COMPARE THE OPTIMAL ALGORITHMS FOR WAVENET APPLICATIONS IN LOAD FORECASTING

Duong Ngoc Hung[1-3], Nguyen Tung Linh[2*], Nguyen Minh Tam[3]
[1]Tien Giang University, [2]Electric Power University, [3]HCM University of Technology and Education

ARTICLE INFO

Received: **21/11/2022**
Revised: **11/4/2023**
Published: **13/4/2023**

KEYWORDS

Harris hawks optimization
Microgrid
HHO algorithm
Wavenet
Short-term load forecasting

ABSTRACT

Power load forecasting is an important issue in microgrid energy management. Accurate load forecasting is urgently required for effective power management for microgrid. This paper considers the evaluation of the effectiveness of applying different optimization algorithms to the proposed Deep Learning Neural Network, which is Wavenet. Models combining optimization algorithms with Wavenet are applied for short-term load forecasting. In order to evaluate the accuracy of the predictive models, this study used optimization algorithms (HHO, Adam, RMSprop, SGD, Adagrad) to calculate the Wavenet network. To perform calculations for the model, we work with the load data set of a microgrid model belonging to the Ho Chi Minh City power grid. The results show that our HHO model outperforms the model based on other optimization algorithms in terms of Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE).

SO SÁNH THUẬT TOÁN TỐI ƯU CỦA MẠNG WAVENET TRONG BÀI TOÁN DỰ BÁO PHỤ TẢI ĐIỆN

Dương Ngọc Hùng[1-3], Nguyễn Tùng Linh[2*], Nguyễn Minh Tâm[3]
[1]Trường Đại học Tiền Giang, [2]Trường Đại học Điện lực, [3]Trường Đại học Sư phạm kỹ thuật TP Hồ Chí Minh

THÔNG TIN BÀI BÁO

Ngày nhận bài: **21/11/2022**
Ngày hoàn thiện: **11/4/2023**
Ngày đăng: **13/4/2023**

TỪ KHÓA

Tối ưu Harris hawrk
Lưới điện nhỏ
Thuật toán HHO
Mạng Wavenet
Dự báo ngắn hạn

TÓM TẮT

Dự báo phụ tải điện là một vấn đề quan trọng trong quản lý lưới điện nhỏ. Dự báo phụ tải chính xác là yêu cầu cấp thiết để quản lý điện hiệu quả cho lưới điện nhỏ. Bài báo này xem xét việc đánh giá hiệu quả của việc áp dụng các thuật toán tối ưu hóa khác nhau sử dụng mạng noron học sâu để giải quyết đó là mạng Wavenet. Các mô hình kết hợp các thuật toán tối ưu hóa với Wavenet được áp dụng cho dự báo phụ tải ngắn hạn. Để đánh giá độ chính xác của các mô hình dự báo, nghiên cứu này đã sử dụng các thuật toán tối ưu hóa (HHO, Adam, RMSprop, SGD, Adagrad) để tính toán mạng Wavenet. Để thực hiện tính toán cho mô hình, phương pháp thực hiện với tập dữ liệu phụ tải của mô hình lưới điện nhỏ thuộc lưới điện Thành phố Hồ Chí Minh. Kết quả cho thấy mô hình HHO hoạt động tốt hơn mô hình dựa trên các thuật toán tối ưu hóa khác về lỗi bình phương trung bình gốc (RMSE) và lỗi tỷ lệ phần trăm tuyệt đối trung bình (MAPE).

DOI: https://doi.org/10.34238/tnu-jst.6956

[*] Corresponding author. *Email: ntlinh1505@gmail.com*

1. Introduction

In the current development trend, technology increasingly occupies a large part of life. And the demand for electricity is increasing rapidly, along with the development of microgrid (MG) [1]. MG models as a small-scale grid, advanced techniques and tools are devised for optimal energy operation [2]. The importance of consumer load forecasting is increasingly emphasized. The problem of short-term load forecasting (STLF) is considered complicated compared to other problems. The accurate short-term forecast results will support the efficient and convenient operation and exploitation of the power system in the area. If the prediction indicates that the storage capacity is insufficient to support future loads, the utility can notify the user of the situation. This ultimately causes them to reduce their electricity usage, as users not only want to pay more for conventional energy, but also want incentives from the authorities.

