
Article:
Laib, O., Khadir, M. and Mihaylova, L. orcid.org/0000-0001-5856-2223 (2019) Toward efficient energy systems based on natural gas consumption prediction with LSTM Recurrent Neural Networks. Energy, 177. pp. 530-542. ISSN 0360-5442

https://doi.org/10.1016/j.energy.2019.04.075

Toward Efficient Energy Systems Based on Natural Gas Consumption Prediction with LSTM Recurrent Neural Networks

Oussama Laib[a], Mohamed Tarek Khadir[a], Lyudmila Mihaylova[b]

[a] *LabGED, Department of Computer Science, Badji Mokhtar University, Po-Box 12, 23000, Annaba, Algeria*

[b] *Department of Automatic Control and Systems Engineering, University of Sheffield, Sheffield, UK*

Abstract

Finding suitable forecasting methods for an effective management of energy resources is of paramount importance for improving the efficiency in energy consumption and decreasing its impact on the environment. Natural gas is one of the main sources of electrical energy in Algeria and worldwide.

To address this demand, this paper introduces a novel hybrid forecasting approach that resolves the two-stage method's deficiency, by designing a Multi Layered Perceptron (MLP) neural network as a nonlinear forecasting model. This model estimates the next day gas consumption profile and selects one of several local models to perform the forecast. The study focuses firstly on an analysis and clustering of natural gas daily consumption profiles, and secondly on building a comprehensive Long Short Term Memory (LSTM) recurrent models according to load behavior.

The results are compared with four benchmark approaches: the MLP neural network approach, LSTM, seasonal time series with exogenous variables models and multiple linear regression models. Compared with these alternative approaches and their high dependence on historical loads, the proposed approach presents a new efficient functionality. It estimates the next day consumption profile, which leads to a significant improvement of the forecasting accuracy, especially for days with exceptional customers consumption behavior change.

Keywords: Hourly Natural gas consumption; Clustering; Time series; Artificial neural network; Long Short Term Memory; Day-ahead forecast.

Acknowledgement

We thank The National Company for Electricity and Gas (Sonelgaz) and the Prospective and analysis department (DGSP) for their assistance, expertise and for providing the research team with the necessary data, which greatly helped obtaining the presented results.

Email addresses: laib@labged.net (Oussama Laib), khadir@labged.net (Mohamed Tarek Khadir), l.s.mihaylova@sheffield.ac.uk (Lyudmila Mihaylova)

1. Introduction

Natural gas consumption forecasting has proven to be one of the most delicate tasks that power system operators face during the last decades. Energy load prediction in Algeria is highly challenging, due to its substantial European contribution in energy supply, on one hand, and the country size, population and the important economic growth on the other hand. Indeed, the country spans on an enormous area of 2,381,741 km^2 in the north of Africa, between the 19-37° north latitude and 9° west and 12° east longitudes. This has a significant consequence on its climatic and customers behavior diversity, making energy load forecasting even more complicated.

Since 1964, Algeria is amongst the biggest natural gas producers and suppliers in the world, and the energetic flux has never been interrupted. Moreover, the national society of electricity and gas in Algeria (SONELGAZ) has to deal with multiple and important market restrictions in several domains. In particular, the reserve, storage and production which is in an increasing rate constitute a challenge in the last years for different reasons such as irregular supply and periods with peak consumption (see Section 3.1.1).

The search of optimality in forecasting is strongly driven by economic reasons, as natural gas resources are scarce. The increased demand and competition require new "eco-friendly" technologies. Since the Algerian economy is relying substantially on hydrocarbons exports with a strong economic growth, the availability of optimal forecasting natural gas consumption models, is therefore of paramount importance.

Energy consumption studies differ according to the forecasting horizon and may be categorized into: *long-term forecasting* (3 to 20 years), *medium-term forecasting* (a month to 2 years) and *short-term forecasting* (30 minutes to a week) where many different methods have been employed to forecast energy consumption on three basis. In the field of natural gas load forecasting, Soldo [1], provides a detailed overview of related works and a classification based on prediction horizons, model paradigms, geographic areas and other criteria. The last two decades witnessed increased investigations aimed at improving the reliability and accuracy an forecasting methods [2]. The research in this area has been conducted by developing three main classes of techniques: statistical time series based techniques (Seasonal Auto-regressive Integrated Moving Average with exogenous inputs (SARIMAX) [3],[4], functional auto-regressive with exogenous variables [5], vector autoregression [6] and extended Kalman filters [7]), non-linear regression methods (Support Vector Regression (SVR) [8], multivariable adaptive regression splines [9], regression trees [10], etc) and computational intelligence based models (Artificial Neural Networks (ANNs) [11],[12], fuzzy logic [13] and support vector machines [14]).

Numerous types of models, suitable for short-term prediction have been developed using a range of tools, approaches and strategies, from linear to nonlinear paradigms including machine learning and neural networks models. The later have been extensively used in the last decades due to their ability to generalize well from a set of clustered input/outputs. Furthermore, Jolanta Szoplik [15] presents results of gas demand forecast in Szczecin, (Poland) using Multi Layered Perceptron (MLP), a type of ANNs. The modeling process consists of considering both calendar and weather factors to build several MLP models, and the one

with higher quality is used to predict the gas consumption on any day of the year and any hour of the day. Feng and Xiaozhong [16] also propose a MLP ANN. In [16] instead of randomly initializing the network weights and thresholds, genetic algorithms are used for this task and the network is trained with an improved Back Propagation (BP) algorithm by including additional momentum and self-adaptive learning rate. Tonkovič et al. [17] develop an MLP and Radial Basis Function (RBF) to model a 24 hours forecasting for the north-east region of Croatia. Taşpınar et al. [3] propose an MLP and RBF ANNs with time series to perform daily natural gas consumption forecasting in some regions of Turkey based on the ambient air temperature, average cloud cover, relative humidity, wind speed and atmospheric pressure as meteorological data.

Since the first successful applications of ANNs to energy load prediction modeling [18],[19], many researchers have explored alternative types of ANNs. Recent works have focused on Recurrent Neural Networks (RNN) and recurrent Long Short Time Memory (LSTM) networks. In the same context, Kong et al. [20] have practically demonstrated that LSTM networks achieve generally better forecasting performance in individual residential households short-term load dataset compared with standard BP ANN and the k-nearest neighbor approach.

In Potočnik et al. [21] a comparison between static and adaptive forecasting models is presented. Various models are studied, including linear models (Multiple Linear Regression (MLR), linear auto-regressive with exogenous inputs and recurrent auto-regressive with exogenous input), neural network models (with one hidden layer feedforward ANN and a RNN) and Support Vector Regression (SVR) models. Calendar effects and forecasting with different granularity are studied by Lusis et al. [10]. MLR, regression trees, SVR and MLP ANNs applied to 30-minutes forecasting intervals, are shown to improve the accuracy of one day-ahead forecasting strategies.

Considering short-term natural gas load forecasting as time series modeling, another noteworthy comprehensive review is presented by Deb et al. [22]. The study analyzes thoroughly the most widely used techniques to forecast historical peak energy time series data. Furthermore, Deb et al. show that hybrid models some of the most efficient techniques for time series energy forecasting, especially for buildings.

Hybrid forecasting techniques have been developed and have practically proven their efficiency by achieving high accurate forecast results compared with single model approaches [23],[24]. They have demonstrated their reliability in solving complex non-linear prediction and control problems. Hernandez et al. [25] propose four MLP models to estimate the peaks and valleys in the electrical load time series, then feed the obtained results together with other variables as inputs into another MLP models to predict the next day's total load. In a similar manner, Krzysztof and Tomasz [26] presente an approach based on classifying the peaks of the time series load using three models with respect to the number of peak levels. ANN, SVM and Random Forest (RF) models are built and tested to accomplish the peaks classification step. In the forecasting step, another ANN, SVM and FR single model is used to forecast the next 24 hours load based on the estimated peak. Following the same context, Ilic et al. [27] present a hybrid structure based on two MLP models, where the first one predicts integrated load value of the next day, the output is then used as input and fed to

the second MLP network in order to obtain the forecasts for each of the 24 hours for the forecasting day. Ghadimi et al. [28] are interested in forecasting the electrical load and price and propose a multistage hybrid forecasting approach that employs three different models. ANN, RBFNN and SVM predict the load obtained from feature selection process, then combine the three forecasts to generate a single unified price/ load forecast with a modified ordered weighted average.

A different, less used, class of hybrid approaches are the two-stage adaptive architectures also known as the divide and conquer approaches. The key idea of the two-stage approaches is: firstly, divide the time series data into several subsets for the purpose of minimizing the non-stationarity recorded in the data and reducing the input space. Secondly, model and predict the time series load in each subset with a separate model. Finally, a chosen technique is applied to integrate the model results. Zhang et al. [29] propose an approach that decomposes the original electricity load into two components, a linear and a non-linear one. Considering the trend as the linear component, the non-linear trend is extracted by subtracting the trend from original electricity load. Furthermore, both components are predicted separately using ARIMA and wavelet ANN, then model results are summed to obtain the final forecasting loads. Substantiating the concept of the superiority of divide and conquer hybrid approaches over the traditional single model ones, Xiao et al. [30] propose a hybrid method to model yearly energy and oil consumption time series which does not include seasonal patterns. The basic idea is to predict the linear trend with an AR model and focuses on the left residual sequence which is the non-linear subseries, then utilises four ensemble prediction models, namely, BP ANN, SVR, GP and RBF ANN. The final decision is made by a special ANN model to establish selective combination forecasting. Thus, both outputs from AR model and the optimal complexity model are added to obtain the final energy consumption value.

In order to perform efficient planning management, it is necessary to best match the daily and weekly seasonality affected by customers' decisions and weather conditions with historical consumption profiles using adequate and accurate models.

1.1. Main Contributions

In response to the challenges that energy systems face this work proposes a novel hybrid approach and evaluates its efficiency. This paper advances the current state-of-the-art methods and has the following main contributions: *i*) an adaptive hybrid architecture is proposed for natural gas consumption forecasting in a large geographic area. *ii*) LSTM models is proposed that can predict efficiently the natural gas consumption. *iii*) a MLP model is proposed to estimate the next day consumption profile.

