Developing a data mining based model to extract predictor factors in energy systems: Application of global natural gas demand

Reza Hafezi [a], AmirNaser Akhavan [b], Mazdak Zamani [c] , Saeed Pakseresht [d] , Shahaboddin Shamshirband [e,f*],

[a] *Futures Studies Research Group, National Research Institute for Science Policy (NRISP), Tehran, Iran.*

[b] *Technology Foresight Group, Department of Management, Science and Technology, Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran.*

[c] *School of Arts and Sciences, Felician University, 262 South Main Street Lodi, New Jersey 07644*

[d] *Director of Research and Technology, National Iranian Gas Company (NIGC), Tehran, Iran.*

[e] *Department for Management of Science and Technology Development, Ton Duc Thang University, Ho Chi Minh City, Vietnam;*
[f] *Faculty of Information Technology, Ton Duc Thang University, Ho Chi Minh City, Vietnam*

[*] *Corresponding Authors email: shahaboddin.shamshirband@tdtu.edu.vn*

Abstract:

Recently natural gas (NG) global market attracted much attention in case it is cleaner than oil, and simultaneously in most regions is cheaper than renewable energy sources. However, price fluctuations, environmental concerns, technological development, emerging unconventional resources, energy security challenges, and shipment are some of the forces that made the NG market more dynamic and complex. From a policy-making perspective, it is vital to uncover demand-side future trends. This paper proposed an intelligent forecasting model to forecast NG global demand, however investigating a multi-dimensional purified input vector. The model starts with a data mining (DM) step to purify input features, identify the best time lags, and to pre-process selected input vector. Then a hybrid artificial neural network (ANN) which equipped with genetic optimizer is applied to set up ANN's characteristics. Among 13 available input features, six features (e.g. Alternative and Nuclear Energy, CO2 Emissions, GDP per Capita, Urban Population, Natural Gas Production, Oil Consumption) selected as the most critical feature via the DM step. Then, the hybrid prediction model is designed to extrapolate the consumption of future trends. The proposed model overcomes competitive models refer to different error based evaluation statistics. Besides, as the model proposed the best input feature set, results compared to the model which used the raw input set, with no DM purification process.

Keywords: *Natural gas demands; Prediction; Energy market; Genetic algorithm; Artificial neural network; Data mining.*

1. Introduction:

The world energy demand increased in the two past decades and even predictions implying the growing trends for the next decades [1-3]. Still, fossil fuels play a critical role in the

energy supply chain due to economic feasibility. Refer to International Energy Agency's (IEA) 2016 report, fossil fuels in the form of liquid fuels, natural gas, and coal contain more than 80% of the world energy consumption [4]. Easiness of utilization, higher performance, compared to traditional energy sources, ease of mobility via land or sea and affordable extraction cost introduced oil and natural gas (NG) as strategic commodities [5, 6]. However, emergent ecological concerns and rethinking of a more peaceful future (sustainable development goals) attracted attention toward climate change challenges (such as greenhouse gases emissions and global warming) [7]. The two non-aligned objectives, on one hand, development and increasing needs for energy supply and on the other hand, global environmental concerns, attracted researchers to study energy systems and develop different plausible future perspectives.

Despite successful efforts, the main problem is still existing, which is defined as “discovering reliable future trends and probable alternative futures in the field of energy systems and uncover the most influencing driving forces to aid energy management process”. This paper is aimed to develop an intelligent learning-based prediction model which is equipped with data mining (DM) techniques to purify and the setup input vector. The DM step is used to select and organize the best input features that represent patterns of future global NG demand trends. Although many previous studies successfully addressed NG global demand prediction problem, we attempt to uncover the most effective driving forces as input features and analyzing how they will affect the objective function (NG global demand prediction). For example, the proposed model studies time relation between input variables and the target variable. So a less-dimension input set is available to policymakers to simplify and experience reliable decision-making process.

As it is impressed by a series of variables and oscillating time series, the NG forecasting problem is a very challenging [8]. These days, massive efforts investigated artificial intelligence (AI) models or integration of several models (hybrid models) for prediction problems to increase the accuracy and the model reliability [9, 10]. Also, numerous notable studies investigated by demand prediction for the case of energy resources [11-16]. The prediction performance of the CDA model overcame compared to the earlier neural networks (NN) and an engineering based model.

Baumeister and Kilian published a research paper to analyze how vector autoregression (VAR) models form policy-relevant forecasting scenarios in the case of an oil market. The model investigates the influence of scenario weights' probability changes to the real-time oil price forecasting [17]. Also, Dilaver et al. investigated NG consumption in Europe to support long-term investments and contracts [18]. They estimated an OECD-Europe NG demand trends with annual time series during the period from 1978 to 2011 by applying a structural time series model (STSM). Finally, three scenario streams developed based on business as usual, high, and low case scenarios.

Li et al. used dynamic system models to create possible outlooks to 2030 for the case of China's NG consumption growth. Then to assess the results accuracy and propose policy recommendations on NG exploration and development of China's NG industry, a scenario

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 June 2019 doi:10.20944/preprints201906.0202.v1

Peer-reviewed version available at Energies 2019, 12, 4124; doi:10.3390/en12214124

analysis step was applied [19]. Also, Suganthi and Samuel provided a comprehensive review of the energy model, which attempted to forecast the demand function [20]. Authors classified prediction models and presented that most of the recent researches contained quantitative models that result in a single future prediction. Models used statistical error functions to estimate, accuracy compared with other comparative models. However, as mentioned above, data-driven models may regret set of effective qualitative variables. In the other hand projecting alternative futures based on qualitative approaches are challenging, especially in the case of validation and moreover, they are extremely affected by the expert group (number of experts and judgment validation). To present a universal review and to dedicate insights about prediction approaches used by previous studies, table 1 summarized models used to address energy consumption prediction problem.

Table 1. Analyzing previous studies, based on their approaches to address energy consumption prediction problem

Approaches		References
Classical computational extrapolation	Time series	[11, 12, 18, 21-30]
	Regression	[15, 31-34]
Econometrics		[35-39]
Expert systems and learning models	Artificial neural network (ANN)	[8, 27, 40-53]
	Genetic programming (GP)	[8, 11, 27, 45, 52, 54-59]
	Ant colony optimization	[60]
	Particle swarm optimization (PSO)	[13, 61]
	Support vector machine (SVM)	[12, 27, 46, 51, 62, 63]
	Fuzzy inference system (FIS)	[8, 31, 49, 57, 64]
Others	Decomposition approach	[65, 66]
	Input-output model	[67, 68]
	Bottom-up model	[69-71]
	Grey method	[13, 24, 68, 72, 73]
	Logistic model	[74]

To dedicate a more detailed understanding of various existed forecasting models table 2 shows the pros and cons of main forecasting methods.

Table2. The major pros and cons of main forecasting methods

Type of Models	Pros & Cons
Classic price modeling/	• Focus on historical data.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 June 2019 doi:10.20944/preprints201906.0202.v1

Peer-reviewed version available at Energies 2019, 12, 4124; doi:10.3390/en12214124

forecasting	• Do not consider jumps and drips of the prices. • Generally, these models were introduced for stock markets. • Do not use the unit root test for time series and econometric methods to estimate their parameters.
Time series models	• Focus on historical data. • Do not cope with extreme jumps and drips. • Do not use feedback loops to dynamically upgrade model adjustment features.
Learning forecasting models	• Focus on historical data. • Generally being able to learn fluctuations and related formerly signals. • Use feedback loops to upgrade model adjustment features dynamically.
Qualitative based forecasting models	• Rarely depend on quantitative forecasting methods. • Donate insights about long-run behaviors of a complex system. • Mostly depend on experts' evaluations, instead of historical time series. • Can dynamically modify input features.

In this paper, we are aimed to propose a learning-based model, which is designed to present a more reliable and relevant input features (driving forces) to initialize a hybrid prediction ANN to equip decision-making process with accurate and reliable forecasts. Following section investigates the proposed methodology and brief descriptions of various steps, then section three is dedicated to presenting the implementation phase and discussing results to show how the proposed methodology overcomes other benchmark models. Finally, section four provides summaries and conclusions.