Due to the superiority of deep learning, this study considers a method proposed in [3], namely a hybrid approach for short-term forecasting of load demand in a typical microgrid (MG). Figure 1 is a combination of a static wavelet packet transform and a feedforward neural network based on the Harris hawks optimization algorithm (HHO). Harris hawks optimization is applied to a feedforward neural network as an alternative training algorithm to optimize the weights and basis of neurons. Considering another approach in the study [4], WaveNet uses dilated causal convolutions and skip-connection to utilise long-term information. This novel type of ML architecture presents various advantages relative to other statistical algorithms.

In addition, many forecasting methods have been proposed by researchers to solve the problem of load forecasting. These approaches are classified as statistical, persistence, machine learning, and hybrid approaches [5]. In the study [6], a multiple linear regression model was applied to predict the hourly load demand. The trial and error method is used by the authors to determine the appropriate structure of the proposed model. A regression-based moving window with an adjustable window size-based method has been presented [7]. The proposed method is compared with a back-propagation neural network (BPNN) to show the effectiveness of the regression-based moving window strategy. Similarly, load forecasting was performed by the authors in [8] for MG using the persistence method. In the study [9], [10], a model based on the Kalman filter was proposed to forecast the short-term load demand of the household.

Based on the references, the machine learning and hybrid methods have some disadvantages such as difficulty in parameter selection and unclear choice of input variables. Therefore, this paper considers implementing a model using Wavenet with Hawks Harris optimization (HHO), and compares it with a number of Wavenet models incorporating other optimization algorithms (including: Adam, RMSprop, SGD and Adagrad), to demonstrate the effectiveness of the technique. The rest of this article is organized as follows. Part 2 provides a detailed description of HHO and Wavenet based on HHO, integrated as an optimization function for the Tensorflow library. Numerical results, images are presented in section 3 and conclusions are provided in the final section.

2. The proposed method

2.1. The Wavenet

Residual model instead of just mapping input data x to a function H(x) output $\hat{y}$, mapping scenario from previous residual block $f\left(x,\{W_i\}\right)$ where W_i is the learned weight and deviation from residual block is considered. Therefore, the output of the residual block can be expressed as:

$$H\left(x\right)=f\left(x,\{W_i\}\right)+x \tag{1}$$

Furthermore, since we use stacked residuals, the output of residuals can be represented as:

$$x_K = x_0 + \sum_{i=1}^{K} f\left(x_{i-1}, W_{i-1}\right) \tag{2}$$

x_K is the output of the residual block K, x_0 is the input of the residual network and $f\left(x_{i-1}, W_{i-1}\right)$ is the output and associated weight of the previous residual blocks.

In addition, connection bypass and port activation are applied to the network to speed up convergence and avoid over-learning. The process of connecting redundancy and bypassing the trigger port is shown in Figure 2.

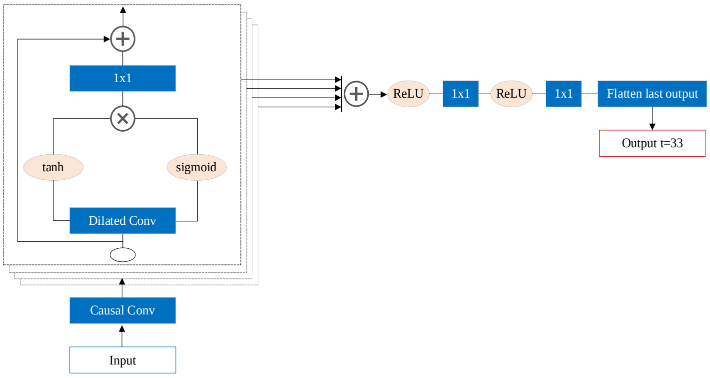

Figure 1. *Overview of the remaining convolution block and gated activation function*

The activation gates are inspired by the LSTM layer where tanh and sigmoid (σ) act as learned filter and learned gate, respectively. The use of controlled activations has been shown to perform better than the use of ReLU activations in time series data [4]. The output of the dilated convolution with rated activations can be expressed as:

$$z = tanh\left(\omega_f * x\right) \odot \sigma\left(\omega_g * x\right) \tag{3}$$

where w_f and w_g are the learned filter and learned gate, respectively.

