وزارة التعليم العالي والبحث العلمي

BADJI MOKHTAR -ANNABA
UNIVERSITY
UNIVERSITE BADJI MOKHTAR
ANNABA

جامعة باجي مختار
- عنابة -

Faculté des Sciences de l'Ingéniorat
Département d'Informatique

THÈSE

Présenté en vue de l'obtention du diplôme de
Doctorat LMD - 3éme cycle

TITRE

Classification et Prévision Multi-termes de la Consommation de Gaz Naturel en Utilisant les Paradigmes de l'Intelligence Artificielle

Domaine: Mathématiques et Informatique
Filière: Informatique
Spécialité: **STIC** (**S**ciences et **T**echnologies de l'**I**nformation et de la **C**ommunication)

Par
Oussama Laib

DIRECTEUR DE THÈSE : Pr. Tarek Mohamed KHADIR

Devant le jury

PRESIDENT :	Seridi Hassina	Prof	U.B.M. ANNABA
EXAMINATEUR :	Farah Nadir	Prof	U.B.M. ANNABA
EXAMINATEUR :	Farou Brahim	Dr	U. GUELMA

Année : 2019/2020

This thesis is lovingly dedicated to:
My mom for her patient and her unconditional constant love that have been sustaining me throughout my whole life.

My dad for his support, advices and belief.

My lovely aunty and her husband for their affection and positive influence in my live.

My brother, my sister, my brother in law and the cutest niece ever.

My wife for beginning an important part of my story.

My friends for making so many ordinary moments, extraordinary.

Acknowledgements

With no doubt, the work on this thesis has been the most challenging endeavor I have ever undertaken so far. I want to thank in the first place my advisor Dr. Mohamed Tarek khadir, for his guidance, patience and encouragement.

I also acknowledge the help and assistance of Dr. Farah Nadir from Computer Science department, University of Badji Mokhtar Annaba.

I must also acknowledge the help and assistance of Dr. Lyudmila Mihaylova from the Department of Automatic Control and Systems Engineering, University of Sheffield.

Many thanks to all my immediate family and friends for their constant encouragement and support.

Without forgetting to acknowledge my colleges from computer science department, Kheire Eddin Farfar, Nadji Ahmed Zakaria and Hamza Frihia for their collaboration and invaluable discussions and for their understanding and encouragement.

We thank The National Company for Electricity and Gas (SONELGAZ) and the Prospective and analysis department (DGSP) for their expertise and for providing the research team with the necessary data, which greatly helped achieving what we have done in this thesis.

ملخص

يمثل التنبؤ باستهلاك الطاقة في الجزائر تحديًا كبيرًا، نظرًا لإسهامها الأوروبي الهام في إمدادات الطاقة، من ناحية، وحجم البلد والسكان والنمو الاقتصادي الجلي من ناحية أخرى. كل هذا وغيره له تأثير كبير على تنوع المناخ وسلوك الزبائن، مما يجعل التنبؤ باستهلاك الطاقة أكثر تعقيدًا. البحث عن المثالية في التنبؤ مدفوع بقوة لأسباب اقتصادية، حيث أن موارد الغاز الطبيعي شحيحة. بما أن الاقتصاد الجزائري يعتمد اعتمادًا كبيرًا على صادرات الهيدروكربونات وبالإضافة الى النمو الاقتصادي المتذبذب، فإن اعتماد نماذج التنبؤ لاستهلاك الغاز الطبيعي أمر بالغ الأهمية. لتلبية هذا الطلب، تقدم الأطروحة التالية ثلاثة مساهمات لمعالجة مشكلة التنبؤ باستهلاك الغاز الطبيعي على المدى القصير والطويل.

أولاً، يتم بحث منهج التنبؤ على مرحلتين، والذي يعتمد في صيغته على: 1) تصنيف البيانات اليومية لاستهلاك الغاز الطبيعي إلى مجموعات مختلفة ذات وتيرة متشابهة باستخدام عدة خوارزميات تصنيفية. 2) تصميم وتدريب عدة نماذج انحدار باستخدام بيانات الغاز الطبيعي للسوق الجزائرية مع اعتماد عوامل خارجية متمثلة في التأثيرات الطقس (درجة الحرارة) والتقويم (يوم الأسبوع، مؤشر الساعة). تتمحور الابتكارية المطروحة الرئيسية في هذا العمل على البحث في تقنيات تصنيف متعددة من أجل تحسين تحليل ودراسة بيانات استهلاك الغاز الطبيعي. ختاما، يتم تلخيص وتقييم تأثير المجموعات التي تم الحصول عليها حسب كل تقنية على دقة التنبؤ.

ثانياً، تقديم منهجية تنبؤ جديدة لحل عيب الطريقة المكونة من مرحلتين، وذلك من خلال تصميم شبكة عصبية اصطناعية متعددة الطبقات كمتابع تنبؤ غير الخطي. يقوم هذا النموذج بتقدير وتيرة استهلاك الغاز في اليوم التالي وتحديد أحد النماذج المتخصصة لتنفيذ التنبؤ. ترتكز الدراسة أولاً على تحليل وتجميع لبيانات الاستهلاك اليومي للغاز الطبيعي ذات وتيرة استهلاك متماثلة، وثانياً على بناء نماذج متكررة شاملة للذاكرة طويلة المدى (LSTM) وفقًا لوتيرة الاستهلاك اليومي كل صنف على حدا. تأكيدا لفعالية المنهجية، تتم مقارنة النتائج مع أربعة نماذج قياسية اخرى، وهي: الشبكة العصبية الاصطناعية MLP، LSTM، نموذج سلسلة زمنية موسمية مع اعتبار المتغيرات الخارجية، ونموذج الانحدار الخطي المتعدد. بالمقارنة مع هذه الأساليب البديلة واعتمادها الكبير على حجم الاستهلاك السابق، تقدم المنهجية المقترحة وظيفة فعالة جديدة. تقدير طبيعة الاستهلاك في اليوم التالي من شأنها أن تؤدي إلى تحسن كبير في دقة التنبؤ، خاصة بالنسبة للأيام ذات سلوك استهلاك استثنائي.

تتمثل المساهمة النهائية أخيرا في تطوير نماذج للشبكات العصبية الاصطناعية من اجل تنبؤ طويل المدى للاستهلاك السنوي للغاز الطبيعي في الجزائر لقطاعات الضغط الثلاثة (الضغط المنخفض والضغط المتوسط والضغط العالي).

الكلمات المفتاحية: الشبكات العصبية الاصطناعية، تصنيف السلاسل الزمنية، التنبؤ باستهلاك اليوم المقابل، الشبكات المتكررة ذات ذاكرة قريبة وبعيدة المدى، التنبؤ طويل المدى، استهلاك الغاز الطبيعي.

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, the following thesis presents three contributions to tackle the natural gas consumption forecasting problem in the short and long term.

Firstly, the two-stage forecasting approach is investigated, which consists of two major phases: 1) it classifies the natural gas consumption daily pattern sequences into different groups with similar attributes. 2) the design and training of multiple autoregressive Gaussian Process models phase is carried out using the Algerian natural gas market data together with exogenous inputs consisting in weather (temperature) and calendar (day of the week, hour indicator) factors. The main novelty in this work consists of the investigation of multiple different clustering techniques for better analysis and clustering of natural gas consumption data. The impact of the obtained clusters, by each technique, is then summarized and evaluated with respect to the prediction accuracy.

Secondly, a novel forecasting approach is introduced to resolve the two-stage method's deficiency, by designing a Multi Layered Perceptron (MLP) neural network as a nonlinear forecasting monitor. 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.

The final contribution consists of developing a Neural Networks approach to predict the annual natural gas consumption in Algeria for the three pressure sectors (low pressure, medium pressure and high-pressure sector). Four main distribution areas constitutes the Algerian distribution company (SONELGAZ), and each one is consists of several distribution divisions (DD). Thus, instead of creating a single neural network model based on an aggregated dataset to estimate a sector consumption. Each DD's consumption forecast is occurred separately by selecting the most influential inputs, then developing its own Multi

Layer Perceptron (MLP) model trained with Levenberg-Marquardt learning algorithm. Finally summing their results to get the total consumption for the sectors.

Keywords: Artificial neural network, Time series classification, Long short term memory, Day-ahead forecasting, Long-term forecasting, Natural gas consumption.

Resumé

Le gaz naturel est l'une des principales sources d'énergie électrique en Algérie et dans le monde. Il est donc impératif d'étudier de comprendre de modéliser et de prédire cette demande. Cette thèse présente trois contributions distinctes traitant les problèmes de prévision de la consommation de gaz naturel à court et à long terme.

En premier lieu, l'approche de prévision à deux étapes est étudiée et implémentée. Elle consiste en deux phases principales : 1) classer les séquences de schémas quotidiens de consommation de gaz naturel en Algérie en différents groupes aux attributs similaires (classes). 2) la conception et la formation de modèles de types processus gaussiens auto-régressifs multiples est réalisée en utilisant les données du marché Algérien du gaz naturel ainsi que les intrants exogènes constitues de la météo (temperature) et du calendrier (jour de la semaine, indicateur d'heure). La principale nouveauté de ce travail consiste en l'investigation de multiples techniques de clustering différentes pour une meilleure analyse et clustering des données de consommation de gaz naturel. L'impact des clusters obtenus, par chaque technique est résumé et évalué en fonction de la précision de la prédiction.

En deuxième lieu, une nouvelle approche de prévision est introduite pour résoudre les déficiences de la méthode en deux étapes, en concevant un perceptron multicouche (MLP) réseaux de neurones en tant que moniteur de prévision non linéaire. Ce modèle estime le profil de consommation journalière de gaz et choisit l'un des modèles locaux pour effectuer la prévision. L'étude porte tout d'abord sur une analyse et une classification des profils de consommation quotidienne de gaz naturel, et ensuite sur la construction d'un modèle récurrent complet de la mémoire à long terme (LSTM) en fonction du comportement de la charge. Les résultats sont comparés à quatre approches de référence : le MLP, LSTM, série temporelle saisonnière avec variables exogènes et modèle de régression linéaire multiple.

Finalement, la dernière contribution consiste à développer une approche de réseaux de neurones pour prédire la consommation annuelle de gaz naturel en Algérie pour les trois secteur de pression (basse pression, moyenne pression et haute pression) à moyen et long termes.

Mots clés: Réseaux de neurones artificiels, Classification des séries temporelles, Réseaux récurrents à mémoire court-terme et long-terme, Prévision à un jour, Prévision à long terme, Consommation de gaz naturel.

Contents

List of Figures

List of Tables

List of Abbreviations

ABC	Artificial Bee Colony.
ANN	Artificial Neural Networks.
ANFIS	Adaptive Neuro-Fuzzy Inference System.
AR	Auto-Regressive.
ARMA	Auto-regressive Moving Average.
ARIMA	Auto-regressive Integrated Moving Average with Exogenous inputs.
BP	Back Propagation.
CART	called Classification and Regression Trees.
CEEMDAN	Complete Ensemble Empirical Mode Decomposition with Adaptive Noise.
CNN	Convolutional Neural Networks.
DP	Dirichlet processes.
ESN	Echo State Network.
FL	Fuzzy logics.
FM	Forecasting Monitor.
GA	Genetic Algorithm.
GEP	Gene Expression Programing.
GM	Grey Model.
GP	Gaussian Process.
GPQR	Gaussian Process Quantile Regression.
GRU	Gated Recurrent Units.

HDBSCAN	Hierarchical Density-Based Spatial Clustering of Applications with Noise.
IEMD	Improved Empirical Mode Decomposition.
LML	Log-Marginal Likelihood.
LSSVM	Least Squares Support Vector Machines.
LSTM	Long-Short Term Memory.
MA	Moving-average.
MAE	Mean Absolute Error.
MAPE	Mean Absolute Percentage Error.
MARS	Multivariable Adaptive Regression Splines.
MF	Membership Functions.
MLP	Multi-Layer Perceptron.
MLR	Multiple Linear Regression.
MOHGP	Mixture of Hierarchical Gaussian Process.
NARX	Nonlinear Auto-Regressive Exogenous
OWA	Ordered Weighted Averaging.
RBF	Radial Basis Function.
RF	Random Forest.
RMSE	Root Mean Square Error.
RNN	Recurrent Neural Networks.
RTRL	Real Time Recurrent Learning.
SARIMA	Seasonal Auto-regressive Integrated Moving Average with Exogenous inputs.
SDAE	Stacked Denoise Autoencoder Based.
SVR	Support Vector Regression.
SVM	Support Vector Machine.
SOFM	Self-Organizing Feature Map.
WNN	Wavelet Neural Networks.
WT	Wavelet Transform.

Chapter 1

Introduction

Contents

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.

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.

1.1 Thesis definition

This thesis investigates the natural gas consumption forecasting for two horizons, long term and short term using a batch of artificial intelligence and statistical techniques. Relating to artificial intelligence techniques, it includes designing two types of ANN: standard feed forward neural networks MLP and the recurrent neural networks Long Short Term Memory (LSTM) by considering various factors that could have an impact on the natural gas consumption. In addition, this thesis is an in depth study of exploring the two-stage hybrid approach known as the divide and conquer approach. Finally, the core of the thesis is based on the two-stage method to design a novel hybrid forecasting approach. The new approach explores the trend and patterns in the consumption data in order to accurately predict the future consumption.

1.2 Research questions

Through the upcoming sections of the thesis, we will try to solve and answer the outlined questions:

1. Why computational intelligence based models for natural gas consumption forecasting?

   Due to the high demand of energy load forecasting, a batch of techniques has been developed to tackle this problem. However, in the recent decades a new type based on artificial intelligence has been proposed, the use of these kind of methods allow to find complex non-linear relationships between considered variables and the energy consumption which could not be afforded by other statistical or mathematical techniques. These methods include artificial neural networks, fuzzy logic, support vector machine, etc.

2. What type of exogenous variables could be used for better prediction accuracy?

There are many variables that may influence the trend of natural gas consumption whether for short or long term. Through this thesis, variety of variables is used like weather factors (wind speed, humidity, nebulosity) and economical factors (oil prices, number of clients, GDP, natural gas price,... etc). Besides another factor representing calendar information (day of the week, is a holiday day, season) is also used for short term forecast only.

3. What are the followed strategies to improve the predictions?

   Since the present thesis is focusing on both long and short term forecasting, two batch of models were applied and investigated to reach the highest accuracy. Also, in order to find the best possible input features combination, multiple feature combinations have been evaluated beginning by choosing the appropriate range of precedent endogenous (load) inputs, and then selecting the most influential exogenous features.

4. How good are the obtained predictions compared with traditional techniques?

   To give the obtained predictions a reliability, the long term results is compared with the Multiple linear regression method. Due to having two contributions on the short term forecasting, the first contribution's which is based on investigating multiple clustering methods to classify the consumption time series is compared with two other proposed methodologies in the state of the art that divides the natural gas consumption dataset. The second contribution approach in the same forecasting horizon is compared with three well known benchmark methods: SARIMAX, MLR and MLP methods.

1.3 contributions

The main contributions included in this thesis can be summarised as follows:

1. Design of several ANNs models to predict the national natural gas consumption in the long term horizon. These models has been used to develop a forecasting software and then to be used by the Algerian national company of electricity and gas (SONELGAZ). The software finally provides a yearly forecasting of the natural gas consumption for a range of 50 years in three main aspects:

  - Forecast regional residential consumptions, divided in distribution offices across the country.
  - Forecast the natural gas consumption for the economical activities sector (industrial, agriculture, residential, etc...). The predictions has been related to the three pressure levels in the Algerian market (High, medium and low).
  - Forecast the natural gas consumption used for electricity production which being generated by three types of turbine power technology: Steam turbines, gas turbine and combined Cycles turbines.

2. Following and adopting the two-stage based approaches to model short term natural gas consumption, an investigation of multiple different clustering techniques is made. For a better analysis and classification of the natural gas consumption, the hourly time series data is split using different kind of clustering methods. After division process is complete, multiple local auto-regressive Gaussian Process models are developed for a specific variation according to each clustering method. Finally, the impact of each clustering method on the constructed Gaussian Process models is assessed in terms of forecasting accuracy.

3. The essential contribution is: proposing a novel hybrid forecasting approach that resolves the two-stage method's crucial deficiency, by designing a Multi-Layered Perceptron Neural Network as a nonlinear forecasting monitor. This model estimates the next day gas consumption profile and selects one of several local Long Short Term Memory recurrent neural network models to perform the forecast.

1.4 Research approach

The foundation of this thesis lies on the natural gas consumption datasets provided by SONELGAZ, as mentioned above, two datasets were used. The first one is composed by 14 yearly points starting from 2000, and measured for three major sectors: residential, industrial and power generation sectors. The second dataset consists of hourly consumption recorded in the year 2014. In addition to natural gas consumption data, for the yearly basis, GDP, oil price, population, clientele size data are also included. Whereas the hourly consumption dataset only includes temperature.

Basically, to answer the research questions outlined above, an extensive researches on the dataset is made based on MLP and LSTM model designed using Keras which is a high level neural networks API written in Python. GP construction and testing is based on the GPy framework also written in python and developed by Sheffield university machine learning group.

1.5 Thesis outline

After presenting in this introductory section the backgrounds and motivations, the research questions that this thesis is trying to tackle, the research approach and the main contributions, the present thesis is structured as follows:

- Chapter 2: State-of-the-art: This chapter presents a full review of the past researches in the literature during the last decades on the natural gas consumption forecasting. More precisely, this chapter covers mostly the realised works on the long and the short term horizons.

- Chapter 3 Artificial neural networks: this chapter presents the basic theory of artificial neural networks, with a deep introduction to two different architectures, the Multi-Layered Perceptron topology and the Long Short Term Memory which is recurrent based neural network. Following full description of the neural networks models, also a detailed explanation of all the occurred calculation to learn and develop these two models is given.

- Chapter 4 is devoted to explore the two-stage approach which consists of classifying the natural gas consumption daily pattern sequences using different non-supervised methods followed by modeling each consumption cluster using Gaussian Processes.

- Chapter 5 is concerned with the proposed forecasting approach which presents a new efficient functionality to estimate the next day consumption profile, and thus leads to a significant improvement of the forecasting accuracy, especially for days with exceptional customers consumption behavior change.

- Chapter 6 is dedicated to study the natural gas consumption for the long term horizon. An analysis of the forecasting results obtained by MLP ANNs compared with linear regression is included and finally presenting the PREVGAZ long term forecasting tool which is based on the obtained results.

Chapter 2

State of the Art

Contents

Energy consumption forecasting studies differ according to the horizon and thus can be categorized into: yearly, monthly, daily and hourly scales, where many different methods have been employed to forecast energy consumption on these four basis. In the field of natural gas load forecasting, Soldo (Soldo 2012) provided an analysis and synthesis of published researches and presents a classification based on prediction horizons, model paradigms, area of study, etc. Regarding the dedicated efforts to improve the reliance and accuracy of future natural gas consumption forecasting, the last two decades witnessed a widespread investigations in that direction (Hong 2013).