2. The methodology of research:

As noted previously, the following research is aimed to expand a data mining based prediction model. Two primary goals targeted: (1) determining features which effectively present trends for NG global demand and (2) identification of time lags to define time relations between input variables and the target variable, (3) developing an adaptive intelligent prediction model that can extrapolate future trends for the global NG demands. Fig 1 conceptually shows designed data mining genetic-neural network (DmGNn) methodology to approach noted goals.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 June 2019 doi:10.20944/preprints201906.0202.v1

Peer-reviewed version available at Energies 2019, 12, 4124; doi:10.3390/en12214124

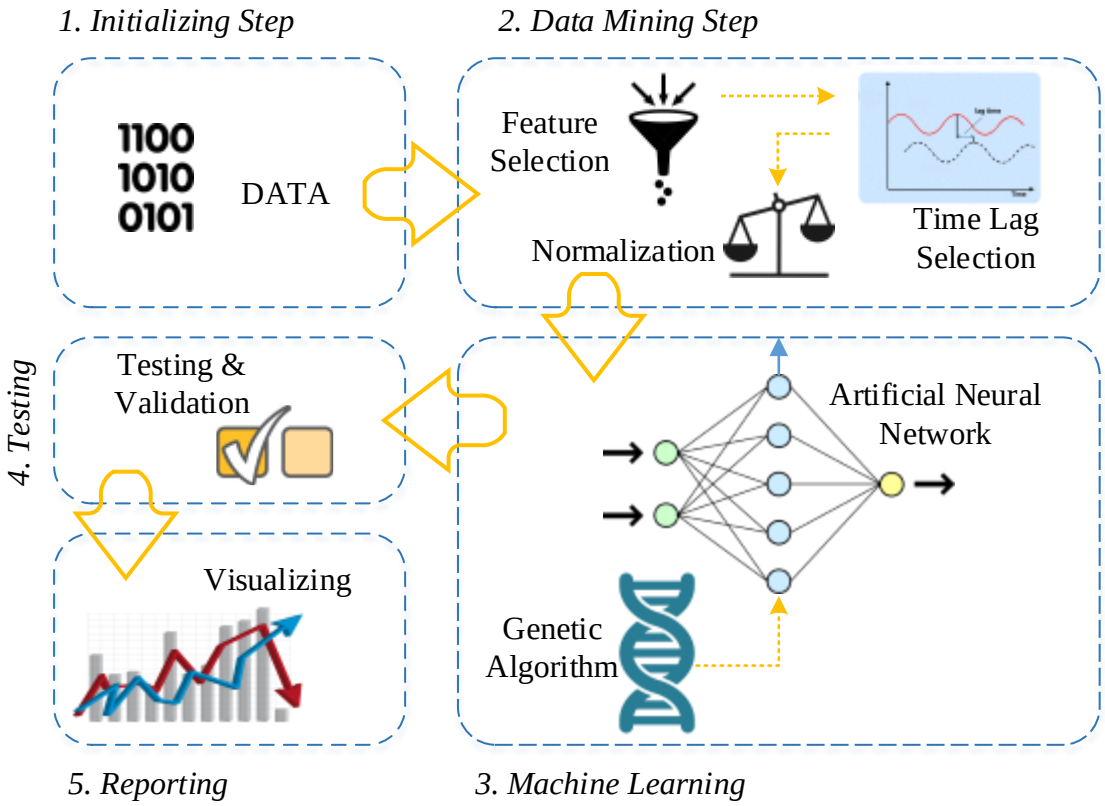


Fig1. The research methodology (by Authors)

The main phases and steps of the proposed methodology discussed as follow:

PHASE 1.

Step1. Data gathering: in this step, previous studies reviewed to detect raw input features. Unlike most of the previously published researches, this paper pursues the maximum approach, means we will gather and use maximum available input data to ensure that the developed model will not neglect a possible solution. In simple words, the proposed methodology does not limit the solution space due to the use of confined input features — o*utput:* input feature set.

PHASE 2.

Step2. Feature selection: this step is designed to select the most relevant subset of the gathered features. The main target is to reduce problem dimensions while preserving all local optimal solutions — o*utput:* refined input feature sub-set.

Step3. Time lag selection: is investigated to study how different time lags for input features may affect forecasting accuracy. This step will study time relation between the input variable and the target variable. *Output:* timed input features sub-set.

Step4. Normalization: different scales of input features may cause in a biased final forecasting model. This step is aimed at reproducing input features but in similar, uniform scales. *Output:* uniformed timed input features sub-set.

PHASE 3.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 June 2019 doi:10.20944/preprints201906.0202.v1

Peer-reviewed version available at Energies 2019, 12, 4124; doi:10.3390/en12214124

Step5. Design of the forecasting model: in this step, an ANN is equipped with a GA in order to optimize the network's characteristics and develop an accurate prediction model. *Output:* prediction framework

PHASE 4.

Step6. Implementation: finalized input features applied to the prediction framework. In this step, the input set divided into two main portions, one to train and other to test the performance of the prediction framework. *Outputs:* adjusted prediction model & obtained extrapolated results.

PHASE 5.

Step7. Validation: this step dedicated to comparing the obtained results of the proposed prediction framework with other benchmark comparative models. *Output:* output/accuracy analysis

To model complex systems (like ours), selecting a robust model architecture is very challenging [75, 76]. The DM approach is selected to handle the complexity of input variables. DM is defined as the process of extracting appealing patterns and deriving knowledge in massive datasets. So, as Han et al. noted: "*the principal dimensions are data, knowledge, applications, and technologies*" [77].

2.1. Data gathering and data pre-processing:

Input data remarkably affect the accuracy and quality of the obtained results. In the case of energy consumption, previous researches investigated different sets of input features to predict energy consumption's upcoming trends. A significant limitation of a prediction model is that it cannot reflect effects of variables which did not exist in the input feature set (those have been neglected). To ensure robustness and the validity of the proposed prediction model, the paper proposes the maximal approach, which means to investigate all available input data and reduce dataset dimension through a DM technique. This approach has the advantage of retaining all signals and trends while simultaneously, the model faces an undeniable challenge that is the increased complexity level due to the large input set which may negatively affect prediction efficiency. In another hand, it is a challenging process to set up strategic decisions based on an extensive collection of parameters/inputs. To handle the noted problem a DM based data pre-processing step is proposed by this paper to examine and purify input features. Table 3 summarizes the most frequently used input features (by other researchers) and the features which were available/accessible online.

Table3. Initialize input features

Title	Unit	Reference(s)	Source
Alternative and Nuclear Energy	% of total energy use	Proposed by authors	World Bank
CO2 Emissions	metric tons per capita	[36]	World Bank
CO2 Emissions	Kt		World Bank
Energy Imports, Net	% of energy use	Proposed by authors	World Bank

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 June 2019 doi:10.20944/preprints201906.0202.v1

Peer-reviewed version available at Energies 2019, 12, 4124; doi:10.3390/en12214124

Fossil Fuel Energy Consumption	% of total	[78]	World Bank
GDP Growth	annual %	[36, 42, 46, 60, 61, 79-87]	World Bank
GDP per Capita	current US$		World Bank
Population Growth	annual %	[33, 39, 40, 60, 61, 79, 81-83, 86-91]	World Bank
Urban Population	Person	[79, 85]	World Bank
Gold Price	10:30 A.M.in London Bullion Market, US$	Proposed by authors	World Bank
Natural Gas Production	Billion cubic meters	Proposed by authors	British Petroleum
Oil Consumption	Million tones	Proposed by authors	British Petroleum
Crude Oil Prices	US dollars per barrel ($2013)	[39, 78, 80, 86, 88, 89, 92]	British Petroleum

In machine learning problems, it is very challenging to select a representative collection of features to build the model [93]. Studying more features (a larger feature set), helps to explore more problem dimensions and to reduce the threat of missing potential solutions, but at the same time it may conclude more computational complexity, learning algorithm confusing and over learning.

DM, as a process, generally contains data cleaning, integration, selection, and transformation to discover patterns, evaluate them, and present the extracted knowledge [77, 94]. In knowledge discovery processes, such as DM, the feature subset selection is very crucial, not only for the insight achieved from determining variables, but also for the upgraded reprehensibility, scalability, and the validity of the constructed models [95]. This research uses a correlation-based feature selection (CFs) algorithm to determine the most relevant input features. CFs was initially proposed by Hall in 1999 [93]. The key idea of CFs is the high correlation rate among features and the prediction class (target variable), yet selected features remain uncorrelated with each other [93]. "Best First" [96] and "Greedy Stepwise" [97] searching methods were applied to the CFs to study input dataset using various searching paradigms. Both of searching methods resulted in the same feature subset which means they support each other. Finally, through 13 representative input features (presented in table 3) 6 input features selected as the model's input, contains: (1) alternative and nuclear energy, (2) CO^2 emissions, (3) GDP per capita, (4) urban population, (5) NG production and (6) oil consumption.

Sometimes important features in a time series dataset show their influence with lags of time. Also, there would be time lags for a policy/decision in the complex energy market. Detecting related lags would assist a prediction model to accurately follow possible fluctuations [76]. At this step, the proposed DmGNn methodology attempts to determine time lags related to finalized feature subset correlated with the target attribute (i.e. NG global demand).

Numerous lag selection approaches exist that contain lag selection as a pre-processing, post-processing, or even as a part of the learning process [98]. Among popular statistical tests based on information criteria pre-processing lag selection methods, Akaike information criteria (AIC), Bayesian information criteria (BIC) and Schwarz Bayesian information criteria (SBIC) are well used [99, 100]. Information criteria methods consider 1 lag (as the minimum number) to p which define intermediate lags. The main hypothesis is to define the lag order p to minimize the following equation:

$$IC(p) = N \ln \hat{\sigma}^2(p) + p\left[f(N) \right] \quad (1)$$

Where $\hat{\sigma}^2(p)$ is defined as the estimated regression variance, related to the sample size and order p of lag structure, and N is the number of observations [101]. $p\left[f(N) \right]$ is the penalty function to increase the order of the model. Different choices of $f(N)$ cause in different information criteria.