Besides the geographical area that has not been studied before, this work focuses in the first place on carefully identified clusters of a daily consumptions. This step known as the consumption profile identification step divides the dataset into several subsets using K-Means in order to reduce the non-stationarity of the time series. Consequently, a number of local models matching the number of clusters are developed in a process known as the modeling phase. Multiple LSTM ANN models are constructed and assessed according to the nature of daily natural gas consumption profiles included in each cluster with respect

to existing exogenous factors, rising the issue of managing and choosing the appropriate model for each forecast. Finally, a forecasting MLP ANN classification model, is designed and trained. After conducting several experiments, one of several local LSTM developed models is selected. The performance of the developed approach is thoroughly compared with a number of benchmark state-of-the-art approaches.

2. Outline of the proposed approach

The developed approach consists of several stages and these are shown on Figure 1. The first stage considers and identifies the customers the consumption profiles. A clustering method is used for the purpose of reducing both of the non-stationarity of the natural gas consumption time series data and the inputs space. The second stage consists in modelling the gas consumption profile. After dividing the data, multiple models are developed according to the nature of each data subset.

This two-stage forecasting approach possesses high learning and prediction capabilities compared with conventional approaches since it is based on a global model and identified clusters, obtained either from the training or test datasets [31]. Despite the effectiveness on minimizing the range of error, one the limitation of this approach is that, there is no certain functional process could determines which model should perform the prediction of the next day, this is the case when attempting to forecast new days which their consumption profile has not been identified.

Besides realizing day type consumption classification and modeling an efficient natural gas consumption with LSTM models, this paper introduces a novel monitoring stage that resolves the two-stage approach's deficiency, by developing a combinatory MLP used as a Forecasting Monitor (FM). The role of the FM-MLP model lies in estimating the next-day consumption profile according to the similarity measures calculated during the clustering phase, then choosing the right local LSTM model to perform the forecasting for next day.

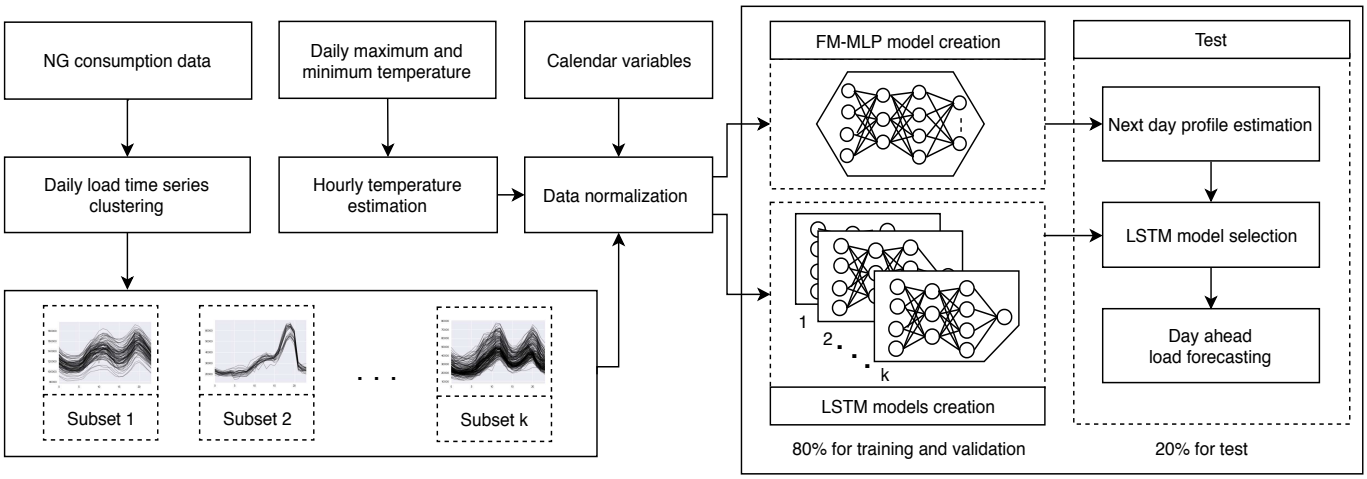


Figure 1: Main steps of the proposed approach.

3. Methodology

3.1. Day consumption profiles classification

3.1.1. Available data

According to the Annual report from the Algerian ministry of Energy [32] , the production of natural gas in Algeria reached 96.6 Bm^3 for the sole year 2017, representing 55% of all primary energy production types.

The total national energy consumption (including losses) reached 59,6 MTep in 2017. This is equivalent to (760 BTh 76 Bm3) reflecting an increase of 2.1% when compared to 2016 dominated by natural gas 37%, electricity 30% and oil product 27% consumptions percentages. This increase, when compared to 2016, is mainly driven by the increase of the electricity consumption (5.5% increase). More than 90% of electricity production in Algeria is thanks to natural gas driven technologies [33]. Natural gas consumption witnesses a 1.4% increase, and GPL (Liquid natural gas, mainly as fuel vehicle for transport purposes) with a 5% increase.

It is clear from the above statistics that energy consumption is due to social economical growth of the country mainly driven by the domestic demand (44%) with a total number of energy subscribers of 9,2 millions against 8,8 millions at the end of 2016. Liquid gas also has a demand increase, mainly driven by GPL/C (liquid gas for vehicles) with an increase of 30% due to the promotional price of $0,076 per liter. Table 1 shows the evolution over the last two years of national natural gas consumption in m^3 and electricity in GWh. The amount of natural gas needed to produce electricity power is also calculated, taking into account the average specific consumption factor of steam and combined cycle turbines (approximately equal to 2.8), which constitute the basis of electricity power generators.

Table 1: Natural gas consumption over the last two years

Product	**Units**	**2016**	**2017**	**Evolution**	
				Quantity	**%**
Natural Gas	10 m^3	22 997	23 311	297	1,4
Electricity	GWh	70 748	75 675	932	5,5
Estimation of natural gas for Electricity production	10 m^3	17 828	19 070	1 242	5,5
Total gas consumption	10 m^3	40 825	42 381	1 539	3,63

If the national demand trend in energy is continues in the next few years, the national production in natural gas will barely suffice to cover it. Consequently, the export capabilities of Algeria by 2030 will be reduced to 10 Bm^3, which could lead to a critical situation for an economy relying on exports of fossil resources [34].

The present approach is validated over Algerian hourly natural gas consumption data for both residential and industrial sectors for 2014 provided by the national electric and gas company SONELGAZ. However, hourly temperature data are missing, with only minimum and maximum daily temperatures provided. In the lack of real hourly temperature measurements, linear and cubic interpolation fails to represent the daily temperature profile dynamics. For this purpose daily hourly temperature profiles are approximated using the

minimum and maximum daily values recorded by the Algerian meteorological office using the equations introduced by Linvill [35]. Daily generated temperature profiles [36] are used as exogenous inputs for the prediction of natural gas load. Figure 2 presents hourly natural gas consumption for the residential and industrial sector and generated daily temperature profiles. Moreover, the load fluctuation through the entire observed year shows a strong dependence with the time of the year (month) as well as the temperature.

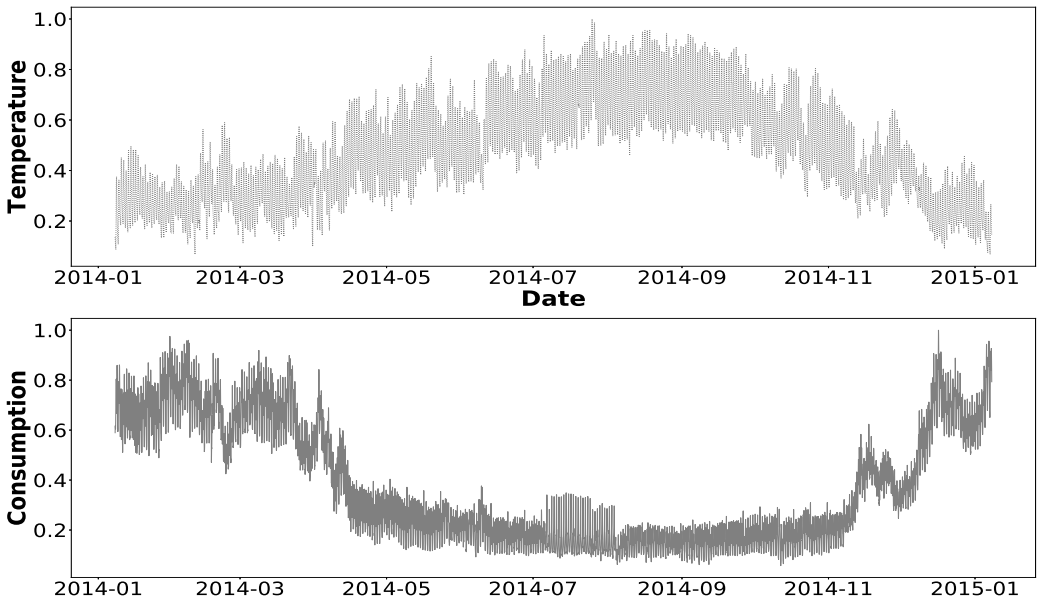


Figure 2: Hourly recorder natural gas consumption and temperature during 2014.

Beside temperatures, natural gas consumption variability over time is dependent on different essential exogenous variables when considering LSTM neural networks modeling. The most influential input variables could be grouped in: historical factors, meteorological factors (such as temperature, wind speed, humidity and sunshine) as well as economic factors (oil prices, number of clients, GDP and natural gas price). Considering short term forecasting, on an hourly basis, it is highly recommended to take into account other types of factors in order to cover the seasonality pattern, usually in the form of one hot encoding inputs (eg.,hour of the day, day of the week and month of the year) [37].

Due to the non-stationarity in historical gas consumption time series, the daily load time series are regrouped using a clustering method. Based on the results obtained from the clustering phase, each day in the dataset is labeled and grouped with the other days that have a similar curve according to the clustering method measures. A local forecasting approach based on multiple models is adopted in [38], where each model is constructed to focus on a single clustered subset. Usually this approach outperforms global model-based forecasting approaches, where a single model handles the forecasting process for the entire output space.