With the final goal of performing efficient one hour to a day ahead forecasting, most of this thesis belongs to the fourth category. Practically, researches in this area have been conducted by developing three classes of models: statistical time series based techniques (Seasonal Auto-regressive Integrated Moving Average with exogenous inputs (SARIMAX) (Taşpınar et al. 2013)(Elamin & Fukushige 2018), functional auto-regressive with exogenous variables (Chen et al. 2018), Vector Autoregressive (García-Ascanio & Maté 2010), extended kalman filter (Fagiani et al. 2015), etc), non-linear regression methods (Support Vector Regression (SVR) (Bai & Li 2016), Multivariable Adaptive Regression Splines (MARS) (Özmen, Yılmaz & Weber 2018), Classification and Regression Trees (CART) (Lusis et al. 2017), etc) and computational intelligence based models (Artificial Neural Networks (ANNs) (Ruiz et al. 2018)(Karimi & Dastranj 2014), Fuzzy Logic (Barak & Sadegh 2016), Gaussian Processes (Maritz et al. 2018), etc).

2.1 Statistical time series based techniques

2.1.1 Linear regression technique

Although the huge number of alternative statistical techniques, this one is among the most utilised. The principal idea of multiple linear regression in its simplest form is to find the relationship between independents and dependent variables. Regardless the complexity of natural gas load pattern, generally, the use of MLR is based on determining how historical load, meteorological conditions, calendar factors, etc as independent variable impacts the future consumption. The MLR model can be defined as:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_n x_n + \varepsilon \tag{2.1}$$

denoting y is the dependent variable, $x_1, ..., x_n$ are n independent variables, $\beta_0, ..., \beta_n$ are the regression coefficients and ε is the error term.

Finally, after fitting the regression coefficients to minimize the difference of observed and predicted loads, and after preparing the future explanatory variables X_f which correlated with the willing-to-forecast load y_f, the final forecasting form is given by:

$$y_f = X_f \beta \tag{2.2}$$

Linear models have been found to be part of modern econometric methods especially on energy consumption time series modeling. Tao Hong et al. (Hong et al. 2010) developed regression models to predict hourly power load. The model training was based on 3 years dataset which includes historical load and temperature, also different other input variety has been tested involving calendar information as well. Saber and Alam (Saber & Alam 2017) presented modeling approach that uses a multivariate regression technique to achieve a week-ahead forecasting with 168 steps.

2.1.2 AR models

Assuming that time series internal structure follows a linear attitude with a combination of previous loads with an inclusion of autocorrelation, trend or seasonal variations. For such time series, an auto-regressive technique can be suitable for its modeling and therefore provide a convenient solution to forecast the future load values. The general form of an AR model with p^{th} order is defined as:

$$L_t = \sum_{i=1}^{p} \phi_i L_{t-1} + \varepsilon \tag{2.3}$$

where ε is white noise or prediction error and ϕ is the model's coefficient. The simplest AR is a model with a first-order AR(1), where only one lagged past load is considered. There are many ways to estimate the coefficients, such as the ordinary least squares procedure (Mbamalu & El-Hawary 1993) or method of moments known as Yule-Walker method (Eshel n.d.).

2.1.3 ARMA models

Firstly introduced by Box-Jenkins in (Box & Jenkins 1994) presenting the basic form for forecasting purposes due to its high level prediction abilities for certain variety of data. From its name, the core of this model is in fact a combination of an autoregressive and a moving average models, in other term, each load point L_t of the time series is linearly explained with lagged points $(L_{t-1}, ..., L_{t-p})$ and previous noise values $(\varepsilon_{t-1}, ..., \varepsilon_{t-q})$. The general form of a moving average model with the q^{th} order MA(q) is given by:

$$L_t = \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-1} \tag{2.4}$$

Using same coefficients from equations. 2.3 and 2.4, the autoregressive moving average model with an order of (p,q) is then written as:

$$L_t = \sum_{i=1}^{p} \phi_i L_{t-1} + \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-1} \tag{2.5}$$

2.1.4 ARIMA models

Basically, ARMA models perform well in case of having stationary time series, where in fact this may arises an issue when it comes to deal with weakly stationary time series. In (Box & Jenkins 1976) Box and Jenkins proposed another model that includes autoregressive, moving average parts as well as a differencing process to transform the time series to be more stationary. Practically, the ARIMA(p, d, q) models requires three parameters: the autoregressive parameters $(\phi_1, ..., \phi_p)$, the moving average parameter $(\theta_1, ..., \theta_p)$ and the number of differencing steps d.

Moreover, in the earlier decades of using statistical techniques to forecast energy consumption, ARMA models knew an extensive use for this respect. Demirel in (Demirel et al. 2012) has successfully employed ARMA model to forecast the natural gas consumption time series in Istanbul. The methodology was based in addition to historical values of consumptions, three other factors were also included such as temperature, natural gas price, and number of consumers. Formally, this yield to a new model called ARMAX, where the 'X' here refers to the independent variables. For the same city, Ervural et al. (Ervural et al. 2016) proposed an ARMA technique that is fitted using Genetic Algorithm (GA) in order to model monthly natural gas consumption time series. The idea was to use GA for an effective identification process to estimate the ARMA parameters.

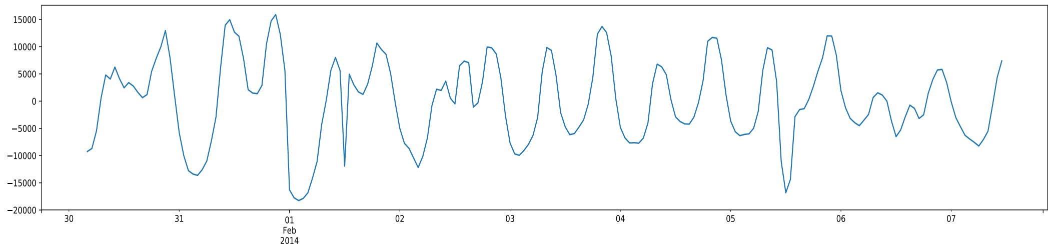
2.1.5 SARIMAX models

Another strong baseline approach, which is an enhanced version of ARIMA models, where the seasonality in time series are taken into account, making it a first class time series technique, introduced by Box and Jenkins (Box 2013) and becoming since, one of the popular methods for short term load forecasting. The multivariate version of SARIMA has enhanced the ability to integrate explanatory (exogenous) variables and adding seasonal terms in order to increase the performance.

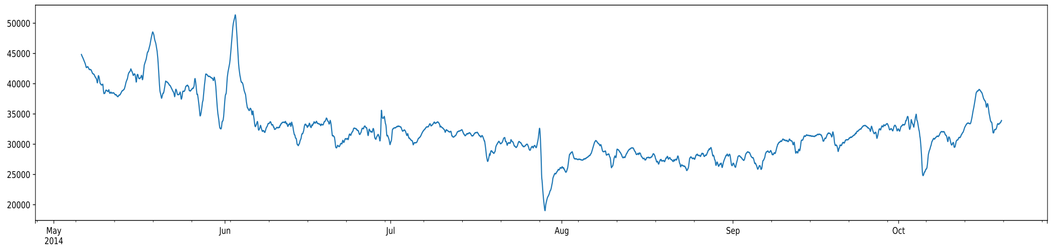
The basic concepts that should be introduced as a prelude for this section, especially for the statistical seasonal time series based models are the concepts of autoregression, autocorrelation, time series stationarity, differentiation, moving average error, etc which could afford a significant understanding of the time series in order to build a suitable forecasting model.

In a time series with trends or seasonality, the value is affected by time, therefore it is called stationary when the value of time series is not dependent on time. For instance, a random white noise is a sample of typical stationary time series. Daily temperature at a location is not stationary as there will be a seasonal trend and it is affected by time. Meanwhile, the noise component of a time series is stationary. Stationary time series do not have any means of being forecasted as they are completely random.

Fig. 2.1a is an example of non-stationary data because of the presence of both trend and seasonality and Fig. 2.1b is an example of stationary data because there is no clear trend or seasonality.



(a)



(b)

Figure 2.1: Two samples of a non-stationary and a stationary time series.

Since time series having trends or a redundancy of some patterns which cause the seasonality, the load are then became more related to time. Seasonal ARIMA can be expressed as the notion 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 and s is the span of repeating the seasonal pattern.

Using the SARIMA ability of modeling the inner seasonal cycles in time series, James has developed three models including SARIMA (Taylor 2010) to perform hourly electricity demand forecasting. The 6 years data allowed to employe a third seasonal cycle besides the intraday and intraweek ones, the use of such data with multiple years motivated the modeling of an annually seasonal cycle which then proven laterally to be more accurate compared to the double seasonal models.

Liu et al. proposed SARIMAX model in comparison study with ANN in (Liu et al. 2014) for short term load forecasting. The study includes two steps, the first one presents a temperature forecasting where ANN had performed better than SARIMAX. In the second step, using the obtained forecast from step number 1, For the order of autoregressive and moving average in the seasonal component, as there is obvious autocorrelation at lag 24, the SARIMAX included a SAR and SMA part with order one at lag 24, and for non-seasonal component, the optimal value of 5 was identified for the order of autoregressive and moving average. The final results shows a superiority of ANN on a day-ahead forecast, whereas, SARIMAX had the best performance on the weekly basis where a 168 steps ahead is performed.

2.1.6 Grey models

New concept has been introduced by Deng in 1982 (Ju-Long 1982) to solve special problems where information could not be sufficiently available, in other term, the actual objective of the grey models theory is to estimate the system behavior with very limited data. The standard form of a Grey system is noted by GM(1,1) which is a first order variable time series prediction model. GM(1,1) model relays on 2 parameters only, a represents the grey developed coefficient and b represents the grey controlled variable. These two parameters are estimated by the least square estimation method following a first-order deferential equation of GM(1,1). The future load value L is then predicted according the following equation:

$$L(t+1) = (1 - e^{a})\left(x(1) - \frac{a}{b}\right)e^{-at} \tag{2.6}$$

For energy consumption time series, specifically for electrical load forecasting, Yao (Yao et al. 2003) proposed a transformed grey-based model with an improved way to find the parameter a in terms of calculation time and prediction accuracy. The proposed Grey model was used to achieve a ver short term forecasting for demand-control of electricity in Taiwan, where each point of the time series were recorded in minutes.

Another application of Grey models was carried out by Qiang et al. in (Zhou et al. 2008). They employed multiple grey models to develop an online hourly forecasting system for building thermal load in Hong Kong. The project had included three prediction modules beside the main one. The first was a regressive module for solar radiation estimation that counts on the forecasted cloud amount, the second and third modules are dynamic grey models to forecast temperature and relative humidity. Finally, a modified GM(1,1) model and convolutional GM(1,1) model are trained and constructed based on the predicted weather conditions information to forecast the next day building thermal load.

An enhanced version of GM was proposed by Jin et al. (Jin et al. 2012) to model hourly power load in China. Basically, the essential of the proposed idea is based on two optimization phases, an external and an internal phases. The internal phase does follow a different modeling feasibility test for better choosing the best performance model and an improved process to estimate the parameter a by adding an extra variable into the parameter calculation equations. The external optimization was to develop three strategies or models for a specific period of time in the day. Thus, the daily load time series division is occurred according valleys and peaks: the midnight segment (from 0:00 to the first peak), morning segment (from the first peak to the first valley), the afternoon segment (from the first valley to the evening peak) and finally, the evening segment. The next load forecasting then achieved after selecting the right strategy comparing the correlation coefficients of the three strategies for each segment.

Another improved GM was proposed by Li et al. (Li et al. 2011) to handle one step ahead short term load forecasting problem in Japan. The proposed model called 3spGM(2, 1) by transforming the basic GM(1,1) from a first-order single variable to a second-order grey model. The main reason for this amendment was to avoid the bad forecasting performance when the original data show a considerable randomness.

2.2 Non-linear regression methods

2.2.1 SVR models

Support Vector Regression (SVR) is a regression formulation of the Support Vector Machines (SVM) proposed in (Vapnik 2000). The theory is well established and is explained in several excellent works [31 33]. The SVR model can be represented by a linear combination of N kernel functions denoted by ϕ_i , weights w_i , and bias b:

$$y(t+1) = \sum_{i=0}^{N} w_i \phi_i(u_i) + b \tag{2.7}$$

Concerning energy time series forecasting, Potočnik et al. (Potočnik et al. 2014), Han Liu et al. (Liu et al. 2004) have used SVR models to perform daily natural gas demand forecasting. The SVM implementation was based on 4 datasets: natural gas consumption dataset recorded on daily basis for the years 2001 and 2002, daily average temperature for the same years, dates of holidays and calendar dataset representing the day of the week. After training a SVM model which include a Radial Basis Function as a kernel function, the results was compared with a combined SOFM and MLP and showed a better performance for SVM over the ANN approach.

Another SVR based application was conducted by Clark and Lou Lim (Clark et al. 2018), where they have applied a SVR to model electrical load time series for short term forecasting. The study was setup to achieve a day ahead prediction in quarter of hours resolution using a SVR model with a Radial Basis Function Kernel.

2.2.2 Multivariable Adaptive Regression Splines models

MARS is a sort of non-linear data driven method, firstly introduced by Friedman (Friedman 1991), as a regression model in the first place, includes various Basis Functions (BFs) called Splines taking the role of being the predictors accordingly with the given original data. Basically, MARS method can be seen as an extension of linear model that automatically models nonlinearities and interactions between variables. The training of MARS models is done through two phases, at first, optimum quantities of FBs and knot locations is determined, and then the least-squares method is applied to fit the model in order to reach the best data approximation utilizing the remaining BFs.

Ayşe et al. (Özmen, Yılmaz & Weber 2018) developed MARS and Conic MARS models to predict the residential natural gas consumption for a region in Ankara. Based on four years daily consumption data, together with meteorological data (average temperature and HDD, average humidity and average of air pressure), gas price and exchange rate of Turkish Liras to USD, both MARS and CMARS were build and compared with ANN and LR model, the comparison was made by predicting consumption for 2013 and where it showed a superiority over the benchmark methods.

Rezaie et al. (Rezaie-Balf et al. 2019) have used MARS models for a different kind of time series. Daily Solar Radiation (DSR) was forecasted in a two stage methodology, at the first step data decomposition was fulfilled with the use of Complete Ensemble EMD with Adaptive Noise (CEEMDAN) method intentionally to overcome the non-stionarity of the time-series, then at the second stage, a MARS, a Self-Adaptive MARS (SaMARS) and a Gene Expression Programing (GEP) models were developed in order to predict the DSR at two stations in South Korea.

2.2.3 Regression tree models

Also called Classification and Regression Trees (CART) proposed by Breiman (Breiman et al. 1984) is a developed form of Decision trees approaches to tackle regression problems such time series forecasting. Dedicating the used of decision trees for continuous data prediction as regression, CART uses the same concept of decision trees which find the relationship between inputs and outputs through a hierarchical structure consisting of nodes, branches and leaves. Unlike classification tree, where it provide discrete class labels or probability distributions, CART method handle the decomposition of the domain into a certain amount of regions in which target function values can be approximated with sufficient accuracy to generate numerical predictions.

Lusisa et al. (Lusis et al. 2017) have studied Regression Trees beside three other techniques (MLR, SVR and MLP) to perform a day-ahead load forecasting for the residential sector with a 30 minute prediction per step. The training of these models was done based on Historical load intra-day profiles, daily and weakly patterns, weather conditions data also been considered. To handle the within time series seasonality, calendar informations were added to the models as binary variables. The test RMSE results was pretty similar for all the models but generally, Regression Trees seems be be significantly the best forecasting model.

Ruiz-Abellón et al. in (Ruiz-Abellón et al. 2018) have made an evaluation study to investigate the performance of multiple Regression Trees based methods to forecast hourly electrical load of a campus university in Cartagena (Spain). Namely, bagging, random forest, conditional forest and boosting, were four methods constructed based on the most commonly used data: historical loads, temperature, calendar variables and other binary variables presenting the special days in the academic context. Finally, effectively achieving a predictions with a horizon of 48 hours.

2.3 Computational intelligence based models

2.3.1 ANN models

As inspired by the function of the human brain that comprise large amount of processing cells called neurons, Artificial Neural Networks models are one of powerful techniques with an intense ability to approximate to any nonlinear phenomena in the data (Gooijer & Kumar 1992). This concept relies in the first place on the connections between neurons which store some sort of information that can be used and retrieved laterally. These connections or what are called 'weights' are identified or determined according to given informations and to the characteristics of the data. Fitting the ANN weights is occur by training the model with a collection of data arranged into pairs, inputs and outputs, and they are calculated through a batch of equations explained in Chapter. 3, Section. 3.1 to reduce the squared difference between the given outputs and the model's outcomes.

ANN 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 (Szoplik 2015) 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 (Yu & Xu 2014) also proposed a MLP ANN. In (Yu & Xu 2014) instead of randomly initializing the network weights and thresholds, genetic algorithms are used then the network is trained with an improved Back Propagation (BP) algorithm by including additional momentum and self-adaptive learning rate. Tonkovič et al. (Tonković et al. 2009) develop an MLP and Radial Basis Function (RBF) to model 24 hours consumption time series for the north-east region of Croatia. Taşpınar et al. (Taşpınar et al. 2013) 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 (Park et al. 1991),(Lee et al. n.d.), many researchers have explored alternative types of ANNs. Recent studies have focused on Recurrent Neural Networks (RNN) and recurrent Long Short Time Memory (LSTM) networks. In the same context, Kong et al. (Kong et al. 2017) 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. Another comparative study was made by Bianchi (Bianchi et al. 2017) by reviewing five different types of RNN, namely, ERNN, LSTM, GRU, NARX and ESN. The prediction performance of these latest models is compared based on some benchmarks and real-world short term forecasting problems. As final conclusions, in terms of training time and complexity, ERNN, LSTM and GRU had the slowest training process, and in terms of prediction accuracy, LSTM and GRU seem to have the best performance in most cases.

2.3.2 Fuzzy logic models

Fuzzy Logics can be described as systems that employs of human expertise through three main conceptual parts: fuzzy rules as a rule base, a database which used linguistically to define the Membership Functions (MF) and a reasoning or a logic mechanism to derive an output through an inference procedure based on the fuzzy rules. More precisely, FL performs a non-linear mapping starting from the input space to the output space through a bunch fuzzy if-then rules, where each rule describes a certain local behavior. Accordingly, the if-then parameters at the first layer of the FL determine the fuzzy region in the input space, and output parameters determine the corresponding output. However, there is no standard procedure to define the MF parameters. Utilizing the ability of learning by ANN, a combination of FL and ANN is proposed by Jang in (Jang et al. 1991) and (Jang 1993) benefiting the advantages of each method to design an Adaptive Neuro-Fuzzy Inference System (ANFIS) structure the uses the fuzzy system to represent information in interpretable way, and the learning ability of ANN for an automatic Fuzzy rules estimation and parameter optimization.

For energy consumption forecasting, Fayaz and Kim (Fayaz & Kim 2018) have designed an ANFIS model to prediction energy consumption in residential buildings. In four steps methodology, staring by data acquisition, preprocessing then prediction and assessment to build three models including ANFIS where a different numbers and types of membership function were tested, finally perform a week-ahead hourly forecasting.

In (Mamlook et al. 2009), Mamlook et al. proposed a Fuzzy Inference system to predict electrical load on hourly basis. In addition to historical data, the study involves three others (temperature, weather and a binary set to determine the day type), and the output is divided into seven fuzzy sets (Very very low, Very low, Low, Normal, High, Very high, Very very high) according to the degree of consumption with respect to the inputs characteristics.

Concerning natural gas time series modeling, Azadeh in (Azadeh et al. 2010) has presented an ANFIS model for this purpose. Since the demand time series was on daily basis, standard historical variables presenting the day of the week, demand of the same day in previous year, demand of a day before and demand of 2 days before were used to form the final structure of the model in order to accurately forecast the Iranian natural gas demand. Comparing the obtained results with ANN, the ANFIS model had the best performance.