A -20 to +20 time lags implemented for each feature versus the target attribute using "Matlab" software. Fig 2 summarizes results of the time lag selection process for selected features, alternative and nuclear energy, and CO2 emissions. For each chart, the vertical axis shows the level of correlation between the correspondence feature and targeted variable while horizontal axis implies different time lags. The order p defines the effective time lag which possess the highest correlation level, according to the chart.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 June 2019 doi:10.20944/preprints201906.0202.v1

Peer-reviewed version available at Energies 2019, 12, 4124; doi:10.3390/en12214124

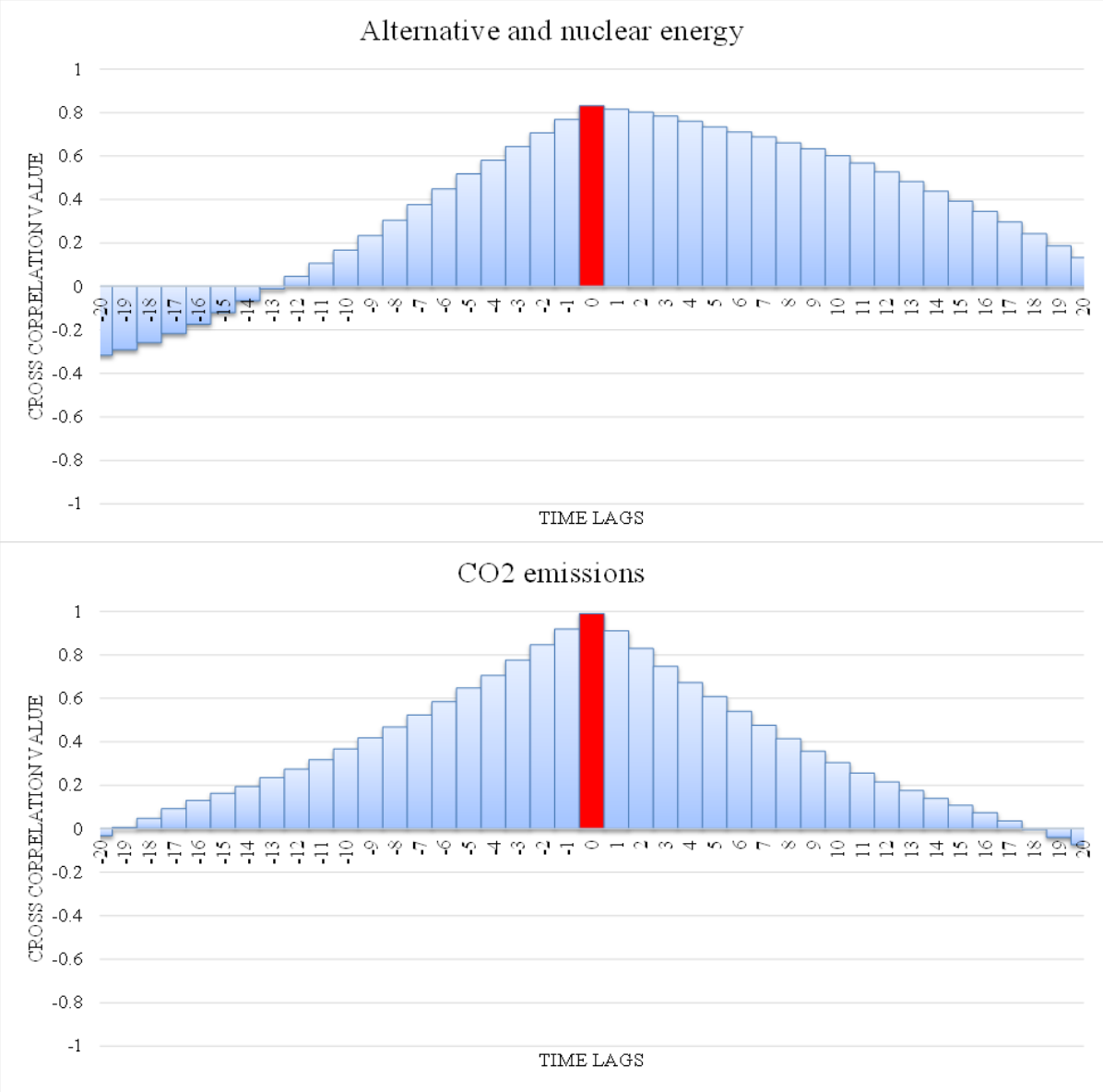


Fig2. Time lag selection results for selected input features (red bars show the time lags that represent higher correlation among the feature versus target attribute for each input feature).

Now, optimum input features are detected. Six selected features are representatives of all 13 identified input features and also the selected subset has been reorganized based on detected time lags.

Although an optimum set of input features have been selected, still input features are asymmetric and the units are different in scales. Data normalization step is investigated to restrain the parameters range influence on the results and adapt values of different features with different domains and scales to a shared scale. The "min-max" normalization method is used to adjust dataset using the following equation:

$$Normalized\ Data = (y(i) - \min\{y\}) / \max\{y\} - \min\{y\} \quad (2)$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 June 2019 doi:10.20944/preprints201906.0202.v1

Peer-reviewed version available at Energies 2019, 12, 4124; doi:10.3390/en12214124

Where $y(i)$ is an i^{th} element in the column and $\min\{y\}$ minimum and $\max\{y\}$ is the maximum of related column's elements.

The next sub-section is dedicated to discussing the forecasting framework.

2.3. Designing the forecasting framework:

2.3.1. Artificial neural network:

Computational intelligence methods such as an artificial neural network (ANNs) [102] are modern paradigms to handle complex optimization problems [103-105]. ANN is organized as a simplified abstract of the biological nervous system to emulate neurons mechanism. A neuron is the computation unit of an ANN. Mathematically a neuron is a function, which aimed at dynamically reduce deviation cost. The mathematical description of a neuron presented as follows:

$$o_j(t) = f\left\{\left[\sum_{i=1}^{n} w_{ij} x_i(t-\tau_{ij})\right] - T_j\right\} \quad (3)$$

Where x_i and o_j respectively are the input and the output at time t, τ_{ij} defines the delay between x_i and o_j. T_j presents the threshold of the j^{th} neuron, while w_{ij} is the connection coefficient from neuron i to neuron j.

An ANN consists of characteristics: the input layer, the hidden layer, the interconnection between different layers, the learning step to find the optimum values of interconnections weights, the transformer function which assigned to produce outputs refer to weighted inputs, the number of neurons performing in each layer and the output layer. Fig 3 schematically presents the architecture of an ANN with a single hidden layer.

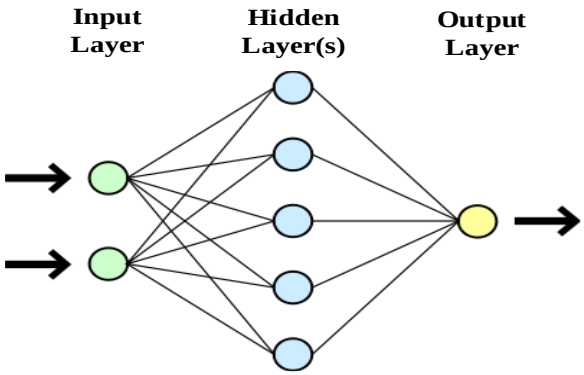


Fig3. A simple artificial neural network

As it has been shown in the fig.3 neurons are deployed in layers. Nodes of layers in row are connected to show interactions and information flow in an ANN. The connection between node i and j defines by the weight w_{ij} and also a bias b_i parameter is assigned to each neuron [106]. To minimize the error at each step (which is known as epoch) an ANN compute and error function and uses an algorithm to reduce the error value.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 June 2019 doi:10.20944/preprints201906.0202.v1

Peer-reviewed version available at Energies 2019, 12, 4124; doi:10.3390/en12214124

An ANN has the ability to be trained in order to build a precise network and minimize the lost function via adjusting w_{ij} weight matrices [76]. So, the performance of learning algorithm will define the performance of the ANN. In this paper, genetic algorithm (GA) is used to equip ANN as the learning algorithm. In the next section, GA procedure is explained briefly.

2.3.2. Genetic algorithm:

Training an ANN is very complex which can directly influence outcomes' quality. Recently, numerous academic studies are presented which applied meta-heuristic and intelligent algorithms (i.e. GA) as learning algorithms [107].

GA is an evolutionary optimization approach developed by Holland in 1975 [108] which acts based on random search procedure. Compared to traditional optimization methods the GA has numerous advantages. For example the algorithm converge to a good, feasible solution faster than other existing traditional methods [11]. Series of computational operators like selection, mutation, and crossover functions are used in a GA to achieve a reliable solution. Fig 4 briefly presents the GA procedure.

