

**A multi-Objective Optimisation Metaheuristic Hybrid technique
for Forecasting the Electricity Consumption of the UAE: A Grey
Wolf Approach**

**Andreas Karathanasopoulos, Chia Chun Lo, Sovan Mitra,
Mohamed Osman, Hans-Jörg von Mettenheim and Slim Skander**

This is the peer reviewed version of the following article: Andreas Karathanasopoulos, Chia Chun Lo, Sovan Mitra, Mohamed Osman, Hans-Jörg von Mettenheim and Slim Skander 2024. A multi-Objective Optimisation Metaheuristic Hybrid technique for Forecasting the Electricity Consumption of the UAE: A Grey Wolf Approach, Journal of Forecasting. Advanced online publication, which has been published in final form at <https://doi.org/10.1002/for.3187>.

This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions. This article may not be enhanced, enriched or otherwise transformed into a derivative work, without express permission from Wiley or by statutory rights under applicable legislation. Copyright notices must not be removed, obscured or modified. The article must be linked to Wiley's version of record on Wiley Online Library and any embedding, framing or otherwise making available the article or pages thereof by third parties from platforms, services and websites other than Wiley Online Library must be prohibited.

The WestminsterResearch online digital archive at the University of Westminster aims to make the research output of the University available to a wider audience. Copyright and Moral Rights remain with the authors and/or copyright owners.

A multi-Objective Optimisation Metaheuristic Hybrid technique for Forecasting the Electricity Consumption of the UAE: A Grey Wolf Approach

By Andreas karathanasopoulos¹ Chia Chun Lo² Mitra Sovan³ Mohamed Osman¹ Hans von Mettenheim⁴ Slim Skander¹

1University of Dubai, Dubai Business School. UAE, Dubai.

Andreas karathanasopoulos: Professor in Finance, email: andreas.kara@hotmail.com

Mohamed Osman: Associate Professor in Economics , email: mosman@ud.ac.ae

Slim Skander : Professor in Finance , email: sslim@ud.ac.ae

2 a.Huatai Futures Co., Ltd, China b. Doctoral Think Tank, Macau, China

Chia Chun Lo: Professor in Finance and risk management, email: clo5@gsu.edu

3 Westminster university, London, United Kingdom.

Mitra Sovan, Associate Professor in finance, email: S.Mitra@westminster.ac.uk

4 Ipag Business University , Paris, France

Hans von Mettenheim, Professor in Finance , email: hj.von-mettenheim@ipag.fr

Abstract

By implementing a multi-objective optimization approach in forecasting, we introduce three optimization models grey wolf optimizer, genetic algorithm, and differential evolution algorithm combined with multilayer perceptron neural networks and support vector machines to predict electricity consumption in the UAE. The hybrid models' accuracy and efficiency were evaluated using various forecasting metrics. This study's contributions are threefold: it is the first to employ such a sophisticated hybrid approach, particularly using the recently introduced grey wolf optimizer, it compares optimization techniques with the established Pearson correlation-based method for dimensionality reduction and it represents one of the most extensive macroeconomic forecasts in the UAE using multi-objective heuristic hybrid optimization methods. Our findings indicate that the grey wolf optimizer significantly outperforms all other models, followed by the genetic algorithm.

Keywords: Macroeconomic forecast, statistical measures, Grey wolf, differential evolution algorithm, genetic algorithm.

Introduction

Currently the area of macroeconomic/financial forecasting is characterized in the literature as unstructured, noisy and non-stationary. Correspondingly, the macroeconomic environment where voluminous literature of forecasting research has been conducted seems to be complex and non-linear. In response to these challenging characteristics there has been significant advances in the development of new techniques on forecasting which utilizes learning machines such as artificial neural

networks and support vector machines. Artificial neural networks and support vector machines have been initially used for time series in a wide range of prediction applications since the early 1980s.

Although a broad literature survey has been conducted on the application of artificial intelligence systems¹, yet most of these studies focus on nonlinear models such as neural networks and support vector machines. In this category of artificial intelligence, the most popular and well-known neural network is the traditional multi-layer perceptron (MLP). In terms of training this neural network as Rumelhart et al. (1986) suggested they use the “back-propagation error” algorithm (BP) which can be described as “a mile-stone” in the area of forecasting. More precisely, even if neural networks have been well established, we are still facing some considerable problems such as: 1) type of input; 2) correct weight optimization; 3) fitness function selection and finally 4) maximum output optimization. Regarding the Support Vector Machines (SVM), it is noteworthy to mention that since the first construction of this model by Vapnik, (2000), there has been significant literature with improved versions of SVMs that provide better statistical results².

In order to overcome these limitations and deficiencies in the literature, we are currently introducing a multi-objective optimization metaheuristic approach that utilizes neural network and support vector machines architectures combined with Genetic Algorithms, Differential Evolution Algorithms and Grey Wolf Optimizers. More notably, we propose fully adaptive architectures that decrease the number of

¹ For details on the use of artificial intelligence see Zhang and Berardi (2001), Hansen and Nelson (2003), Karathanasopoulos et al (2011), (2012), (2014), (2015a) and (2016).

² For instance see the papers by Smola and Scholkopf (1998), Burges (1998), Vapnik and Cortes (1995), Osuna (1997), Pontil and Verri (1997), Wu et al. (2009), and Huang et al (2005).

parameters that practitioners need to optimize while on the other hand improving the forecasting accuracy of neural networks and support vector machines.