2.2. Harris hawks optimization

Harris's Hawks Optimzation (HHO) is an optimization algorithm inspired by hawks' hunting methods in the wild, this is an algorithm in the metaheuristic branch of nature inspired (nature-based). This method is highly customizable due to its high customization ability, suitable for many reality problems (if we can make the input data suitable for the input form of the method), makes for a relatively high density of use after it was first introduced in 2019 by Heidari and colleagues in the prestigious Journal for Future Generations Computation. However, another method has also appeared under the name Harris hawks algorithm, an advanced variant of the gray wolf optimization (GWO) method, which adds the ability to optimize problems requiring ask multiple variables to be considered. In the framework of this paper, the method mentioned is only capable of optimizing one variable, with the underlying mathematical rules different from the above method. The falconry method is divided into three stages: exploration, move from exploration to exploitation, and exploit.

2.2.1. Exploration phase

In the exploration phase, the algorithm is inspired by the observations and decision-making of hawks. Where t is the current iteration, Yrand is a randomly selected hawks from the population, Yprey is the position of the prey, and UB and LB are the upper and lower bounds of the search space. Other side, r_1; r_2; r_3; r_4; and q are random numbers from 0 to 1. Y_m is the average position of the falcon, calculated in Eq (1):

$$Y_m(t) = \frac{1}{N}\sum_{i=1}^{N} Y_i\,(t) \tag{4}$$

Where, N is the total number of group members and Y_i represents the position of each hawk at each iteration respectively.

2.2.2. Phase of transition from exploration to exploitation

The second phase of HHO is the transition from exploration to exploitation.

During this phase, the prey's energy gradually decreases due to the process of running and chasing. The retained energy fraction of the prey is described as follows:

$$\mathrm{E} = 2\mathrm{E}_0 * \left(1 - \frac{\mathrm{t}}{\mathrm{T}}\right) \tag{5}$$

Where, E is the escape energy of the prey, E0 is determined from -1 to 1 (indicating the initial energy level of the prey at each iteration), and T is the maximum number of iterations, respectively. Note that the value of E defines the basis for the exploration-to- exploitation transition. When the value of tense the exploration phase occurs and when the value of tense the exploitation phase begins.

2.2.3. Exploitation phase

The third stage in HHO is called exploitation. During this stage, the Harris hawk makes a surprise pounce to capture the selected prey. There are four mounting patterns in this stage to capture selected prey. Here, let q be the probability of successful strike of the hawk, q will depend on the position of the selected hawk in each search relative to other hawks in the flock and on the position of the prey.

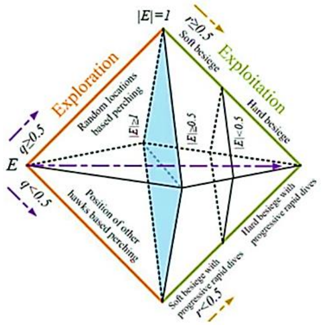

Figure 2. *General diagram of the HHO method [1]*

After identifying the target in the exploration phase, the hawks will perform a round of prey. Let r be the probability of the prey successfully escaping (r < 0.5) and unsuccessfully escaping (r >= 0.5) before the hawks' encirclement plan. In the wild, because their prey is always looking for a way out, the falcons will choose between a soft besiege and a hard besiege depending on the prey's E. Since the prey's energy stat is finite and decreases with each cycle, the hawks will increase in intensity in each siege through each corresponding cycle. We convention that when abs(E) >= 0.5, the hawks will do a light catch, and when abs(E) < 0.5, a hard catch will be done. After catching, the optimization value will be expressed as the position of the prey when the hawks perform a surprise pounce to finish. Due to the randomness of the method, all four cases corresponding to different levels of r and E are considered.

2.3. Other used optimization algorithm

- *Stochastic Gradient Descent (SGD)*

Figure 3 presents stochastic that is a variation of Gradient Descent. Instead of after each epoch we will update the weight (Weight) once, in each epoch with N data points we will update the weight N times. Looking at it on the one hand, SGD will reduce the speed of 1 epoch. However, looking at it the other way, SGD will converge very quickly after only a few epochs. The SGD formula is similar to GD but works on each data point.