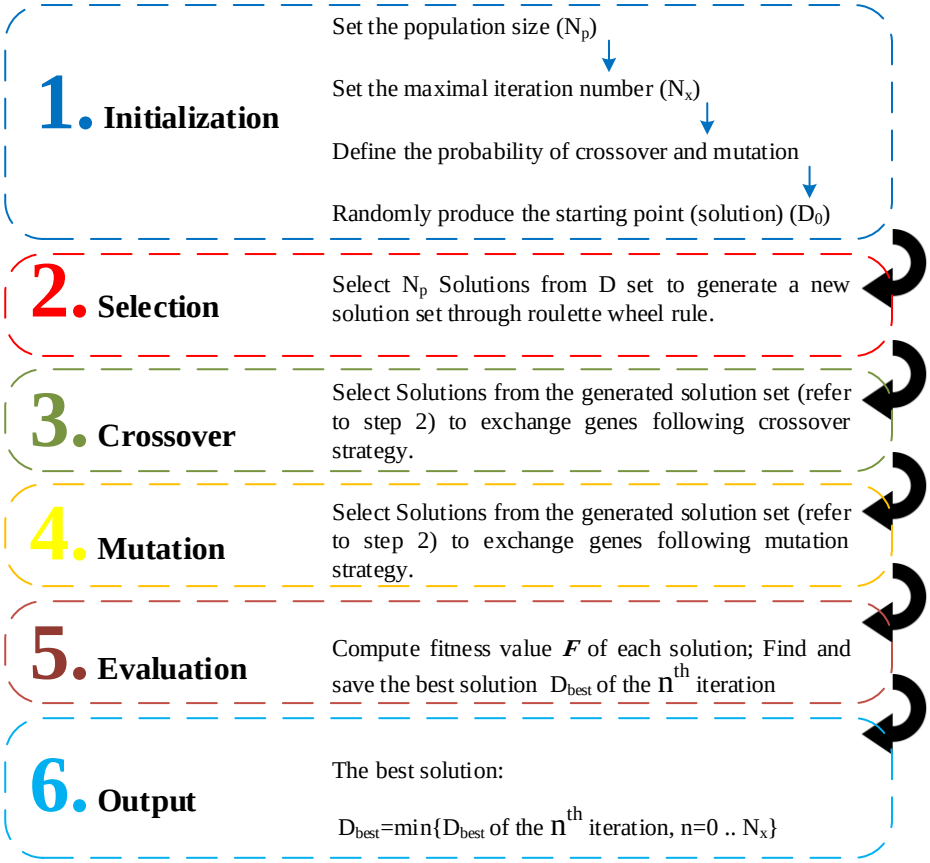


Fig4. The genetic algorithm (GA) procedure

2.3.3. Genetic neural network:

In this paper, weights and thresholds of the ANN are updated by a GA. For this purpose, input vectors transformed to a genetic gene in the format of the chromosome. Then, the initial population is formed from the randomly generated chromosome. Now values of the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 June 2019 doi:10.20944/preprints201906.0202.v1

Peer-reviewed version available at Energies 2019, 12, 4124; doi:10.3390/en12214124

optimization algorithm such as selection, crossover, and mutation rates can be set to design the algorithm. The fitness function is the reciprocal of the quadratic sum of the difference between predicted and real values [109]. Roulette wheel selection is used to select a new individual, then two chromosomes are exchanged via crossover operation to generate a new individual. Finally, mutation step is applied to avoid premature convergence.

Equipping an ANN with a GA could save training time and improve the precision of the forecasting model [109]. Fig 5 schematically shows the flowchart of the presented GNN.

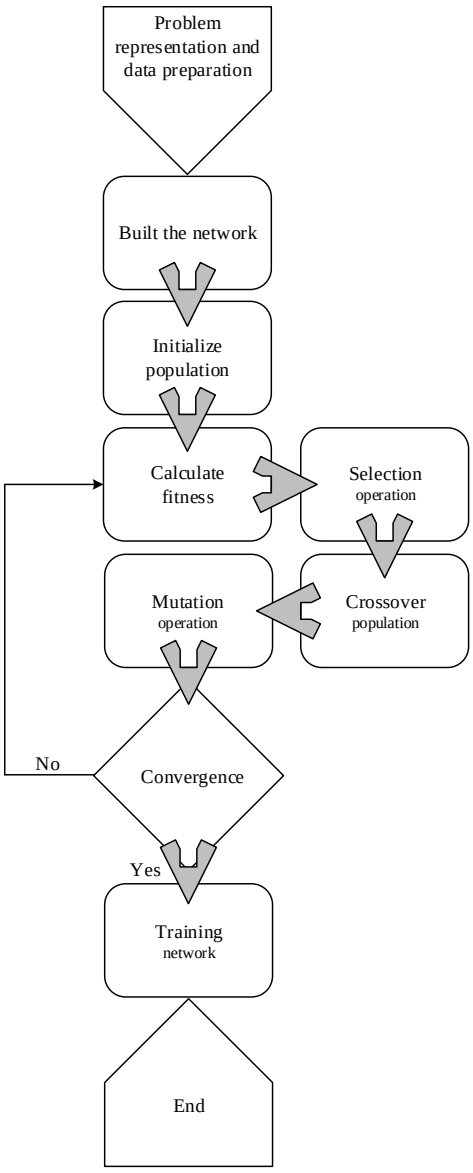


Fig5. Flowchart of a GNN

Next section is dedicated to present the architecture of ANN, which is the basic framework of the developed forecasting model.

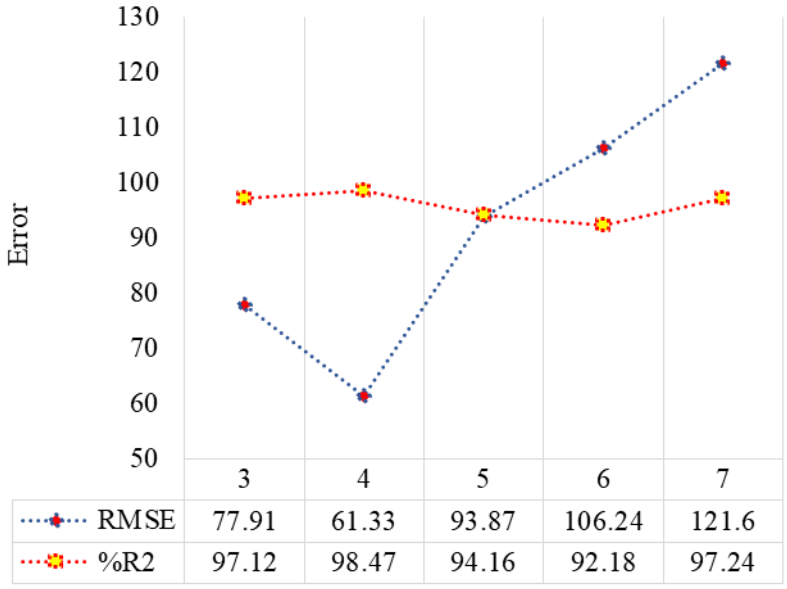
2.3.4. The architecture of the ANN:

This research targeted to present an accurate NG demand predictions, so the selected features were inputted at the initiatory layer (input layer) of the designed ANN. A single hidden layer

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 June 2019 doi:10.20944/preprints201906.0202.v1

Peer-reviewed version available at Energies 2019, 12, 4124; doi:10.3390/en12214124

network was designed to perform the prediction so the model contains a three-layer architecture. Fig6 shows the performance of a three-layered NN for three, four, five, six and seven neurons in the hidden layer. Four neurons were used for the hidden layer as it returns the best performance among other tested number of neurons (see fig6).



	3	4	5	6	7
RMSE	77.91	61.33	93.87	106.24	121.6
%R2	97.12	98.47	94.16	92.18	97.24

Fig6. Performance of DmGNn for different numbers of neurons (A: R2 statistic for the different number of DmGNn neurons; B: RMSE statistic for the different number of DmGNn neurons)

As it has been represented in fig 6, based on the R^2 and root mean square error (RMSE) statistics, four number of neurons the proposed data mining genetic-neural network (DmGNn) model performs better than other examined set.