The first stage of the proposed approach is to classify samples of historical natural gas load segments from the year 2014 with: $H_{2014} = \{d_0, d_1, ..., d_{364}\}$ into k clusters containing

identical daily load time series, where each historical segment (day x) is represented by a 24 hours consumption vector $d_x = \{l_0, l_1 ... l_{23}\}$ with l_h being the consumption at hour h. From a statistical point of view the problem in this case is considered as an unsupervised time series classification problem where the number k is unknown. To achieve this challenge, two questions should be answered: which clustering technique should be chosen and, if appropriate, how many clusters should be created. Nevertheless, There are more than a unique answer to both questions as it is dependent on sensitive factors. In the light of previous conducted work on the Algerian natural gas consumption dataset [39], where three different clustering methods were applied to non-supervisely classify the load profiles, namely: K-means, Mixture of Hierarchical Gaussian Process (MHGP) method [40], which combines Gaussian processes (GPs) approach to model the time series and Dirichlet processes (DPs) to perform clustering. The third is a Hierarchical Based Clustering with Noise (HDBSCAN) [41]. The approach is a non parametric method similar to MHGP. By looking at enhancing the overall forecast quality through, firstly, data segmentation investigating several techniques, K-means seems to be the most suitable one for the present dataset. In addition, K-means proved their efficiency when it comes to clustering large databases, including Algerian electrical data load in [36] a time series very much similar to the present one and will be used here.

3.1.2. K-Means clustering

A centroid-based clustering K-means is applied to find the appropriate grouping of daily consumption profiles, giving $C = \{c_1, \ldots, c_k\}$ denoting some partition of H_{2014} into k disjoint nonempty clusters, and $n(c)$ denoting the number of data points assigned to a given cluster c. To determine the structure of day type, the k-means method follows an iterative algorithm of two steps [42] where every data observation d_i is assigned to its nearest cluster centroid μ_i to form the clusters, then adjust the means $(\mu_1, \ldots, \mu_k)$ so that the total within-cluster sum of squared φ is minimized given by the distortion in Equation (1), where D is the Euclidean distance between each point to its centroid. This process keeps converging until the assignments are unchanged so the means remains unmoved when updated

$$\varphi(C) = \sum_{c=1}^{K} \sum_{i \in c} D(d_i, \mu_i) \tag{1}$$

Beginning with a random selection of k means, the data points assignment occurs according:

$$c_i = \underset{k}{argmin}\{D^2(\mu_k - d_i)\} \tag{2}$$

where c_i is the index of the cluster which the point d_i is assigned to. The means are updated based on the new partition through:

$$\mu_c = \frac{1}{n(c)} \sum_{i \in c} d_i \tag{3}$$

3.2. Artificial neural network for natural gas forecasting

Recently, there has been a significant interest in the development of neural network algorithms for various applications. This has successfully brought ANNs on the top of numerous fields when it comes to solve non-linear problems. Because of the ability to approximate any measurable non-linear function to a certain desired degree of accuracy (Cybenko theorem [43]), ANNs are proven powerful universal approximators [44]. For function approximation, two major types of ANNs are used: feed-forward and recurrent architectures. Nowadays, Convolutional Neural Networks (CNN) are the most prominent model in the feed-forward variations, where they are destined and applied excessively for computer vision and pattern recognition tasks [45]. However, CNNs are highly profitable when the problem characteristics are not well defined as they are extracted in the first phase of the training algorithm. This is not the case when it comes to a well defined function approximation like in natural gas load forecasting, where the regression input vector (characteristic vector) is known a priori. On the other hand, another type of RNN is LSTM, firstly introduced by Hochreiter and Schmidhuber where they demonstrated in [46] that LSTM can help solving many unsolved tasks using standard learning algorithms for RNN, especially for sequence processing tasks.

3.2.1. MLP for daily consumption profile estimation

In this particular part of the study, a MLP is developed to estimate the next day consumption profile, playing the role of FM, by properly selecting a LSTM model to handle the day ahead natural gas consumption prediction. The FM-MLP model is characterized by a network of an input layer, two hidden layers and an output layer. The number of neurons in the input layer is equal to the number of the selected variables, where a different input sample sizes are considered to examine the effect of their selection on the estimation accuracy. The output layer consists of numerous neurons with respect to the recognized daily consumption profiles (number of clusters). The MLP topology is presented in the right side of Figure 3, where the neurons output is forwarded to the successive layer, the output of the j^{th} unit is computed as follows:

$$y_j = f\left(z_j\right) \qquad \text{where} \qquad z_j = \sum_{i=1}^{I} w_{ji} y_i + w_{j0} \tag{4}$$

where w_{ji} is the weighted connection between the i^{th} neuron to the j^{th} neuron in the next layer, and $w_{j,0}$ corresponds to the bias which is considered as an external neuron input. Here, I denotes the number of neurons in the source layer and f represents the activation logarithmic sigmoidal function used in this study:

$$\text{sigmoid}: \qquad f(x) = \frac{1}{1+e^{-x}} \tag{5}$$

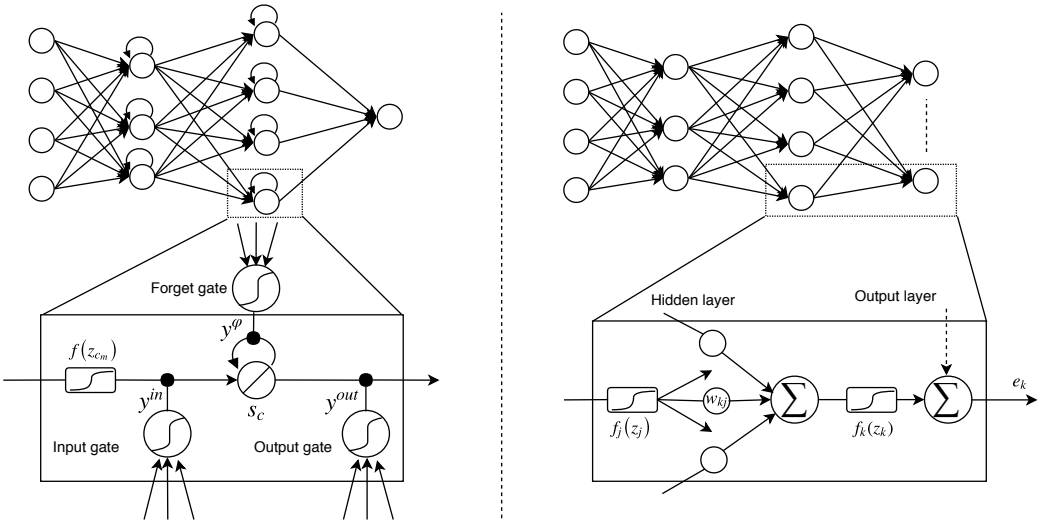


Figure 3: An LSTM memory block with one cell (left side). An MLP network structure with information processing through neurons (right side).

MLP training is performed using the BP algorithm, and is accomplished through adjusting the gradient weights to improve the ANN performance, in other term, minimizing the sum squared error which is calculated by:

$$E(t) = \frac{1}{2}\sum_{k=1}^{K} e_k^2 = \frac{1}{2}\sum_{k=1}^{K}(d_k - o_k)^2 \tag{6}$$

where E denotes the summed square error, k is the output layer neurons count, d_k and o_k represent the k^{th} target output and the network output respectively.

During the online training, the network weights are updated every time a data pattern is processed. The update process follows the idea of taking a step in the direction of the steepest descent, which is the direction of the negative gradient, to reach a minimum [47]. Thus, the weight update is given by:

$$w_{ji}(t+1) = w_{ji}(t) + \Delta w_{ji}(t) \tag{7}$$

and

$$\Delta w_{ji}(t) = -\mu\frac{\partial E}{\partial w_{ji}} = \mu\delta_j y_i \tag{8}$$

In summary, applying the chain rule using the learning rate μ which is defined in the beginning of the training cycle, and thus, controls the training speed and stability of the network [48], the gradient of the cost function with respect to weights is then:

$$\text{for weights in the output layer:} \quad \delta_k = 2f_k'(z_k)e_k \tag{9}$$

$$\text{for weights in the hidden layer:} \quad \delta_j = 2f_j'(z_j)\sum_{k=1}^{K}\delta_k w_{kj} \tag{10}$$

3.2.2. LSTM for natural gas consumption forecasting

RNNs are inherently different compared to MLP. Even if the forward pass remains the same, with the exception that activations arise at the hidden layer from both the external input of the neuron itself and the previous time-step hidden layer activations. This way the network can establish the temporal correlations between previous information and the current state.

As stated before in section. 1, LSTMs have been found to outperform other traditional RNNs on tasks involving long time lags, their architecture permits to perform many more successful runs and much faster learning compared with Real Time Recurrent Learning (RTRL), Back Propagation Through Time, Recurrent Cascade-Correlation and Elman Nets [46]. Once daily consumption profiles are regrouped, several LSTM models are trained to learn the data for each cluster, the FM-MLP takes charge of selecting the right LSTM model in order to handle the forecasting task, which means predicting all hourly loads for the corresponding cluster. Typically, LSTM differs from the standard ANNs with its hidden layer units, where the summation units are replaced instead, by memory blocks. As shown in the left side of Figure 3, that illustrates an LSTM memory block with a single cell, each block contains one or more self-connected memory cells and three sources of inputs: the input z_m^{in}, output z_m^{out} and the cell itself z_m^{φ}. Each source is squashed with an activation function known as a gate that provides continuous regulators of write, read and reset operations for the cells. Consequently, the calculation of y_m in Equation (4) is replaced with the following sequence of equations

$$z_{c_m}(t) = \sum_{m=0}^{k} w_{c_m j} y_j(t-1) \tag{11}$$

where $z_{c_m}(t)$ is the network cell input which is firstly calculated during each forward pass, then

$$z_m^{in}(t) = \sum w_{mj}^{in} y_j(t-1) \quad ; \quad y_m^{in}(t) = f_m^{in}(z_m^{in}(t)) \tag{12}$$

$$z_m^{\varphi}(t) = \sum w_{mj}^{\varphi} y_j(t-1) \quad ; \quad y_m^{\varphi}(t) = f_m^{\varphi}(z_m^{\varphi}(t)) \tag{13}$$

$$z_m^{out}(t) = \sum w_{mj}^{out} y_j(t-1) \quad ; \quad y_m^{out}(t) = f_m^{out}(z_m^{out}(t)) \tag{14}$$

through the current study m refers to memory block with only one cell c_m, see [49] for details. Moreover, s_{c_m} indexes the cell state of the m^{th} memory block which updated according to Equation 15

$$s_{c_m}(t) = y_m^{\varphi}(t) s_{c_m}(t-1) + y_m^{in}(t) f(z_{c_m}(t)) \tag{15}$$

with

$$s_{c_m}(0) = 0. \tag{16}$$

LSTM network training is a fusion of BP for output units and gate weights, and quietly modified and truncated version of RTRL for input weights, input gates and forget gates. Gradient truncates after one time-step and not by the flow of activation around the recurrent connection, thereby mitigating the gradient vanishing and exploding problem [50]. This eases

the implementation of the algorithm, which is an important property for tasks such as time series prediction [51]. For a full derivation of the algorithm see [52].