The rest of the paper is organized as follows: Section two presents the literature review which focuses on the three forecasting optimizers while section three outlines the dataset and the two input selection processes. In section four we present all the nonlinear models while section five reports and documents the statistical results. Section six concludes the paper.

2. Literature review

In the last decade there has been a voluminous literature on neural networks and support vector machines showing their superior accuracy and efficiency over other linear and nonlinear techniques. Zhang and Berardi (2001), Hansen and Nelson (2003), Abdelaziz et al (2014), Karathanasopoulos et al (2011), (2012), (2014), (2015a) and (2016) have succinctly indicated that forecasting different stock instruments that utilize neural networks and support vector machines have significantly outperformed linear models such as autoregressive moving average model, naïve strategy and random walk model. Furthermore, the last few years have witnessed the marriage of linear and nonlinear models. The reason for these hybrid combinations is that the new output is takes the advantages of both models which entails the significant improvements in forecast performance.

Nag and Mitra (2005) were the first to apply a succesful hybrid model combining neural network and genetic algorithm for modelling the daily foreign exchange rates. Moreover, significant results were reported by Karathanasopoulos et al (2013) and (2015b) who were trying to create hybrid models combining support vector

machines with genetic algorithms and neural networks with partial swarm optimizers in terms of forecasting different financial markets.

The contribution and uniqueness of this paper is the combination of the two previous generic stochastic models with Genetic Algorithms (GA), Differential Evolution (DE) and Grey Wolf Optimizers (GWO). It's the first time we have adopted such a sophisticated approach of combining multilayer perceptron networks and support vector machines with so many optimizers especially with the grey wolf optimizers. One early noteworthy research output came from Kennedy and Eberhart (1995) who indicated that hybrid algorithms combined with neural networks optimize the model parameters, the structure and the training procedure. Subsequently, Zhang et al. (2007), Karathanasopoulos et al (2010), Irani and Nasimi (2011) and Karathanasopoulos et al (2013) concurred by underscoring that combining optimizers with neural networks will always improve the forecasting power of the main model.

Finally, there are some optimizers which are less used such as the Grey Wolf Optimizer which was developed by Mirjalili et al. in 2014. Essentially, the GWO method came from the social behavior of grey wolf, also known as *Canis Lupus*, which belongs to *Canidae* group. This algorithm consists of three main stages: the first stage is the social hierarchy; second stage is the encircling of the prey, while the third stage is the hunting and the attacking of the prey. Similarly, like any other metaheuristic algorithm, exploitation and exploration are also two important features in GWO which are known as attacking of the prey and search for the prey. Even though, we don't have sufficient literature on this methodology, yet this new entry optimizer has been proven to be competitive and better than the existing ones³.

³ See Mirjalili et al. (2014).

On the other hand, genetic algorithms (GA) and Differential evolution optimizers (DE) have been utilized successfully in many daily solving problems. Onwubolu and Davendra (2006), Qu et al. (2012) and Wang et al. (2012) have efficaciously applied GAs and DE algorithms to select the initial connection weights and thresholds of neural network. Vesterstrom and Thomsen (2004), have persuasively argued that the DE algorithm performs better than other popular GA intelligent algorithm, based on 34 widely used benchmark functions. Likewise, Wang et al. (2012) and Cai and Wang (2013), have recognized that DE can also use individual local information and population to search for the optimal solution. Conclusively, Onwubolu and Davendra (2006), Wang et al. (2012) and Qu et al. (2012), state that DE and GA are the best evolutionary algorithms in a variety of benchmarking models.

3. Analysis of the data and input selection

The data used in this paper has been downloaded from Thomson Reuters. The observations are coming in monthly basis and give us the advantage to forecast one year ahead the electricity consumption of the UAE. Before proceeding with our optimization models, we first feed our models with a series of inputs correlated to the main index we forecast. Analytically for the electricity consumption forecast we use the general UAE index, five- and ten-year Sukuk bonds, crop production index, food production index, export index, import index, exchange rate pegged in dollars, Dow jones index, gross domestic product index, urbanization index, investment index and the consumer price index. Finally, in order to train our hybrid models, we divided the dataset into training period and out of sample period. The total number of observations starts from 1990 till 2018. There is no available information after 2019 as there is a consistent one years' delay from the data providers in the UAE. More importantly, the data have been clustered into different groups as reported in table 1.

Table 1: Total dataset divided in training and validation periods.

Name of Period	Trading Days	Beginning	End
<i>Total Dataset</i>	336	1/1/1990	30/12/2018
<i>Training Dataset</i>	216	1/1/1990	30/12/2009
<i>Validation Set</i>	120	1/1/2010	30/12/2018

The above dataset has been finalized in the way that it's presented in table 1 after checking different unsuccessful datasets. The number of different combinations exceeds 1000, focusing on good statistical performance in the in-sample and out-of-sample period. With regards to the inputs, we use 20 different variables and after two selections, we categorically keep the ones that significantly improve the performance of our combined/hybrid models. In table 2 we present all the input variables for the models.