3. Outputs and Results:

As mentioned before this paper is aimed at developing a forecasting model to accurately forecast global NG demand. Here, historical behavior of the global NG demand during 1965 to 2013 period (billion cubic meters) is gathered via www.bp.com. Now the model is designed and it can be used to project future trends for NG global consumption. For this reason, 40 historical annual fundamental time series data (from 1665 to 2004) are investigated as a learning set. The forecasting period contains 9 annual values for NG global demand prediction problem (from 2005 to 2013). Ten iterations have been investigated for the proposed DmGNn model. Fig 7 presents projections (average for 10 iterations) resulted by the DmGNn models.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 June 2019 doi:10.20944/preprints201906.0202.v1

Peer-reviewed version available at Energies 2019, 12, 4124; doi:10.3390/en12214124

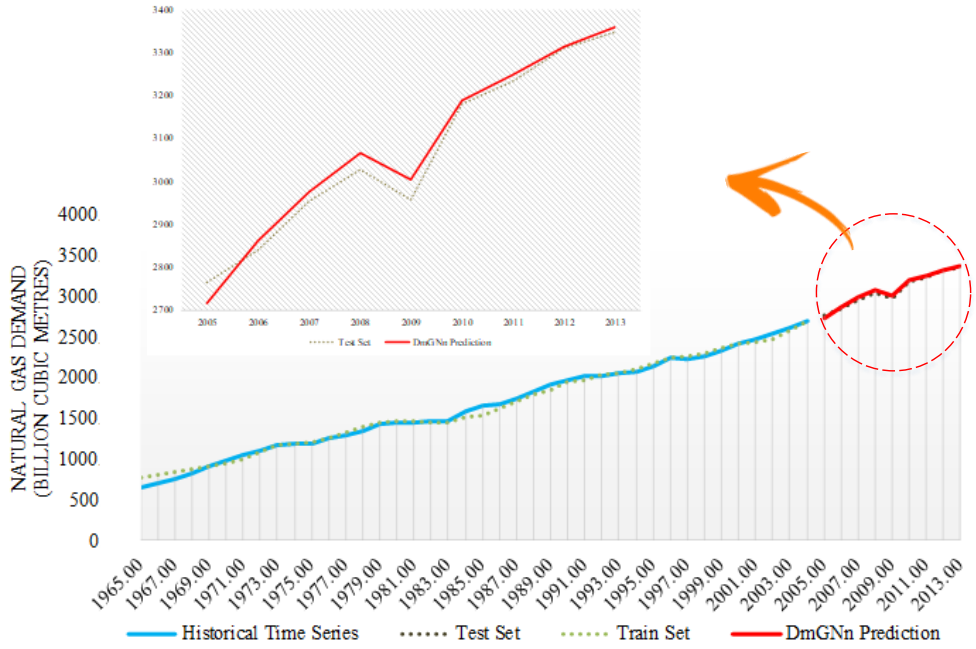


Fig7. Performance of the proposed DmGNn model for the training and testing data sets

Learning models were extensively applied in the case of NG demand predictions [49, 110]. Some competitive prediction models were selected to compare outputs of the proposed model and analysis of the accuracy. Adaptive Neuro-Fuzzy Inference Systems (ANFIS) [111-113] and a set of classical well-known neural network based techniques such as: Radial Basis Function Neural Network (RBF) [114, 115], Multi-Layered Perceptron (MLP) [116, 117] and Generalized Regression Neural Network (GRNN) [118-120] are nominated and optimized (through trial and error processes) to prove the accuracy of the proposed DmGNn model through a comparison study.

To evaluate different models, a set of mathematical criteria organized to measure prediction performance. A relatively large set of validity indicators support the justification of a model usage [8]. These statistics are summarized in table 4 (where y_i refers to real historical value and f_i presents forecasting value).

Table4. Used mathematical criteria to evaluate forecast errors

Error Title	Abbreviation	Formula
R-squared	R2	$R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$ *
Mean Absolute Error	MAE	$MAE = \frac{1}{n}\sum_i \lvert f_i - y_i \rvert$
Mean Absolute Percentage Error	MAPE	$MAPE = \frac{100}{n}\sum_i \left\lvert \frac{f_i - y_i}{y_i} \right\rvert$
Mean Bias Error	MBE	$MBE = \frac{1}{n}\sum_i (f_i - y_i)$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 June 2019 doi:10.20944/preprints201906.0202.v1

Peer-reviewed version available at Energies 2019, 12, 4124; doi:10.3390/en12214124

Root Mean Square Error	RMSE	$RMSE = \sqrt{\frac{\sum(y_i - f_i)^2}{n}}$

* Where, $SS_{tot} = \sum_i (y_i - \bar{y})^2$ and $SS_{res} = \sum_i (y_i - f_i)^2 = \sum_i (errors)^2$.

Each model ran for 10 times and the average of outputs was calculated. Table 5 presents the performance of the proposed and competitive models refer to statistics introduced in table 4.

Table5. Statistical errors for each prediction model

Models	Characteristics	R^2	MAE	MAPE	MBE	RMSE
DmGNn	Number of Neurons= 4; Maximum generation= 100; Cross Over Probability=0.8; Mutation Probability= 0.05;	**0.9847**	**52.19**	**1.69**	13.54	**61.33**
MLP	Maximum Epochs= 200; Train Parameter Goal= 1e-7;	0.8241	115.59	3.80	-44.85	145.61
ANFIS	FIS Generation Approach: FCM; Number of Clusters= 10; Partition Matrix exponent= 2;	0.8494	63.45	1.89	21.31	84.31
RBF	Spread Value= 0.17;	0.0018	308.64	10.42	-308.64	366.51
GRNN	Spread Value= 1;	0.9864	127.63	4.17	**-4.03**	142.12

As it is shown in table 5 the proposed DmGNn significantly outperforms other competitive models. The pattern of the absolute error for each model is shown in Fig 8, which represents how various forecasting models behave along the test period. As it is shown the proposed DmGNn outperforms other benchmark forecasting models (with lower absolute error value for forecasting period) and resulted in a robust forecast series (unlike other forecasting models DmGNn's forecast errors showed a low swing pattern).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 June 2019 doi:10.20944/preprints201906.0202.v1

Peer-reviewed version available at Energies 2019, 12, 4124; doi:10.3390/en12214124

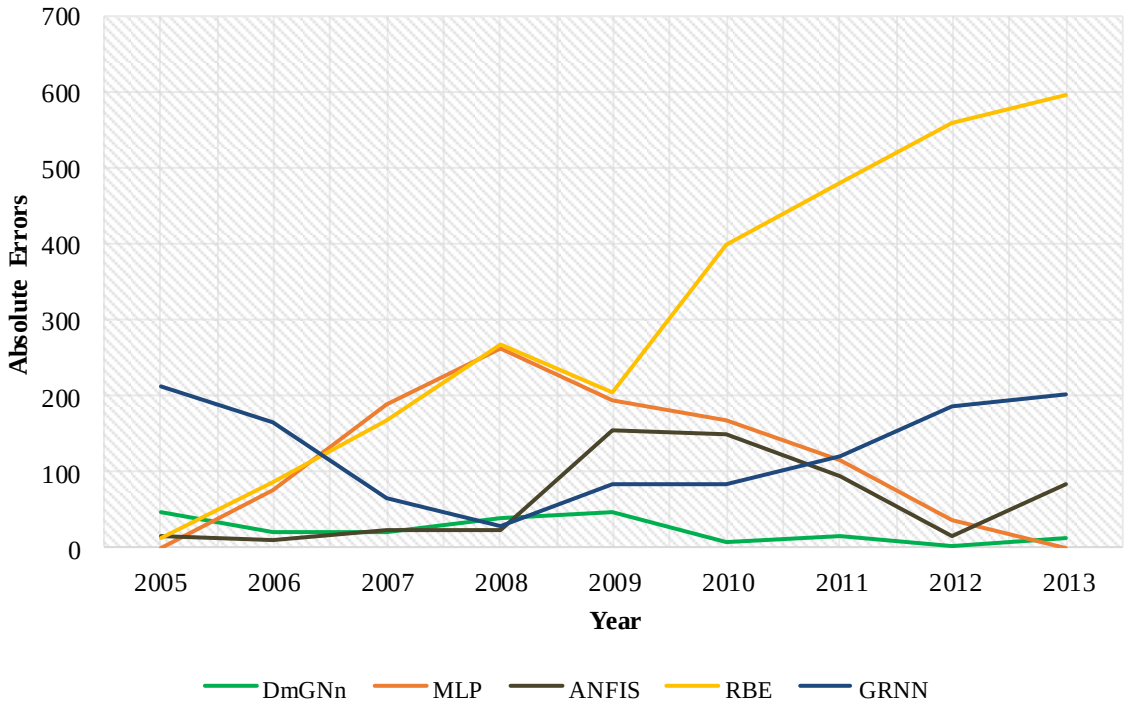


Fig8. Absolute error for each forecasting model along the testing period

Yet table 5 and fig 8 showed how the proposed GNN model overcame other benchmark models. To show the efficiency of the data mining phases, both pre-processed and raw data were applied to the design forecasting model. Fig 9 and table 6 compared the results.