For the output units, weight changes via gradient descent given by Equations (7), (8) and (9), output gate weights are also obtained by standard back propagation

$$\Delta w_{mj}^{out}(t) = \mu \delta_m^{out}(t) y_j(t) \tag{17}$$

$$\delta_m^{out}(t) \stackrel{tr}{=} f'^{out}_m(z_m^{out}(t)) \Bigg(\sum s_{c_m}(t) \sum w_{kc_m} \delta_k(t) \Bigg) \tag{18}$$

here $\stackrel{tr}{=}$ represents error truncation.

$$\frac{\partial s_{c_m}(t)}{\partial w_{mj}^{in}} \stackrel{tr}{=} \frac{\partial s_{c_m}(t-1)}{\partial w_{mj}^{in}} y_m^{\varphi}(t) + f(z_{c_m}(t)) f'^{in}_m(z_m^{in}(t)) y_j(t-1) \tag{19}$$

$$\frac{\partial s_{c_m}(t)}{\partial w_{c_m j}} \stackrel{tr}{=} \frac{\partial s_{c_m}(t-1)}{\partial w_{c_m j}} y_m^{\varphi}(t) + f'(z_{c_m}(t)) y_m^{in}(t) y_j(t-1) \tag{20}$$

$$\frac{\partial s_{c_m}(t)}{\partial w_{mj}^{\varphi}} \stackrel{tr}{=} \frac{\partial s_{c_m}(t-1)}{\partial w_{mj}^{\varphi}} y_m^{\varphi}(t) + s_{c_m}(t-1) f'^{\varphi}_m(z_m^{\varphi}(t)) y_j(t-1) \tag{21}$$

internal state error $e_{s_{c_m}}$ is calculated separately for each memory cell in order to calculate weights changes

$$e_{s_{c_m}}(t) \stackrel{tr}{=} y_m^{out}(t) \Bigg(\sum w_{kc_m} \delta_k(t) \Bigg) \tag{22}$$

Weights corresponds to connections to the input, the cell and the forget gates are updated using the partials from Equations 19, 20 and 21:

$$\Delta w_{c_m}(t) = \mu e_{s_{c_m}}(t) \frac{\partial s_{c_m}(t)}{\partial w_{c_m j}} \tag{23}$$

$$\Delta w_{mj}^{in}(t) = \mu e_{s_{c_m}}(t) \frac{\partial s_{c_m}(t)}{\partial w_{mj}^{in}} \tag{24}$$

$$\Delta w_{mj}^{\varphi}(t) = \mu e_{s_{c_m}}(t) \frac{\partial s_{c_m}(t)}{\partial w_{mj}^{\varphi}} \tag{25}$$

3.3. Experiments

3.3.1. Clustering

In order to use the K-means clustering method it is necessary to initialize the number of clusters k in advance. Many criteria could be used to find the optimal count of k whether by considering compactness, separation or both in terms of summation or ratio. On one hand, as clustering evaluation indices based on within and between-cluster distances, the Silhouette index determines the optimal number of clusters by maximizing the value of

the pairwise difference of between and within-cluster distances [53], the Calinski-Harabasz criterion validates the clustering by calculating the average between- and within-cluster sum or squares [54], the Davies-Bouldin index determines the best partition by choosing the minimum ratio of within-cluster scatter to between-cluster separation [55]. On the other hand, some criteria focus only on one aspect, such as the R-squared index that validates the clustering by measuring the homogeneity level between clusters [56]. Since the main purpose of the clustering phase in this study is regrouping the similar load profiles, it is more significant to validate the clusters by looking at the total within-cluster sum of squares without considering the between-cluster separation. After conducting several clustering processes with different count of groups, and according to the Elbow method [57], 3 clusters was selected as the best choice. In addition to these three identified classes (c_1, c_2 and c_3), the Algerian natural gas load behaves differently and the consumption patterns are sensibly different to nominal consumption in special periods such as the month of Ramadan and the special period of national and religious holidays labelled c_4 and c_5 respectively.

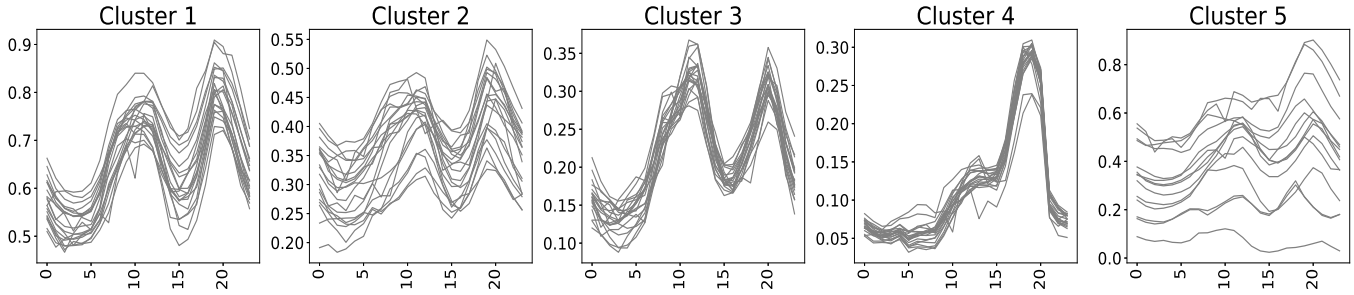


Figure 4: Average load for the obtained clusters.

Once the clustering process is accomplished, 365 daily time series $d_x[l_0...l_{23}]$ in the dataset H_{2014} are labeled with its correspondent cluster c_i. Figure 4 shows 20 daily load samples from each cluster.

Table 2: Season's day count per cluster

Seasons	**Winter**	**Spring**	**Summer**	**Autumn**
Cluster 1	82	0	0	18
Cluster 2	2	15	0	27
Cluster 3	0	74	64	40
Cluster 4	0	0	29	0
Cluster 5	4	4	1	5

These numbers represent days included in each cluster.

Table 2 shows how days included in each cluster spread across all four seasons. It can clearly be seen, that cluster c_1 covers most of the the winter observations with some days belonging to autumn. In addition, cluster c_2 is composed of two weeks from the spring season and about a month of autumn days. Cluster c_3 contains the rest of observation of the year with the exception of special days and periods. In order to complete the classification, cluster c_4 is a subset of c_3 representing 29 days of the month of Ramadan. For being a group of holidays only, cluster c_5 lays on a very limited count of observation across the year.

3.3.2. Neural networks for natural gas consumption modeling

After daily consumption profiles identification is performed, the natural gas consumption profiles data is then divided into three randomly drawn subsets: the first set, contains 60% of the input data is used for training to adjust the model parameters. The second set representing 20% of the input data, is used for the validation to ensure avoiding overfitting and the remaining 20% are saved and used for tests in order to assess the quality of the proposed approach. The evaluation using the first two sets was carried out by each model, representing each cluster, separately, while the third set is used to test the performance of the FM-MLP and LSTMs models combination.

In addition to historical and exogenous attributes, additional types of calendar information is included in the experiments. The first attribute is a day indicator identifying the day of the week for each day forecast. This could help distinguishing between working days and holidays and distinguishing between the first and last days of the week. Figure 5 shows the variety of the daily average load per week on the generated clusters. The second calendar variable is an hourly indicator which captures the periodic pattern during the day. For representing long term changes, a monthly indicator was used.

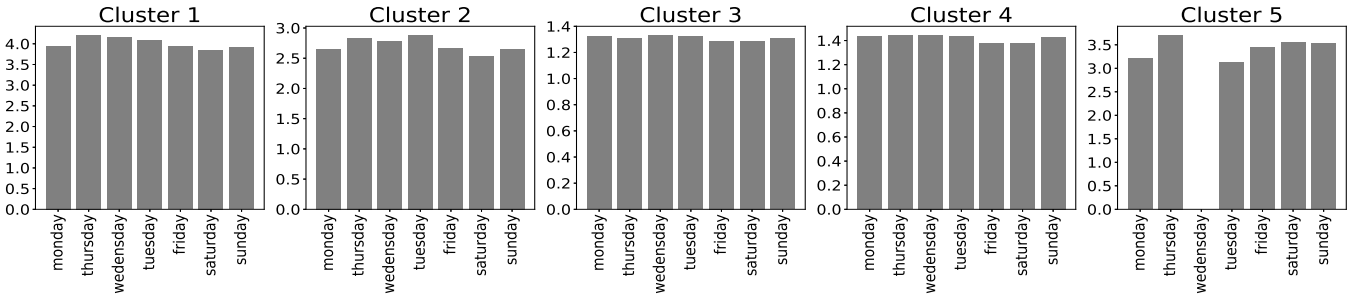


Figure 5: The daily average natural gas consumption per week.

Dealing with two distinct tasks leads to the necessity of using different measures to evaluate the model's quality. The consumption profiles classification is formulated as a multi-type classification problem. The accuracy of the FM-MLP model is characterised by the number of the correct predictions cp divided by p the total number of predictions made, multiplied by 100 to represent it in a percentage:

$$Accuracy = \frac{cp}{p}\ 100 \tag{26}$$

As sigmoidal activation function is adopted for the output from the FM-MLP model which means that the model outcome is a probability between 0 and 1. In order to access these results precisely, a cross entropy loss or log loss is also used by comparing the outputs o_i with the targets d_i according:

$$log\ loss = \frac{1}{l}\sum_{i=1}^{l}[\ d_i log(o_i)\ \ + (1-o_i)\ log(1-o_i)\ ] \tag{27}$$

In order to check the accuracy level of the forecasting models, the Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) criteria

$$MAPE = \frac{100\%}{N} \sum_{i=1}^{N} \left| \frac{a_i - p_i}{a_i} \right| \tag{28}$$

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |a_i - p_i| \tag{29}$$

$$RMSE = \sqrt{\sum_{i=1}^{N} \frac{(a_i - p_i)^2}{N}} \tag{30}$$

are applied as a benchmark calculated according to the following formulas [58], where N denotes the number of predicted samples, a_i and p_i are the actual and predicted values respectively.