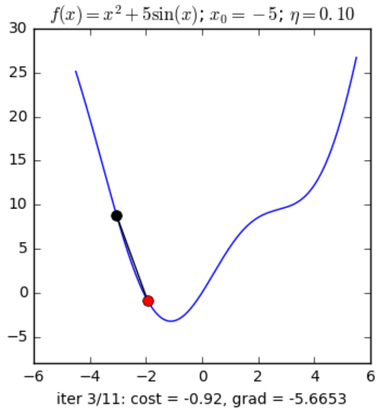

Figure 3. *Optimization algorithm simulation result - Stochastic Gradient Descent*

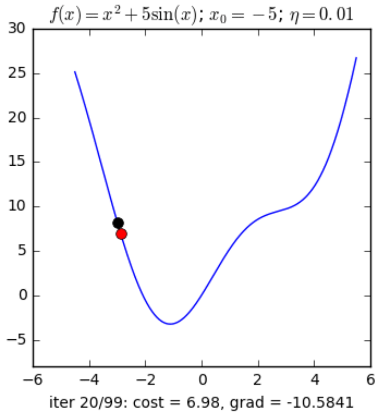

Figure 4. *Optimization algorithm simulation result - Adagrad*

Looking at the two pictures above, we see that SGD has a pretty zigzag path, not as smooth as GD. It's easy to understand because 1 data point cannot represent the whole data. Here, GD has a limitation for large databases (several million data), it becomes cumbersome to calculate the derivative over the entire data through each loop. Besides, GD is not suitable for online learning. Online learning, continuously updated data, each time we add data, we have to recalculate the derivative on the entire data, resulting in a long computation time, the algorithm is no longer online. So SGD was born to solve that problem, because each time new data is added, it only needs to be updated on that 1 data point, suitable for online learning.

- *Adagrad*

Figure 4 is unlike previous algorithms where the learning rate is almost the same during training (learning rate is constant), Adagrad considers the learning rate as a parameter. That is, Adagrad will let the learning rate change after every time t.

Where: n : constant gt : gradient at time t

ϵ : error avoidance coefficient (divided by sample equals 0)

G : is a diagonal matrix where each element on the diagonal (i,i) is the square of the path parameter vector function at time t.

- *RMSprop*

RMSprop solves the Adagrad descending learning rate problem by dividing the learning rate by the average of the square of the gradient. The most obvious advantage of RMSprop is that it solves Adagrad's slow learning rate problem (the problem of decreasing learning rate over time will cause training to slow down, possibly leading to freezing).

The RMSprop algorithm can give a solution that is only the local minimum, but not the global minimum like Momentum. Therefore, people will combine both Momentum algorithms with RMSprop to give an Adam optimal algorithm.

- *Adam*

As mentioned in Figure 5, Adam is a combination of Momentum and RMSprop. If explained in terms of physical phenomena, Momentum is like a ball going downhill, and Adam is like a very heavy ball with friction, so it easily crosses the local minimum to the global minimum and when it reaches the global minimum, it does not take a long time to oscillate back and forth around the target because it has friction so it is easier to stop.

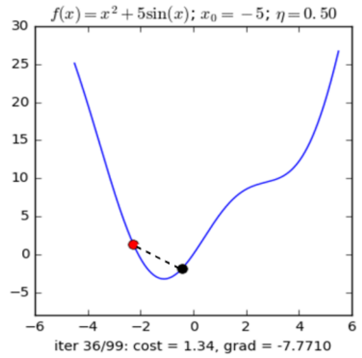

Figure 5. *Optimization algorithm simulation result - Adam*

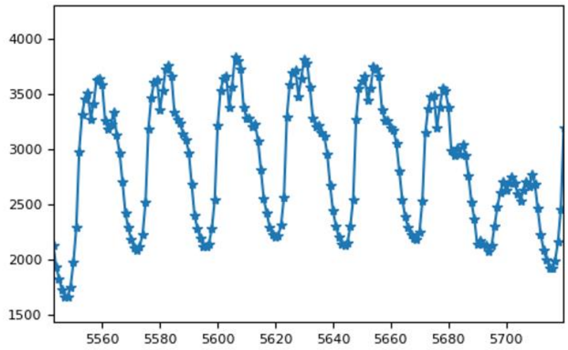

Figure 6. *Data of a power load area for 1 week in Ho Chi Minh City*

2.4. Harris hawks optimization with Wavenet

The input data is processed through the top section, where the information learned from the previous steps, in this case the hyperparameters, has been processed through the HHO algorithm. This information is passed through three main gates, the forget gate, the input port, and the output port, respectively, to determine the state of the loaded or cleared state of the hidden state. After configuring the structure of the Wavenet Network, the set of Wavenet weights will be adjusted by a training algorithm to minimize errors. Therefore, the optimization algorithms applied to train Wavenet achieve high accuracy and are compared with each other. In this work, the search agents are encoded as vectors in the interval [-1, 1].