Table 2: Explanatory variables for Electric consumption in UAE

Number	Variable	Lag
1	Electricity Consumption (UAE)	1
2	Dubai Financial Market General Index (UAE)	1
3	Dubai Financial Market General Index (UAE)	2
4	7-year maturity Sukuk bond (UAE)	1
5	5-year maturity Sukuk bond (UAE)	1
6	4-year maturity Sukuk bond (UAE)	1
7	Crop production index (UAE)	1

8	Crop production index (UAE)	2
9	Food production index (UAE)	1
10	Export index (UAE)	1
11	Export index (UAE)	3
12	Import index (UAE)	1
13	Import index (UAE)	2
14	Exchange rate Pegged in USD	1
15	Consumer Price index (UAE)	1
16	Consumer Price index (UAE)	4
17	Dow Jones index	1
18	Urbanization index (UAE)	1
19	Gross domestic product (UAE)	1
20	Gross domestic product (UAE)	2

The forecast of the electricity consumption consists of total electricity generated annually plus imports and minus exports, expressed in kilowatt-hours. The discrepancy between the amount of electricity generated and/or imported and the amount consumed and/or exported is accounted for as a loss in transmission and distribution.

3.1 EXAMINING THE RELATION BETWEEN THE ELECTRICITY CONSUMPTION INDEX AND THE OTHER FINANCIAL INDICES.

At the outset, in order to examine the relationship between the electricity consumption index and the other macroeconomic/financial indices, we first compute the correlation between the electricity consumption index with the value of the previous day's closing price of the other financial indices. The results of this exercise is presented in table 3.

Table 3: Pearson correlations of the variables with the electricity consumption index

Financial Indices	Pearson correlations (two tailed) with electricity consumption index
Dubai Financial Market General Index (UAE)	0.067
Dubai Financial Market General Index (UAE)	0.045
7-year maturity Sukuk bond (UAE)	0.090
5-year maturity Sukuk bond (UAE)	0.078
4-year maturity Sukuk bond (UAE)	0.003
Crop production index (UAE)	-0.006
Crop production index (UAE)	-0.008
Food production index (UAE)	-0.071
Export index (UAE)	-0.023
Export index (UAE)	0.009
Import index (UAE)	0.081
Import index (UAE)	-0.045
Exchange rate Pegged in USD	-0.007
Consumer Price index (UAE)	0.023
Consumer Price index (UAE)	0.012
Dow Jones index	0.005
Urbanization index (UAE)	-0.001
Gross domestic product (UAE)	0.055

Gross domestic product (UAE)	0.069
------------------------------	-------

In general, the Pearson's correlation essentially measures the linear association between two variables. The values of the correlation coefficient range from -1 to +1, and the sign of the correlation coefficient indicates the direction of the relationship. The absolute value of the correlation coefficient indicates the strength, with larger values indicating stronger relationships. The significance level (or p-value) is the probability of obtaining results as extreme as the one observed. If the significance level is very small (for example 0.05 which is the default significance level used by most statistical packages like SPSS or R) then the correlation is significant, and the two variables are linearly correlated. On the other hand, if the significance level is relatively large (e.g., 0.50) then the correlation is not significant, and the two variables are not linearly correlated. These conclusions encourage us to continue investigating the impact of all the 19 variables in the electricity consumption index.

3.2 Inputs for both neural networks

The inputs that are selected in the best execution of the Optimized SVM and MLP models with GAs are presented in table 4.

Table 4: Input Selection with GA

Number	Variable	Lag
1	Dubai Financial Market General Index (UAE)	1
2	Dubai Financial Market General Index (UAE)	2
3	7-year maturity Sukuk bond (UAE)	1
4	5-year maturity Sukuk bond (UAE)	1

5	4-year maturity Sukuk bond (UAE)	1
6	Consumer Price index (UAE)	1
7	Consumer Price index (UAE)	4
8	Dow Jones index	1
9	Gross domestic product (UAE)	1
10	Gross domestic product (UAE)	2

Similarly, the inputs that are selected in the best execution of the Optimized SVM and MLP models combined with DE algorithms are presented in table 5.

Table 5: Input Selection with DE

Number	Variable	Lag
1	5-year maturity Sukuk bond (UAE)	1
2	4-year maturity Sukuk bond (UAE)	1
3	Consumer Price index (UAE)	1
4	Consumer Price index (UAE)	4
5	Dow Jones index	1
6	Gross domestic product (UAE)	1

Likewise, the inputs that are selected in the best execution of the Optimized SVM and MLP models combined with Grey Wolf Optimizer are presented in table 6.

Table 6: Input Selection with GWO

Number	Variable	Lag
1	Dubai Financial Market General Index (UAE)	1
2	Dubai Financial Market General Index (UAE)	2
3	7-year maturity Sukuk bond (UAE)	1
4	5-year maturity Sukuk bond (UAE)	1
5	4-year maturity Sukuk bond (UAE)	1
6	Consumer Price index (UAE)	1
7	Consumer Price index (UAE)	4
8	Import index (UAE)	1

4. Forecasting Models

4.1 Multi-Layer Perception

The most popular model in the family of neural networks is the multilayer perceptron (MLP) which has been developed by Rumelhart et al. (1986) in order to solve a non-linear problem. A MLP is a feedforward network consisting of three layers, the input layer, one or more hidden layers and the output layer. The main objective of the MLP model is to reduce through the backpropagation function the square error. It is well

known that one major application area of neural networks is the forecasting one. Aladag et al (2010), Zhang et al. (2007) and Dunis et al. (2010), have adopted the MLP networks in their papers to perform the forecasting of different stock markets. They concluded that MLP networks in comparison to other methodologies is more accurate, reliable and efficient. Similarly, other research findings by Weigend et al. (1990), Weigend et al. (1992), Sharda et al. (1994), and Cottrell et al. (1995), Zhang et al. (1998), and more recently Aladag et al. (2010) have indicated that the MLP network is not only comparatively easy to build but also fast to run. Moreover, a significant advantage of adopting this model is in its incredible forecasting performance in the real-world time series.