In order to achieve a better calculation of network weights, using an appropriate approach to normalize the natural gas consumption data is critical before feeding it to the LSTM networks. The outputs and inputs data are normalized to an interval between [0, 1]. Hence, a value of x is normalized to x' by computing:

$$x' = \frac{x}{x_{max}}, \tag{31}$$

where x' is the new value, x is the old value and x_{max} is the largest variable value in the year.

3.3.3. FM-MLP model

Due to the superiority of the four layers feed-forward network over the network containing only three layers for the training of input-target pairs highlighted in [59], a four layer topology is adopted in the current study. All experiments are performed with this four layers network structure in order to estimate the next day consumption profile. Because of the prior knowledge of last two added consumption profiles (included in clusters 4 and 5), there will be no need for their estimation. Therefore, the FM-MLP will have only 3 outputs representing the first 3 clusters.

The overall classification model structure and configuration is summarized in Table 3

During the training stage of the proposed FM-MLP, multiple variables are taken into account: lagged consumption values $[l_1, l_2, \ldots l_{24}]^{i-1}$ and $[l_1, l_2, \ldots l_{24}]^{i-2}$ that represents the 48 hourly loads preceding the estimated day d_i, the maximum and minimum temperature of the forecast day forecast. Calendar characteristics are also considered during the classification procedure and used in term of one-hot-encoding [0,1][1].

[1]Day of the week information consists in a 7 dimension vector with a values of 1 for the day d_i to forecast and zeros elsewhere, the month indicator is also used in the same manner with a 12 attributes vector.

Table 3: Summary of used inputs, FM-MLP structure and learning parameters

Inputs			**Structure**	**Configuration**		
Lagged load	Temperature	Calendar variables		Activation	Learning rate	Epochs
$[l_1, ..., l_{24}]^{i-1}$	$tempMax^i$	day of week $[0, 1]_7$	20 nodes in 1^{st} hidden layer	sigmoid	0.001	1000
$[l_1, ..., l_{24}]^{i-2}$	$tempMin^i$	month of year $[0, 1]_{12}$	10 nodes in 2^{nd} hidden layer			

Table 4 shows the results obtained by the FM-MLP during training, validation and test phases. Based on the selected inputs, the FM-MLP shows a very powerful ability to estimate the next day consumption profile where only a single miss-classed estimation case has been recorded during the test phase.

Table 4: Accuracy, log loss and number of wrongly estimated profiles obtained on training, validation and test process

Results	**Training**			**Validation**			**Test**		
	accuracy	*log loss*	*miss estimated*	*accuracy*	*log loss*	*miss estimated*	*accuracy*	*log loss*	*miss estimated*
FM-MLP	100%	1.59 10^{-6}	0	100%	4.58 10^{-5}	0	98.43%	5.90 10^{-2}	1

3.3.4. LSTM models

To model the natural gas forecasting, several LSTM models representing the number of clusters were trained with the same variables used in the consumption profiles classification step. Enabling the model to capture the conditional dependencies between successive hourly consumptions over time, several experiments with different time lags was achieved as presented in Table 5. After evaluating various combinations of 1, 2, 4, 6, 8, 10 and 12 lagged load, it is noticeable that the best lag length windows varies from a cluster to another.

Table 5: Various combinations of 1, 2, 4, 6, 8, 10 and 12 time lags were evaluated using training sets

Models	**1 lag**	**2 lags**	**4 lags**	**6 lags**	**8 lags**	**10 lags**	**12 lags**
LSTM 1	7.14	7.64	7.55	7.95	7.71	7.04	**5.66**
LSTM 2	**10.61**	10.97	12.31	12.25	13.06	11.78	10.77
LSTM 3	9.90	10.65	10.83	10.81	10.43	11.77	**9.41**
LSTM 4	7.94	7.84	8.94	8.05	8.90	8.66	**7.77**
LSTM 5	15.53	17.77	16.66	19.41	18.91	**13.86**	16.46

Bold underlined values represent the best performance for each model.

Beside the selected time lags and based on the autocorrelation analysis, two more previous hourly loads were added: (l_{t-24} and l_{t-168}) representing the load at the same time of the previous day and the last week load at the same hour. This will help the model to keep real input values even for the case of multi step ahead forecasting. As weather information, actual estimated hourly temperatures, maximum and minimum temperature were used ($temp_t, tempMax_i$ and $tempMin_i$ respectively). In addition to calendar information that

has been used to train the FM-MLP, an hour indicator is also used, totalizing 60 possible input combination for LSTM models construction.

Another experiment was conducted to determine the most appropriate inputs. In order to enhance the prediction accuracy, these inputs are varied depending on each cluster's nature. However, the experiments are initialized using an input vector containing only the lagged load values. Then, as indicated in Table 6, other variables were added sequentially to observe their impact on the model's prediction accuracy.

Table 6: MAPE according to input combinations on training set

Experiments	**Input variables**					**Models**				
	Lagged load	**Tempe-rature**	**Day Ind**	**Hour Ind**	**Month Ind**	**LSMT1**	**LSMT2**	**LSMT3**	**LSMT4**	**LSMT5**
Exp 1	√	√				6.02	7.64	9.10	7.46	10.62
Exp 2	√		√			6.22	9.90	9.99	6.90	12.70
Exp 3	√			√		4.60	5.76	8.25	6.23	61.42
Exp 4	√				√	6.84	12.32	9.74	9.81	9.26
Exp 5	√	√	√			5.79	5.86	7.19	6.81	7.21
Exp 6	√	√		√		3.65	**3.97**	6.93	5.84	**3.21**
Exp 7	√	√	√	√		**3.52**	2.65	**5.78**	**4.40**	2.54
Exp 8	√	√	√	√	√	2.13	5.03	7.65	4.24	60.66

Bold underlined values represent the best performance for each model.

Despite the good performance achieved during the training process, generalization and good performances on test datasets are not straightforward. The network weights are optimized according to the training set, therefore, ANN models may experience overfitting during the learning process, causing poor generalization and thus degraded forecasting performances on test datasets [60]. This phenomena leads to choose the input combination used in (Exp 7) for LSTM 1 and LSTM 4 as well as inputs vector used in (Exp 6) for LSTM 2 and LSTM 5. As an example of overfitting cases, the MAPEs in test for LSTM 1 in (Exp 7) and (Exp 8) are 3.62% and 4.58% respectively, the input combination in (Exp 8) will thus be ignored.

The ANN topology represents the number of neurons per layer, the number of layers and how these neurons are connected. Generally, the neurons are arranged into one or two layers besides the input and the output layers [61]. Furthermore, finding the optimal number of hidden layers and neurons number within these layers must be decided precisely to construct a more accurate model, less sensitive to overfitting. However, there is no explicit method to determine these parameters. Hence, in addition to finding the most influential inputs for the forecasting, random search is used. Because a grid experiment with a fine-enough resolution for optimization would be prohibitively expensive, James Bergstra and Yoshua Bengio [62] recommended random search to find better models in most cases and in less computational time. By performing thirty experiments over the specified parameter values for each model, Table 7 shows different combinations of number of hidden neurons and values of learning rate to find the most successful model configuration.

Table 7: Parameters used in different experimental set-ups with 1000 epochs

Parameter	**Range**
1^{st} hidden layer nodes	(20 - 100)
2^{nd} hidden layer nodes	(20 - 50)
Learning rate	(10^{-3} - 0.01)

4. Results and Discussion

4.1. Models forecasting results

For the current forecasting problem, the performance across the most successful experiments related to the LSTMs architecture and the learning rate rule are presented in Table 8

Table 8: Parameters obtained by the random search optimization

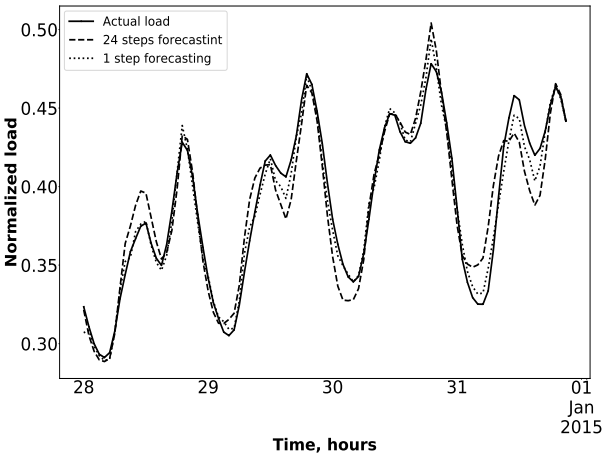
Models	**hidden layer size**	**Learning rate**
LSTM 1	(20 - 10)	0.001
LSTM 2	(20 - 10)	0.008
LSTM 3	(50 - 30)	0.002
LSTM 4	(90 - 30)	0.003
LSTM 5	(20 - 10)	0.001

After selecting the most convenient inputs with the best model topology, determined through several epochs and ANN configurations, Table 9 reports the MAPE, MAE and RMSE for each LSTM model. Moreover, the model's performance on test set was evaluated based on the FM-MLP consumption profile estimations, thus, the MAPE, MAE and RMSE were obtained by the assigned LSTMs for all 5 clusters.