Data set:

To perform simulations for the proposed method, the dataset used is from a power load area in Ho Chi Minh City (Figure 6).

Raw data before being included in the proposed method will be processed, including steps such as checking and replacing null values based on information about surrounding values, separating the dataset into training data and validation data as well as normalization of input data. The data used for the study were obtained from January 11, 2010 to December 23, of which data from January 11, 2010 to December 31, 2017 was used for training and the rest January 1, 2018 to December 23, 2018 were used for verification. In the data normalization process, the fundamental equation used is:

$$z_i = \frac{x_i - \min(x)}{\max(x) - \min(x)} \tag{17}$$

x = x1, ..., xn and zi are the i-th normalized data.

Proposed Model:

The problem of short-term load forecasting, through the number of previous studies, is said to be a complex problem. In this study, the proposed model is a Wavenet model with hyperparamters optimized by the HHO algorithm. In which, each wavenet neuron input is the optimal value from the data set that has been processed through the HHO algorithm. Inspired by the success of Santhadevi D. et al.'s model in dealing with continuous time data series during malware filtering [11], a similar model is used for load forecasting.

Figure 7 shows the HHO integration for optimized calculation for Wavenet network. The weight optimization parameters are calculated by the HHO algorithm for all layers of the Wavenet network. These parameters are included for the training network with 70% dataset and the testing network with 30% dataset.

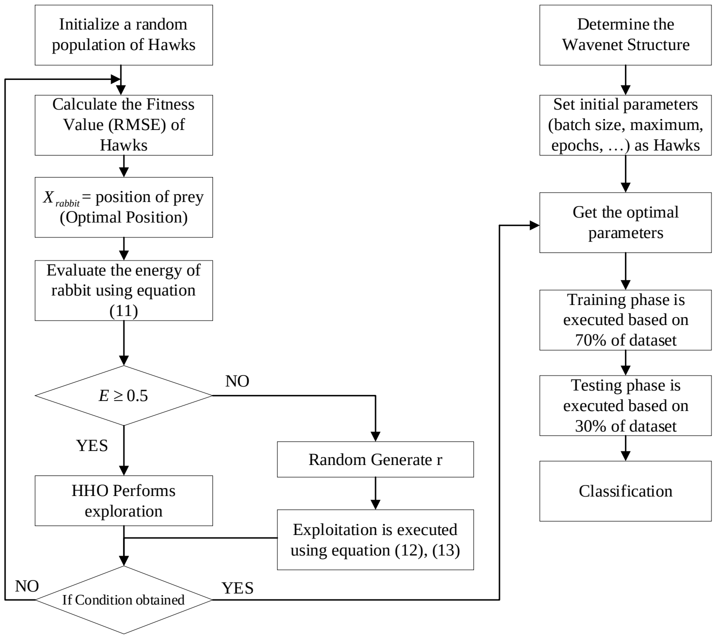

Figure 7. *Flowchart diagram of the HHO-Wavenet Algorithm*

3. Simulation results

3.1. Model performance rating

To evaluate the model performance, we compare this model with previous deep learning based models which performed very well in the case of STLF. Those models are referenced from Model 1 using the SGD-Wavenet network, Model 2 using Adagrad-Wavenet and Model 3 incorporating the Adam-Wavenet layer; Model 4 uses the RMSprop-Wavenet network. The configuration of each comparison model was identical to the published papers.

During the testing phase, all models were evaluated with three commonly used distinct metrics, the base mean square error (RMSE) and the mean absolute percentage error (MAPE). MAPE is identical to MAP but it uses the ratio of difference to actual load while RMSE is another metric that tends to have higher values than others. The higher the value is because of the indices, the worse the performance of the model.

3.2. Load forecasting

Table 1 shows the performance of the models mentioned above. The proposed model performs better than other models in most of the criteria, in which the RSME and MAPE coefficients clearly show the superiority of the proposed method. In which, the proposed model gives much lower figures than the compared methods (at least % compared to the next method).