4.2 Support Vector Machines

Support vector machines (SVM) are traditional stochastic processes that can be applied to classification and regression problems which are essentially based on statistical learning theory. This approach is considered as an extension to nonlinear models and was developed by Vapnik 2000. Classification problems with the use of SVMs have been solved by Smola and Scholkopf (1998), Burges (1998), Vapnic and Cortes (1995), Osuna (1997), Pontil and Verri (1997), Wu et al. (2009), and Huang et al (2005). More specifically in classification problems we try to find an optimal ground in between those two classes providing the bigger margin area in between them. Mathematically, we will have a quadratic programming problem where the number of variables is equal to the number of observations. Currently scientists have created more advanced versions of support vector machines that successfully solve numerous problems. For instance, SVM have been applied in many prediction problems in finance and economics and to this effect, there is significant research conducted due to their forecasting power such as Ince and Trafalis (2008), Cao and Tay (2003), Huang et al. (2005), Kim (2003),

Karathanasopoulou et al. (2014), Karathanasopoulos et al. (2015a) and Karathanasopoulos et al. (2015b).

4.3 Evolutionary Optimization

In the next Sections, we present a brief description of the first two optimization techniques followed by a thorough exposition of the grey wolf optimizer since it is relatively new. For a complete discussion of GA, we follow Holland (1995), and for DE we follow the paper published by Das and Suganthan in (2011). Likewise, for the grey wolf optimizer, we follow Mirjalil et al. (2014), who published their work on GWOs. Finally, it's worth mentioning that all the fitness functions are calculating the mean squared error.

4.4 Genetic Algorithm

Genetic algorithm is an optimization algorithm inspired by the principle of natural selection⁴. They have provided encouraging results in problems where the search space is big, complicated and there is no available mathematical analysis. At the outset, a population of candidate solutions is generated. Consequently, these solutions which are called chromosomes are optimized through a number of generations and two genetic operators characterized as crossovers and mutations. The chromosomes consist of genes which are the optimizing parameters. In each generation all chromosomes are evaluated through a fitness function and subsequently the more fitted chromosomes are selected to survive. Afterwards a sub-sample of the survivors is selected to evolve through one-point crossover or mutation.

⁴ For a detailed exposition see Holland (1995).

In sum, one-point crossover creates two offsprings from every two parents. The probability of an individual to be selected as a parent is called crossover probability and a crossover point c_x is then selected randomly. The two offsprings are made by both concatenating the genes that precede c_x in the first parent with those that follow and include c_x in the second parent. Thenceforth, the two offsprings replace their parents in the population. Mutation places random values in randomly selected genes with a certain probability termed as mutation probability. The mutation operator helps by avoiding local optima and exploring a larger surface of the search space. In this study the crossover and mutation probabilities are set to 0.9 and 0.1 respectively.

Crossover is used with the expectation that new chromosomes will have good parts of old chromosomes and conceivably the new chromosomes will be better. However, it is categorically recommended to leave some part of the population to survive till next generation. This is the primary reason why a high but not equal to one crossover probability was used. The selection step is based on the Holland's (1995) guidelines. The initial population is set to 30 while the termination criteria is either the algorithm reach the 200th generation or the average fitness across the current population is less than 5% away from the best fitness in the last five generations.

4.5 Differential Evolution

According to Das and Suganthan (2011), the differential evolution (DE) algorithm is a stochastic real parameter optimization algorithm. Similar to the GA, the DE is an evolutionary algorithm based on a sequence of iterations (generations) where individuals are updated through selection, mutation and crossover operators. The mutation operator perturbs population individuals with the scaled differences of

randomly selected and distinct population members. In DE the potential solutions are represented as strings of continuous variables comprising of feature and parameter variables. Discrete values are generated through rounding the continuous values and from the outset, a random population is generated. The algorithm for every individual X_i of the population, selects three other random and distinct individuals ($(X_{1,i}, X_{2,i}, X_{3,i})$). Subsequently a donor vector is generated:

$$V_i = X_{1,i} + F * (X_{2,i} - X_{3,i}) \quad [1]$$

where F is the mutation scale factor. Inevitably it follows that when a binomial crossover operator is applied which combines every member of the population X_i with its corresponding donor vector V_i will consistently produce a trial vector U_i :

$$U_i(j) = \begin{cases} V_i(j) & , \text{ if } (rand_{i,j}[0,1] \leq Cr) \\ X_i(j) & , \text{ otherwise} \end{cases} \quad [2]$$

In the following juncture, all trial vectors are evaluated and if the value of the fitness function is lower than the one of the population X_i , then the trial vector takes its position in the population. Hence, this insinuates profoundly that the population is only getting better through these iterations. The parameter F controls the size of the differentiation quantity which is going to be applied to a candidate solution from the mutation operator while Cr determines the number of genes which are expected to change in a population member.