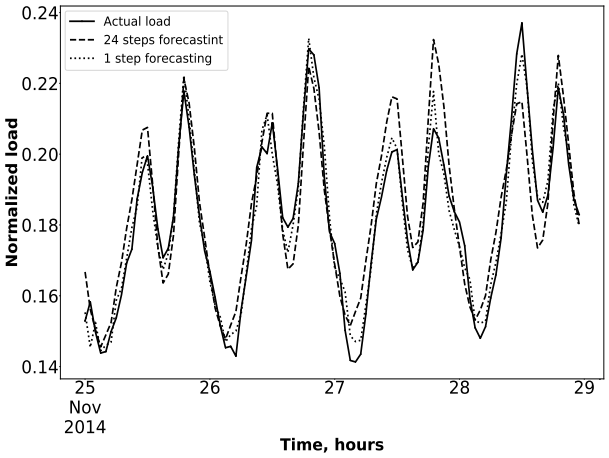
Table 9: MAPE and MAE of each cluster on the test sets

Models	**MAPE**				**MAE**				**RMSE**			
	1 step ahead		**24 steps ahead**		**1 step ahead**		**24 steps ahead**		**1 step ahead**		**24 steps ahead**	
	Train	**Test**	**Train**	**Test**	**Train**	**Test**	**Test**	**Train**	**Train**	**Test**	**Train**	**Test**
Cluster 1	1.16	1.32	3.52	3.62	0.0040	0.0045	0.0119	0.0125	0.0060	0.0157	0.0060	0.0158
Cluster 2	1.65	1.93	3.97	4.48	0.0034	0.0040	0.0086	0.0094	0.0048	0.0113	0.0054	0.0120
Cluster 3	2.88	3.40	5.78	6.88	0.0024	0.0032	0.0051	0.0062	0.0033	0.0071	0.0043	0.0087
Cluster 4	2.55	3.32	4.40	5.02	0.0018	0.0025	0.0035	0.0040	0.0036	0.0059	0.0025	0.0045
Cluster 5	2.21	3.83	3.21	5.11	0.0028	0.0087	0.0051	0.0113	0.0041	0.0063	0.0102	0.0133

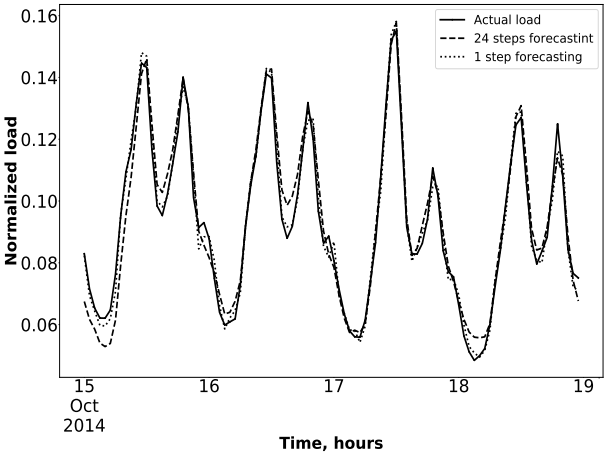
Figures 6a and 6b clearly show a high similarity between actual and forecast load on the two clusters that cover the period of winter and the period of spring and autumn. In contrast, regardless the sort of inputs used to train the LSTM models and their structure, and due to the high non-linearity level in the natural gas consumption variation recorded in the summer, as presented in Figure 6c, LSTM 3 forecast could not perfectly reflects the natural gas consumption. As shown in Figures 6d and 6e, the corresponding LSTM models were not able to accurately estimate the daily peaks which consequently increase the MAPE, MAE and RMSE calculated on these two clusters.



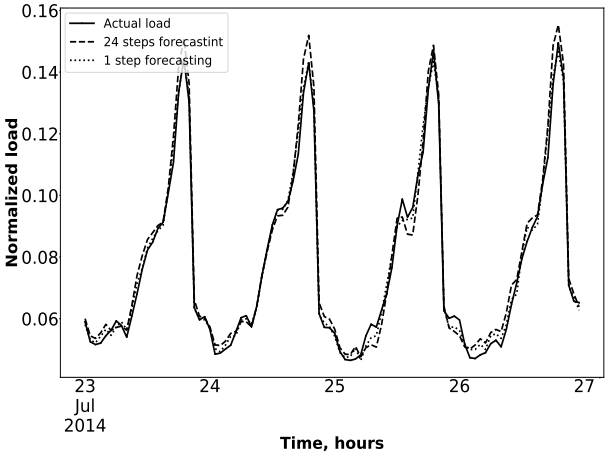
(a) Cluster 1



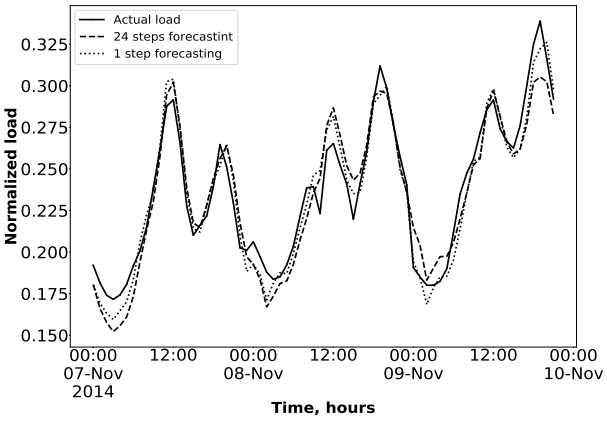
(b) Cluster 2



(c) Cluster 3



(d) Cluster 4



(e) Cluster 5

Figure 6: 24 hour forecast samples from each cluster test set

4.2. Used benchmark models

To validate the enhancement of the proposed forecasting framework, the performance is compared with several strong state-of-the-art forecasting techniques, namely BP-ANN, LSTM-RNN, SARIMAX and MLR.

4.2.1. MLP-ANN and LSTM-RNN

The first considered benchmark models are global neural networks, with two considered types: a feed forward multi-layered ANN and the other is a LSTM RNN. These two proposed models have the same ANN structure containing 36 inputs (1, 24 and 168 lagged load, temperature and one-hot-encoding [0,1] variables to indicate the hour of the day and the day of the week). The ANNs structure is composed of two hidden layers with 20 neurons in the first and 10 in the second.

4.2.2. SARIMAX

Another important baseline approach is an extension of ARIMA models, which is a first class time series technique, introduced by Box and Jenkins [63]. The multivariate version of SARIMA has enhanced the ability to integrate explanatory (exogenous) variables and to add seasonal terms in order to increase the performance.

Seasonal ARIMA can be shown as SARIMA$(p, d, q), (D, D, Q)s$ where p, d and q referred to the order of autoregressive terms, the order of differencing and the number of moving averages terms in the non-seasonal and seasonal components of the model respectively. the term s represents the span of repeating the seasonal pattern which is equal to 24 in the current study referring to 24 hourly load per day.

Practically, most of time series are non-stationary and need to be transformed into a stationary ones before being used to regulate the model's parameters. Since the actual dataset as a load time series includes two types of seasonality, which is daily and weekly seasonality, two differencing were applied to obtain a stationary series, one at 24 lags and the second at 168 lags respectively. Also, a first order differencing was applied as well to overcome the trend in the load time series besides the differencing at seasonal lags. According to sample autocorrelation function (ACF) and partial autocorrelation function (PACF) for determining p and q variables, it is more evident to choose SARIMAX$(1,1,1)(1,1,1)_{24}$, SARIMAX$(2,1,1)(1,1,1)_{24}$, SARIMAX$(3,1,1)(2,1,1)_{24}$ and SARIMAX$(6,0,1)(2,0,1)_{24}$, whereas the last model presented the lowest MAPE among the candidate models.

4.2.3. Multiple linear regression

As for statistical techniques, MLR basis on estimating the unknown values of a variable from a set of other related known values. These models represent the gas consumption load as a linear function of several dependent and independent variables. However, despite the well pronounced uniform general trends of the data, non-linearities are still present and are difficult to capture. For instance, a MLR function provides and estimates load y at time t and n predictor variables $x_i(i : 1,\ldots, n)$ by:

$$y_t = \sum_{i=1}^{n} \theta_i x_i + \epsilon \tag{32}$$

where $\theta_i(i:1,\ldots,n)$ are regression parameters and ϵ is an error term.

4.3. Comparison and discussion

The benchmark models were trained using a set formed by combining the 5 training sets used for the LSTM models construction. Therefore, comparing their forecasting quality will be based on the 5 sets used to assess the LSTM models performance. Moreover, Each cluster deserves a particular consideration, and due to the high distinction in the recognized clusters over the year, weighted arithmetic mean is applied instead of the ordinary mean to calculate an average MAPE, MAE and RMSE that represents the overall error. Equation (33) expresses the weighted average of an error ($\overline{E}$), where w_i is the number of days counted in each cluster c_i.

$$\overline{E} = \frac{1}{N}\sum_{i=1}^{5}(\ w_i E_i\ ) \tag{33}$$

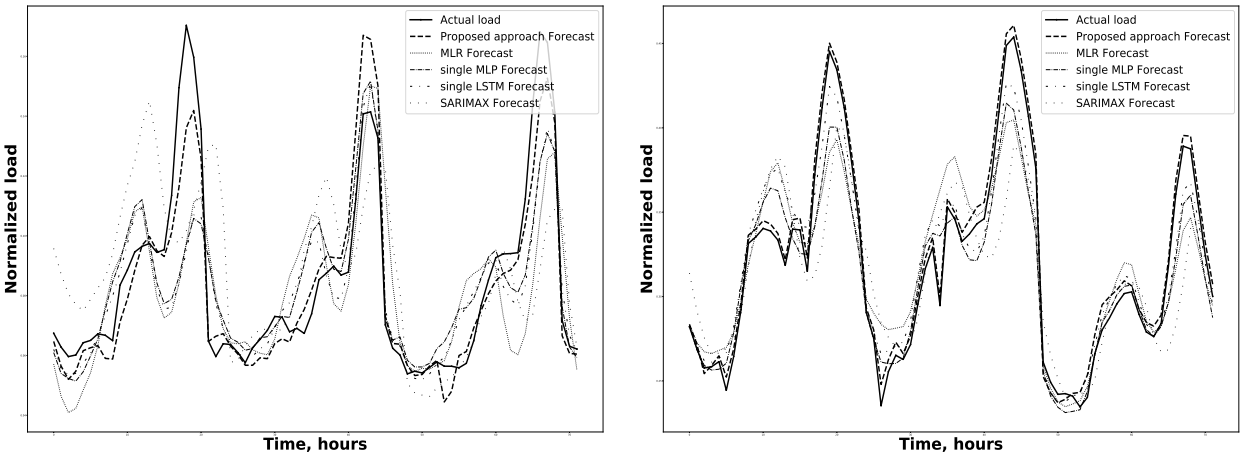
Table 10: Performance of the proposed approach compared with MLP-ANN, LSTM-RNN, SARIMAX and MLR for one day ahead NG forecasting