Table 1. *Forecast results*

Model	RMSE	MAPE (%)
SGD-Wavenet	49.78	1.32
Adagrad-Wavenet	47.64	1.35
Adam-Wavenet	45.26	1.21
RMSprop-Wavenet	25.09	0.63
HHO-Wavenet	**16.57**	**0.51**

According to the results shown in Table 1, the forecasting model using Adam-Wavenet gives the highest error results. The Adagrad-Wavenet and SGD-Wavenet models give better MAPE errors of 1.69% and 0.58%, respectively. In the RMSprop-Wavenet model, the MAPE has dropped to 0.51. But the RMSprop-Wavenet is still pretty high with an RMSE of 95.93. The proposed HHO algorithm is optimized for Wavenet networks, the results are significantly improved with MAPE of 0.3% and RMSE of 56.13.

Furthermore, the forecast results are shown graphically as shown in Figures 8-12. Each figure shows actual data (blue) and forecast data (red). The proposed method (HHO-Wavenet) shows better accuracy when shown on the graph, the forecast results and the actual data almost coincide. Other methods have large errors, the graph shows the difference between two large data.

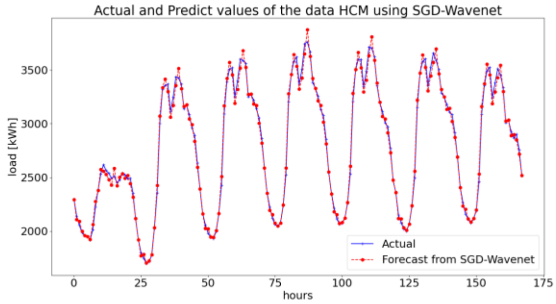

Figure 8. *Graph showing actual value with the proposed method of combining SGD-Wavenet network*

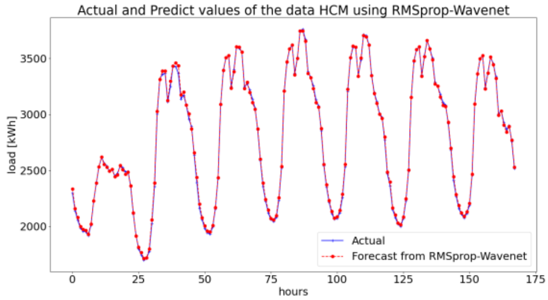

Figure 9. *Graph showing actual value with the proposed method of combining basic RMSprop-Wavenet*

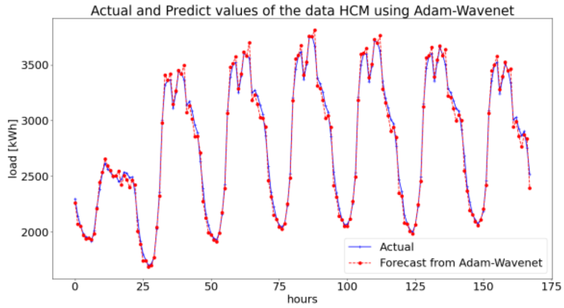

Figure 10. *Graph showing actual value with the proposed method of combining basic Adam-Wavenet*

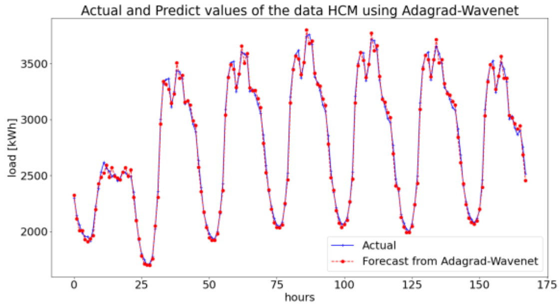

Figure 11. *Graph showing actual value with the proposed method of combining basic Adagrad - Wavenet*