Panapakidis and Dagoumas also tackled the same problem (Panapakidis & Dagoumas 2017), where they proposed a hybrid model based on ANFIS to model daily natural gas consumption time series. The model was a combination of 4 techniques: firstly uses Wavelet Transform (WT) to decompose the time series into a set of sub-series, then, uses GA to optimize the ANFIS parameter according to each sub-series. Finally, the ANFIS output is used as input beside other exogenous variables to acquire the final prediction.

2.3.3 Gaussian Process models

Another machine learning method for regression , the Gaussian Process Regression (GPR) models have been successfully applied to many different areas such as electricity load forecasting (Lourenco & Santos 2010) (Alamaniotis et al.

2014), wind power forecasting (Chen et al. 2013), groundwater level time-series modeling (Raghavendra & Deka 2016), etc.

Gaussian Process models are considered as a collection of random variables that predict consumption C_t at time t for a given input x_t. Assuming that f is a latent function, which provides the values for each data point according to:

$$C_t = f(t) + \alpha \tag{2.8}$$

where $\alpha \sim N(0; \sigma^2)$ is a Gaussian noise with a zero mean and a variance σ^2. Noting that a Bayesian inference is performed and hence the posterior predictive distribution of f can be written as follows:

$$f(x) \sim GP(m(x), k(x, x')) \tag{2.9}$$

where $f(x)$ is the real process to model, x and x' are two different points. Here $m(x)$ is the mean value which is equal to zero in this case and $k(x, x')$ is the kernel function. The hyper-parameters (called the length-scale) can be varied to increase or reduce the correlation between points and consequentially the smoothness of the resulting function.

Depending on the hyper-parameters of the kernel function, predictions are correlated with already observed values that have been recently observed. However, the influence of different variables on gas consumption is defined by the hyper-parameters of the covariance function which can be derived by maximizing the marginal likelihood. The log-marginal likelihood (LML) is defined as:

$$log\ p(X, \theta) = -\frac{1}{2}C^T K^{-1} C - \frac{1}{2}log|K| - \frac{n}{2}log 2\pi \tag{2.10}$$

where the first term is the data-fit, the second term is a complexity penalty and the last term is a normalizing constant with n being the number of training samples.

Prakash et al. (Prakash et al. 2018) have investigated GPRs potentials to forecast energy consumption on two horizons, long-term horizon on daily basis and short-term horizon on hourly basis. Various covariance functions were tested on three different datasets with a variety in data size, prediction horizon and load type. For the short term forecasting, GPR models with a linear kernel with noise were used based on historical consumption and temperature values to predict the next 30 minute electrical load, cooling load and lightning load. In the other hand, for long-term horizon, GPR models with a combination of

Matern and linear kernels are used to predict the previously mentioned energy consumptions for one to five days-ahead.

Lourenço and Santos (Lourenço & Santos 2012) presented another application of GP for short-term load forecasting. The study was to examine the GPR efficiency in a regression problem basing on endogenous data only. Different input vectors including different set of lagged values were tested, and the most accurate GPR model was selected to predict the electrical load for the next 24 hours.

2.4 Hybrid approaches

Hybrid forecasting techniques have been developed and have practically proven their efficiency by achieving high accurate forecast results compared with single model approaches (Fan & Chen 2006), (Chou & Tran 2018),(Du et al. 2018). They have demonstrated their reliability in solving complex non-linear predictions and control problems. Hernandez et al. (Hernández et al. 2013) proposed hybrid approach consisted of 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 model to predict the next day's total load. In a similar manner, Krzysztof and Tomasz (Krzysztof & Tomasz 2017) presented 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 then 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 as well, Ilic et al. (Ilic et al. 2012) presented 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 forecasted day. Ghadimi et al. (Ghadimi et al. 2017) have interested in forecasting the electrical load and price and proposed 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. (Zhang et al. 2018) proposed 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's results are summed to obtain the final forecast. Substantiating the superiority of the concept of divide and conquer hybrid approaches over traditional single model ones, Xiao et al. (Xiao et al. 2018) proposed 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 focus on the left residual sequence which is the non-linear sub-series, then utilise a group of prediction models, namely, BP ANN, SVR, GP and RBF ANN. The final decision is made by a special ANN model that establish a selective combination forecasting. Thus, both outputs from AR model and the optimal complexity model are aggregated to obtain the final energy consumption value.

2.5 Discussion and conclusion

Despite the very successful application of the above-mentioned techniques in time series modeling, especially with energy consumption problems, each method proved to have certain advantages and disadvantages which mainly affected by its characteristics and the nature of the case-of-study. For this reason, in the following section we will try to put the light on each technique's weaknesses and strengths.

When it comes to statistical methods, both MLR, AR and MA techniques provide very simple computation and calculation models, and they are even able to perform well with a limited observations. Although, AR or MA models provide poor results compared with to sophisticated techniques when dealing with non-linear problems. On the other hand, ARIMA is considered as a universal approximator, and it can fit to any time series due to the regressing and averaging elements. However, selecting and identifying the best model's parameters is complex and time consuming process. Furthermore, regarding time series, ARIMA models are not suitable for a long-term prediction where many steps ahead are performed.

One of the strengths of Grey systems is that it can approximate even in low information environments due to its ability to solve problems with very limited observations. As explained above, the Grey models are easy to be applied for time series prediction and also easy to be trained and constructed. As weaknesses, Grey models seem to straggle with random or noisy patterns which make it not useful to forecast time series that include special curves.

As computational intelligence methods, ANN seems to be more advantageous than other statistical and regression methods, particularly, ANN can non-linearly map the input/output relationship without any prior knowledge, and this gives ANN more flexibility and efficiency in energy time series modeling. However, some precautions are needed to be heeded when utilising ANN, firstly with weights values initialisation which could cause a bad convergence, secondly with the apprehensive of falling in a local minima, and thirdly with keeping balance between obtaining a well-generalized model and an overfitted one. Concerning GPR models, it is seen that they have the same advantages like ANN such as dealing with non-linear problems, able to carry out long-term time series forecasting with a limited set of data. The same as ANN, finding the best kernel and the optimum parameters can be more exhaustive during the calculation process especially when dealing with huge datasets.

For non-linear regression methods, both Decision Tree and Fuzzy Systems can clearly show the relationship between input and output variables through the splitting conditions and allow to visualize the structure of the problem. The main advantages of this approach are the flexibility of the model (efficient for linear and non-linear problems), no preprocessing or transformations on the data are required, and provide very accurate predictions. However, this type of models exhibits the problem of recognizing the random or noise pattern in the time series, which defectively effects their performance.

The performance of conducted researches that have been proposed to examine real-world data is shown in Table. 2.1 including the forecasting horizon next to the recorded MAPE.

Table 2.1: A Review of most commonly used models for energy load prediction conducted for various countries

Models	**Model type**	**Data**	**Prediction horizon**	**Accuracy MAPE(%)**	**Ref.**
MLR	MLR	Hourly electricity load (USA)	24 hours	4.55	(Hong et al. 2010)
	MLR	Hourly electricity load	168 hours	3.99	(Saber & Alam 2017)
ARIMA	ARMAX	Daily natural gas consumption (Istanbul)	1 day	0.19	(Demirel et al. 2012)
	GA-ARMA	Monthly natural gas consumption (Istanbul)	1 month	0.13	(Ervural et al. 2016)
	SARIMA	Month electricity load (Turkey)	12 months	2.60	(Bozkurt et al. 2017)
	SARIMAX	Month electricity load Abu Dhabi (UAE)	24 hours	1.58	(Bozkurt et al. 2017)
Grey	GM(1,1)	Hourly solar radiation (Hong Kong)	24 hours	7.34	(Zhou et al. 2008)
	T-3spGM(2,1)	Hourly electricity load (Japan)	24 hours	0.85	(Li et al. 2011)
	GM(1,1)	Hourly electricity load (China)	24 hours	4.55	(Jin et al. 2012)
	SGM(1,1)	Hourly electricity load (China)	24 hours	6.13	(Wang et al. 2018)
CART	XGBoost	Hourly electricity load in campus university (Spain)	24 hours	8.14	(Ruiz-Abellón et al. 2018)
SVR	SVR	Hourly natural gas load in Xi'an (China)	7 days	1.62	(Liu et al. 2004)
	SVRM	Hourly electricity load	24 hours	4.09	(Clark et al. 2018)
	LS-SVM	30-min wind speed in Penglai (China)	3 steps	10.62	(Du et al. 2018)
		30-min electricity load (Singapore)	3 steps	0.68	
		30-min electricity price in New South (Wales)	3 steps	2.13	
MARS	MARS	Hourly energy prices (Canada)	168 hours	11.8	(Zareipour et al. 2006)
	CMARS	Daily natural gas consumption (Ankara)	1 day	0.05	(Hong et al. 2010)
ANN	MLP	Daily natural gas consumption (Turkey)	1 day	0.81	(Taşpınar et al. 2013)
	MLP	Hourly natural gas consumption (Poland)	24 hours	5.6	(Szoplik 2015)
	MLP	Hourly natural gas consumption (Croatia)	24 hours	9.36	(Bozkurt et al. 2017)
	DELM	Hourly residential buildings energy consumption	168 hours	5.70	(Fayaz & Kim 2018)
	LSTM	Hourly smart meter data (Wales)	12 hours	8.64	(Kong et al. 2017)
Fuzzy	Fuzzy logic	Hourly electricity load (Jordan)	24 hours	<5	(Mamlook et al. 2009)
	ANFIS	Daily natural gas consumption (Iran)	1 day	0.02	(Azadeh et al. 2010)
	ANFIS	Daily natural gas consumption (Greece)	24 hours	3.63	(Panapakidis & Dagoumas 2017)
GP	GPR	Hourly electricity load (Portugal)	24 hours	<1.5	(Lourenço & Santos 2012)
	GPQR	Hourly electricity load	1 hour	3.22	(Lauret et al. 2012)
	GPQR	Hourly electricity load	1 hour	1.73	(Yang et al. 2018)

In real-world problems which are even more complex to be handled with a single model, whether it is a statistical or a machine learning technique, capturing the complexity in energy load time series is highly challenging. Therefore, many researchers decided to try utilising new sort of approaches by designing hybrid models to take advantages of different abilities in a time. However, it is still more difficult to find which models to combine for the time series problem modeling. As an example, one of the famous combination approaches is the ANN-ARIMA, this approach allows to exercise the ANN's ability of modeling the non-linear components in energy time series and similarly with the ARIMA's ability of modeling linear component. Consequently, a strong model is developed to effectively improve the forecasting accuracy. In the same way, other different models could be combined for this respect. Table. 2.2 presents a review of several different combinations of hybrid approaches for energy consumption time series forecasting.

Table 2.2: Review of hybrid energy forecasting approaches.

Models combination	**Novelty**	**Ref.**
ILP-HLP	Build 2 MLP models where the first MLP(ILP) predicts the total load of the next day then feeds it to the second MLP(HLP) to obtain the forecast for each 24 of that day.	(Ilic et al. 2012)
ARIMA-ANFIS	Develop 5 different MLP models to estimate the peaks and valleys values in the day then use them to perform a 24 hours power load forecasting.	(Hernández et al. 2013)
WT-ANN-ANFIS	Begins by predicting hourly power loads using WT and ANN, then improve the output with the use of WT and ANFIS.	(Hooshmand et al. 2013)
ANN-GA	Construct a ANN model with the use of GA to optimise its topology and parameters	(Karimi & Dastranj 2014)
FFNN-RBFNN-ANFIS	Develop these three different models to forecast daily heating energy consumption. Finally, use three way to combine each model's output to obtain the overall forecast.	(Jovanović et al. 2015)
ARIMA-ANN	Uses ARIMA to predict the periodicity and linear component in the hourly power load time series, and constructs an ANN model to predict the tendency and non-linear component.	(Zhuang et al. 2015)
ANN-RBFNN-SVM-OWA	Build three separate models, ANN, RBFNN and SVM then use the OWA method to fuse each model's prediction to obtain the final smart grid power load forecast.	(Ghadimi et al. 2017)
ANN-ABC	Propose a MLP model trained by the ABC for daily natural gas consumption forecasting.	(Akpinar et al. 2017)
IEMD-ARIMA-WNN	Extract the linear and non-linear component from the electricity load time series using IEMD, then build ARIMA and WNN models to model each component.	(Zhang et al. 2018)
CEEMD-WOA-LSSVM	Uses CEEMD method for time series decomposition, then estimates the LSSVR parameters with WOA to increase the prediction efficiency of the LSSVR model.	(Du et al. 2018)
(c,l)-LSTM+CNN	Design a hybrid network consisted of c LSTM models with l layers then followed by a CNN to model daily power load demand.	(Kim et al. 2019)
K-means-MLP-LSTM	design a MLP model as a nonlinear forecasting monitor to estimate the next day gas consumption profile then selects one of several local LSTM models to perform the forecast.	(Laib et al. 2019)
Two-stage approach	Divid natural gas consumption data into sub-sets based on ambient temperature then construct multiple LR models accordingly to forecast the next day consumption.	(Baldacci et al. 2016)
	Classify the daily electrical load segments into 3 profiles with respect to certain degrees of the load peak during the day, then develop three ANN models to handle each load profile's forecast.	(Krzysztof & Tomasz 2017)
	Split the electricity load according to load profiles based on K-means clustering method, then, several methods were build for each cluster.	(Shchetinin 2018)
	Use K-means to classify the similar load time series profiles then design a SDAE to predict the electricity load for each load profile.	(Farfar & Khadir 2018)
	Use different clustering methods to classify the similar load time series profiles to observe the impact of each method on the prediction accuracy of developed GPR models.	(Laib et al. 2018)

Chapter 3

Artificial Neural Networks

Contents

ANNs are an attempt to find a mathematical representation for information processing in biological systems. In 1943, McCulloch and Pitt (McCulloch & Pitts 1943) brought to light the term of "artificial neural network" in a way to use the learning capabilities of the human brain in order to make machines learn. From the aspect of practical applications on non-linear regression, neural networks have the ability to control the non-linear relationships between the desired loads and the related factors that could affect the load in a direct or in an indirect way.

In fact, the first ANNs version which is Mcculoch-Pitts models lacked procedure for learning, which considered as a crucial problem in order to be used for artificial intelligence. However, a research psychologist Rosenblatt has focused his works on the computational ability of the network units that constitutes a single layer networks (Rosenblatt 1957), and employ it with the perceptron training rule to model the basic OR, AND and NOT functions, where the classes are linearly separable. In other word, to make computers able to deal with formal logical reasoning which should litterally solve artificial intelligence problems. The essential work of Rosenblatt was basically to find a way that makes the perceptrons learn. However, continuing the Donald Hebb's work (Hebb 1962) where he inspired the idea of knowledge and learning that occurs in the human brain via changing synapses between neurons, stated by Hebb: "When an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased". In contract, Rosenblatt's perceptron idea was not following the exact Hebb's idea, so instead, he created a full intuitive and simple learning approach, where for each set of input-target samples if the perceptron output is too small, then increase the weights, otherwise decrease the weights of the output is higher than the target. Thus, this gave birth to machine learning where he built a computer that can learn to classify simple shapes and generate output to categorize the inputs with (true/1 or false/2) which actually known as classification.

Nonetheless, the Rosenblatt's perceptron design and learning schema declined with a disappointing analysis presented by Minsky and Papert in 1969 by publishing their skepticism on the limitations of Perceptrons (Minsky & Papert 2017), especially when it comes to learn the simple boolean function XOR due to its non linear separability. Furthermore, it may possibly solve such functions if a networks consisting in more layers is considered. However, at this level there were no available training algorithm dedicated to multi-layers architectures especially the one proposed by Rosenblatt, and he argued in his book that the learning rule outlined for perceptrons became useless when the case is related to multiple layer network. Besides, mentioning that a multi-layer network is required to represents a simple non-linear functions and that there is no viable way to train this kind of networks, Minsky has successfully convinced most of the artificial intelligent world that this is the dead-end of neural networks.

In 1986, Rumelhart, Hinton and Williams came out with one of the most used training algorithms for multi-layer networks until this day, (Rumelhart et al. 1988) their publication introduced the generalized delta rule for learning by back-propagation. In the same year, the same group also published another paper (Rumelhart et al. 1985) which was addressed especially for the problems discussed by Minsky in his famous book (Minsky & Papert 2017), so this revived the idea of designing a multi-layer neural network that could be trained to tackle complex non-linear problems. In 1989, (Hornik et al. 1989) another paper were published which considered essentially as a mathematical prove that a multiple layer allows neural networks theoretically to model any function like XOR, and therefore will be able to predict, classify and recognize patterns.

However, The focus of this chapter is on neural networks as efficient models for long and short term natural gas forecasting. Particularly, this part will mostly discuss two specific classes of neural networks that have been proven to have the greatest value on machine learning, namely the multilayer perceptron as feed forward network and the long short term memory as recurrent network, we then explain the detailed steps of the learning algorithm. Finally, discus the application results of these different neural network architectures.

3.1 Multi-layer perceptrons

A multi-layer feed-forward neural network is a set of interconnected processing units consisted in one input layer, one or more hidden layers and one output layer. This shape allows the captured informations injected in the input layer to be transferred through hidden layers until it reach the output layer in one way direction only (no feedback loops). The network parameters is then adjusted with a repeated disclosure to input/output training patterns. During the training process, the networks error is obtained by calculating the deference between the network output and the desired response, the network parameters are then repeatedly adjusted until the optimum state is reached where the network output is becoming close enough to the supervised output.

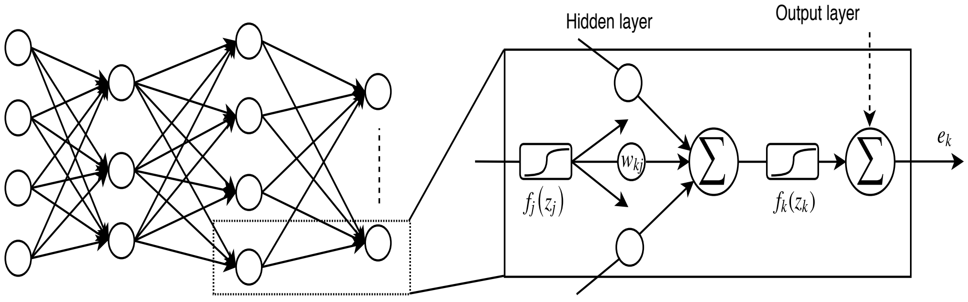


Figure 3.1: The MLP network structure with information processing through neurons.

Basically, there are two types of feed-forward architectures: single hidden layer structure and two hidden layers structures. In both, the activity of the input neurons is restricted on representing unprocessed informations only, whereas the role of the hidden neurons in the case of one or two layers networks is affected by the input signal and the connections weights between input and hidden neurons, the output reaction is therefore relies on the activations of the hidden neurons multiplied with the hidden-output connections weights. As been described in the introduction of this chapter, the Minsky's proposition pointed that in order to handle a non-linear tasks, the multi-layer architecture networks would definitely provide more computational efficiency compared with the single-layer structure networks which was practically proven on the XOR problem.

Technically, Figure. 3.1 illustrates a multi-layer perceptron network with D input unites, C output units and several hidden unites, which all arranged in layers. The j^{th} units in layer l computes the output according the following formula

$$y_j = \theta(z_j) \tag{3.1}$$

where

$$z_j = \sum_{i=1}^{K} w_{ji} y_i + w_{j0} \tag{3.2}$$

where w_n is the weighted connection between the n^{th} neuron to the i^{th} neuron in the next layer, and $w_{i,0}$ corresponds to the bias which considered as an external input to the neuron. Here, K denotes the number of neurons in the source layer. The θ denotes the activation function.

