculations on a qualitative basis. The analysis of natural gas consumption data is a multidimensional time series analysis, which often implies a large number of dynamic characteristics, and the mechanism of each factor cannot be described quantitatively, so the forecasting of natural gas consumption is a complex non-linear problem,

and the accuracy of multidimensional time series forecasting is still a great challenge, and the development of high-precision multidimensional time series forecasting analysis methods is of great significance [2]. Box and Jenkins and Hannan were the first to propose the classical linear multidimensional time series analysis model, the controlled autoregressive integrating moving average ($CARMA$) [3][4], Boker and Keviczky gave their fixed-order F-test discriminant method [5]. Deng Zili and Guo Yixin further developed the F-test discriminant for order, suborder and time lag, forming a complete identification of the structure of the $CARMA$ model and giving its simplified form, the controlled autoregressive (CAR) model with controlled terms [6]. and artificial neural networks (ANN) with good nonlinear approximation capability [7]. However, the selection of metrics in these traditional forecasting methods is arbitrary and the requirements for parameter selection and correction are high. Vapnik's support vector machine (SVM) [8], which is based on statistical learning theory, is the fastest growing machine learning method and was originally used for pattern recognition (SVC), with the introduction of the ε-insensitive loss function, it has been extended to be used for nonlinear time series analysis or nonlinear regression analysis (SVR) [9][10][11]; SVR is based on structural risk minimisation, which better solves the problems of small samples, nonlinearity, overfitting, dimensional disaster and office minimum, etc., with excellent generalisation and propagation ability In addition, it well solves the problems faced by the above methods, and has good performance in practical applications. However, SVR still has some limitations in practical applications. First, the selection of its kernel functions is empirical. Second, the selection of the corresponding parameters of some SVR programs and their kernel functions is also empirical [12]. Third, like $ANNs$, SVR models are less interpretable [13]. The limitations of SVR are overcome, or partially overcome, by dynamic selection of kernel functions [14][15] and forced screening of variables and giving the relative order of importance of variables in the order of elimination to give partial interpretability to the SVR. Thus, a non-linear multi-dimensional time series forecasting method based on $LSVR$, incorporating time series analysis and regression analysis, is developed for forecasting natural gas consumption in European countries, and the accuracy of the forecast can be guaranteed.

2. Design of the Forecasting Model

In this section, the detailed mathematical model of the $LSVR$ (Lagrangian Support Vector Regression) with $Sorensen$ kernel and the GWO (Grey Wolf Optimizer) used in this paper will be presented in **Section 2.1** and **Section 2.2,** respectively. And the complete multi-step forecasting model based on NAR (Nonlinear Auto-Regressive) model as out-of-sample holdout validation will be shown in **Section 2.3**.

2.1 Lagrangian Support Vector Regression with $Sorensen$ kernel

The standard SVR formulation is a constrained, quadratic optimization problem, written in matrix form is as follows:

$$\min \frac{1}{2} w^t w + \upsilon(e^t \xi + e^t \xi^*) \quad s.t. \begin{cases} y - Aw - be \le \varepsilon e + \xi \\ Aw + be - y \le \varepsilon e + \xi^* \end{cases} \tag{1}$$

where $\xi_i, \xi_i^* \ge 0 \; for \; i = 1, 2, \dots, m$, ξ and ξ^* are the slack variables, t represents the transpose of the matrix, matrix $A \in R^{m \times n}$, $LSVR$ has made two changes on the basis of SVR:

(1) Change ξ and ξ^* in 1-norm to be the square of 2-norm, which makes make it unnecessary for the slack variable to be greater than 0.
(2) Add b^2 to $w^t w$ in **Eq(1)**.

Thus the $LSVR$ can be formulated as the following form:

$$\min \frac{1}{2}(w^t w + b^2) + \frac{\upsilon}{2}\sum_{i=1}^{m}(\xi_i^2 + \xi_i^{*2}) \quad s.t. \begin{cases} y_i - A_i w - b \le \varepsilon + \xi_i \\ A_i w + b - y_i \le \varepsilon + \xi_i^* \end{cases} \tag{2}$$

where ε and υ are the input parameters.