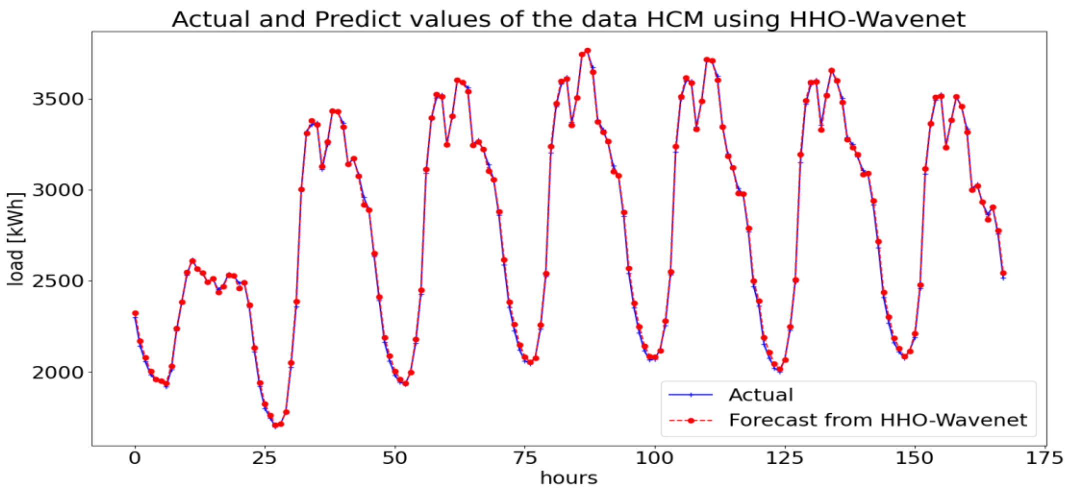

Figure 12. *Graph showing actual value with the proposed method of combining basic HHO-Wavenet*

4. Conclusion

It can be seen from the above results that when using the HHO algorithm as an optimizer for the Wavenet network, the network's performance has been significantly improved, with outstanding MAPE and RMSE parameters (table 1), compared to applying other optimization algorithms (including: Adam, RMSprop, SGD, Adagrad). However, a current drawback of the study is that the algorithm used is more computationally intensive and time consuming than the compared methods. One limitation of only one possible value is optimized when the falcon algorithm is used in its current form. In the future, we will focus our research on expanding the algorithm to be able to optimize multiple values at the same time as well as reduce the computational resource consumption of the method, while keeping or improve program performance.

REFERENCES

[1] K. Yu, Q. Ai, S. Wang, J. Ni, and T. Lv, "Analysis and optimization of droop controller for microgrid system based on small-signal dynamic model," *IEEE Transactions on Smart Grid*, vol. 7, no. 2, 2016, Art. no. 695e705.

[2] U. B. Tayab and Q. M. Humayun "Enhanced droop controller for operating parallelconnected distributed-generation inverters in a microgrid," *J. Renew Sustain Energy*, vol. 10, no. 4, 2018, Art. no. 045303.

[3] U. B. Tayab, Z. Ali, Y. Fuwen, L. Junwei, and M. Kashif "Short-term load forecasting for microgrid energy management system using hybrid HHO-FNN model with best-basis stationary wavelet packet transform," *Journal Elsevier Energy*, 2020, doi: 10.1016/j.energy.2020.117857.

[4] F. D. Rueda, J. D. Suárez, and A. R. Torres, "Short-Term Load Forecasting Using Encoder-Decoder WaveNet Application to the French Grid," *Energies*, vol. 14, 2021, Art. no. 2524, doi: 10.3390/ en14092524.

[5] D. Niu, Y. Wang, and D. D. Wu, "Power load forecasting using support vector machine and ant colony optimization," *Expert Syst. Appl.*, vol. 37, pp. 2531–2539, 2010.

[6] G. F. Fan, Y. H. Guo, J. M. Zheng, and W. C. Hong, "Application of the Weighted K-Nearest Neighbor Algorithm for Short-Term Load Forecasting," *Energies*, vol. 12, 2019, Art. no. 916.

[7] Y. Dong, Z. Zhang, and W. C. Hong, "A hybrid seasonal mechanism with a chaotic cuckoo search algorithm with a support vector regression model for electric load forecasting," *Energies*, vol. 11, 2018, Art. no. 1009.

[8] R. Zhang, Z. Y. Dong, Y. Xu, K. Meng, and K. P. Wong, "Short-term load forecasting of Australian National Electricity Market by an ensemble model of extreme learning machine," *IET Gener. Transm. Distrib*., vol. 7, pp. 391–397, 2013.

[9] M. Ghofrani, M. Ghayekhloo, and A. Arabali, "Ghayekhloo, A. A hybrid short-term load forecasting with a new input selection framework," *Energy*, vol. 81, pp. 777–786, 2015.

[10] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, and Y. Zhang, "Short-term residential load forecasting based on LSTM recurrent neural network," *IEEE Trans. Smart Grid*, vol. 10, pp. 841–851, 2017.

[11] B. Janet and D. Santhadevi, "Deep Long Short Term Memory with Harris Hawk Optimization: Enhanced Deep Learning Framework for Malware Detection at the Edge of the Iot System," *Research Square*, August 2nd, 2021, doi: 10.21203/rs.3.rs-713566/v1.