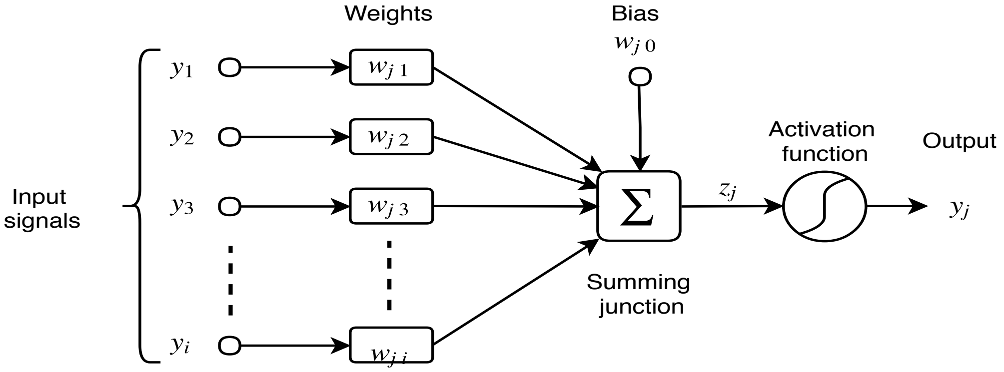


Figure 3.2: Illustration of a Nonlinear neuron, labeled j.

3.1.1 Activation function

In order to compute and limit the amplitude range y_j for each neuron of the multi-layer perceptron, a continuously differentiable activation function is required to be associated to the neurons. Basically, There are two identified types of activation function namely, threshold functions and sigmoidal functions.

1. Threshold function: The output in this type as illustrated in Figure. 3.3a is given by:

$$Threshold(z_j) = \begin{cases} 1 \; if \; z_j \geq 0 \\ 0 \; if \; z_j > 0 \end{cases}$$

2. Sigmoidal non-linearity: Unlike threshold units, sigmoidal units are so sensitive when the sum of inputs z_j is near 0, consequently, it makes of this type of functions being strongly recommended for constructing a multi-layer perceptron. There are two forms of the sigmoidal function to identify:

   (a) Logistic function: An example of the sigmoidal functions taking a S-shape, transforming the sum of input values into a range between [0,1].

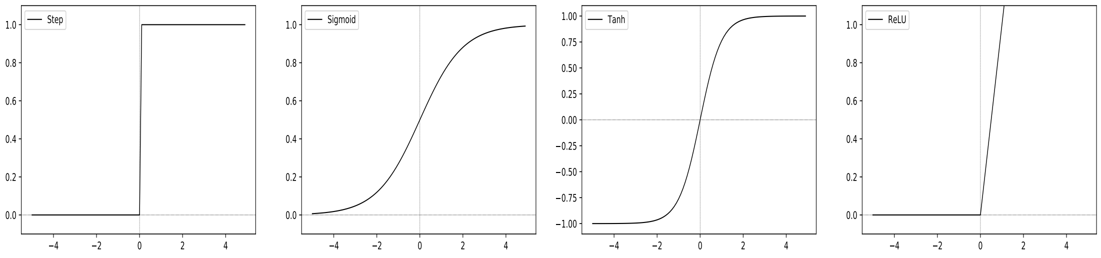
$$\sigma(x) = \frac{1}{1 + e^{-z_j}} \tag{3.3}$$

   (b) Hyperbolic tangent function: This activation function also has a S-shape where the output value lies in the rage of $0 \leq y_j \leq 1$. According to this function, it saturates to -1 or 1 when z_j becomes highly negative or highly positive respectively.

$$tanh(z_j) = \frac{2}{1 + e^{-2z_j}} - 1 \tag{3.4}$$

- In addition to the these activation functions, with less frequently used is the Rectified (ReLU) activation function, also known as the rectified linear unit and is a very simple function whereas each value's amplitude $z_j < 0$ is truncated at 0 and keep it as it is if $z_j > 0$ as expressed bellow:

$$ReLU(z_j) = \begin{cases} 0 \; if \; z_j < 0 \\ z_j \; otherwise \end{cases}$$



(a) Threshold (b) Sigmoid (c) Tanh (d) ReLU

Figure 3.3: Activation functions

The output amplitude is therefore can take many forms, theoretically, there is no strategy to find the correct activation function that should be used for a specific condition, in most cases, an empirical process is recommended in order to properly select the activation function for a given task.

3.2 Network training

3.2.1 Supervised learning

Considering learning with teacher which refers to the supervised learning paradigm, figure. 3.4 shows a diagram of this form of learning. The basic concept of neural network supervised training is to be guided by an external knowledge from the environment, and then be translated and represented by a set of input-output samples. These build-in knowledge samples is then used as training vector for the neural network providing the optimum desired actions or responses to be performed by the neural network according the entered inputs. The neural network parameters are then updated according both influences of the inputs vector and the produced error signal which is the difference between the desired response and the actual output of the network. After adjusting the parameters in an iteratively manner, the emulation of network behavior with the teacher is presumed, thus, once the neural network becomes tuned to and all the available knowledge from the environment are properly learned, the network is left to deal with the environment completely by itself.

3.2.2 Unsupervised learning

For this paradigm of the unsupervised learning also called self-organized learning, there is no external teacher to oversee the training process, thus, there are

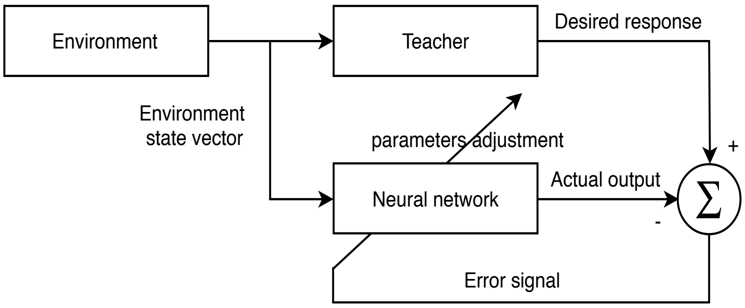


Figure 3.4: Illustration of the supervised learning paradigm.

no need for desired targets to be relied to. Instead, as indicated in figure. 3.5, the training is mainly concerned in finding the appropriate network parameters with respect to the similarities in the data. After optimizing the parameters according a specific task-independent measure the network will be able to create new classes based on an internal representation of the input data.

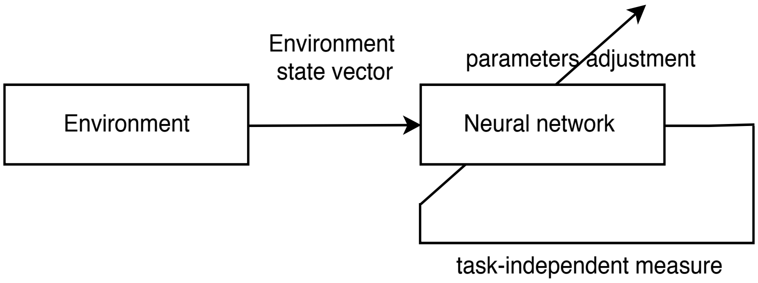


Figure 3.5: Illustration of the unsupervised learning paradigm.

3.2.3 Back-propagation learning

Illustrating the first learning process for feed forward network in its simplest cases Figure xx, a neuron k is the only computational unit in the output layer, representing the only neural network output signal, y_k is compared with the desired response or the target output $d_k(n)$ to compute the error $e_k(n)$ at the n^{th} time step according the following definition

$$e_k(n) = d_k(n) - y_k(n) \tag{3.5}$$

In an iterative process of adjusting the neuron k 's synaptic weights, the corrective adjustment is applied to minimize the gap between the network output and the desired response. This process occurred by adopting a cost function also called an index of performance $E(n)$ instead of using the standard distance $e_k(n)$ defined as:

$$E(n) = \frac{1}{2}e_k^2(n) \tag{3.6}$$

Most obviously, it is necessary to decide how to follow the error gradient. The simplest method, known as steepest descent or gradient descent which originally considered by (Rumelhart et al. 1986), is to find derivative of the loss function with respect to the network weights, then update the weights by taking a small fixed-size step in the direction of the negative error gradient of the loss function according the formula:

$$w_{ji}(t+1) = w_{ji}(t) + \Delta w_{ji}(t) \tag{3.7}$$

and

$$\Delta w_{ji}(t) = -\mu \frac{\partial E}{\partial w_{ji}} \tag{3.8}$$

where $\mu \in [0,1]$ is the learning rate. This computation is repeated to reach some stopping criteria such as falling into a local minima where it fails to reduce the loss or accomplishing a given number of steps.

Local minima phenomena is considered as one of most crucial problems that faces the gradient descent method while updating weights. To avoid getting trapped in this problem, (Plaut et al. 1986) had introduced a momentum factor m that effectively allows the network to ignore the insignificant features in the error, and therefore, improve the selected final state of the network.

$$\Delta w_{ji}(t+1) = -\mu \frac{\partial E}{\partial w_{ji}} + \beta \Delta w_{ji}(t) \tag{3.9}$$

$\beta \in [0,1]$ also chosen in the range of [0,1].

The gradients could be calculated and defied over the entire training dataset (batch learning), over small subsets (sequential learning) or over every individual training example (online learning) where this latest tends to be more efficient then both sequential and batch learning in the case of dealing with big amount of data with significant redundancy (Lecun et al. 1998). In the same aspect, stochasticity in online learning according to (Lecun et al. 1998) also can help preventing the network from falling into the local minima due to the diversity of the loss function for each training sample by randomly organizing these samples before using it in the training process. Another stochastic online learning was proposed by (Schraudolph 2002) is stochastic meta-descent that shows a fast convergence and a considerable improved results for different tasks.

Since the goal of this section is to find an efficient calculation procedure for evaluating the gradient of the network error function E which known as the error back-propagation technique, it should be mentioned that the term back-propagation, in addition to describing the MLP training process using gradient descent applied to sum of squares error function, it could also refer to the MLP networks architecture. However, it will be more helpful to recognize two distinct stages of the backward pass. In the first stage represents the backward propagation of the errors through the network, where the second stage consists of adjusting the weights using the calculated derivatives which can be achieved using a variety of optimization methods.

3.2.4 Error function derivatives

Now considering the evaluation of the $E(n)$ regarding the weight w_{ji} related to unit j, where the output of each neuron depends the particular input pattern. In order to find the chain rule for partial derivative, it is essential to consider z_j due to the inclusion of w_{ji} as formulated in equation. 3.2. Hence, we note

$$\frac{\partial E}{\partial w_{ji}} = \frac{\partial E}{\partial z_j}\frac{\partial z_j}{\partial w_{ji}} \tag{3.10}$$

Now we introduce

$$\delta_j = \frac{\partial E}{\partial z_j} \tag{3.11}$$

using the activations received from the neurons in the previous layer i, we can write

$$\frac{\partial z_j}{\partial w_{ji}} = y_i \tag{3.12}$$

Substituting equations (3.11) and (3.12) into (3.10), we then obtain

$$\frac{\partial E}{\partial z_j} = \delta_j y_i \tag{3.13}$$

3.2.5 Update process

Much obviously now from the equation. 3.13 shows that the derivative is obtained simply by multiplying values of output end δ by the input end activation. Therefore, in order to evaluate the derivatives, the value of δ differs with respect to the which layer and weights the derivative is related to. Thus, for the output neurons:

$$\delta_k = d_k - y_k \tag{3.14}$$

Concerning the units in the hidden layer, to evaluate δ with the use of the chain rule for partial derivative

$$\delta_j = \frac{\partial E}{\partial z_j} = \sum \frac{\partial E}{\partial z_k}\frac{\partial z_k}{\partial z_j} \tag{3.15}$$

using the fact that loss function depends only on each hidden unit k through its influence on the output units. Differentiating 3.1 and 3.2 and substituting into 3.15 gives

$$\text{for weights in the output layer:} \quad \delta_k = 2f_k'(z_k)e_k \tag{3.16}$$

$$\text{for weights in the hidden layer:} \quad \delta_j = 2f_j'(z_j)\sum_{k=1}^{K}\delta_k w_{kj} \tag{3.17}$$

3.3 Recurrent neural networks

The conceptual dissimilarity that differs the recurrent neural networks from the standard feedforward networks relies on the flow of the activation signals through numerous units within each layer, where the recurrent topologies allow a cyclical connections

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.

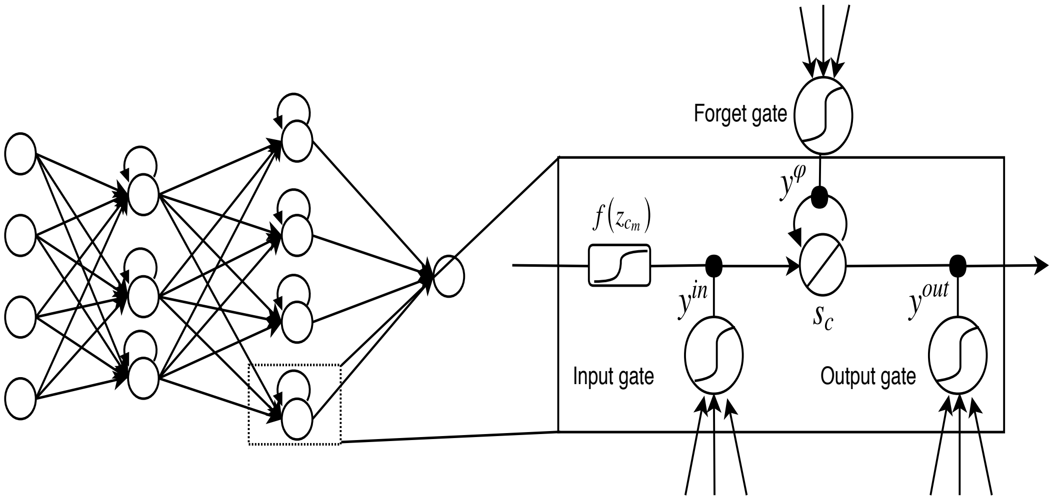


Figure 3.6: LSTM memory block with one cell.

3.3.1 Forward pass

As stated before in section. 2 , LSTMs have been found to outperform other traditional RNNs on tasks involving long time lags, its architecture permits to 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 (Hochreiter & Schmidhuber 1997). Typically, LSTM differs from the standard ANNs with its hidden layer units where the summation units are replaced instead, by memory blocks. Each block contains one or more self-connected memory cells and three sources of inputs: the input z^{in}_m, output z^{out}_m and the cell itself z^{φ}_m, 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. Illustration of an LSTM memory block with a single cell is provided in Figure. 3.6. More precisely, this means that y_m in equation (3.2) is replaced with the following sequence of equations

$$z_{c_m}(t) = \sum_{m=0}^{k} w_{c_m j} y_j(t-1) \tag{3.18}$$

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

$$z^{in}_m(t) = \sum w^{in}_{mj} y_j(t-1) \quad ; \quad y^{in}_m(t) = f^{in}_m(z^{in}_m(t)) \tag{3.19}$$

$$z_m^{\varphi}(t) = \sum w_{mj}^{\varphi} y_j(t-1) \quad ; \quad y_m^{\varphi}(t) = f_m^{\varphi}(z_m^{\varphi}(t)) \tag{3.20}$$

$$z_m^{out}(t) = \sum w_{mj}^{out} y_j(t-1) \quad ; \quad y_m^{out}(t) = f_m^{out}(z_m^{out}(t)) \tag{3.21}$$

through the current study m refers to memory block with only one cell c_m, see (Gers et al. 2002) for details. Moreover, s_{c_m} indexes the cell state of the m^{th} memory block which updated according to equation. 3.22

$$s_{c_m}(t) = y_m^{\varphi}(t) s_{c_m}(t-1) + y_m^{in}(t) f(z_{c_m}(t)) \tag{3.22}$$

with

$$s_{c_m}(0) = 0 \tag{3.23}$$

3.3.2 Backward pass

LSTM network training is a fusion of BP for output unites 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 (Hochreiter et al. 2001). This eases the implementation of the algorithm, which is an important property for tasks such as time series prediction (Graves et al. 2005). For a full derivative of the algorithm see (Gers et al. 2000).

For the output units, weight changes via gradient descent given by equations (3.7), (3.8) and (3.16), output gate weights are also obtained by standard back propagation

$$\Delta w_{mj}^{out}(t) = \mu \delta_m^{out}(t) y_j(t) \tag{3.24}$$

$$\delta_m^{out}(t) \overset{tr}{=} f'^{out}_m(z_m^{out}(t)) \Big(\sum s_{c_m}(t) \sum w_{kc_m} \delta_k(t) \Big) \tag{3.25}$$

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

$$\frac{\partial s_{c_m}(t)}{\partial w_{mj}^{in}} \overset{tr}{=} \frac{\partial s_{c_m}(t-1)}{\partial w_{mj}^{in}} y_m^{\varphi}(t) + f(z_{c_m}(t)) f'^{in}_m(z_m^{in}(t)) y_j(t-1) \tag{3.26}$$

$$\frac{\partial s_{c_m}(t)}{\partial w_{c_m j}} \overset{tr}{=} \frac{\partial s_{c_m}(t-1)}{\partial w_{c_m j}} y_m^{\varphi}(t) + f'(z_{c_m}(t)) y_m^{in}(t) y_j(t-1) \tag{3.27}$$

$$\frac{\partial s_{c_m}(t)}{\partial w^{\varphi}_{mj}} \overset{tr}{=} \frac{\partial s_{c_m}(t-1)}{\partial w^{\varphi}_{mj}} y^{\varphi}_m(t) + s_{c_m}(t-1) f'^{\varphi}_m(z^{\varphi}_m(t)) y_j(t-1) \tag{3.28}$$

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

$$e_{s_{c_m}}(t) \overset{tr}{=} y^{out}_m(t) \Big(\sum w_{kc_m} \delta_k(t) \Big) \tag{3.29}$$

3.3.3 Update process

Weights corresponds to connections to the input, the cell and the forget gates are updated using the partials from Equations 3.26, 3.27 and 3.28:

$$\Delta w_{c_m}(t) = \mu e_{s_{c_m}}(t) \frac{\partial s_{c_m}(t)}{\partial w_{c_{mj}}} \tag{3.30}$$

$$\Delta w^{in}_{mj}(t) = \mu e_{s_{c_m}}(t) \frac{\partial s_{c_m}(t)}{\partial w^{in}_{mj}} \tag{3.31}$$

$$\Delta w^{\varphi}_{mj}(t) = \mu e_{s_{c_m}}(t) \frac{\partial s_{c_m}(t)}{\partial w^{\varphi}_{mj}} \tag{3.32}$$

3.4 Conclusion

Neural networks modeling on this type of modeling has been increasingly attracted by non-specialist researchers, especially from the world of energy. Looking for some powerful machine learning tools to replace the expert systems and mathematical models which were not particularly successful for the energy markets, and therefore it could represent a promising alternative.

In conclusion, the predictive abilities of this application were found to be better than those of statistical systems. This application has not got beyond the prototype stage because of the problems of confidence and trust. Perhaps a natural development would be to couple this system to an expert system in order to automate the processing of market information.

Chapter 4

The two stage forecasting approach review

Contents

4.1 Introduction

The variety of customer profiles, the high dependence on seasonal and climate aspects, together with the actual gas consumption limit the maximum accuracy that classical single model prediction approaches (Potočnik et al. 2008) can provide. To overcome this limitation there are techniques that rely on multiple models. Multiple models are often combined with the divide-and-conquer approaches for solving such complex problems (Potočnik & Govekar 2010).