To solve the convex quadratic problem above, introducing two Lagrange multiplies $u_1 = (u_{11},\ u_{12}, \dots, u_{1m})^t$ and $u_2 = (u_{21},\ u_{22}, \dots, u_{2m})^t$, the Lagrangian Function L can be obtained as follows:

$$\begin{aligned} L = {} & \frac{1}{2}(w^t w + b^2) + \frac{\upsilon}{2}\sum_{i=1}^{m}(\xi_i^2 + \xi_i^{*2}) \\ & + \sum_{i=1}^{m} u_{1i}(y_i - A_i w - b - \varepsilon - \xi_i) \\ & + \sum_{i=1}^{m} u_{2i}(A_i w + b - y_i - \varepsilon - \xi_i^*) \end{aligned} \tag{3}$$

The optimality condition is that the partial derivative of L with respect to the original variable is 0, we can obtained the solution of **Eq(3)**:

$$w = A^t(u_1 - u_2) and\ b = e^t(u_1 - u_2) \tag{4}$$

and the dual problem can be written as the minimization problem:

$$\min \frac{1}{2}[(u_1 - u_2)^t(A^t A + ee^t)(u_1 - u_2)] + \frac{1}{2\upsilon}(u_1^t u_1 + u_2^t u_2) - y^t(u_1 - u_2) + \varepsilon e^t(u_1 + u_2) \tag{5}$$

The linear $LSVR$ is the method to output a approximate function $f(\cdot)$, based on the **Eq(4),** the linear regression estimation function is given as:

$$f(x) = [x \quad 1]\begin{bmatrix} A^t \\ e^t \end{bmatrix}(u_1 - u_2) \tag{6}$$

Define an augmented matrix $D = [A \quad e]$, the **Eq(5)** can be equally expressed as:

$$\min \frac{1}{2}[u_1^t \quad u_2^t] M \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} - p^t \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \tag{7}$$

where

$$M = \begin{bmatrix} \frac{1}{\upsilon} + DD^t & -DD^t \\ -DD^t & \frac{1}{\upsilon} + DD^t \end{bmatrix} \tag{8}$$

and

$$p = \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} = \begin{bmatrix} y - \varepsilon e \\ -y - \varepsilon e \end{bmatrix} \tag{9}$$

The linear $LSVR$ in **Eq(7)** can be extend to nonlinear model with kernel matrix K. In this paper, we used the Sorensen kernel which expressed as the following form:

$$K(u, v) = \frac{2u \cdot v}{\|u\|_2^2 + \|v\|_2^2} \tag{10}$$

where $u \cdot v$ is a inner product of the two vectors.
Replacing DD^t by $K = K(D, D^t)$ in **Eq(7)**, for any $x \in R^n$, the kernel regression estimation function $f(\cdot)$ is obtained to be of the following form:

$$f(x) = K([x^t \quad 1], D^t)(u_1 - u_2) \tag{11}$$

2.2 Grey Wolf Optimizer

The GWO algorithm is inspired the unique hunting and hierarchy behavior of the grey wolves. Grey wolves have a very strict social dominant hierarchy, in order to mathematically model the hierarchy in GWO, the best solution is α, the second and the third best solutions are β and δ, and the rest of the candidate solutions called ω. The hunting behavior is divided into two stages: Encircling and hunting for prey. The specific mathematical modeling steps for the two stages are described below.

Encircling: The Grey wolves encircle the prey first when they hunt. The encircle behavior modeled as follows:

$$\vec{\boldsymbol{D}} = \left|\vec{\boldsymbol{C}} \cdot \vec{\boldsymbol{X}}_{\boldsymbol{p}}(\boldsymbol{l}) - \vec{\boldsymbol{X}}(\boldsymbol{l})\right| \tag{12}$$

$$\vec{\boldsymbol{X}}(\boldsymbol{l}+\boldsymbol{1}) = \vec{\boldsymbol{X}}_{\boldsymbol{p}}(\boldsymbol{l}) - \vec{\boldsymbol{A}} \cdot \vec{\boldsymbol{D}} \tag{13}$$

where l represents the current iteration, $\vec{X}_p$ is the position of the prey, $\vec{X}$ is the position of a grey wolf, and $\vec{A}$ and $\vec{C}$ are coefficient vectors, which calculated as follows:

$$\vec{A} = 2\vec{a} \cdot \vec{r}_1 - \vec{a} \tag{14}$$

$$\vec{\boldsymbol{C}} = \boldsymbol{2} \cdot \vec{\boldsymbol{r}}_{\boldsymbol{2}} \tag{15}$$

where $\vec{r}_1, \vec{r}_2$ are random vectors in $[0, 1]$, the components of $\vec{a}$ are linearly decreased from 2 to 0 during the course of iterations.