Table 6. Statistical errors for different input vectors (raw versus processed)

Input protocol	R^2	MAE	MAPE	MBE	RMSE
Processed data	**0.9847**	**52.19**	**1.69**	13.54	**61.33**
Raw data	0.9679	79.96	2.61	**2.66**	94.21

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 June 2019 doi:10.20944/preprints201906.0202.v1

Peer-reviewed version available at Energies 2019, 12, 4124; doi:10.3390/en12214124

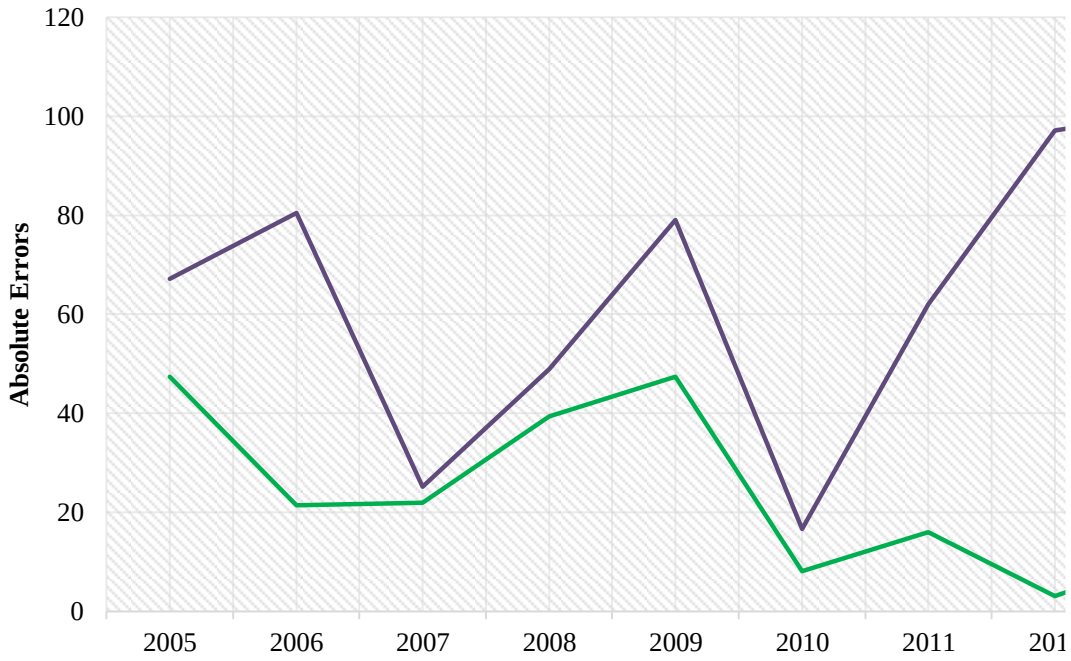


Fig9. Absolute error for each different input protocols along the testing period

4. Conclusion:

Energy is a major topic both in practice and theory which many researchers investigated issues related to energy sectors and industries. The international energy supply system is characterized by a complicated combination of technological, social, economic and political elements. Predicting and planning for future global energy market is an interesting and simultaneously a challenging subject in both research and practical investment projects. Thus accurate prediction of energy demand is critical to developing future policies, modify current plans and evaluate potential strategies. This paper primary targeted to provide an accurate and robust prediction model to predict the global natural gas demands. In other hand authors aimed at introducing a process which reduces problem space dimensions to define the most relevant features which affect NG future consumption trends. So policymakers can monitor and manipulate NG market refers to extracted features.

In order to investigate maximum feasible solutions and to prevent missing any potential optimal solution, all available input features were gathered based on the literature review and related online dataset survey. Input features would define the model structure and support the accuracy of the output results. Although, increasing in the number of input variables may cause computational complexity and reducing interpretability of the results. Instead, a large number of input features expands solution space and consequently reduces the probability of ignoring appropriate answers. A feature selection step is proposed and is implemented to reduce the dataset dimensions while guarantees that the prediction model will explore all optimal solutions. Finally, 6 input features were selected among 13 primary input features. The feature selection approach guarantees to investigate all solution space using a limited set of input features. Then possible time lags among input features versus the targeted attribute (NG global demand) were studied and subsequently applied to the refined input set.

Investigating suitable time lags will cause in a more accurate and rational prediction model, which guarantees synchronization between input features and the target attribute at *t* time step. Finally, a neural network framework is developed which equipped using a genetic algorithm to optimize the network's characteristics aimed to predict future NG global demands.

Four comparative models are investigated to study the performance of the proposed data mining genetic-neural network (DmGNn) model. The proposed DmGNn model outperforms other benchmark models refer to 5 different error statistics. Based on the R2 statistic the DmGNn track real testing set fluctuations very well (only missed about 2%). Moreover, to distinct how the proposed pre-processing step affects the model accuracy, DmGNn model compared to a single GNn (without pre-processing step). As shown the proposed pre-processing step improves predictions both in term of accuracy and reliability (robustness). Moreover, based on the interpretative capability index, the DmGNn dedicates a more clear vision about future trends since it uses a smaller input dataset. A limited input feature set enables decision makers to design responsive policies/strategies/actions as they aware of attributes affecting the global NG demands.

The proposed DmGNn is characterized by high flexibility, universal operation, learning ability and low requirements for computation resources. As a result, it can be used by decision makers and market participants who face a complex environment.

References:

1. BP, *Statistical Review of World Energy*. 2016.
2. IEA, *World Energy Outlook*. 2016.
3. EIA, *Annual Energy Outlook*. 2017.
4. EIA, *Annual Energy Outlook*. 2016.
5. Hafezi, R., A. Akhavan, and S. Pakseresht, *Projecting plausible futures for Iranian oil and gas industries: Analyzing of historical strategies.* Journal of Natural Gas Science and Engineering, 2017. **39**: p. 15-27.
6. Alipour, M., et al., *A new hybrid fuzzy cognitive map-based scenario planning approach for Iran's oil production pathways in the post-sanction period.* Energy, 2017. **135**: p. 851e864.
7. Hafezi, R., M. Bahrami, and A.N. Akhavan, *Sustainability in development: rethinking about old paradigms.* World Review of Science, Technology and Sustainable Development, 2017. **13**(2): p. 192-204.
8. Panapakidis, I.P. and A.S. Dagoumas, *Day-ahead natural gas demand forecasting based on the combination of wavelet transform and ANFIS/genetic algorithm/neural network model.* Energy, 2017. **118**: p. 231-245.
9. Kazemi, S., et al., *A hybrid intelligent approach for modeling brand choice and constructing a market response simulator*. Knowledge-Based Systems, 2013. **40**: p. 101-110.

10. Júnior, S.E.R. and G.L. de Oliveira Serra, *A novel intelligent approach for state space evolving forecasting of seasonal time series.* Engineering Applications of Artificial Intelligence, 2017. **64**: p. 272-285.
11. Ervural, B.C., O.F. Beyca, and S. Zaim, *Model Estimation of ARMA Using Genetic Algorithms: A Case Study of Forecasting Natural Gas Consumption.* Procedia-Social and Behavioral Sciences, 2016. **235**: p. 537-545.
12. Zhu, L., et al., *Short-term natural gas demand prediction based on support vector regression with false neighbours filtered.* Energy, 2015. **80**: p. 428-436.
13. Xu, N., Y. Dang, and Y. Gong, *Novel grey prediction model with nonlinear optimized time response method for forecasting of electricity consumption in China.* Energy, 2016.
14. Bianco, V., F. Scarpa, and L.A. Tagliafico, *Scenario analysis of nonresidential natural gas consumption in Italy.* Applied Energy, 2014. **113**: p. 392-403.
15. Baldacci, L., et al., *Natural gas consumption forecasting for anomaly detection.* Expert Systems with Applications, 2016. **62**: p. 190-201.
16. Zavanella, L., et al., *Energy demand in production systems: A Queuing Theory perspective.* International Journal of Production Economics, 2015. **170**: p. 393-400.
17. Baumeister, C. and L. Kilian, *Real-time analysis of oil price risks using forecast scenarios.* 2011.
18. Dilaver, Ö., Z. Dilaver, and L.C. Hunt, *What drives natural gas consumption in Europe? Analysis and projections.* Journal of Natural Gas Science and Engineering, 2014. **19**: p. 125-136.
19. Li, J., et al., *Forecasting the growth of China's natural gas consumption.* Energy, 2011. **36**(3): p. 1380-1385.
20. Suganthi, L. and A.A. Samuel, *Energy models for demand forecasting—A review.* Renewable and sustainable energy reviews, 2012. **16**(2): p. 1223-1240.
21. Aras, H. and N. Aras, *Forecasting residential natural gas demand.* Energy Sources, 2004. **26**(5): p. 463-472.
22. Erdogdu, E., *Natural gas demand in Turkey.* Applied Energy, 2010. **87**(1): p. 211-219.
23. Gori, F., D. Ludovisi, and P. Cerritelli, *Forecast of oil price and consumption in the short term under three scenarios: Parabolic, linear and chaotic behaviour.* Energy, 2007. **32**(7): p. 1291-1296.
24. Kumar, U. and V. Jain, *Time series models (Grey-Markov, Grey Model with rolling mechanism and singular spectrum analysis) to forecast energy consumption in India.* Energy, 2010. **35**(4): p. 1709-1716.
25. Akkurt, M., O.F. Demirel, and S. Zaim, *Forecasting Turkey's natural gas consumption by using time series methods.* European Journal of Economic and Political Studies, 2010. **3**(2): p. 1-21.
26. Sen, P., M. Roy, and P. Pal, *Application of ARIMA for forecasting energy consumption and GHG emission: A case study of an Indian pig iron manufacturing organization.* Energy, 2016. **116**: p. 1031-1038.