This chapter investigates several divided-and-conquer approaches for the purposing of forecasting the natural gas consumption. Inspired from (Farfar

& Khadir 2018), This approach is based on splitting the Algerian natural gas hourly consumption of 2014 into multiple subsets using different kinds of clustering methods. The dataset division is made by regrouping the daily pattern sequence of the 24 hours load using three powerful methods. After the division process, multiple local auto-regressive Gaussian Process (AR-GP) models are developed for a specific variation of similar daily curves according to each clustering method. Finally, the results obtained by several non-supervised classifiers are compared over the problem for load consumption prediction.

Fig. 4.1 presents the computation procedure for two-stage approach review on gas consumption prediction.

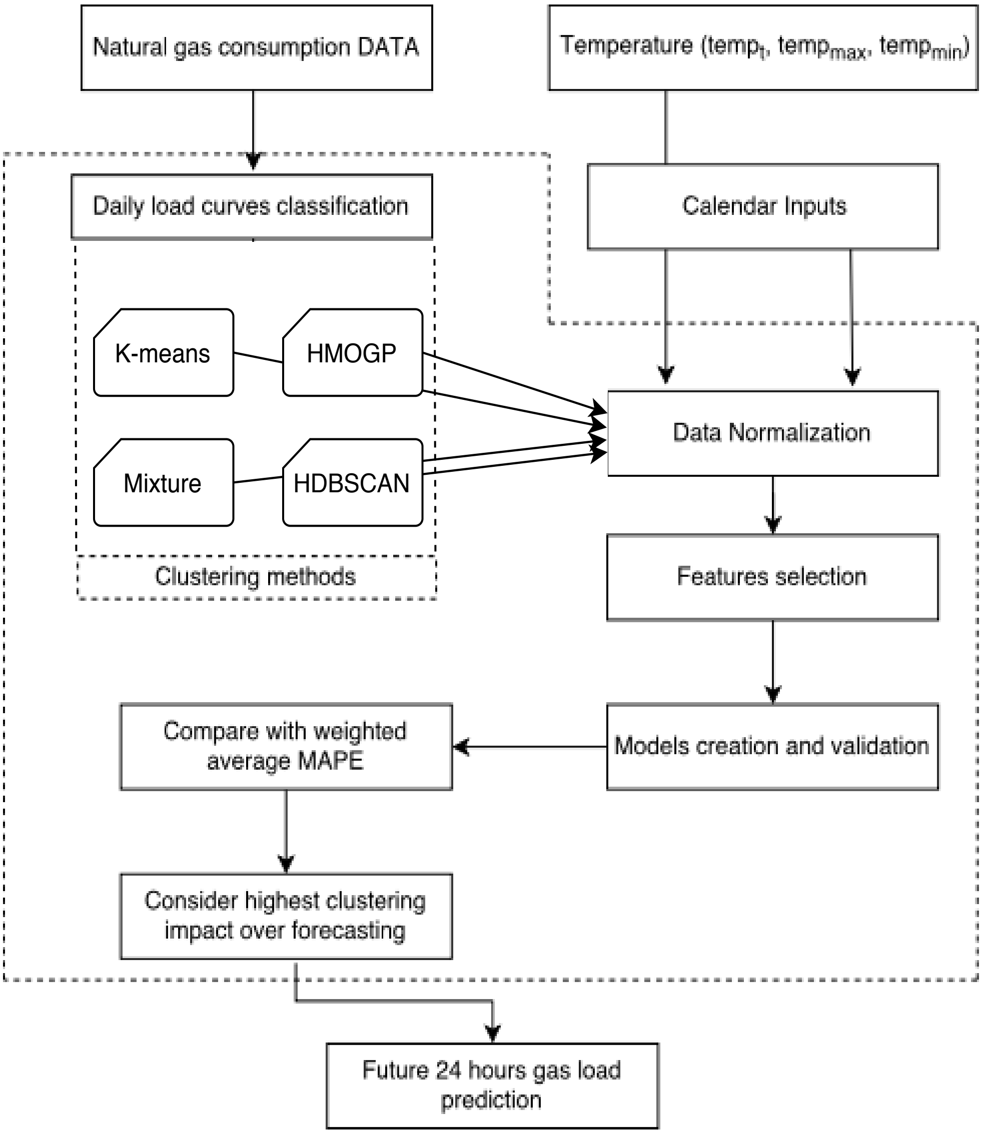
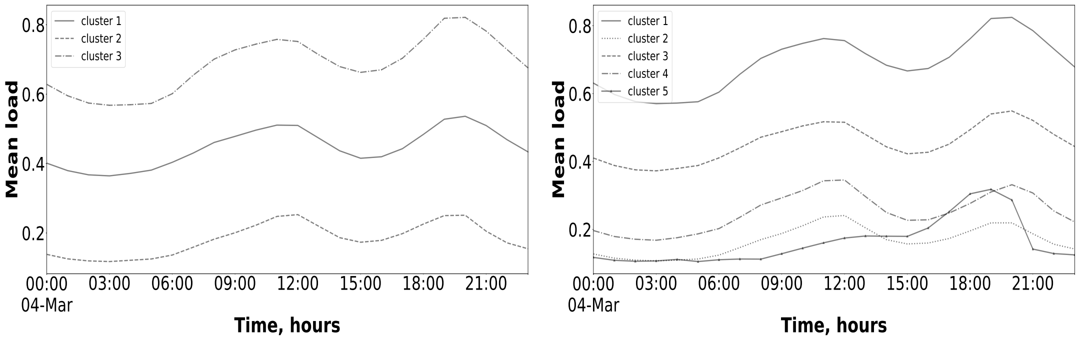


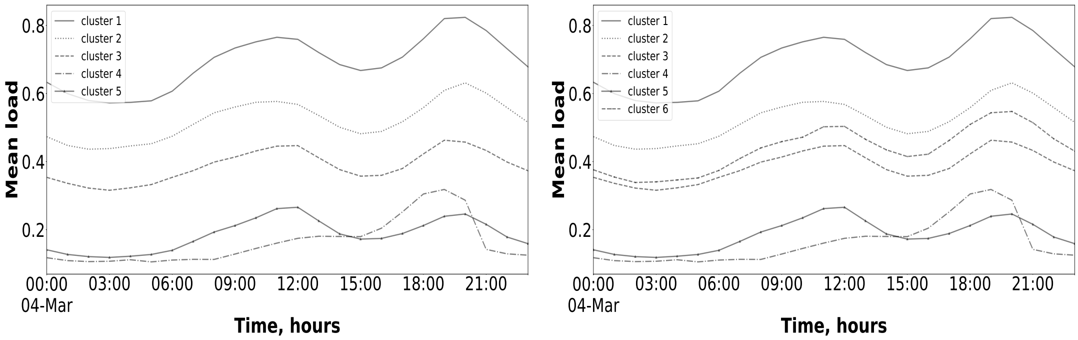
Figure 4.1: Main steps of the proposed framework.

4.2 Daily load curves classification



(a) KMeans result

(b) HMOGP result



(c) Mixed of KMeans and HDBCAN result

(d) HDBSCAN result

Figure 4.2: Average load for clusters obtained by each method.

The first step of the proposed approach is to classify samples of historical segments $H_{2014}\big[D_0 \dots D_{364}\big]$ into K clusters containing identical daily load curves, where each historical segment (day x) is represented in 24 hours consumption vector $D_x = \{C_0, C_1 ... C_{23}\}$ with C_h is the consumption at a specific hour h. As the number K is unknown in this case, from a statistical point of view this issue is considered as an unsupervised curves classification problem. To group the daily consumption curves, different techniques are applied.

4.2.1 K-Means

First, a centroid based KMeans is used, with the need of specifying the number of cluster K (Fixed in this case to 3: Winter, Summer and autumn-spring). It is supposed that there is a daily consumption and meteorological variables correspondence.

4.2.2 HDBSCAN

Secondly, a hierarchical density-based clustering method is used. The current method was firstly introduced by Campello et. al in (Campello et al. 2013) and (Hensman et al. 2013), where it improves the DBSCAN method by transforming it into a hierarchical clustering algorithm. Thus, it generates a complete density-based clustering hierarchy from which a simplified hierarchy composed only of the most significant clusters can be easily extracted. This method requires only one parameter which represents the minimum size of the cluster.

4.2.3 MOHGP

Another non-parametric clustering method is used, but unlike the HDBSCAN the mixture of hierarchical Gaussian Process is specialized on structural time series. MOHGP is proposed by James Hensman in (Campello et al. 2015) which is a combination of two Bayesian non-parametric algorithms, where it combines Gaussian processes (GPs) approach to model time-series and Dirichlet processes (DPs) to perform clustering.

4.2.4 Mixture of K-means and HDBSCAN

The last used clustering technique is a result of two combined methods (HDBSCAN and KMeans): where it keeps the 3 clusters obtained by the KMeans and add another two clusters which get recognized by the HDBSCAN. There is a significant difference between the obtained clustering results. The reason is due to the fact that these clusters represents two different time periods of the year: the first is the period of the Ramadan and the second one is the period of national and religious holidays. During these holiday periods the consumption patterns are unique.

4.3 Gaussian Process Regression method for time-series modeling

Once daily curves are regrouped, an AR-GP model is trained to learn the data for each cluster. Hence, every model handles the forecasting task for all hourly load in the corresponding cluster. By each GP model a single value C_t is predicted depending essentially on the following inputs: First, the previous lagged observations $C_{t-1}, C_{t-24}, C_{t-168}$ which represent the consumption of the previous hour, the same hour of the previous day and the same hour of the previous week. Then, meteorological factors corresponding to the temperature T_t, the maximum T_{max} and the minimum T_{min} temperature of the day are considered.

Because of the dependence of the performance of the GPs on the chosen kernel, a radial basis function (RBF) kernel is adopted, also known as the squared exponential provides an expressive kernel to model smooth functions. The RBF kernel is given by:

$$k(x, x') = \ exp\Bigg(\ -\frac{1}{2} d\Big(\frac{x}{l}, \frac{x'}{l}\Big)^2 \Bigg) \tag{4.1}$$

The dataset is separated then into two partitions, the first partition (70 %) is for the fitting and optimizing of the GP's hyper-parameters, the rest (30 %) is for the test to evaluate the model quality.

In order to achieve a better estimation of hyper-parameters of covariance functions, using an appropriate approach to normalize the time series data is critical before feeding it to the GP model. The outputs and inputs data are normalized to an interval between [0, 1]. Hence, a value of X is normalized to X' by computing:

$$X' = \frac{X}{X_{max}} \tag{4.2}$$

where X' is the new value, X is the old value and X_{max} is the largest consumption value in the year.

4.4 Experimental results

4.4.1 Clustering

In order to make a clear visualization of the classification process, the next Figures (4.2a , 4.2b, 4.2c, 4.2d) show the mean daily load curve in each cluster obtained by the four clustering methods. At the end of the clustering, 365 daily curves of in the dataset H_{2014} are all labeled with its correspondent cluster K_x.

$$H_{2014} \begin{bmatrix} D_0[C_0 ... C_{23}] & K_x \\ \vdots & \vdots \\ D_{364}[C_0 ... C_{23}] & K_x \end{bmatrix}$$

Table 4.1: Season's day count per cluster

Clustering Method	**Cluster ID**	**Seasons**			
		Winter	***Spring***	***Summer***	***Autumn***
KMeans	C 1	85	0	0	18
	C2	3	19	0	30
	C 3	0	74	94	42
HMOGP	C 1	82	3	0	18
	C 2	0	31	65	40
	C 3	6	15	0	25
	C 4	0	44	0	7
	C 5	0	0	29	0
HDBSCAN	C 1	79	3	0	18
	C 2	5	8	0	4
	C 3	4	0	0	23
	C 4	0	0	29	0
	C 5	0	74	64	40
	C 6	4	4	1	5
Mixture	C 1	82	0	0	18
	C 2	2	15	0	27
	C 3	0	74	64	40
	C 4	0	0	29	0
	C 5	4	4	1	15

These numbers represent days included in each cluster.

In order to be able to compare results obtained in each clustering method in Table. 4.1, the clusters are related to a season which represented the most number of days, this is determined by the seasons of the year. The comparison of results is therefore based on the labeled seasons.

4.4.2 Features selection

The most appropriate features must be determined in order to enhance the prediction accuracy. The investigation started using only 3-dimensional input vector containing the historical load $(C_{t-1}, C_{t-24}, C_{t-168})$. Furthermore, other features were added sequentially as indicated in Table. 4.2 to observe its impact on estimating one hour ahead.

To measure the error in estimated loads, the models are evaluated based on the Mean Absolute Percentage Error calculated according to the following formula:

$$MAPE = \frac{100\%}{N} \sum_{i=1}^{N} \left| \frac{A_i - P_i}{A_i} \right| \tag{4.3}$$

where N denotes the number of predicted samples, A_i and P_i are the actual and predicted values respectively

Table 4.2: Mean Absolute Percentage Error according to features combination

Experiments	**Exp 1**	***Exp 2***	***Exp 3***	***Exp 4***	***Exp 5***	***Exp 6***
Historical Inputs	✓	✓	✓	✓	✓	✓
Temp		✓			✓	✓
Day Ind			✓			✓
Hour Ind				✓	✓	✓
Kmeans method						
C1 MAPE	3.48	3.89	3.55	2.41	2.25	**2.05**
C2 MAPE	4.54	4.94	4.39	3.16	**2.90**	2.68
C3 MAPE	7.64	6.18	7.31	5.79	**5.88**	4.84
HDBSAN method						
C1 MAPE	5.16	4.93	4.81	4.27	**3.01**	3.10
C2 MAPE	6.24	4.68	4.54	2.94	2.66	**3.28**
C3 MAPE	8.29	6.00	6.08	5.49	4.06	**4.00**
C4 MAPE	7.91	6.60	7.43	6.94	6.24	**5.78**
C5 MAPE	8.25	5.68	6.09	6.42	3.57	**4.23**
C6 MAPE	10.12	7.83	18.34	10.02	**3.48**	1.59
HMOGP method						
C1 MAPE	5.26	5.13	4.87	4.37	3.41	**3.21**
C2 MAPE	8.35	5.89	7.63	5.90	3.96	**3.58**
C3 MAPE	6.45	4.38	5.40	4.72	**4.07**	4.20
C4 MAPE	8.68	7.70	8.19	7.44	**6.84**	6.23
C5 MAPE	8.25	5.68	6.09	6.42	3.57	**4.23**
Mixed method						
C1 MAPE	5.16	4.93	4.81	4.27	3.37	**3.07**
C2 MAPE	8.76	5.37	5.66	6.15	**3.92**	3.51
C3 MAPE	7.99	6.68	7.50	7.03	6.32	**5.70**
C4 MAPE	8.25	5.68	6.09	6.42	3.57	**4.23**
C5 MAPE	10.12	7.83	18.34	10.02	**3.48**	1.59

Because of the complexity of the load time series, the experiments show that every time the AR-GP injected with exogenous variables, it provides a significantly better forecasting results. The correlation of natural gas consumption with temperature is -0.70 which means that there is significant relevance between the two variations. Moreover, adding the temperature variables reduced the MAPE in all clusters especially in the winter period. The unexpected improvement is in the summer period which leads to the fact of the AR-GP is not influenced by the temperature as an indicating value for hotness or coldness but influenced by value that indicates the period in the day which corresponds the load.

Apart from using historical and exogenous attributes, the experiments also involved two different kinds of calendar variations. The first is a daily indicator, to identify the day of the week which related to the predicted load. Identifying

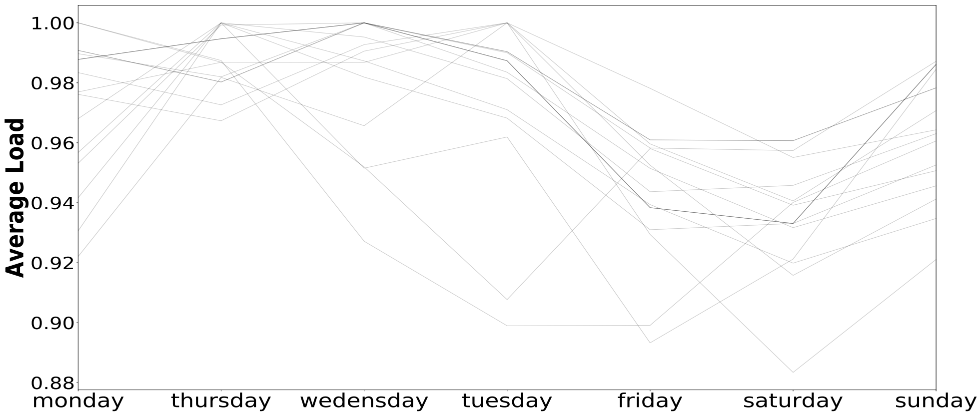


Figure 4.3: The daily average natural gas consumption per week.

the day for which forecast is performed can help the model to distinguish between working days and holidays and also to recognize the first day of the week from the last ones. Fig. 4.3 shows the variety of daily average load per week on the generated clusters. The second calendar inputs is an hourly indicator which is considered as a very strong intraday periodic pattern.

Despite the effectiveness of the models on the training process, generalization and good performances on test and validation data is not straightforward. There are two main causes that occur the inconsistency of the model's performance through training and testing sets: the first cause is when the influence of an exogenous factor doesn't cover the entire period of the correspondent cluster, like in the case of KMeans clusters: C2 and C3, the best input combination is (Exp 5) because the error during test is lower than in (Exp 6) 4.5%, 4.2% respectively. The hyper-parameters of the kernel are optimized during the fitting of the AR-GP by maximizing the LML. As the LML have a multiple local optima, this means that the model may fall in the over-fitting phenomena, which is the second cause that makes the covariance function will tend to have a poor predictive performance on the test unlike on training where is it the case in (Exp 6) with HDBSCAN C5, the MAP-Errors in training and test are (1.59% and 395.72% respectively) thus, the feature combination in (Exp 6) will be ignored.

4.4.3 Load forecasting results

After selecting the most convenient features, Table. 4.3 reports the mean absolute percentage error for AR-GP based on the proposed clustering methods.

Injecting predicted load in every iteration is the main concept for the stepwise forecasting. Furthermore, AR-GP models are extremely sensitive to the historical consumption, thus, a non-accurate estimation will definitely lead to a very bad performance along the rest of the 24 hours ahead, and this is the case when the AR-GP model is constructed basing on a load prior to the desired one (C_{t-1}). Meanwhile, we should note that under some circumstances an AR-GP deteriorates with respect to the prior load that could alter its performance. Therefore, the previous load will not be used in the forecasting process. Table. 4.3 obviously expresses seven cases where the prior load is ignored (C3), (C3, C4, C5), (C2, C4) and (C4) in KMeans, HMOGP, HDBSCAN and Mixture clusters respectively. Consequently, the 1 step forecasting error will be exactly the same as the 24 steps forecasting error.