Hunting: The hunt of the prey always guided by the α wolf, sometimes, the β and δ wolf participate in the hunt. Assume that the α, β and δ have better knowledge of the prey in the abstract space. Therefore, save the best three solutions obtained so far, and update other search agents' position (including ω) according to the position of the best search agents. This progress is modeled as follows:

$$\vec{X}_{1,2,3} = \vec{X}_{\alpha,\beta,\delta} - \vec{A}_{1,2,3} \cdot \vec{D}_{\alpha,\beta,\delta} \tag{16}$$

$$\vec{X}(l+1) = \frac{\vec{X}_1 + \vec{X}_2 + \vec{X}_3}{3} \tag{17}$$

where $\vec{X}_{1,2,3}$ represents $\vec{X}_1$, $\vec{X}_2$ and $\vec{X}_3$, $\vec{X}_{\alpha,\beta,\delta}$ represents $\vec{X}_\alpha$, $\vec{X}_\beta$ and $\vec{X}_\delta$.

2.3 Complete multi-step forecasting strategy based on NAR model

In this section, the NAR model which can transform the original time series to supervised learning dataset will be given in **Section 2.3.1**, the out-of-sample holdout validation scheme presented in **Section 2.3.2**, and the complete algorithm flow will be shown in **Section 2.3.3**.

2.3.1 Nonlinear Auto-Regressive model for multi-step forecasting

A machine learning model is a class of models with high-dimensional inputs, Preferably without using raw time series data as input. The Nonlinear Auto-Regressive (NAR) model is a type of model that reconstructs the original data set based on phase space reconstruction and lag methods. Given a Univariate times series data $U = \{u_1, u_2, u_3, \dots, u_n\}$, phase space reconstruction of U yields a new dataset Φ in **Eq(18)** for supervised learning, the new dataset (matrix: Φ) can be expressed as follows:

$$\Phi = \begin{pmatrix} u_1 & u_2 & \dots & u_\tau & u_{\tau+1} \\ u_2 & u_3 & \dots & u_{\tau+1} & u_{\tau+2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ u_{n-\tau} & u_{n-\tau+1} & \dots & u_{n-1} & u_n \end{pmatrix} \tag{18}$$

Where τ is the time lag.

2.3.2 out-of-sample holdout validation

In this paper, we use the out-of-sample holdout validation to validate the best hyperparameters of the model. Different with conventional *k-fold* cross-validation in machine learning models, the out-of-sample holdout validation requires only one validation on the validation set, *k-fold* cross-validation will disrupt the order of the data set when performing multiple cross-validation, but there is a certain irrationality in the validation of the time series model. Because it is unreasonable to validate past data with future data.

Split the reconstructed data set according to the ratio of about *8:1:1* according to the sequence, the split datasets are training, validation, and test sets, respectively. First, the default hyperparameters of $LSVR$ are used to train the model on the training set, then GWO will optimize the model hyperparameters on the validation set, and the final multi-step forecasting process will be performed on the test set.

$$MAPE = min\frac{1}{n}\sum\left|\frac{u_j - \widehat{u_j}}{u_j}\right| \times 100\% \tag{19}$$

Throughout the process, we use $MAPE$ which presents in **Eq(19)** as an indicator to evaluate model performance. The smaller the $MAPE$, the better the model performance.

2.3.3 Complete multi-step forecasting model

The completed multi-step forecasting model will be shown in this part. The detailed model flow chart is shown in **Figure 1**, and the specific model prediction process is divided into the following five steps:

(1) Phase space reconstruction of the original time series U into a supervised learning dataset using NAR model.
(2) Split the dataset in a ratio of about $8:1:1$.
(3) Train the $LSVR$ model on the training set with default hyperparameters.
(4) With the minimum $MAPE$ as the goal, use GWO for optimization on the validation set.
(5) Multi-step prediction on test set.