		Pro.app		**MLP**		**LSTM**		**SARIMAX**		**MLR**	
		Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
MAPE	cluster 1	3.52	3.62	3.97	5.36	3.25	5.49	4.37	3.39	4.69	4.48
	cluster 2	3.97	4.48	4.69	5.26	3.91	5.34	6.11	5.03	7.55	5.67
	cluster 3	5.78	6.88	6.62	6.84	6.30	6.38	9.27	7.12	9.94	9.12
	cluster 4	4.40	5.02	8.79	7.77	8.11	7.48	9.89	11.13	19.20	23.30
	cluster 5	3.21	5.11	7.71	8.55	6.98	10.88	10.47	8.68	8.64	10.06
	average	**4.73**	**5.48**	5.87	6.38	5.34	6.27	7.64	6.22	8.89	8.59
MAE	cluster 1	0.0119	0.0125	0.0133	0.0187	0.0112	0.0192	0.0152	0.0115	0.0160	0.0153
	cluster 2	0.0086	0.0094	0.0103	0.0110	0.0088	0.0110	0.0144	0.0104	0.0176	0.0121
	cluster 3	0.0051	0.0062	0.0059	0.0061	0.0054	0.0059	0.0076	0.0064	0.0082	0.0088
	cluster 4	0.0035	0.0040	0.0075	0.0057	0.0068	0.0056	0.0079	0.0074	0.0155	0.0155
	cluster 5	0.0051	0.0113	0.0138	0.0194	0.0119	0.0231	0.0187	0.0205	0.0164	0.0218
	average	**0.0072**	**0.0083**	0.0078	0.0106	0.0087	0.0107	0.0109	0.0088	0.0123	0.0119
RMSE	cluster 1	0.0157	0.0158	0.0175	0.0223	0.0149	0.0227	0.0202	0.0156	0.0212	0.0193
	cluster 2	0.0113	0.0120	0.0141	0.0142	0.0118	0.0137	0.0191	0.0128	0.0224	0.0151
	cluster 3	0.0071	0.0087	0.0081	0.0080	0.0075	0.0079	0.0111	0.0085	0.0110	0.0116
	cluster 4	0.0059	0.0045	0.0109	0.0071	0.0104	0.0071	0.0127	0.0090	0.0208	0.0189
	cluster 5	0.0063	0.0133	0.0194	0.0223	0.0164	0.0257	0.0272	0.0259	0.0218	0.0258
	average	**0.0092**	**0.0108**	0.0140	0.0147	0.0122	0.0154	0.0180	0.0143	0.0194	0.0181

Bold underlined values represent the best average performance.

According to the results presented in Table 10, the proposed approach had the lowest average MAPE of 24 steps ahead compared with the four benchmark methods. Additionally, the weighted average MAPE obtained by the proposed approach on test is 5.87%, which is an improvement over the global MLP and LSTM, SARIMAX and MLR by 16.42%, 14.41%,13.50% and 56.75% respectively. Despite the high performance of global ANN approaches (MLP and LSTM), it is noticed that the proposed forecasting approach made a significant improvement on the forecast quality especially over the test dataset.

Considering only the forecasting performance, the proposed method has the minimum error including for special days when consumption curves change, this is evident in the case



(a) Ramadan month (b) National holidays

Figure 7: Comparison between actual and predicted natural gas consumption with the use of the proposed and benchmark approaches

of national holidays or when the customers behavior changes due to occasional reasons. Moreover, all benchmark approaches show a weakness on these specific periods due to the high dependency of the previous loads. However, multiple LSTMs models are relatively accurate compared to a global model approach. Figure 7 presents the advantage of the proposed approach upon other approaches on the national holidays and the beginning of the month of Ramadan.

5. Conclusion

The main contribution of this work consists in the developed two-stage Forecasting Monitoring Multi Layered Perceptron (FM-MLP) approach. After partitioning the Algerian natural gas hourly consumption data by applying K-means according to the consumption profiles, the FM-MLP is then trained based on the dataset segmentation for the purpose of estimating the next day consumption profile. The FM-MLP decides which LSTM recurrent model will handle best its forecast. The approach achieves 98.34% accuracy of classification of the natural gas consumption profile on the real test dataset.

Based on the obtained groups that contain similar consumption profiles and by using calendar and temperature information, multiple LSTM recurrent networks are constructed according to each cluster data. The performance of the developed approach is carefully validated and evaluated based on the gas consumption profile forecast made by the FM-MLP. The weighted average of the MAPE, MAE and RMSE obtained by the assigned LSTM models on the testing set are 5.48%, 0.0083 and 0.0108, respectively.

In order to validate the obtained results, the MLP ANN, LSTM RNN, SARIMAX and MLR models are employed. Comparing the predicted results with each benchmark method in terms of MAPE, MAE and RMSE, the developed approach shows improved accuracy. The superiority is especially pronounce in the periods with irregular gas consumption, such

as during holidays and special days periods where daily natural gas consumption posses a quite unique curve that could not be captured using a single general approach.

References

[1] B. Soldo, Forecasting natural gas consumption, Applied Energy 92 (2012) 26–37.
URL https://doi.org/10.1016/j.apenergy.2011.11.003

[2] W.-C. Hong, Intelligent Energy Demand Forecasting, Springer London, 2013.
URL https://doi.org/10.1007/978-1-4471-4968-2

[3] F. Taşpınar, N. Çelebi, N. Tutkun, Forecasting of daily natural gas consumption on regional basis in Turkey using various computational methods, Energy and Buildings 56 (2013) 23–31.
URL https://doi.org/10.1016/j.enbuild.2012.10.023

[4] N. Elamin, M. Fukushige, Modeling and forecasting hourly electricity demand by SARIMAX with interactions, Energy 165 (2018) 257–268.
URL https://doi.org/10.1016/j.energy.2018.09.157

[5] Y. Chen, W. S. Chua, T. Koch, Forecasting day-ahead high-resolution natural-gas demand and supply in Germany, Applied Energy 228 (2018) 1091–1110.
URL https://doi.org/10.1016/j.apenergy.2018.06.137

[6] C. García-Ascanio, C. Maté, Electric power demand forecasting using interval time series: A comparison between VAR and iMLP, Energy Policy 38 (2) (2010) 715–725.
URL https://doi.org/10.1016/j.enpol.2009.10.007

[7] M. Fagiani, S. Squartini, L. Gabrielli, S. Spinsante, F. Piazza, Domestic water and natural gas demand forecasting by using heterogeneous data: A preliminary study, in: Advances in Neural Networks: Computational and Theoretical Issues, Springer International Publishing, 2015, pp. 185–194.
URL https://doi.org/10.1007/978-3-319-18164-6_18

[8] Y. Bai, C. Li, Daily natural gas consumption forecasting based on a structure-calibrated support vector regression approach, Energy and Buildings 127 (2016) 571–579.
URL https://doi.org/10.1016/j.enbuild.2016.06.020

[9] A. Özmen, Y. Yılmaz, G.-W. Weber, Natural gas consumption forecast with MARS and CMARS models for residential users, Energy Economics 70 (2018) 357–381.
URL https://doi.org/10.1016/j.eneco.2018.01.022

[10] P. Lusis, K. R. Khalilpour, L. Andrew, A. Liebman, Short-term residential load forecasting: Impact of calendar effects and forecast granularity, Applied Energy 205 (2017) 654–669.
URL https://doi.org/10.1016/j.apenergy.2017.07.114

[11] L. Ruiz, R. Rueda, M. Cuéllar, M. Pegalajar, Energy consumption forecasting based on elman neural networks with evolutive optimization, Expert Systems with Applications 92 (2018) 380–389.
URL https://doi.org/10.1016/j.eswa.2017.09.059

[12] H. Karimi, J. Dastranj, Artificial neural network-based genetic algorithm to predict natural gas consumption, Energy Systems 5 (3) (2014) 571–581.
URL https://doi.org/10.1007/s12667-014-0128-2

[13] S. Barak, S. S. Sadegh, Forecasting energy consumption using ensemble ARIMA–ANFIS hybrid algorithm, International Journal of Electrical Power & Energy Systems 82 (2016) 92–104.
URL https://doi.org/10.1016/j.ijepes.2016.03.012

[14] K. Gajowniczek, T. Zbkowski, Two-stage electricity demand modeling using machine learning algorithms, Energies 10 (10) (2017) 1547.
URL https://doi.org/10.3390/en10101547

[15] J. Szoplik, Forecasting of natural gas consumption with artificial neural networks, Energy 85 (2015) 208–220.
URL https://doi.org/10.1016/j.energy.2015.03.084

[16] F. Yu, X. Xu, A short-term load forecasting model of natural gas based on optimized genetic algorithm

and improved BP neural network, Applied Energy 134 (2014) 102–113.
URL https://doi.org/10.1016/j.apenergy.2014.07.104

[17] Z. Tonković, M. Zekič-Sušac, M. Somolanji, Predicting natural gas consumption by neural networks, Tehniki vjesnik 16 (3) (2009) 51–61.