Table 4.3: Forecasting MAPE based on the clustering approaches

Method	**Cluster (season)**	**1 step training MAPE**	**1 step test MAPE**	**24 steps training MAPE**	**24 steps test MAPE**
KMeans	C1 winter	1.33	1.34	2.05	3.33
	C2 sp & au	1.88	1.99	2.90	4.25
	C3 summer	5.88	7.10	5.88	7.10
HMOGP	C1 winter	1.33	1.50	3.21	5.42
	C2 spring	1.73	2.24	3.96	5.00
	C3 autumn	4.20	7.25	4.20	7.25
	C4 summer	6.84	7.32	6.84	7.32
	C5 ramadan	4.23	6.72	4.23	6.72
HDBSCAN	C1 winter	1.26	1.48	3.10	4.31
	C2 spring	3.28	4.53	3.28	4.53
	C3 autumn	2.09	1.97	4.00	4.11
	C4 summer	3.82	3.42	5.78	7.35
	C5 ramadan	4.23	6.27	4.23	6.27
	C6 sp-days	2.31	2.72	3.48	5.56
Mixture	C1 winter	1.26	1.48	3.07	4.31
	C2 sp & au	1.70	2.15	3.92	4.81
	C3 summer	3.24	3.31	5.70	4.72
	C4 ramadan	4.23	6.27	4.23	6.27
	C5 sp-days	2.31	2.72	3.48	5.56

sp & au: spring and autumn
sp-days: special days

Each cluster deserved a particular consideration, and Because of the high distinction in the generated daily load curves by each clustering approach, there is no general indication of how effective a daily load curves classification procedure can be given without applying a relevant comparative evaluation.

The weighted arithmetic mean is applied instead of the ordinary mean to calculate an average MAPE that represents the error for a given clustering method.

$$\overline{MAPE} = \frac{1}{N} \sum_{i=1}^{N} \left(\frac{\sum_{j=1}^{4} w_j MAPE_i}{\sum_{j=1}^{4} w_j} \right) \tag{4.4}$$

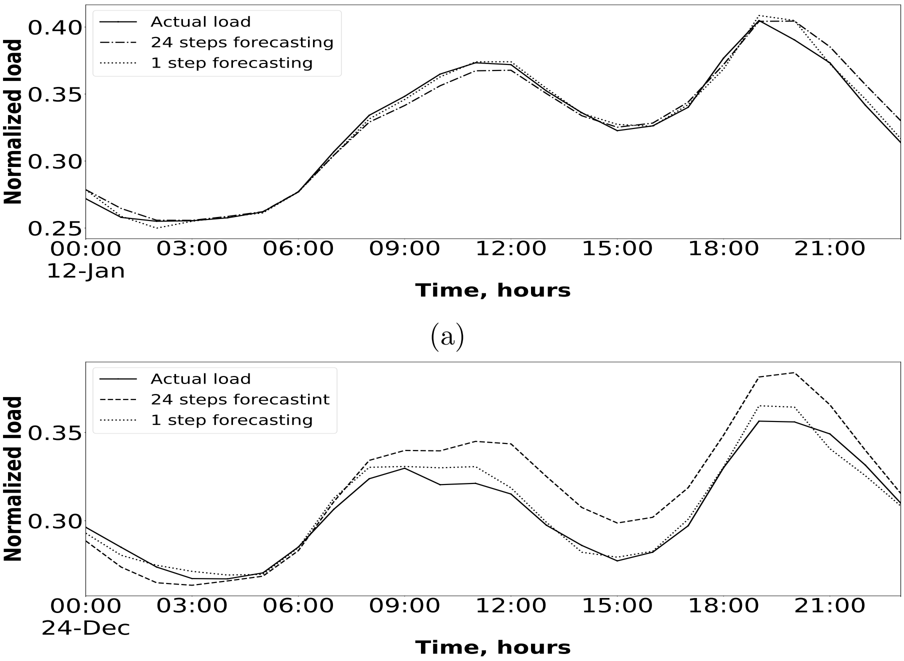
Equation (4.4) expresses the weighted average of the mean absolute error ($\overline{MAPE}$), where w_j is the number of days per season j (winter, spring, summer and autumn) counted in each cluster i.

A summarized comparison between actual and forecast gas consumption in terms of mean absolute percentage error shown in Table. 4.4. The results indicate that the AR-GP performs much better on the Mixture method amongst all other clustering methods on the test period. The reason of choosing the Mixture method is because of the performance stability of AR-GP through training and testing sets compared to its performance on KMeans, HMOPG and HDBSCAN clusters. Additionally, the $\overline{MAPE}$ obtained from the Mixture method clusters, evaluated on test set is 4.77%, which is an improvement over the KMeans, HMOPG and HDBSCAN by 16.53%,19.83% and 25.91% respectively.

Table 4.4: $\overline{MAPE}$ based on the clustering approaches

Approach	**Train** $\overline{MAPE}$	**Test** $\overline{MAPE}$
KMeans	4.37%	5.63%
HMOGP	4.20%	5.82%
HDBSCAN	4.58%	6.19%
Mixture	4.56%	4.77%

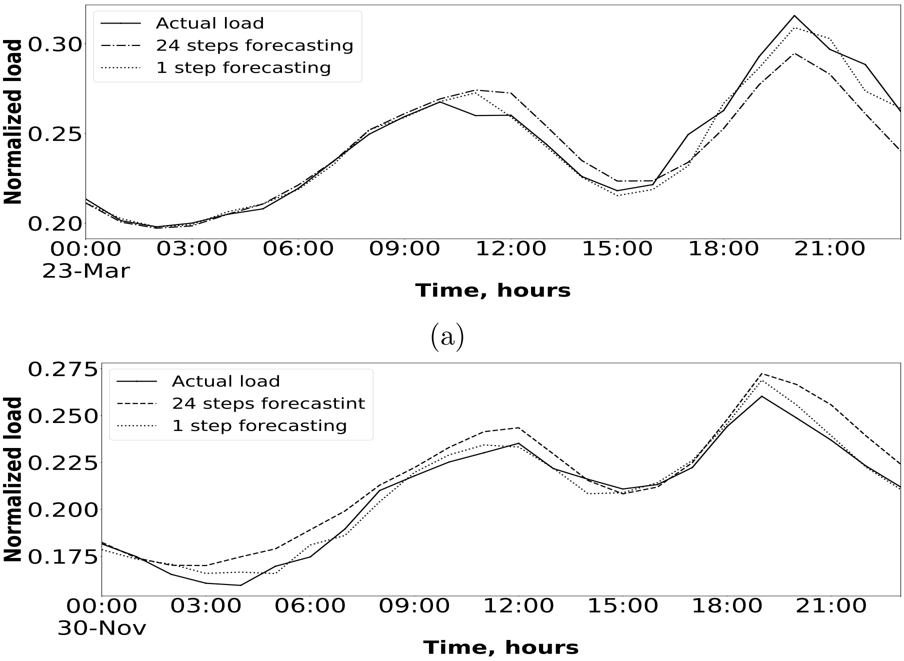
The results of the forecasts for gas demand for the Algerian market reported with the use of Mixture clustering method clusters are illustrated in Figures. 4.4, 4.5, 4.6, 4.7 and 4.8. The labeled figures (a) and (b) presented the results through learning and test period respectively.



(a)

(b)

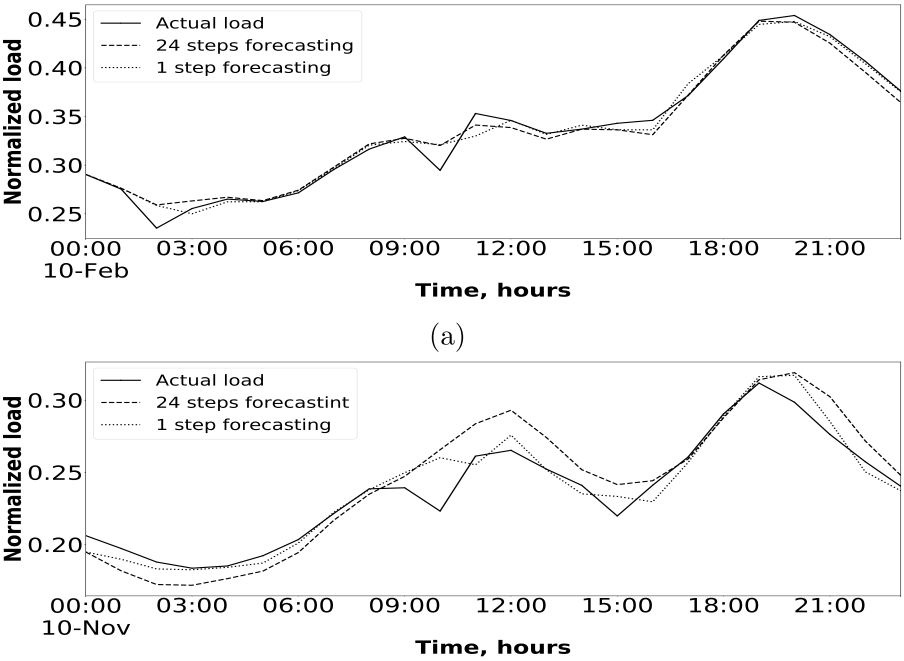
Figure 4.4: 24 hour forecast through training and testing winter period



(a)

(b)

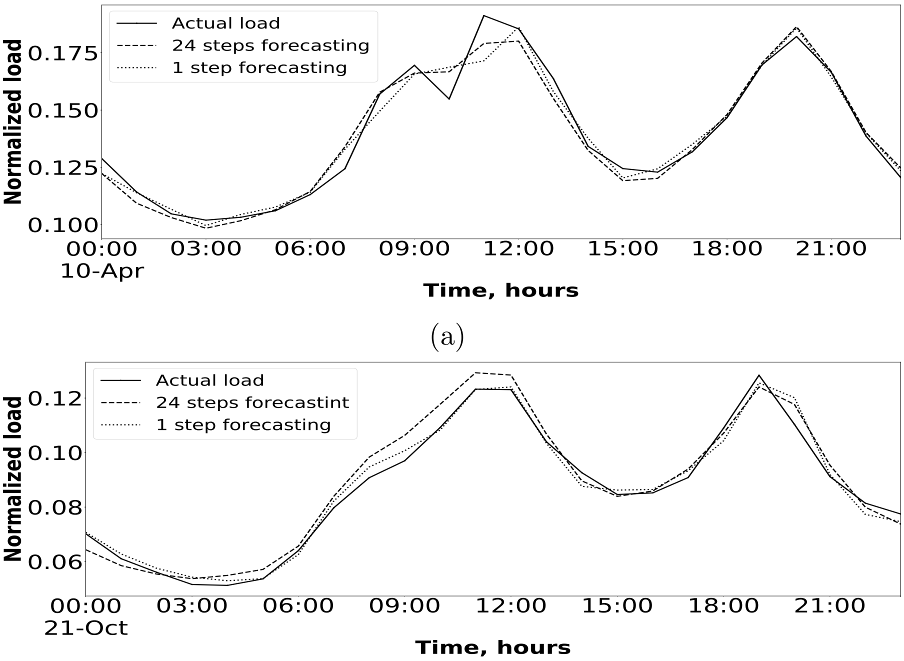
Figure 4.5: 24 hour forecast through training and testing Spring and Autumn period



(a)

(b)

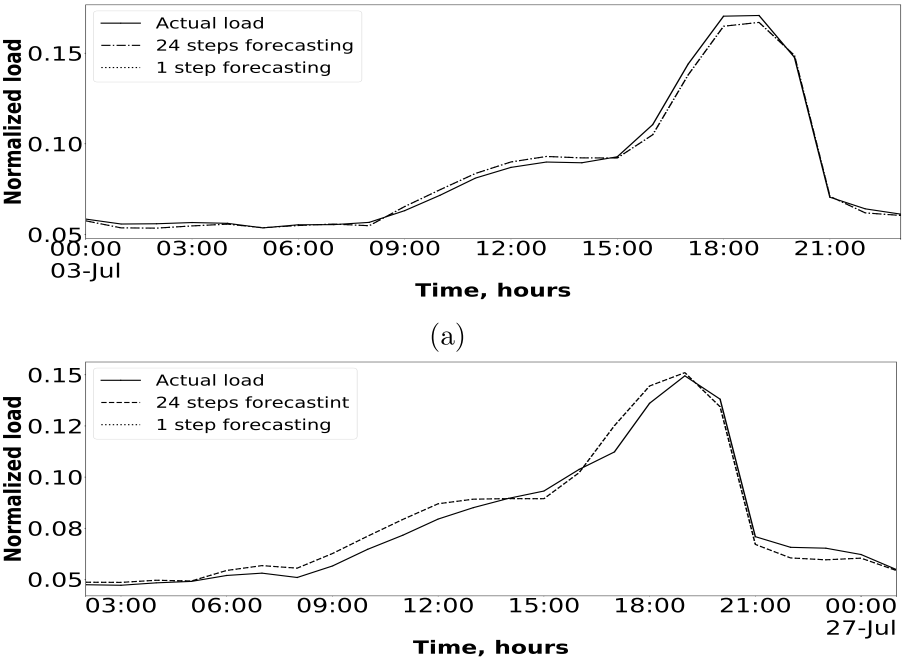
Figure 4.6: 24 hour forecast through training and testing special days period



(a)

(b)

Figure 4.7: 24 hour forecast through training and testing summer period



(a)

(b)

Figure 4.8: 24 hour forecast through training and testing ramadan period

As a test-bench experiment, a noise can be added to the inputs to cover different kinds of uncertainties in the measurement. Three levels of simulated random noise were added to inputs vectors: 1%, 3% and 5%. Because of the prior data normalization from [0,1], the inputs noise values was randomly generated from -0.01, -0.03 and -0.05 to 0.01, 0.03 and 0.05 respectively according the noise percentage. Occurring prediction with this uncertainties should definitely lead to an increase in the $\overline{MAPE}$. Unexpectedly, AR-GPs models show a very powerful ability of handling the noisy inputs, where even 5% added noise did not inadequately effect the prediction accuracy and results error increasing by 8% only and barely increases after 1% of noise is added. Fig. 4.9 illustrates the $\overline{MAPE}$ increase with regard to the noise level.

To evaluate the results obtained by the proposed approach, a comparison with another two divide-and-conquer approaches are conducted. The first method was developed in (Franco & Fantozzi 2015) is considered in the preliminary analysis for energy forecasting, where the dataset is split according to the holiday (i.e., Friday, Saturday and other holidays), to working days, to pre-holidays and to special days.

A second approach proposed by Mustafa Akpinar in (Akpinar & Yumusak 2013), splits the data into six monthly subsets shown in Table 4.5.

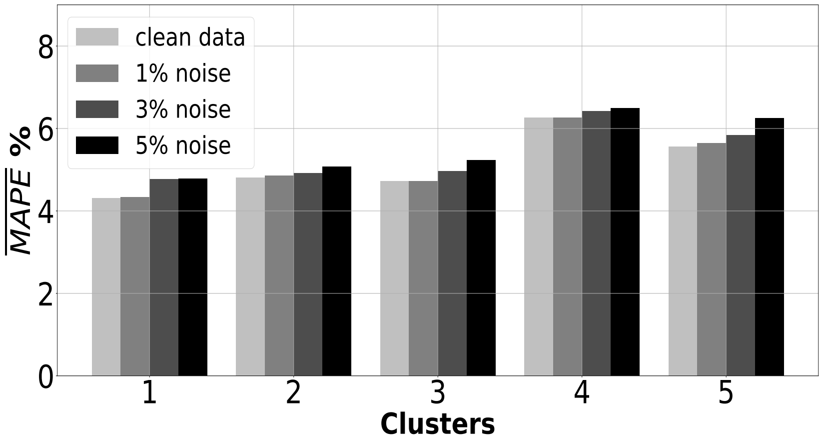


Figure 4.9: GP models performance on test dataset with respect to noise level.

Table 4.5: Models and Months

Model Name	**Data Months**
Model 1	January, February
Model 2	March
Model 3	April
Model 4	June, July
Model 5	August, September, October
Model 6	November, December

From comparison results shown in Table 4.6, it can be noticed that the proposed method gives a very accurate forecast for practical needs even when compared with the two benchmark methods.

Table 4.6: Average prediction $\overline{MAPE}$ for every cluster according to each approach

Clusters	**Training $\overline{MAPE}$ %**			**Test $\overline{MAPE}$ %**		
	MIX	APP1	APP2	MIX	APP1	APP2
Cluster 1	3.92	5.08	4.76	4.31	7.82	6.18
Cluster 2	3.92	4.83	8.63	4.81	6.16	10.68
Cluster 3	5.70	2.72	4.03	4.72	5.56	9.87
Cluster 4	4.23	5.65	10.09	6.27	4.90	11.48
Cluster 5	3.48	/	6.91	5.56	/	6.50
Cluster 6	/	/	5.72	/	/	5.60
Average	**4.56**	5.25	6.95	**4.77**	5.96	8.13

MIX: prediction based on Mixture clustering method.
APP1, APP2: error based on first and second benchmark approaches.

4.5 conclusion

This chapter investigates the practical aspects of development of forecasting the Algerian natural gas consumption. There is a strong correlation of the natural gas load with meteorological elements which is mainly represented by temperature for the residential sector and physical and statistical factors like season of the year, day of the week for the industrial sector. Load data of 2014 is analyzed and clustered using different clustering methods in order to classify daily load profiles according to the similarity measures of each clustering method.

Based on the grouped daily load curves and using temperature with calendar inputs, multiple models construction are conducted by several experiments to determine the most influential factor. Forecasting results of 2014 are summarized and expressed in mean absolute percentage error. The average calculated $\overline{MAPE}$ on training and test datasets is 4.56% and 4.77%, which was achieved using mixture of KMeans and HDBSCAN method.

Classifying load curves into a huge amount of groups or adopting many different models does not necessarily improve the forecasting results, even in the case of using powerful clustering techniques. Contrarily, properly segmented and classified clusters can enhance the overall quality of the developed models considerably.

Chapter 5

Novel forecasting approach based on a monitoring stage

Contents

5.1 Introduction

In response to the challenges that energy systems face, in this chapter we propose a novel approach and evaluate its efficiency. This chapter advances the current state-of-the-art methods and has the following main contributions:

- An adaptive architecture is proposed for natural gas consumption forecasting in a large geographic area.

- LSTM models is proposed that can predict efficiently the natural gas consumption.

- 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.

5.2 Methodology

The developed approach consists of several stages, shown in Figure 5.1. The first stage considers and identifies the customers 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 (Alvarez et al. 2011). 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 chapter 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.

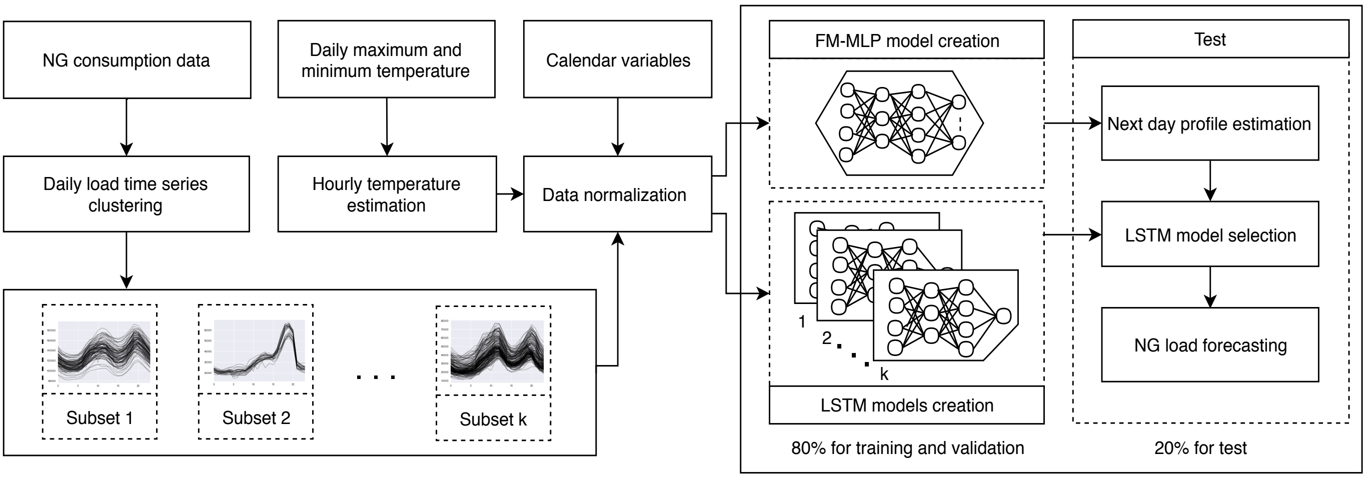


Figure 5.1: Main steps of the proposed approach.