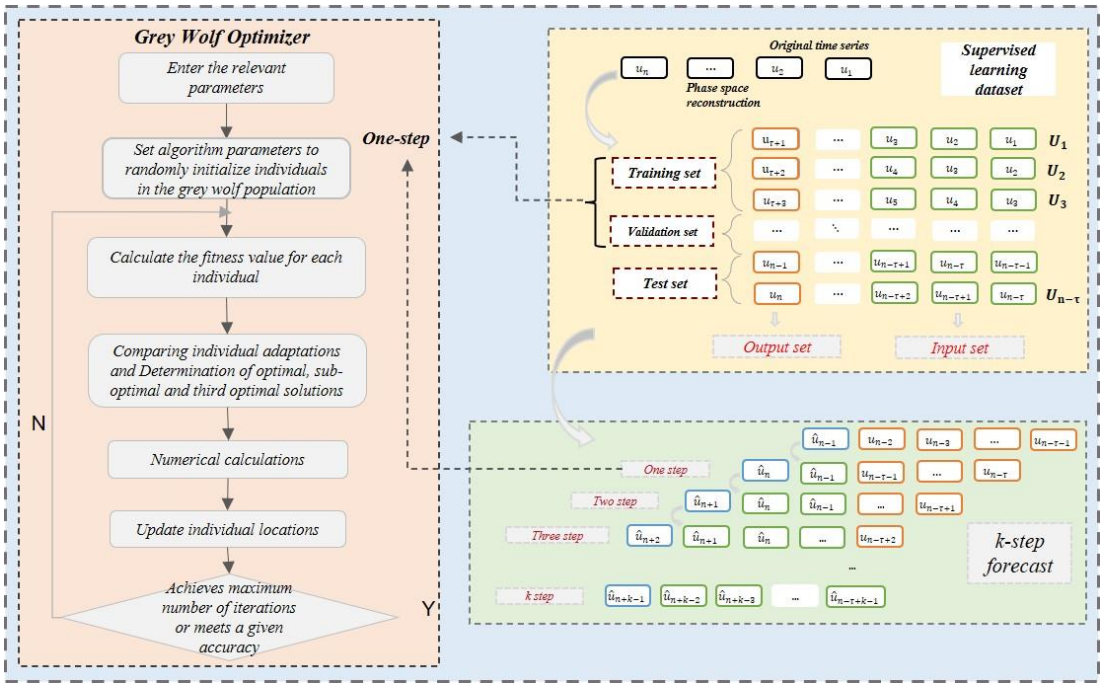

Figure 1 Complete algorithm flow

3. Dataset Description

In this paper, the dataset we used is from the publicly available European natural gas consumption (NGC) dataset in the Eurostat(https://ec.europa.eu/eurostat), which collects monthly NGC data from Jan 2014 to May 2022 for a total of 100 months. Total monthly natural gas consumption for a total of 27 countries within Europe. Use the first eighty points to train the LSVR model, use the next ten points to find the optimal hyperparameters of the model, and use the last ten points for multi-step forecasting. The original dataset is shown in **Figure 2**.

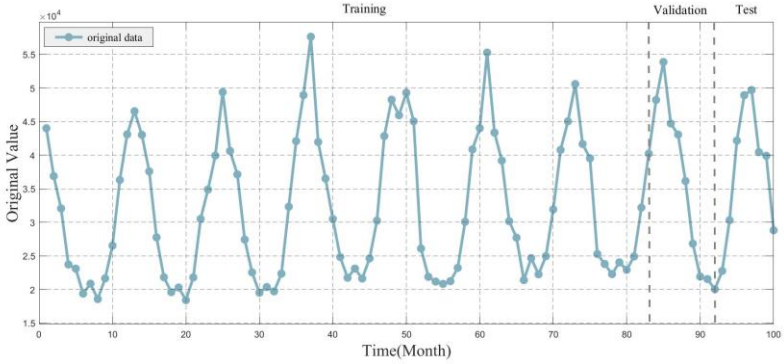

Figure 2 Original dataset

4. Multi-Step Forecasting Results and Discussion

In this section, we compared $GWO\text{-}LSVR$ with five machine learning models with high generalization performance at three different time lags. These five models include SVR similar to $LSVR$, tree models Random Forest (RF), Light Gradient Boosting Machine ($LightGBM$), and Extreme Gradient Boosting ($XGBoost$), neural network model Multilayer Perceptron (MLP). The modeling and optimization process of these five models is the same as $GWO\text{-}LSVR$. The detailed comparison results are shown in **Section 4.1**. The impact of different optimization algorithms on the results is discussed in **Section 4.2**.