[18] D. Park, M. El-Sharkawi, R. Marks, L. Atlas, M. Damborg, Electric load forecasting using an artificial neural network, IEEE Transactions on Power Systems 6 (2) (1991) 442–449.
URL https://doi.org/10.1109/59.76685

[19] K. Lee, Y. Cha, C. Ku, A study on neural networks for short-term load forecasting, in: Proceedings of the First International Forum on Applications of Neural Networks to Power Systems, IEEE.
URL https://doi.org/10.1109/ann.1991.213492

[20] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, Y. Zhang, Short-term residential load forecasting based on LSTM recurrent neural network, IEEE Transactions on Smart Grid 10 (1) (2017) 841–851.
URL https://doi.org/10.1109/tsg.2017.2753802

[21] P. Potočnik, B. Soldo, G. Šimunović, T. Šarić, A. Jeromen, E. Govekar, Comparison of static and adaptive models for short-term residential natural gas forecasting in croatia, Applied Energy 129 (2014) 94–103.
URL https://doi.org/10.1016/j.apenergy.2014.04.102

[22] C. Deb, F. Zhang, J. Yang, S. E. Lee, K. W. Shah, A review on time series forecasting techniques for building energy consumption, Renewable and Sustainable Energy Reviews 74 (2017) 902–924.
URL https://doi.org/10.1016/j.rser.2017.02.085

[23] S. Fan, L. Chen, Short-term load forecasting based on an adaptive hybrid method, IEEE Transactions on Power Systems 21 (1) (2006) 392–401.
URL https://doi.org/10.1109/tpwrs.2005.860944

[24] P. Du, J. Wang, W. Yang, T. Niu, Multi-step ahead forecasting in electrical power system using a hybrid forecasting system, Renewable Energy 122 (2018) 533–550.
URL https://doi.org/10.1016/j.renene.2018.01.113

[25] L. Hernández, C. Baladrón, J. Aguiar, L. Calavia, B. Carro, A. Sánchez-Esguevillas, J. Sanjuán, Á. González, J. Lloret, Improved short-term load forecasting based on two-stage predictions with artificial neural networks in a microgrid environment, Energies 6 (9) (2013) 4489–4507.
URL https://doi.org/10.3390/en6094489

[26] G. Krzysztof, Z. Tomasz, Two-stage electricity demand modeling using machine learning algorithms, Energies 10 (10) (2017) 1–25.
URL https://doi.org/10.3390/en10101547

[27] S. Ilic, S. Vukmirovic, A. Erdeljan, F. Kulic, Hybrid artificial neural network system for short-term load forecasting, Thermal Science 16 (suppl. 1) (2012) 215–224.
URL https://doi.org/10.2298/tsci120130073i

[28] N. Ghadimi, A. Akbarimajd, H. Shayeghi, O. Abedinia, A new prediction model based on multi-block forecast engine in smart grid, Journal of Ambient Intelligence and Humanized Computing.
URL https://doi.org/10.1007/s12652-017-0648-4

[29] J. Zhang, Y.-M. Wei, D. Li, Z. Tan, J. Zhou, Short term electricity load forecasting using a hybrid model, Energy 158 (2018) 774–781.
URL https://doi.org/10.1016/j.energy.2018.06.012

[30] J. Xiao, Y. Li, L. Xie, D. Liu, J. Huang, A hybrid model based on selective ensemble for energy consumption forecasting in China, Energy 159 (2018) 534–546.
URL https://doi.org/10.1016/j.energy.2018.06.161

[31] F. M. Alvarez, A. Troncoso, J. C. Riquelme, J. S. A. Ruiz, Energy time series forecasting based on pattern sequence similarity, IEEE Transactions on Knowledge and Data Engineering 23 (8) (2011) 1230–1243.
URL https://doi.org/10.1109/tkde.2010.227

[32] M. of Energy, Bilan Energetique National, http://www.energy.gov.dz/francais/uploads/MAJ_2018/Stat/Bilan_Energtique_National_2017_edition_2018.pdf (2017).

[33] P. monde, Algrie - Production d'lectricit - partir de gaz naturel (% de la production totale) — Statistiques, http://perspective.usherbrooke.ca/bilan/servlet/BMTendanceStatPays?codeTheme=6&codeStat=EG.ELC.NGAS.ZS&codePays=DZA&optionsPeriodes=Aucune&codeTheme2=6&codeStat2=x&codePays2=DZA&optionsDetPeriodes=avecNomP&langue=fr (2019).

[34] N. Grim, SONATRACH : De moins en moins de gaz exporter, https://www.algerie-eco.com/2018/01/22/sonatrach-de-de-gaz-a-exporter/ (2018).

[35] D. E. Linvill, Calculating chilling hours and chill units from daily maximum and minimum temperature observations, HortScience 25 (1) (1990) 14–16.
URL http://hortsci.ashspublications.org/content/25/1/14.full.pdf

[36] K. E. Farfar, M. T. Khadir, A two-stage short-term load forecasting approach using temperature daily profiles estimation, Neural Computing and Applications.
URL https://doi.org/10.1007/s00521-017-3324-x

[37] D. Sebalj, J. Mesaric, D. Dujak, Predicting natural gas consumption a literature review, in: Proceedings of 28th Central European Conference on Information and Intelligent Systems (CECIIS '17), Varadin, Croatia, 2017, pp. 293–300.

[38] L. Zhu, M. Li, Q. Wu, L. Jiang, Short-term natural gas demand prediction based on support vector regression with false neighbours filtered, Energy 80 (2015) 428–436.
URL https://doi.org/10.1016/j.energy.2014.11.083

[39] O. Laib, M. T. Khadir, L. Mihaylova, A gaussian process regression for natural gas consumption prediction based on time series data, in: Proceedings of the 2018 21st International Conference on Information Fusion (FUSION), 2018, pp. 55–61. doi:10.23919/ICIF.2018.8455447.

[40] R. J. G. B. Campello, D. Moulavi, A. Zimek, J. Sander, Hierarchical density estimates for data clustering, visualization, and outlier detection, ACM Transactions on Knowledge Discovery from Data 10 (1) (2015) 1–51. doi:10.1145/2733381.

[41] J. Hensman, N. D. Lawrence, M. Rattray, Hierarchical bayesian modelling of gene expression time series across irregularly sampled replicates and clusters, BMC Bioinformatics 14 (1) (2013) 252. doi:10.1186/1471-2105-14-252.

[42] D. J. C. MacKay, Information Theory, Inference & Learning Algorithms, Cambridge University Press, New York, NY, USA, 2002.

[43] G. Cybenko, Approximation by superpositions of a sigmoidal function, Mathematics of Control, Signals and Systems 2 (4) (1989) 303–314.
URL https://doi.org/10.1007/BF02551274

[44] K. Hornik, M. Stinchcombe, H. White, Multilayer feedforward networks are universal approximators, Neural Networks 2 (5) (1989) 359–366.
URL https://doi.org/10.1016/0893-6080(89)90020-8

[45] W. Liu, Z. Wang, X. Liu, N. Zeng, Y. Liu, F. E. Alsaadi, A survey of deep neural network architectures and their applications, Neurocomputing 234 (2017) 11–26.
URL https://doi.org/10.1016/j.neucom.2016.12.038

[46] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Computation 9 (8) (1997) 1735–1780.
URL https://doi.org/10.1162/neco.1997.9.8.1735

[47] D. E. Rumelhart, G. E. Hinton, R. J. Williams, Learning representations by back-propagating errors, Nature 323 (6088) (1986) 533–536.
URL https://doi.org/10.1038/323533a0

[48] C. M. Bishop, Pattern Recognition and Machine Learning (Information Science and Statistics), Springer-Verlag, Berlin, Heidelberg, 2006.

[49] F. A. Gers, N. N. Schraudolph, J. Schmidhuber, Learning precise timing with LSTM recurrent networks, Journal of Machine Learning Research 3 (2002) 115–143.
URL http://nic.schraudolph.org/pubs/GerSchSch02.pdf

[50] S. Hochreiter, Y. Bengio, P. Frasconi, Gradient flow in recurrent nets: the difficulty of learning long-term dependencies, in: J. Kolen, S. Kremer (Eds.), Field Guide to Dynamical Recurrent Networks, IEEE Press, 2001.

[51] A. Graves, S. Fernández, J. Schmidhuber, Bidirectional lstm networks for improved phoneme classification and recognition, in: W. Duch, J. Kacprzyk, E. Oja, S. Zadrożny (Eds.), Artificial Neural Networks: Formal Models and Their Applications – ICANN 2005, Springer Berlin Heidelberg, Berlin, Heidelberg, 2005, pp. 799–804.

[52] F. A. Gers, J. A. Schmidhuber, F. A. Cummins, Learning to forget: Continual prediction with LSTM, Neural Comput. 12 (10) (2000) 2451–2471.
URL http://dx.doi.org/10.1162/089976600300015015

[53] P. J. Rousseeuw, Silhouettes: A graphical aid to the interpretation and validation of cluster analysis, Journal of Computational and Applied Mathematics 20 (1987) 53–65.
URL https://doi.org/10.1016/0377-0427(87)90125-7

[54] T. Calinski, J. Harabasz, A dendrite method for cluster analysis, Communications in Statistics - Theory and Methods 3 (1) (1974) 1–27.
URL https://doi.org/10.1080/03610927408827101

[55] D. L. Davies, D. W. Bouldin, A cluster separation measure, IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-1 (2) (1979) 224–227.
URL https://doi.org/10.1109/tpami.1979.4766909

[56] S. Sharma, Applied Multivariate Techniques, Wiley, 1995.
URL https://books.google.co.uk/books?id=6iURRAAACAAJ

[57] R. L. Thorndike, Who belongs in the family?, Psychometrika 18 (4) (1953) 267–276.
URL https://doi.org/10.1007/bf02289263

[58] F. Shaikh, Q. Ji, Forecasting natural gas demand in China: Logistic modelling analysis, International Journal of Electrical Power & Energy Systems 77 (2016) 25–32.
URL https://doi.org/10.1016/j.ijepes.2015.11.013

[59] S. Tamura, M. Tateishi, Capabilities of a four-layered feedforward neural network: four layers versus three, IEEE Transactions on Neural Networks 8 (2) (1997) 251–255.
URL https://doi.org/10.1109/72.557662

[60] I. V. Tetko, D. J. Livingstone, A. I. Luik, Neural network studies. (1). Comparison of overfitting and overtraining, Journal of Chemical Information and Modeling 35 (5) (1995) 826–833.
URL https://doi.org/10.1021/ci00027a006

[61] E. Fiesler, Neural network classification and formalization, Computer Standards & Interfaces 16 (3) (1994) 231–239.
URL https://doi.org/10.1016/0920-5489(94)90014-0

[62] J. Bergstra, Y. Bengio, Random search for hyper-parameter optimization, J. Mach. Learn. Res. 13 (1) (2012) 281–305.
URL http://dl.acm.org/citation.cfm?id=2503308.2188395

[63] G. Box, Box and Jenkins: Time Series Analysis, Forecasting and Control, booktitle = A Very British Affair, Palgrave Macmillan UK, 2013, pp. 161–215.
URL https://doi.org/10.1057/9781137291264_6