5.2.1 Experiments

5.2.1.1 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 (Rousseeuw 1987), the Calinski-Harabasz criterion validates the clustering by calculating the average between- and within-cluster sum or squares (Calinski & Harabasz 1974), the Davies-Bouldin index determines the best partition by choosing the minimum ratio of within-cluster scatter to between-cluster separation (Davies & Bouldin 1979). 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 (Sharma 1995). 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 (Thorndike 1953), 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.

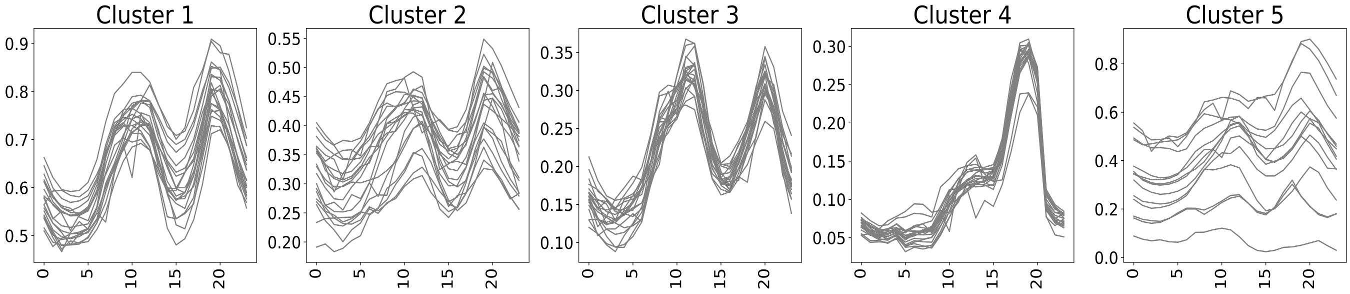


Figure 5.2: Average load for the obtained clusters.

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 5.2 shows 20 daily load samples from each cluster.

Table 5.1: 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 5.1 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.

5.2.1.2 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 5.3 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.

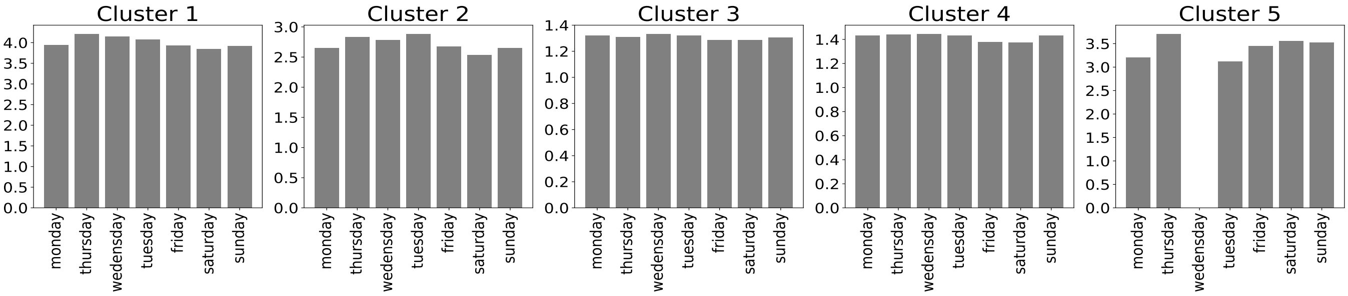


Figure 5.3: The daily average natural gas consumption per week.

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:

$$Accuracy = \frac{cp}{p}\, 100 \tag{5.1}$$

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:

$$log\ loss = \frac{1}{l}\sum_{i=1}^{l}[\ d_i log(o_i)\ \ + (1-o_i)\ log(1-o_i)\ ] \tag{5.2}$$

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

$$MAPE = \frac{100\%}{N}\sum_{i=1}^{N}\left|\frac{a_i - p_i}{a_i}\right| \tag{5.3}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|a_i - p_i| \tag{5.4}$$

$$RMSE = \sqrt{\sum_{i=1}^{N}\frac{(a_i - p_i)^2}{N}} \tag{5.5}$$

are applied as a benchmark calculated according to the following formulas (Shaikh & Ji 2016), 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:

$$x' = \frac{x}{x_{max}}, \tag{5.6}$$

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

5.2.1.3 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 (Tamura & Tateishi 1997), 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 5.2

Table 5.2: 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			

During the training stage of the proposed FM-MLP, multiple variables are taken into account: lagged consumption values $[l_1, l_2, \dots l_{24}]^{i-1}$ and $[l_1, l_2, \dots l_{24}]^{i-2}$ 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].

Table 5.3 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 one single miss-classed estimation case has been recorded during the test phase.

Table 5.3: 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

---

[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.

5.2.1.4 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 Section 5.2.1.3. 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 5.4. 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 5.4: 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	**<u>5.66</u>**
LSTM 2	**<u>10.61</u>**	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	**<u>9.41</u>**
LSTM 4	7.94	7.84	8.94	8.05	8.90	8.66	**<u>7.77</u>**
LSTM 5	15.53	17.77	16.66	19.41	18.91	**<u>13.86</u>**	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 5.5, other variables were added sequentially to observe their impact on the model's prediction accuracy.

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

Table 5.5: MAPE according to input combinations on training set

Experiments	**Input variables**					**Models**				
	Lagged load	**Temperature**	**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.

(Tetko et al. 1995). 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 (Fiesler 1994). 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 (Bergstra & Bengio 2012) 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 5.6 shows different combinations of number of hidden neurons and values of learning rate to find the most successful model configuration.

Table 5.6: 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)

5.3 Results and Discussion

5.3.1 Model's 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 5.7

Table 5.7: Parameters obtained by the random search optimization

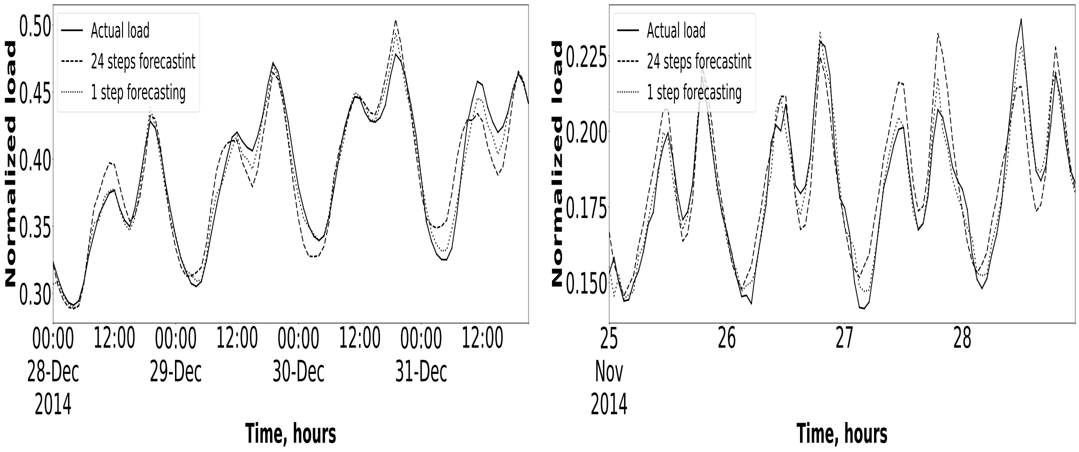
Models	**hidden layer size**	**Learning rate**
LSTM 1	(20 - 10)	0.001
LSTM 2	(20 - 10)	0.008
LSTM 3	(50 - 30)	0.002
LSTM 4	(90 - 30)	0.003
LSTM 5	(20 - 10)	0.001

After selecting the most convenient inputs with the best model topology, determined through several epochs and ANN configurations, Table 5.8 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.

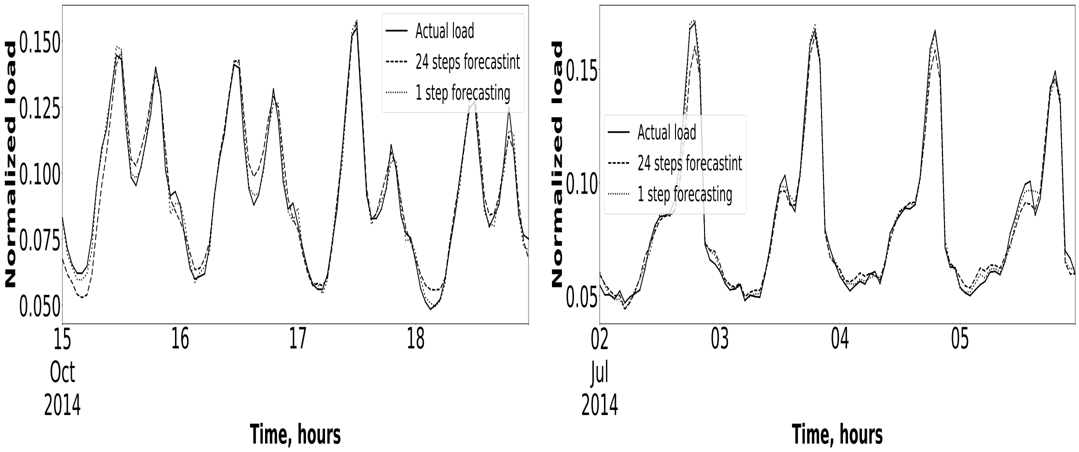
Table 5.8: MAPE, MAE and RMSE of each cluster on the test sets

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

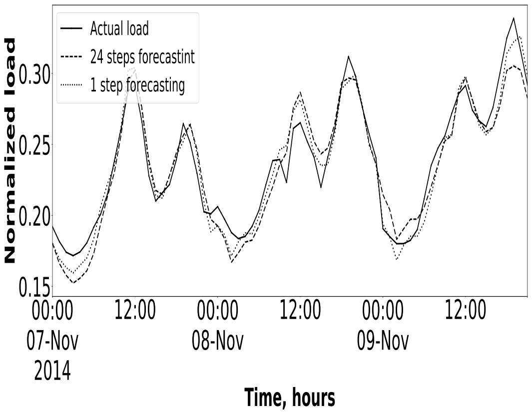
Figures 5.4a and 5.4b clearly show a high similarity between actual and forecast load on the two clusters that cover the period of winter and the period of spring and autumn. In contrast, regardless the sort of inputs used to train the LSTM models and their structure, and due to the high non-linearity level in the natural gas consumption variation recorded in the summer, as presented in Figure 5.4c, LSTM 3 forecast could not perfectly reflects the natural gas consumption. As shown in Figures 5.4d and 5.4e, 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 5.4: 24 hour forecast samples from each cluster test set

5.3.2 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.

5.3.2.1 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.

5.3.2.2 SARIMAX

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.

5.3.2.3 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:

$$y_t = \sum_{i=1}^{n} \theta_i x_i + \epsilon \tag{5.7}$$

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

5.3.3 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 (5.8) expresses the weighted average of an error ($\overline{E}$), where w_i is the number of days counted in each cluster c_i.

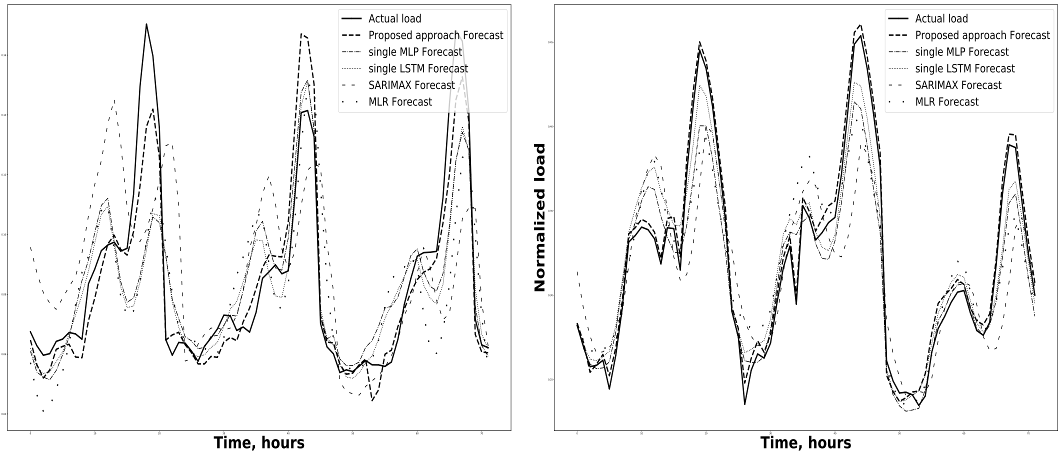
$$\overline{E} = \frac{1}{N} \sum_{i=1}^{5} (\ w_i E_i\ ) \tag{5.8}$$

Table 5.9: 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 5.9, 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.



(a) Ramadan month

(b) National holidays

Figure 5.5: Comparison between actual and predicted natural gas consumption with the use of the proposed and benchmark approaches

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 of national holidays or when the customer's 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 5.5 presents the advantage of the proposed approach upon other approaches on the national holidays and the beginning of the month of Ramadan.

5.4 Conclusion

The main contribution of this chapter 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, MLP ANN, LSTM RNN, SARIMAX and MLR models are designed and used as comparative benchmarks. 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.

Chapter 6

Long term forecasting

Contents

6.1 Introduction

Unlike short-term, long term forecasting horizon is mainly affected by economic factors rather than weather conditions. To improve the forecasting accuracy for this horizon, additional factors was mentioned in a variety of papers such as (gross national product, gross domestic product, population, number or households, oil price (Kermanshahi & Iwamiya 2002)(Ma & Li 2010), residual fuel price, petroleum product price, coal price, manufacturing output, manufacturing energy-weighted output, manufacturing capacity utilization (Behrouznia et al. 2010), etc.)

The main idea behind this work is to employ several ANN models in order to achieve a detailed long-term forecasting for the Algerian natural gas consumption. Each model estimates a specific Distribution Division (DD), and is trained according to its evolution characteristics. The common and basic technique for national level forecasting, such as in our case is to adopt one model and a single dataset for the entire national Market. However, in such level, realizing a forecasting model is considered a very complex process, therefore in this chapter, we propose a standard methodology in which we exploit every single detail available in our dataset rather than taking a vague overview especially for such applied area level.

6.2 Natural gas consumption and factors selection

6.2.1 Yearly natural gas consumption data

In this study, consumption data of natural gas covers the period between 2000 and 2014 are provided by the national natural gas distribution company in Algeria SONELGAZ. Furthermore, Data are collected in three major sectors: low-pressure sector, medium- pressure sector and high-pressure sector. Figure. 6.1 illustrates the natural gas consumption data on the three different sectors.

In the beginning of 2000, the company adopted a new organization scheme, creating four main distribution areas: SONELGAZ distribution branches, namely: the west, east, center and Algiers branches abbreviated as SDO, SDE, SDC and SDA respectively. Each general directorate is constituted by several distribution divisions counting-up to 40 DD. The aim of this work is achieving an accurate and reliable forecasting results using Artificial Neural Networks models for low, medium and high-pressure consumption. In other term to estimate the total natural gas consumption in for every sector, all existing divisions load needs to be treated and forecasted.

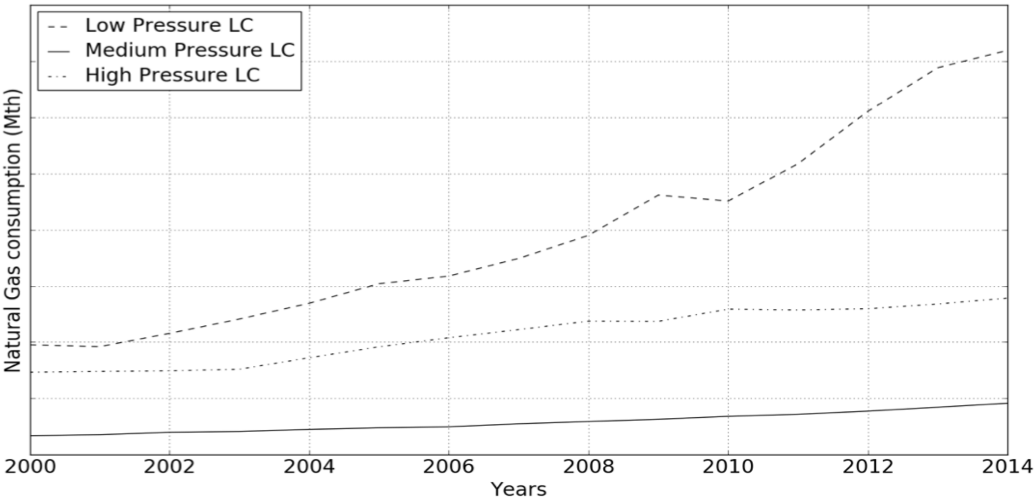


Figure 6.1: Natural gas consumption averages for high, medium and low pressure sectors.

6.2.2 Exogenous inputs selection

This process considered as an import design stage in order to achieve prediction accuracy. Before applying the chosen prediction methods and strategies, factors and inputs that have a significant effect on natural gas demand must be determined. On an annual basis forecasting horizon, GDP, population, number of costumers and the yearly consumption average are taken due to its impact on this time horizon (Kermanshahi & Iwamiya 2002)(Ma & Li 2010).

In order to determine which exogenous features are needed for the models creation processes, the correlation of precedents input variables with the natural gas consumption load are computed. Figure. 6.2 illustrates the average correlation result of natural gas load with the available factors.

The next tables: Table. 6.1 , Table. 6.2 and Table. 6.3 present the selected exogenous factors that have a strong correlation with the natural gas consumption for each DD, and which used in both: our main method and the linear one.