In this study, we adhere to the guidelines provided by Qin et al. (2009). They recommend setting the mutation factor (F) between 0.4 and 1.0, with 0.5 being a common default to strike a balance between exploration and exploitation. The crossover probability (CR) is typically chosen between 0.7 and 0.9 to maintain diversity while ensuring effective convergence. We set the initial population size to 30. Similar to

Genetic Algorithms (GA) and Particle Swarm Optimization (PSO), the termination criterion is a mix of the maximum number of iterations (set to 200) and a convergence criterion. The convergence criterion is met when the mean population fitness remains within 5% of the best population fitness for more than five consecutive iterations. Additionally, strategies like parameter adaptation or tuning based on preliminary runs or meta-optimization techniques may be employed to determine the optimal settings for specific problems. Fine-tuning these parameters is essential as they greatly affect the algorithm's performance and convergence behavior.

4.6 Grey Wolf Optimizer

The Grey wolf optimizer is regarded as a top predator, which makes them first in the food pyramid. In this optimization methodology, we have four hierarchies applied to the grey wolf population which are termed as alpha, beta, delta and omega. In the first level we have alpha group which consists of male and female grey wolves that are responsible for decision making such as hunting. In the second level we have the beta group which is responsible for helping the alpha group in their decision making or any other activities. The beta group can be male or female and would be the best candidate for replacement of the alpha members if one of the alpha participants for instance passes away or becomes old. The third level is the delta group which has to submit solutions

to alpha and beta groups. Delta group still has dominant position in comparison to the omega group. In the last level, we have the omega group which consists of scouts, sentinels, elders, hunters and caretakers. This group is ranked last in the hierarchy and plays the role of a scapegoat. In sum, we can represent the model mathematically by following these three steps:

4.6.1 Social Hierarchy

In GWO, the fittest solution is represented by alpha (α), followed by the second and third best solutions namely beta (β) and delta (δ) respectively. Meanwhile, the rest of the candidate solutions are considered as omega (ω). The hunting (optimization) is guided by α , β and δ while the ω follows the preceding three groups.

4.6.2 Encircling Prey

During the hunting missions, the wolves tend to encircle their prey and this activity can be captured by using the following equation:

$$\vec{D} = |\vec{C} \cdot \vec{X}_p(t) - \vec{X}(t)| \quad [3]$$

$$\vec{X}_{(t+1)} = \vec{X}_p(t) - \vec{A} \cdot \vec{D} \quad [4]$$

t =current iteration,

A and C= coefficient vectors,

X_p position vector of the prey and

X= position vector of the grey wolves.

For vectors A and C is calculated as follows:

$$\vec{A} = 2\vec{a} \cdot \vec{r}_1 - \vec{a} \quad [5]$$

$$\vec{C} = 2 \cdot \vec{r}_2 \quad [6]$$

where components of $a!$ are linearly decreased from 2 to 0 over the course of the iterations. Meanwhile, r_1 and r_2 are random vectors in the range of [0,1].

4.6.3 Hunting

Habitually the hunting activities are guided by the alpha group. However, both the beta and delta group might also be involved in these hunting activities sporadically. In GWO, the alpha group is the fittest candidate solution while the beta and delta groups are essentially the experts about the potential location of the prey. Thus, the first three best solutions obtained so far are saved while the other agents (including omegas) are induced to update their positions based on the position of the best search agents. This can be modeled in the following equations:

$$\vec{D}_\alpha = |\vec{C}_1 \cdot \vec{X}_\alpha - \vec{X}|, \vec{D}_\beta = |\vec{C}_2 \cdot \vec{X}_\beta - \vec{X}|, \vec{D}_\delta = |\vec{C}_3 \cdot \vec{X}_\delta - \vec{X}| \quad [7]$$

$$\vec{X}_1 = \vec{X}_\alpha - \vec{A}_1 \cdot (\vec{D}_\alpha), \vec{X}_2 = \vec{X}_\beta - \vec{A}_2 \cdot (\vec{D}_\beta), \vec{X}_3 = \vec{X}_\delta - \vec{A}_3 \cdot (\vec{D}_\delta) \quad [8]$$

$$\vec{X}_{(t+1)} = \frac{\vec{X}_1 + \vec{X}_2 + \vec{X}_3}{3} \quad [9]$$

5. Empirical results

5.1 Statistical Performance

In tables 7,8 and 9, we present the statistical performance in the out-of-sample period of all the multi-objective optimizing models. To assess this out of sample data set

performance we apply the following criteria which include the Mean Absolute Error (MAE), the mean absolute percentage error (MAPE) the Root Mean Squared Error (RMSE) and the correct directional change (CDC). The results indicate that the lower output implies better forecasting accuracy of the model. (After request we can provide the in-sample results).

Parameters of the classification models (MLP and SVM were) were tuned using 10-fold internal cross-validation in the training dataset. Regarding the optimization methods, default values were used for most of their parameters while to allow a direct comparison the same number of individuals and maximum generation number was used for all of them (population size: 1000, maximum number of generations: 10000).