27. Fagiani, M., et al., *A review of datasets and load forecasting techniques for smart natural gas and water grids: Analysis and experiments.* Neurocomputing, 2015. **170**: p. 448-465.
28. Shi, G., D. Liu, and Q. Wei, *Energy consumption prediction of office buildings based on echo state networks.* Neurocomputing, 2016. **216**: p. 478-488.
29. Zhang, W. and J. Yang, *Forecasting natural gas consumption in China by Bayesian model averaging.* Energy Reports, 2015. **1**: p. 216-220.
30. Taşpınar, F., N. Celebi, and N. Tutkun, *Forecasting of daily natural gas consumption on regional basis in Turkey using various computational methods.* Energy and Buildings, 2013. **56**: p. 23-31.
31. Dalfard, V.M., et al., *A mathematical modeling for incorporating energy price hikes into total natural gas consumption forecasting.* Applied Mathematical Modelling, 2013. **37**(8): p. 5664-5679.
32. Bianco, V., F. Scarpa, and L.A. Tagliafico, *Analysis and future outlook of natural gas consumption in the Italian residential sector.* Energy Conversion and Management, 2014. **87**: p. 754-764.
33. Gorucu, F., *Evaluation and forecasting of gas consumption by statistical analysis.* Energy Sources, 2004. **26**(3): p. 267-276.
34. O'Neill, B.C. and M. Desai, *Accuracy of past projections of US energy consumption.* Energy Policy, 2005. **33**(8): p. 979-993.
35. Adams, F.G. and Y. Shachmurove, *Modeling and forecasting energy consumption in China: Implications for Chinese energy demand and imports in 2020.* Energy economics, 2008. **30**(3): p. 1263-1278.
36. Ramanathan, R., *A multi-factor efficiency perspective to the relationships among world GDP, energy consumption and carbon dioxide emissions.* Technological Forecasting and Social Change, 2006. **73**(5): p. 483-494.
37. Lu, W. and Y. Ma, *Image of energy consumption of well off society in China.* Energy Conversion and Management, 2004. **45**(9): p. 1357-1367.
38. Hunt, L.C. and Y. Ninomiya, *Primary energy demand in Japan: an empirical analysis of long-term trends and future CO 2 emissions.* Energy Policy, 2005. **33**(11): p. 1409-1424.
39. Iniyan, S., L. Suganthi, and A.A. Samuel, *Energy models for commercial energy prediction and substitution of renewable energy sources.* Energy Policy, 2006. **34**(17): p. 2640-2653.
40. Sözen, A., E. Arcaklioğlu, and M. Özkaymak, *Turkey's net energy consumption.* Applied Energy, 2005. **81**(2): p. 209-221.
41. Ermis, K., et al., *Artificial neural network analysis of world green energy use.* Energy Policy, 2007. **35**(3): p. 1731-1743.
42. Sözen, A. and E. Arcaklioglu, *Prediction of net energy consumption based on economic indicators (GNP and GDP) in Turkey.* Energy policy, 2007. **35**(10): p. 4981-4992.

43. Sözen, A., Z. Gülseven, and E. Arcaklioğlu, *Forecasting based on sectoral energy consumption of GHGs in Turkey and mitigation policies.* Energy Policy, 2007. **35**(12): p. 6491-6505.
44. Geem, Z.W. and W.E. Roper, *Energy demand estimation of South Korea using artificial neural network.* Energy policy, 2009. **37**(10): p. 4049-4054.
45. Forouzanfar, M., et al., *Modeling and estimation of the natural gas consumption for residential and commercial sectors in Iran.* Applied Energy, 2010. **87**(1): p. 268-274.
46. Ekonomou, L., *Greek long-term energy consumption prediction using artificial neural networks.* Energy, 2010. **35**(2): p. 512-517.
47. Rodger, J.A., *A fuzzy nearest neighbor neural network statistical model for predicting demand for natural gas and energy cost savings in public buildings.* Expert Systems with Applications, 2014. **41**(4): p. 1813-1829.
48. Aydinalp-Koksal, M. and V.I. Ugursal, *Comparison of neural network, conditional demand analysis, and engineering approaches for modeling end-use energy consumption in the residential sector.* Applied Energy, 2008. **85**(4): p. 271-296.
49. Aramesh, A., N. Montazerin, and A. Ahmadi, *A general neural and fuzzy-neural algorithm for natural gas flow prediction in city gate stations.* Energy and Buildings, 2014. **72**: p. 73-79.
50. Szoplik, J., *Forecasting of natural gas consumption with artificial neural networks.* Energy, 2015. **85**: p. 208-220.
51. Soldo, B., et al., *Improving the residential natural gas consumption forecasting models by using solar radiation.* Energy and buildings, 2014. **69**: p. 498-506.
52. Izadyar, N., et al., *Intelligent forecasting of residential heating demand for the District Heating System based on the monthly overall natural gas consumption.* Energy and Buildings, 2015. **104**: p. 208-214.
53. González-Romera, E., M.Á. Jaramillo-Morán, and D. Carmona-Fernández, *Forecasting of the electric energy demand trend and monthly fluctuation with neural networks.* Computers & Industrial Engineering, 2007. **52**(3): p. 336-343.
54. Lee, Y.-S. and L.-I. Tong, *Forecasting energy consumption using a grey model improved by incorporating genetic programming.* Energy Conversion and Management, 2011. **52**(1): p. 147-152.
55. Ozturk, H.K., et al., *Estimating petroleum exergy production and consumption using vehicle ownership and GDP based on genetic algorithm approach.* Renewable and Sustainable Energy Reviews, 2004. **8**(3): p. 289-302.
56. Yu, F. and X. Xu, *A short-term load forecasting model of natural gas based on optimized genetic algorithm and improved BP neural network.* Applied Energy, 2014. **134**: p. 102-113.
57. Askari, S., N. Montazerin, and M.F. Zarandi, *Forecasting semi-dynamic response of natural gas networks to nodal gas consumptions using genetic fuzzy systems.* Energy, 2015. **83**: p. 252-266.
58. Kovačič, M. and B. Šarler, *Genetic programming prediction of the natural gas consumption in a steel plant.* Energy, 2014. **66**: p. 273-284.

59. Mousavi, S.M., E.S. Mostafavi, and F. Hosseinpour, *Gene expression programming as a basis for new generation of electricity demand prediction models.* Computers & Industrial Engineering, 2014. **74**: p. 120-128.
60. Toksarı, M.D., *Ant colony optimization approach to estimate energy demand of Turkey.* Energy Policy, 2007. **35**(8): p. 3984-3990.
61. Ünler, A., *Improvement of energy demand forecasts using swarm intelligence: The case of Turkey with projections to 2025.* Energy Policy, 2008. **36**(6): p. 1937-1944.
62. Paudel, S., et al., *A relevant data selection method for energy consumption prediction of low energy building based on support vector machine.* Energy and Buildings, 2017. **138**: p. 240-256.
63. Amasyali, K. and N. El-Gohary, *Building Lighting Energy Consumption Prediction for Supporting Energy Data Analytics.* Procedia Engineering, 2016. **145**: p. 511-517.
64. Azadeh, A., et al., *A neuro-fuzzy algorithm for improved gas consumption forecasting with economic, environmental and IT/IS indicators.* Journal of Petroleum Science and Engineering, 2015. **133**: p. 716-739.
65. Sun, J., *Energy demand in the fifteen European Union countries by 2010—: A forecasting model based on the decomposition approach.* Energy, 2001. **26**(6): p. 549-560.
66. Tao, Z., *Scenarios of China's oil consumption per capita (OCPC) using a hybrid Factor Decomposition–System Dynamics (SD) simulation.* Energy, 2010. **35**(1): p. 168-180.
67. Alcántara, V., P. del Río, and F. Hernández, *Structural analysis of electricity consumption by productive sectors. The Spanish case.* Energy, 2010. **35**(5): p. 2088-2098.
68. Liu, X., B. Moreno, and A.S. García, *A grey neural network and input-output combined forecasting model. Primary energy consumption forecasts in Spanish economic sectors.* Energy, 2016. **115**: p. 1042-1054.
69. Huang, Y., Y.J. Bor, and C.-Y. Peng, *The long-term forecast of Taiwan's energy supply and demand: LEAP model application.* Energy policy, 2011. **39**(11): p. 6790-6803.
70. Shabbir, R. and S.S. Ahmad, *Monitoring urban transport air pollution and energy demand in Rawalpindi and Islamabad using leap model.* Energy, 2010. **35**(5): p. 2323-2332.
71. Rout, U.K., et al., *Energy and emissions forecast of China over a long-time horizon.* Energy, 2011. **36**(1): p. 1-11.
72. Wang, Z.-X. and D.-J. Ye, *Forecasting Chinese carbon emissions from fossil energy consumption using non-linear grey multivariable models.* Journal of Cleaner Production, 2017. **142**: p. 600-612.
73. Zeng, B. and C. Li, *Forecasting the natural gas demand in China using a self-adapting intelligent grey model.* Energy, 2016. **112**: p. 810-825.
74. Shaikh, F. and Q. Ji, *Forecasting natural gas demand in China: Logistic modelling analysis.* International Journal of Electrical Power & Energy Systems, 2016. **77**: p. 25-32.