4.1 Analysis of the multi-step forecasting results

In order to quantitatively analyze the performance of the model, the $MAPE$ of multi-step forecasting is used as the evaluation standard. In all experiments, for a more comprehensive comparison of these models, three different lags (that is, τ=3, τ=4, τ=5) were chosen. In the multi-step forecasting process, we predict 5 steps forward under the three time lags, and calculated the MAPE of each step.

Applying the proposed model to the forecast of monthly NGC for all countries on a European scale. The detailed MAPE results shown in **Table 1**. It can be plainly seen that from the table, the proposed $GWO\text{-}LSVR$ model yields the minimal $MAPE$ almost all cases. The numerical result of its $MAPE$ is from 5.844% to 11.622%. Only when τ=4, $GWO\text{-}XGBoost$ has better $MAPE$ than $GWO\text{-}LSVR$ at the fourth step. But in this case, the results of the proposed model are better than the other four models. Detailed MAPE results also shown in **Figure 3**.

Table 1 $MAPE$(%) of the forecasting models

	Steps	$LSVR$	SVR	RF	$LightGBM$	$XGBoost$	MLP
$\tau = 3$	step1	***5.844***	6.966	7.728	7.444	8.913	15.372
	step2	***9.504***	10.193	12.301	10.167	10.637	25.778
	step3	***9.138***	14.148	11.190	13.419	12.804	45.638
	step4	***11.199***	15.376	11.424	17.868	12.938	56.474
	step5	***11.418***	14.914	15.193	17.450	14.291	64.856
$\tau = 4$	step1	***6.368***	7.264	7.548	8.286	6.628	10.296
	step2	***8.069***	9.955	9.414	12.249	8.892	14.273
	step3	***9.632***	12.218	10.142	16.000	10.217	24.397
	step4	11.622	15.345	11.927	19.467	***10.870***	31.967
	step5	***10.932***	14.487	13.120	21.004	11.538	41.182
$\tau = 5$	step1	***6.114***	28.309	6.823	6.799	7.746	7.746
	step2	***7.874***	28.077	9.863	9.729	9.403	9.403
	step3	***10.058***	26.557	10.972	11.534	12.287	12.287
	step4	***11.532***	27.004	12.557	12.585	14.067	14.067
	step5	***11.118***	30.294	13.497	11.588	15.163	15.163

Figure 4 is when τ=5, the output value of the machine learning model after data reconstruction, a total of 95 points, the comparison between the predicted value of the 6 models and the original value. It can be seen that the effect of SVR is a bit worse, and there is a phenomenon of underfitting in the training set, lead to poor prediction results in the test set. MLP has a certain overfitting phenomenon, over-learning the information of the data set,

resulting in very complicated results in the training set. The best performance on the training set is GWO-$XGBoost$, but his prediction results are slightly worse than GWO-$LSVR$.

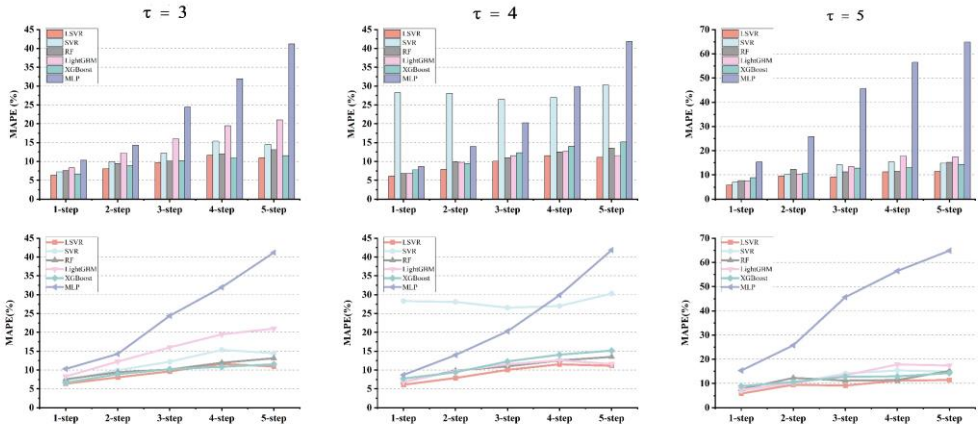

Figure 3 $MAPE$ results for a 5-step forecast with 3 time lags

To sum up, the proposed GWO-$LSVR$ hybrid model can get very good forecasting results whether it is the training set or the test set, and the obtained $MAPE$ is also the best among all comparison models, with strong generalization performance and robust.