Table 7: Statistical performance of electric Consumption with GA selected inputs (out of sample period). Metrics are provided as an average of 10 iterations.

Forecast	MLP	SVM	MLP-DE	MLP-GA	MLP-GWO	SV- DE	SV-GA	SV- GWO
MAE	0.0340	0.0420	0.030	0.0287	0.0270	0.0298	0.0320	0.0245
MAPE	344.67%	312.34%	250.23%	230.12%	196.56%	263.28%	245.56%	201.87%
RMSE	0.0372	0.0365	0.0392	0.0356	0.0321	0.0398	0.0345	0.0296
CDC	50.01%	50.54%	51.87%	53.78%	54.87%	56.87%	59.89%	62.98%

As can be gleaned from table 7, the optimizing models that combine with the GWO have significantly outperformed all the other models in our empirical exercises.

Table 8: Statistical performance of electric Consumption with DE selected inputs (out of sample period). Metrics are provided as an average of 10 iterations.

Forecast	MLP	SVM	MLP-DE	MLP-GA	MLP-GWO	SV-DE	SV-GA	SV-GWO
MAE	0.0560	0.0589	0.0456	0.0560	0.0389	0.0410	0.0512	0.0400
MAPE	256.56%	210.34%	153.34%	163.76%	152.86%	155.91%	168.62%	145.52%
RMSE	0.0345	0.0337	0.0389	0.00356	0.0345	0.0295	0.0245	0.0220
CDC	48.76%	50.98%	52.98%	54.78%	55.51%	61.87%	67.87%	64.32%

The results reported in Table 8 indicate that the GWO optimizer in both combined models provide better statistical accuracy than all the other optimizing models and also it is noteworthy to recognize that all the three optimizers have outperformed the single traditional models.

Table 9: Statistical performance of electric Consumption with GWO selected inputs (out of sample period). Metrics are provided as an average of 10 iterations.

Forecast	MLP	SVM	MLP-DE	MLP-GA	MLP-GWO	SV- DE	SV-GA	SV- GWO
MAE	0.0250	0.0290	0.0210	0.0218	0.0169	0.0198	0.0231	0.0113
MAPE	187.54%	155.34%	148.32%	140.23%	137.87%	147.98%	145.76%	126.91%
RMSE	0.0384	0.0362	0.0267	0.0256	0.0245	0.0201	0.0183	0.0190
CDC	47.87%	46.89%	50.12%	52.12%	55.86%	60.43%	62.78%	63.12%

The results in Table 9 again corroborate the superior performance of the GWO optimizer with respect to all the other optimizing models by providing the best statistical accuracy. Moreover, it's quite important to demonstrate that the GWO picks the best inputs in our learning machines relative to all other optimizing models as all the results in table 9 have revealed since they are slightly better than those in tables 7 and 8. Statistically significant changes were inferred by performing pairwise comparisons of their metrics with students t-test (p-value threshold: 0.05).

6. Conclusion

In this paper we culminate our efforts in investigating the efficiency of support vector machines and multi-layer perceptrons combined with three optimizers in the task of forecasting electricity consumption of the UAE.

Despite the existence of various research works trying to deploy heuristic optimization techniques to optimize the feature subset to be used by classification models for financial forecasting, there still does not exist any established solution in this field. Moreover, the problem of forecasting electricity consumption in UAE is not that has not been extensively studied with these techniques and thus prior knowledge on which features are more predictive is limited. Moreover, the lack of theoretical background on these heuristic optimization techniques necessitates an experimental comparative study to decide which ones is the best performing one. In the present study, in addition to established techniques for dimensionality reduction, such as GAs, we explored additional heuristic methods which are not so widely used, such as the GWO, and interestingly it was found that GWO was the best performing one combining its good exploration and exploitation characteristics in searching the search space.

The statistical performance of our multi-objective optimization models is estimated throughout the period of 1990-2018, using the last seven years for out-of-sample testing. The results indicate that the SVM-GWO and the MLP-GWO have significantly outperformed all the other optimization models. Hence, it's noteworthy to mention that DE optimizers chronicled the second-best performance when combined with MLP and SVM. A salient feature that has emerged from our empirical analysis is that the input selection criteria came after applying two techniques. Both of the input selection mechanisms have chosen consistently the best inputs and have optimized the forecasts. Finally, this paper denotes that the generic models have consistently underperformed relative to the optimization models. As such the forecasting models adopted in this paper would provide worthwhile information for all the entities engaged in the production and distribution of electricity in the UAE, so that they can adapt production and distribution capacities, especially in the months when consumption reaches a maximum value such as during the summer months.

The process of training and evaluation models for forecasting electricity consumption is an offline process and thus computational complexity is not essential. All methods were run on the same workstation using a i7 Intel Processor and 8GB RAM memory. For all methods the training process required on average in between 5-8h depending on whether or not termination criteria were reached before the maximum number of generations reached. Despite the relatively high training computational cost, the average runtime of a trained model to predict the electricity consumption for a day in the validation set was less than 1msec.