75. FazelZarandi, M.H., E. Hadavandi, and I.B. Turksen, *A Hybrid Fuzzy Intelligent Agent-Based System for Stock Price Prediction.* International Journal of Intelligent Systems, 2012. **27**(11): p. 1-23.
76. Hafezi, R., J. Shahrabi, and E. Hadavandi, *A bat-neural network multi-agent system (BNNMAS) for stock price prediction: Case study of DAX stock price.* Applied Soft Computing, 2015. **29**: p. 196-210.
77. Han, J., J. Pei, and M. Kamber, *Data mining: concepts and techniques*. 2011: Elsevier.
78. Arsenault, E., et al., *A total energy demand model of Québec: Forecasting properties.* Energy Economics, 1995. **17**(2): p. 163-171.
79. Tolmasquim, M.T., C. Cohen, and A.S. Szklo, *CO 2 emissions in the Brazilian industrial sector according to the integrated energy planning model (IEPM).* Energy Policy, 2001. **29**(8): p. 641-651.
80. Intarapravich, D., et al., *3. Asia-Pacific energy supply and demand to 2010.* Energy, 1996. **21**(11): p. 1017-1039.
81. Raghuvanshi, S.P., A. Chandra, and A.K. Raghav, *Carbon dioxide emissions from coal based power generation in India.* Energy Conversion and Management, 2006. **47**(4): p. 427-441.
82. Mackay, R. and S. Probert, *Crude oil and natural gas supplies and demands up to the year AD 2010 for France.* Applied energy, 1995. **50**(3): p. 185-208.
83. Parikh, J., P. Purohit, and P. Maitra, *Demand projections of petroleum products and natural gas in India.* Energy, 2007. **32**(10): p. 1825-1837.
84. Nel, W.P. and C.J. Cooper, *A critical review of IEA's oil demand forecast for China.* Energy Policy, 2008. **36**(3): p. 1096-1106.
85. Zhang, M., et al., *Forecasting the transport energy demand based on PLSR method in China.* Energy, 2009. **34**(9): p. 1396-1400.
86. Dincer, I. and S. Dost, *Energy and GDP*. International Journal of Energy Research, 1997. **21**(2): p. 153-167.
87. Kankal, M., et al., *Modeling and forecasting of Turkey's energy consumption using socio-economic and demographic variables.* Applied Energy, 2011. **88**(5): p. 1927-1939.
88. Suganthi, L. and T. Jagadeesan, *A modified model for prediction of India's future energy requirement.* Energy & Environment, 1992. **3**(4): p. 371-386.
89. Suganthi, L. and A. Williams, *Renewable energy in India—a modelling study for 2020–2021*. Energy policy, 2000. **28**(15): p. 1095-1109.
90. Ceylan, H. and H.K. Ozturk, *Estimating energy demand of Turkey based on economic indicators using genetic algorithm approach.* Energy Conversion and Management, 2004. **45**(15): p. 2525-2537.
91. Canyurt, O.E. and H.K. Ozturk, *Application of genetic algorithm (GA) technique on demand estimation of fossil fuels in Turkey*. Energy Policy, 2008. **36**(7): p. 2562-2569.

92. Persaud, A.J. and U. Kumar, *An eclectic approach in energy forecasting: a case of Natural Resources Canada's (NRCan's) oil and gas outlook.* Energy policy, 2001. **29**(4): p. 303-313.
93. Hall, M.A., *Correlation-based feature selection for machine learning*. 1999, The University of Waikato.
94. Wang, J., Y.-I. Lin, and S.-Y. Hou, *A data mining approach for training evaluation in simulation-based training.* Computers & Industrial Engineering, 2015. **80**: p. 171-180.
95. Kim, Y., W.N. Street, and F. Menczer, *Feature selection in data mining.* Data mining: opportunities and challenges, 2003. **3**(9): p. 80-105.
96. Rich, E. and K. Knight, *Artificial intelligence.* McGraw-Hill, New, 1991.
97. Freitag, D. *Greedy attribute selection*. in *Machine Learning Proceedings 1994: Proceedings of the Eighth International Conference*. 2014. Morgan Kaufmann.
98. Parras-Gutierrez, E., et al., *Short, medium and long term forecasting of time series using the L-Co-R algorithm.* neurocomputing, 2014 **128**: p. 433–446.
99. Winker, P., *Optimized multivariate lag structure selection.* Computational Economics, 2000. **16**: p. 87-103.
100. Shahrabi, J., E. Hadavandi, and S. Asadi, *Developing a hybrid intelligent model for forecasting problems: Case study of tourism demand time series.* Knowledge-Based Systems, 2013. **43**: p. 112-122.
101. Burnham, K.P. and D.R. Anderson, *Multimodel inference: understanding AIC and BIC in Model Selection.* Sociological Methods and Research, 2004. **33**: p. 261-304.
102. Rumelhart, D.E., J.L. McClelland, and P.R. Group, *Parallel distributed processing: Explorations in the microstructures of cognition. Volume 1: Foundations*. 1986, The MIT Press, Cambridge, MA.
103. Hadavandi, E., J. Shahrabi, and Y. Hayashi, *SPMoE: a novel subspace-projected mixture of experts model for multi-target regression problems.* Soft Computing, 2016. **20**(5): p. 2047-2065.
104. Hadavandi, E., J. Shahrabi, and S. Shamshirband, *A novel Boosted-neural network ensemble for modeling multi-target regression problems.* Engineering Applications of Artificial Intelligence, 2015. **45**: p. 204-219.
105. Kourentzes, N., *Intermittent demand forecasts with neural networks.* International Journal of Production Economics, 2013. **143**(1): p. 198-206.
106. Teixeira, J.P. and P.O. Fernandesa, *Tourism Time Series Forecast -Different ANN Architectures with Time Index Input.* Procedia Technology, 2012. **5**: p. 445-454.
107. Yadav, A.K. and S. Chandel, *Solar radiation prediction using Artificial Neural Network techniques: A review.* Renewable and Sustainable Energy Reviews, 2014. **33**: p. 772-781.
108. Holland, J.H., *Genetic algorithms.* Scientific american, 1992. **267**(1): p. 66-72.
109. Zhang, J., et al., *Prediction of LBB leakage for various conditions by genetic neural network and genetic algorithms.* Nuclear Engineering and Design, 2017. **325**: p. 33-43.

110. Biswas, M.R., M.D. Robinson, and N. Fumo, *Prediction of residential building energy consumption: A neural network approach.* Energy, 2016. **117**: p. 84-92.
111. Anemangely, M., A. Ramezanzadeh, and B. Tokhmechi, *Shear wave travel time estimation from petrophysical logs using ANFIS-PSO algorithm: A case study from Ab-Teymour Oilfield.* Journal of Natural Gas Science and Engineering, 2017.
112. Zendehboudi, A., X. Li, and B. Wang, *Utilization of ANN and ANFIS models to predict variable speed scroll compressor with vapor injection.* International Journal of Refrigeration, 2017. **74**: p. 473-485.
113. Abdi, J., et al., *Forecasting of short-term traffic-flow based on improved neurofuzzy models via emotional temporal difference learning algorithm.* Engineering Applications of Artificial Intelligence, 2012. **25**(5): p. 1022-1042.
114. Fath, A.H., *Application of radial basis function neural networks in bubble point oil formation volume factor prediction for petroleum systems.* Fluid Phase Equilibria, 2017.
115. Mohammadi, R., S.F. Ghomi, and F. Zeinali, *A new hybrid evolutionary based RBF networks method for forecasting time series: a case study of forecasting emergency supply demand time series.* Engineering Applications of Artificial Intelligence, 2014. **36**: p. 204-214.
116. Heidari, E., M.A. Sobati, and S. Movahedirad, *Accurate prediction of nanofluid viscosity using a multilayer perceptron artificial neural network (MLP-ANN).* Chemometrics and Intelligent Laboratory Systems, 2016. **155**: p. 73-85.
117. Hafezi, R. and A.N. Akhavan, *Forecasting Gold Price Changes: Application of an Equipped Artificial Neural Network.* AUT Journal of Modeling and Simulation, 2018.
118. Park, J. and K.-Y. Kim, *Meta-modeling using generalized regression neural network and particle swarm optimization.* Applied Soft Computing, 2017. **51**: p. 354-369.
119. Hu, R., et al., *A short-term power load forecasting model based on the generalized regression neural network with decreasing step fruit fly optimization algorithm.* Neurocomputing, 2017. **221**: p. 24-31.
120. Lotfinejad, M.M., et al., *A Comparative Assessment of Predicting Daily Solar Radiation Using Bat Neural Network (BNN), Generalized Regression Neural Network (GRNN), and Neuro-Fuzzy (NF) System: A Case Study*. Energies, 2018. **11**(5): p. 1188.