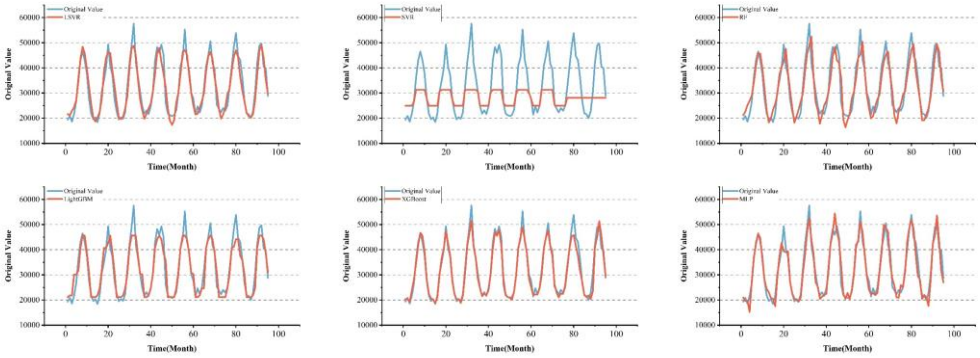

Figure 4 Predict data when τ=5

4.2 Discussion

In this section, we discuss the impact of different swarm intelligence optimization algorithms on the performance of $LSVR$ models. Similar to GWO, WOA is also widely used to solve the problems of complex nonlinear systems. The following are the specific discussion results.

Under three time lags, the 5-step forecasting is performed respectively, and the detailed $MAPE$ results generated by GWO-$LSVR$ and WOA-$LSVR$ are listed in **Table 2**. Under the combination of three time lags and five forecasting steps (15 cases in total), WOA produces only two results that are slightly better than GWO.

Both WOA and GWO can produce relatively small $MAPE$ results in the first and second steps, but from the third step onwards, the results produced by WOA are larger than those of GWO, and the gap gradually increases. The smallest difference is τ=4 at the first step, which is only

0.187%, and the largest difference is τ=4 at the fifth step, with a difference of 2.451%. The detailed results presented in **Figure 5**.

Table 2 $MAPE$ results of GWO-$LSVR$ and WOA-$LSVR$

optimizer	Lag	step1	step2	step3	step4	step5
GWO	$\tau = 3$	***5.844***	9.504	***9.138***	***11.199***	***11.418***
	$\tau = 4$	6.368	***8.069***	***9.631***	***11.622***	***10.932***
	$\tau = 5$	***6.114***	***7.874***	***10.058***	***11.532***	***11.118***
WOA	$\tau = 3$	6.661	***9.157***	10.433	13.576	13.261
	$\tau = 4$	***6.181***	8.366	11.279	13.776	13.383
	$\tau = 5$	6.317	9.350	11.970	13.816	12.614