Despite the promising results of using heuristic optimization techniques for selection the inputs of classification models for forecasting the electricity consumption our study still possesses some limitations. First of all, the performance of the models is

benchmarked in UAE and the fact that Grey Wolf optimization is performing best is not evidence that this approach would be better in forecasting electricity consumption in other countries. Moreover, electricity consumption might be affected by many factors, as well as events, accidents, and proceedings in the international political landscape, making thus a more holistic approach a necessity to increase the predictive performance of the models. Finally, the proposed method used heuristic optimization to optimize the feature subset of the classification models but does not optimize their parameters which are still optimized with internal cross validation. An interesting future direction includes therefore the expansion of the proposed optimization method to optimize on parallel the parameters of the classification model.

References

- Aladag, C.H, Egrioglu, E., and Yolcu U., (2010) Forecast combination by using artificial neural networks, *Neural Processing Letters*, 32 (3), 269–276.
- Abdelaziz, B. A., M. Amer ,H El Baz (2014) - An Epsilon Constraint Method for selecting Indicators for use in Neural Networks for Stock Market Forecasting,*Information Systems and Operational Research* ,52,3,0315-5986, 116-125
- Burges, C. J. C., (1998), “A Tutorial on Support Vector Machines for Pattern Recognition”, *Data Mining and Knowledge Discovery*, Volume 2, pp.1-43, Kluwer Academic Publishers, Boston.
- Broomhead, D., and Lowe, D., (1988), Multivariate functional interpolation and adaptive networks. *Complex Systems*, 2, pp.321-355
- Cai, Y., and Wang, J., (2013) Differential evolution with neighborhood and direction information for numerical optimization,” *IEEE Transactions on Cybernetics*, 43, (6), pp. 2202–2215
- Cortes C., and Vapnik, V., (1995), “Support Vector Networks”, *Machine Learning*, 20,pp. 273-297,
- Cottrell, M.,Girard, B., Girard,Y., Mangeas, M., and C. Muller, C., (1995) Neural modeling for time series: a statistical stepwise method for weight elimination, *IEEE Transactions on Neural Networks*, 6(6), 1355–1364
- Das, S., and Suganthan, P, N., (2011) Differential evolution: A survey of the state-of-the-art. *IEEE Transactions on Evolutionary Computation*, 15(1), pp:4-31
- Diebold, F. and R. Mariano (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics* 13, pp. 253–263
- Dolvin, S, D., (2010), S&P ETFs: Arbitrage Opportunities and Market Forecasting, *The Journal of Index Investing Summer*, 1, (1), pp. 107-116
- Elman, J. L., (1990) Finding structure in time. *Cognitive Science*, 14, pp.179–211.
- Elman, J.L., (1991) Distributed representations, simple recurrent networks, and grammatical structure. *Machine Learning*, 7, pp. 195–224.
- Kennedy, J., & Eberhart, R. C. (1995) Particle swarm optimization. In *Proceedings of the IEEE international conference on neural networks IV* (pp. 1942–1948).
- Kennedy J., and R. C. Eberhart, (1997a), Particle Swarm Optimization, *Proceedings of International Conference on Neural Networks*, Orlando, pp. 1942-1948
- Kennedy, J., & Eberhart, R. C., (1997b), A discrete binary version of the particle swarm algorithm. In *Proceedings of the conference on systems, man, and cybernetics* pp. 4104–4109
- Hansen, P. and A. Lunde (2010), An oxtm software package for multiple comparisons..

- Hansen, P., A. Lunde, and J. Nason (2011), The model confidence set. *Econometrica*, 79, 453–497
- Hansen, J. V., and Nelso, D., (2003), Forecasting and recombining time-series components by using neural networks, *Journal of the Operational Research Society*, 54, (3), pp. 307–317
- Holland, J. H., (1995). *Hidden Order: How Adaptation Builds Complexity* Addison-Wesley, Redwood City, CA
- Huang, W., Nakamori, Y., and Wang, S. Y., (2005), Forecasting stock market movement direction with support vector machine, *Computers & Operations Research*, 32, 10, pp. 2513–2522.
- Irani, R., & Nasimi, R., (2011), Evolving neural network using real coded genetic algorithm for permeability estimation of the reservoir, *Expert Systems with Applications*, 38, (8), pp. 9862–9866.
- Kaastra, I. and Boyd, M., (1996), Designing a Neural Network for Forecasting Financial and Economic Time Series, *Neurocomputing*, 10, pp. 215-236
- Karathanasopoulos, A., Theofilatos, K., Leloudas, P., and Likothanassis, S., (2010), Modelling the ASE20 Greek index using artificial neural networks combined with genetic algorithms, In *Proceeding of the 20th International Conference on Artificial Neural Networks*
- Karathanasopoulos, A., Dunis, C., and Laws, J., (2011), Modelling and trading the Greek stock market with mixed neural network models, *Applied Financial Economics* 21 (23), pp. 1793-1808
- Karathanasopoulos, A., Laws, J., Sermpinis, G., and Dunis, C., (2012), Forecasting and trading the EUR/USD exchange rate with Gene Expression and Psi Sigma Neural Networks , *Expert systems with applications* , 39 (10,) pp. 8865–8877.
- Karathanasopoulos, A., Sermpinis, G., Theofilatos, K., Georgopoulos, E., Dunis, C. (2013a), “Forecasting Foreign Exchange Rates with Adaptive Neural Networks Using Radial-Basis Functions and Particle Swarm Optimization.” *European Journal of Operational Research*, 225(3), pp. 528-540.
- Karathanasopoulos, A., Sermpinis, G., Laws, J. and Dunis, C., (2014), Modelling and Trading the Greek Stock Market with Gene Expression and Genetic Programming Algorithms, *Journal of forecasting*, 33 (8), pp. 596-610.
- Karathanasopoulos, A., Stasinakis, C., Sermpinis, G., and Theofilatos, K., (2015a), Modelling, forecasting and trading the EUR exchange rates with hybrid rolling genetic algorithms—Support vector regression forecast combinations, *European Journal of operation research*, 247 (3) pp. 831-846.
- Karathanasopoulos, A., Sermpinis, G., Stasinakis, C., and Theofilatos, K., (2015b), Forecasting US Unemployment with Radial Basis Neural Networks, Kalman Filters and Support Vector Regressions, *Computational economics*, pp1-19.
- Karathanasopoulos, A., Theofilatos, k., Serpins G., Dunis, C., and Mitra, S., (2016a), Stock market prediction using evolutionary support vector machines: an application to the ASE20 index, *The European Journal of Finance* 22 (12), pp. 1145-1163
- Karathanasopoulos, A., (2016b) Modelling and trading the English stock market with novelty optimization techniques *Economics and Business Letters* 5 (2), pp.50-57