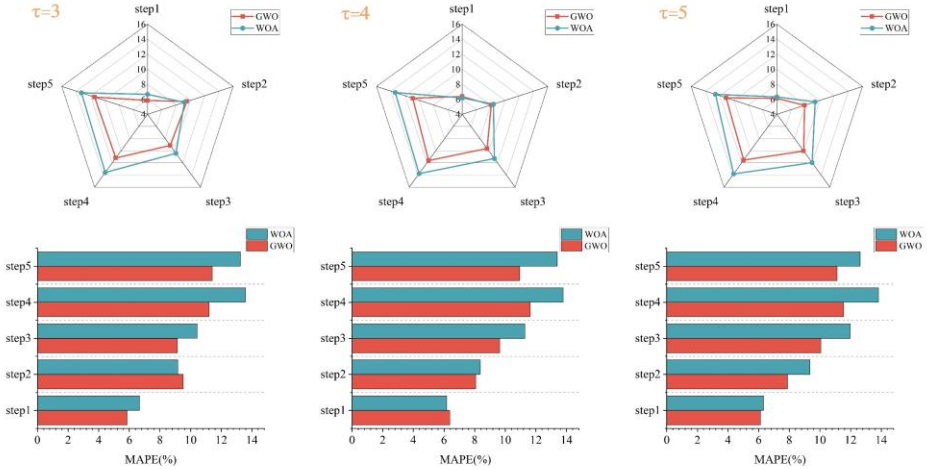

Figure 5 $MAPE$ results yield by GWO-$LSVR$ and WOA-$LSVR$

5. Conclusion

Under the influence of the Russia-Ukraine war, there are energy tensions in Europe, especially the shortage of natural gas. Accurate forecasting of natural gas consumption is necessary.

This paper uses the Lagrangian Support Vector Regression model with the $Sorensen$ kernel, combined with the Grey Wolf Optimizer and Nonlinear Auto-Regressive model, for multi-step forecasting of monthly natural gas consumption, based on the out-of-sample holdout validation. The proposed model was applied to forecast the total monthly natural gas consumption of 27 countries within Europe. Comparing the model with SVR, RF, $LightGBM$, $XGBoost$, and MLP's five combined models based on GWO. It's find that the GWO-$LSVR$ has the best generalization performance and most robust of all hybrid models. The $MAPE$ yield by GWO-$LSVR$ of each step at three time lags are from 5.844% to 11.622%. It also discussed the difference of the GWO and WOA, and find that GWO can better optimize the model hyperparameters in most cases. It can be concluded that the proposed GWO-$LSVR$ model can be used to accurately predict natural gas consumption, and has strong generalization performance and robustness.

REFERENCES

[1] K. Tang, H. Li, and Z. Qian, “Forecasting Study of Natural Gas Consumption by Combined Models Based on LASSO and WOA”, ARJOM, vol. 19, no. 1, pp. 23-31, Jan. 2023.

[2] Yang H-M, Pan C-S, Bai W. A review of time series forecasting methods[J]. Computer Science, 2019, 46(1): 21-28.

[3] Box Q E P, Jenkins G M. Time series analysis: forecasting and control［M］. San Francisco: Holden-day Press,1970.

[4] Hannan E J. The estimation of the order of an ARMA process［J］. Ann. Statist., 1980,8(5): 1071-1081.

[5] Boker J, Keviczky L. Structural properties and structure estimation of vector difference equations［J］. Int. J. Control, 1982, 36(3): 461-476.

[6] Deng Zili, Guo Xinxin. Dynamic system analysis and its applications [M]. Shenyang: Liaoning Science and Technology Press, 1985, 31-130.

[7] Wang Xiaodong, Huang Kun. A neural network approach to natural gas consumption forecasting[J]. Natural Gas Exploration and Development, 2007, 30(3): 70-72.

[8] V Vapnik. The nature of statistical learning theory［M］. New York: Springer Verlag Press, 1995.

[9] Ping-Feng Pai, Wei-Chiang Hong. Support vector machines with simulated annealing algorithms in electricity load forecasting［J］. Energy Conversion and Management, 2005, 46: 2669-2688.

[10] Ping-Feng Pai, Chih-Sheng Lin. A hybrid ARIMA and support vector machines model in stock price forecasting［J］. Omega, 2005, 33(6):497-505.

[11] J L Rojo-Álvarez, G Camps-Valls, M Martínez-Ramón, et al. Support vector machines framework for linear signal processing［J］. Signal Processing, 2005, (85):2316-2326.

[12] Mei Hu, Liang Guizhao, Zhou Yuan, Li Zhiliang. A study of support vector machines for quantitative constitutive relationship modeling [J]. Science Bulletin, 2005, 50(16):1703-1708.

[13] Cheng Jun-Sheng, Yu De-Suke, Yang Y. A support vector regression machine based approach to the Hilbert Huang transformation endpoint effect problem [J]. Journal of Mechanical Engineering, 2006, 42(4): 23-31.

[14] Wang, T. H., Chen, J. T.. A review of research on the selection of kernel functions[J]. Computer Engineering and Design, 2012, 33(3): 1181-1186.

[15] Feng, K.-H. Comparison of SVM classification kernel functions and parameter selection[J]. Computer Engineering and Applications, 2011, 47(3): 123-124.