- Marshall, B., Nhut N., and Visaltanachoti, N., (2013), ETF Arbitrage: Intraday Evidence, *Journal of Banking & Finance*, 37, pp. 3486-3498.
- Nag, A. K., and Mitra A., (2005), Multiple outlier detection in multivariate data using self-organizing maps, *Computational Statistics*, 20 (2), pp. 245-264
- Osuna, E. E., Freund R., and Girosi, F., (1997), "Support Vector Machines: Training and Applications", Technical Report, Massachusetts Institute of Technology, Artificial Intelligence Laboratory, AI Memo No. 1602
- Onwubolu, G., Davendra, D., (2006), Scheduling flow shops using differential evolution algorithm. *European Journal of Operational Research*, 171, pp. 674–679
- Pesaran, M., and Timmermann, A., (1992), A simple nonparametric test of predictive performance, *Journal of Business and Economic Statistics*, 10, pp.461-465
- Pontil M., and Verri, A., (1997), Properties of Support Vector Machines, Technical Report, Massachusetts Institute of Technology.
- Qin, A. K., Huang, V. L., and Suganthan, P. N., (2009), Differential evolution algorithm with strategy adaptation for global numerical optimization. *IEEE Transactions on Evolutionary Computation* 13, pp.398–417.
- Qu, B. Y., Suganthan, P. N., and Liang, J., (2012), Differential evolution with neighborhood mutation for multimodal optimization," *IEEE Transactions on Evolutionary Computation*, 16(5), pp. 601–614
- Rumelhart, D. E., Hinton, G. E., and Williams, R. J., (1986), Learning representation by backpropagating errors, *Nature*, 322, pp. 533–536
- Shapiro, A. F., (2000), 'A Hitchhiker's Guide to the Techniques of Adaptive Nonlinear Models', *Insurance, Mathematics and Economics*, 26, 2, 119-132.
- Sharda, R., (1994), Neural networks for the MS/OR analyst:An application bibliography, *Interfaces*, 24(2), pp. 116–130
- Smola A. J., and Scholkopf, B., (1998), A Tutorial on Support Vector Regression, *NEUROCOLT2 Technical Report Series*, NC2-TR-1998-030
- Tenti, P., (1996), Forecasting Foreign Exchange Rates Using Recurrent Neural Networks, *Applied Artificial Intelligence*, 10(6), pp. 567-581
- Vesterstrom, J., and. Thomsen, R., (2004), A comparative study of differential evolution, particle swarm optimization and evolutionary algorithms on numerical benchmark problems. *Proceedings on Congress on Evolutionary Computation*, 2, pp.1980-1987.
- Wang, L., He, J., and Zeng, Y. R., (2012), A differential evolution algorithm for joint replenishment problem using direct grouping and its application, *Expert Systems*, 29(5), pp. 429–441
- Weigend, A. S., Huberman, B. A., and Rumelhart, D. E., (1990), Predicting the future: A connectionist approach, *International Journal of Neural Systems*, 1, 193–209.
- Weigend, A. S., Huberman, B. A., and Rumelhart, D. E., (1992), Predicting sunspots and exchange rates with connectionist networks. In: Casdagli, M., Eubank, S. (Eds.), *Nonlinear Modeling and Forecasting*. Addison-Wesley, Redwood City, CA, pp. 395–432

Xie, J., Wu, J., and Qian, Q., (2009), Feature Selection Algorithm Based on Association Rules Mining Method. ACIS-ICIS2009, pp.357-362.

Zhang, G., Patuwo, B. E., and Hu, Y. M., (1998), Forecasting with artificial neural networks: The state of the art. International Journal of Forecasting, 14, 35-62.

Zhang, G. P., and Berardi, V. L., (2001), Time series forecasting with neural network ensembles: an application for exchange rate prediction Journal of the Operational Research Society, 52(6), pp. 652–664

Zhang, J. R., J. Zhang, J., Lok, T. M., and Lyu M. R., (2007), A hybrid particle swarm optimization–back-propagation algorithm for feedforward neural network training", Elsevier Applied Mathematics and Computation, 185(2), pp. 1026-1037.