Table 6.1: Selected exogenous factors for the low-pressure sector

Exogenous factor	**Distribution Divisions**
Number of clients	All except: DD28
Annual average consumption	DD2, DD28
Population	All except: DD13, DD19, DD23, DD28
GDP	DD5, DD6, DD8, DD12, DD14, DD15, DD16, DD17, DD18, DD20, DD21, DD22, DD27, DD35, DD38

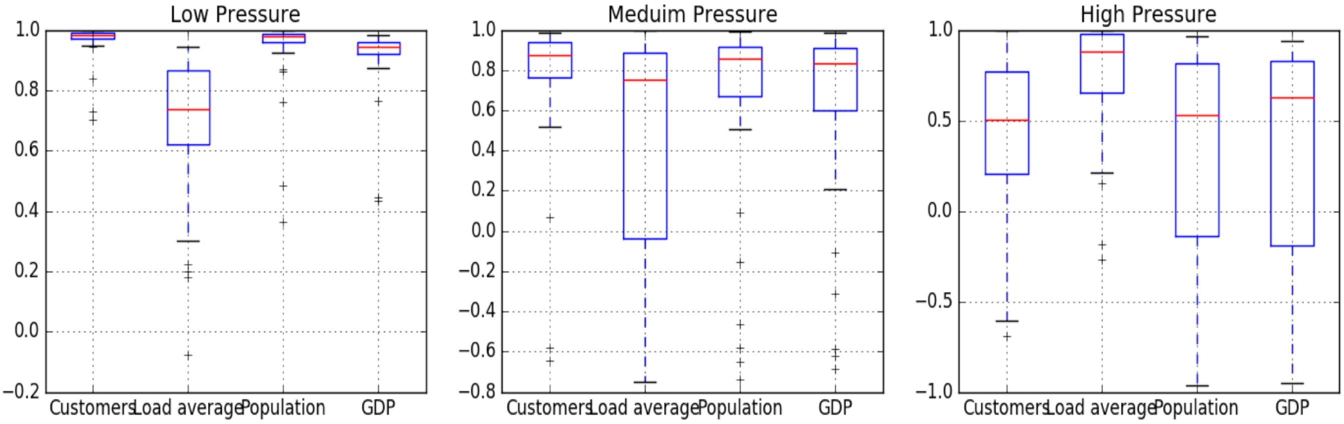


Figure 6.2: correlation between customer's number, yearly average consumption, population and GDP with the each sector load.

Table 6.2: Selected exogenous factors for the medium-pressure sector

Exogenous factor	**Distribution Divisions**
Number of clients	All except: DD7, DD10, DD31, DD32, DD37 and DD40.
Annual average consumption	DD1, DD6, DD7, DD9, DD10, DD13, DD16, DD19, DD21, DD22, DD23, DD28, DD31, DD32, DD33, DD36, DD38, DD39 and DD40.
Population	All except: DD1, DD5, DD6, DD9, DD13, DD16, DD19, DD21, DD23, DD28, DD30, DD31, DD32, DD36, DD38, DD39 and DD40.
GDP	DD2, DD3, DD4, DD11, DD12, DD14, DD15, DD18, DD20, DD25, DD26, DD33, DD34 and DD37.

Table 6.3: Selected exogenous factors for the high-pressure sector

Exogenous factor	**Distribution Divisions**
Number of clients	DD1, DD4, DD14, DD16, DD17, DD18, DD22, DD23, DD29, DD30, DD31, DD32 and DD37.
Annual average consumption	All except: DD2, DD12, DD18, DD21, DD22, DD29, DD30, DD33, DD34, DD35, DD36, DD38 and DD39.
Population	DD2, DD3, DD5, DD12, DD15, DD21, DD24, DD33, DD34, DD37 and DD38.
GDP	DD6, DD12, DD15, DD16, DD21, DD26, DD27, DD33, DD34, DD37 and DD38.

6.3 Methodology

The current work is based on utilising Multi-Layer Perceptron models with two hidden layer architectures that use sigmoidal units.

$$\hat{G} = \omega^T sig(Vx_k + \beta) \tag{6.1}$$

In order to train models, the datasets are divided into a training set $n1$ and test set $n2$, where we used the first 13 years of the input vector to adjust the weights ω^T and biases β on the models, and we keep the last 2 years vector to confirm the prediction accuracy of the models.

The general form of the network is:

$$\hat{G}_k = f(G_{k-1}, E_k, ..., E_k) \tag{6.2}$$

where k denotes the yearly discrete time index, $\hat{G}_k$ the estimated natural gas consumption, G_{k-1} the natural gas load on the previous year and $E_k, ..., E_k$ are the selected exogenous factors for the each model.

Concerning the learning algorithm, there are several ones exist to solve the non-linear least squares problem such as back-propagation (e.g. with an adaptive learning rate and momentum term), quasi-Newton or conjugate gradient algorithms, etc (Suykens et al. 1996). In this study the Levenberg-Marquardt method is used summarised in the following equation:

$$\underset{\omega,V,\beta}{min}\frac{1}{2n_1}\sum_{k=1}^{n1}(\hat{G}-G)^2 \tag{6.3}$$

The network architecture applied for prediction process is schematically presented in Figure. 6.3.

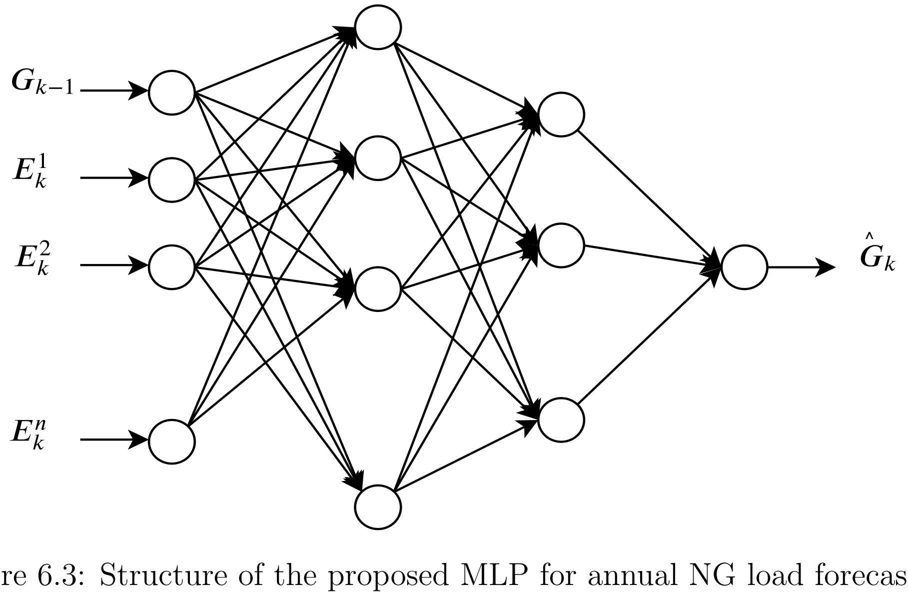


Figure 6.3: Structure of the proposed MLP for annual NG load forecasting.

6.4 Results and Discussion

6.4.1 Multiple MLP Method

Following the methodology of using multiple MLP models, where the prediction of the NG consumption for every single DD is been handled by one model. As been discussed above in the introduction section, each model takes as inputs the past amount of yearly gas consumption for its correspondent DD together with the selected exogenous factors detailed in Section. 6.2.2.

After training and validation process, we estimate the load for all existing DD on a specific pressure sector, and then summing it all to obtain the total consumption.

6.4.2 Single MLP Method

In this method, we developed an only one MLP model for each consumption pressure sector is developed. Table. 6.4 shows the selected input features for the three adopted models.

Table 6.4: the selected factors for the models creation process corresponds to each consumption pressure sector

Exogenous inputs	**EX1**	**EX2**	**EX3**	**EX4**
Low pressure (LP) model	✓		✓	✓
Medium pressure (MP) model	✓		✓	✓
High pressure (HP) model		✓		

Where ex1 represents the number of costumers, ex2 represents the annual average consumption, ex3 represents the population and ex4 represents the GDP factor.

The ANN model's Hyper parameters are presented in Table. 6.4.

Table 6.5: Summary of used inputs, FM-MLP structure and learning parameters

Models	**Input layer No. of neurons**	**No. of Hidden layers**	
		First layer	Second layer
LP model	4	7	5
MP model	4	7	5
HP model	2	4	3

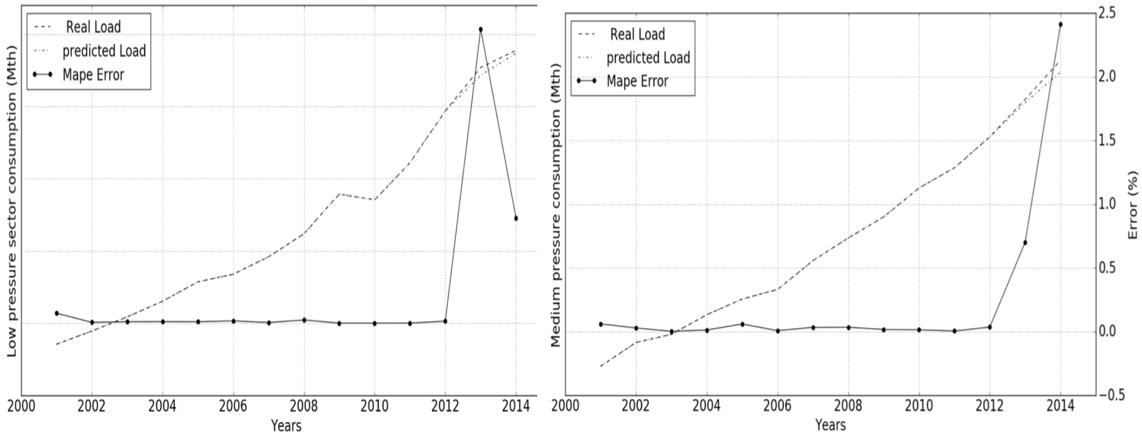
6.4.3 Linear Method

In order to validate and evaluate the enhancement in performance given by ANN methods, a benchmark approach presented in the multiple linear regression, in which each DD trend is explained by its historical value (G_{k-1} and exogenous factors as detailed in Equation. 6.4 (Farfar et al. 2015)

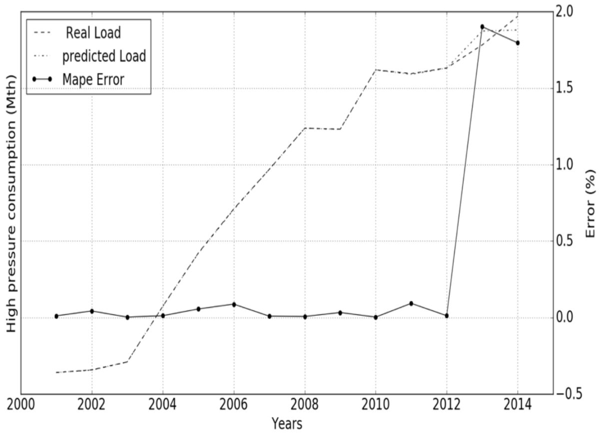
$$\hat{G}_k = \theta_0 G_{k-1} + \theta_1 E_k^1 + ... + \theta_n E_k^n + \varepsilon \tag{6.4}$$

where: $\theta_0 ... \theta_n$ are the model parameters and ε is the white noise.

6.5 Results



(a) Low Pressure sector forecast. (b) Medium Pressure sector forecast.



(c) High Pressure sector forecast.

Figure 6.4: the actual and MLP fitted values for Low, Medium and High pressure sector.

The forecasted results obtained by neural and linear models for natural gas consumption are presented below in Table. 6.6. The mean absolute percentage error (MAPE) is used as a prediction accuracy measure.

Table 6.6: Comparison of forecasting measurement MAPE errors between Linear and Neuronal Approach.

Sectors		Low pressure sector			Medium pressure sector			High pressure sector		
Methods		Linear (%)	Multiple MLP (%)	Single MLP (%)	Linear (%)	Multiple MLP (%)	Single MLP (%)	Linear (%)	Multiple MLP (%)	Single MLP (%)
Training	2001	3.2	0.069	0.459	0.866	0.061	0.113	0.462	0.011	0.031
	2002	0.144	0.006	0.103	0.487	0.029	0.075	3.091	0.043	0.025
	2003	0.563	0.01	0.228	0.184	0.003	0.048	1.957	0.003	0.048
	2004	0.643	0.011	0.043	0.912	0.013	0.201	0.641	0.012	0.044
	2005	0.597	0.01	0.222	0.129	0.06	0.147	1.202	0.056	0.021
	2006	0.504	0.016	0.151	0.347	0.008	0.101	0.42	0.088	0.02
	2007	0.831	0.005	0.185	0.955	0.034	0.183	3.166	0.009	0.012
	2008	0.805	0.022	0.206	1.33	0.035	0.136	1.756	0.007	0.018
	2009	1.239	0.001	0.062	0.481	0.017	0.143	0.935	0.033	0.009
	2010	0.947	0.001	0.149	0.589	0.015	0.265	2.806	0.003	0.008
	2011	0.89	0.001	0.151	0.271	0.005	0.335	1.454	0.093	0.027
	2012	4.238	0.014	0.388	0.535	0.036	0.285	1.69	0.012	0.012
Test	2013	7.659	2.037	1.485	1.539	0.703	1.506	5.78	1.901	0.922
	2014	6.281	0.726	2.495	4.308	2.415	3.826	2.057	1.795	2.603

Table. 6.6 shows very accurate results obtained by the multiple MLP method, where its calculated MAPE through training and test set in low, medium and high pressure sector: 0.0138%, 0.026%, 0.030% respectively for training dataset, and: 1.382%, 1.559%, 1.848% for test dataset. On the other side the classic method with a single MLP for each sector results: 0.187%, 0.170% and 0.015% through training, and 1.991%, 2.667% and 1.763% for test dataset. Therefore, this fairly demonstrated that the proposed method gave more accurate estimation compared with the standard MLP and the linear regression methods.

The multiple MLP method shows consistently accurate results and excellent performance across all training dataset period from 2001 to 20012, and an acceptable result on the test dataset concerning the years 2013 and 2014. Figure. 6.4 shows comparison between real and estimated consumption load along with the MAPE value of each year.

6.6 PREVGAZ-DZ

The best Linear and Nonlinear ANN models obtained for the long term forecasting have been integrated as the core of a long term natural gas forecasting tool named PREVGAZ-DZ and develloped for SONELGAZ.

6.7 Yearly natural gas consumption forecast

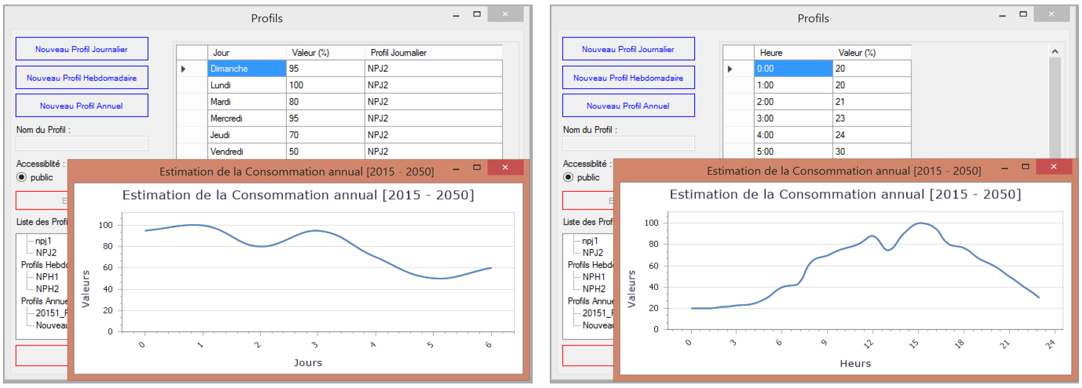
The developed software gives the chance to choose between two different type of models, MLR based models and MLP based models, where the first ones are less efficient compared to the ANN models as discussed in Section. 6.5.

PREVGAZ-DZ provides a forecast for the national market for a range of 50 years in three main aspects. Within the software a branch of models are designed and thus could be categorised as follows:

- Firstly, several MLP and MLR models with respect to the number of distribution offices which counts 40 ones across the country and constitute 4 distribution divisions for the residential sector.
- Secondly, several MLP and MLR models to predict the national natural gas consumption for the economical activities sector which constituted of few principal departments (industrial, agricultural, residential, etc...).
- The predictions has been related to the three pressure levels in the Algerian market (High, medium and low).
- Since the software deals with three pressure sectors, the natural gas consumption related to each DD or SAE department in both low and medium pressure sectors is represented by one separate MLP and MLR model. Whereas in high pressure sector and according to the company demand and the natural gas market hierarchy, a single MLP and MLR model takes the charge of predicting the global consumption for this sector.

In addition to the natural gas consumption forecasting, some other features also been offered and are listed bellow:

- Forecast the natural gas consumption used for electricity production which being generated by three types of turbine power technology: Steam turbines, gas turbine and combined Cycles turbines.
- provide an annual consumption profile according to an existing load matrix using a three profile approach concept. This matrix calculation is thus based on daily, weekly and annual consumption profiles where type of the season, type of the week, type of the day (working day, holidays and pre-holidays) are all considered. Figures.

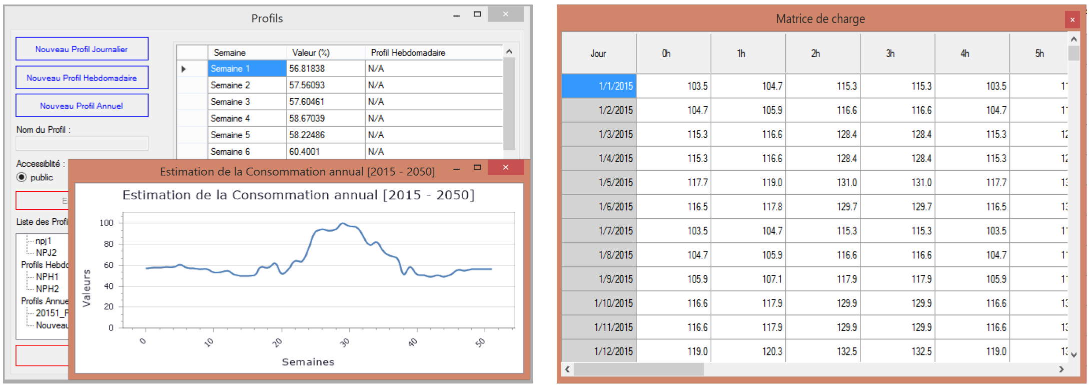


	Jour	Valeur (%)	Profil Journalier
▶	Dimanche	95	NPJ2
	Lundi	100	NPJ2
	Mardi	80	NPJ2
	Mercredi	95	NPJ2
	Jeudi	70	NPJ2
	Vendredi	50	NPJ2

(a) Daily Profile setup.

	Heure	Valeur (%)
▶	0:00	20
	1:00	20
	2:00	21
	3:00	23
	4:00	24
	5:00	30

(b) Weekly profile setup.



	Semaine	Valeur (%)	Profil Hebdomadaire
▶	Semaine 1	56.81838	N/A
	Semaine 2	57.56093	N/A
	Semaine 3	57.60461	N/A
	Semaine 4	58.67039	N/A
	Semaine 5	58.22486	N/A
	Semaine 6	60.4001	N/A

(c) Annual Profile setup.

Jour	0h	1h	2h	3h	4h	5h
1/1/2015	103.5	104.7	115.3	115.3	103.5	11
1/2/2015	104.7	105.9	116.6	116.6	104.7	11
1/3/2015	115.3	116.6	128.4	128.4	115.3	12
1/4/2015	115.3	116.6	128.4	128.4	115.3	12
1/5/2015	117.7	119.0	131.0	131.0	117.7	13
1/6/2015	116.5	117.8	129.7	129.7	116.5	13
1/7/2015	103.5	104.7	115.3	115.3	103.5	11
1/8/2015	104.7	105.9	116.6	116.6	104.7	11
1/9/2015	105.9	107.1	117.9	117.9	105.9	11
1/10/2015	116.6	117.9	129.9	129.9	116.6	13
1/11/2015	116.6	117.9	129.9	129.9	116.6	13
1/12/2015	119.0	120.3	132.5	132.5	119.0	13

(d) Sample of an annual load matrix.

Figure 6.5: the actual and MLP fitted values for Low, Medium and High pressure sector.

6.8 PREVGAZ-DZ functionalities

As a first impression when the software is launched, we can see seven principal tabs, where each includes specific functionalities.

6.8.1 Home tab

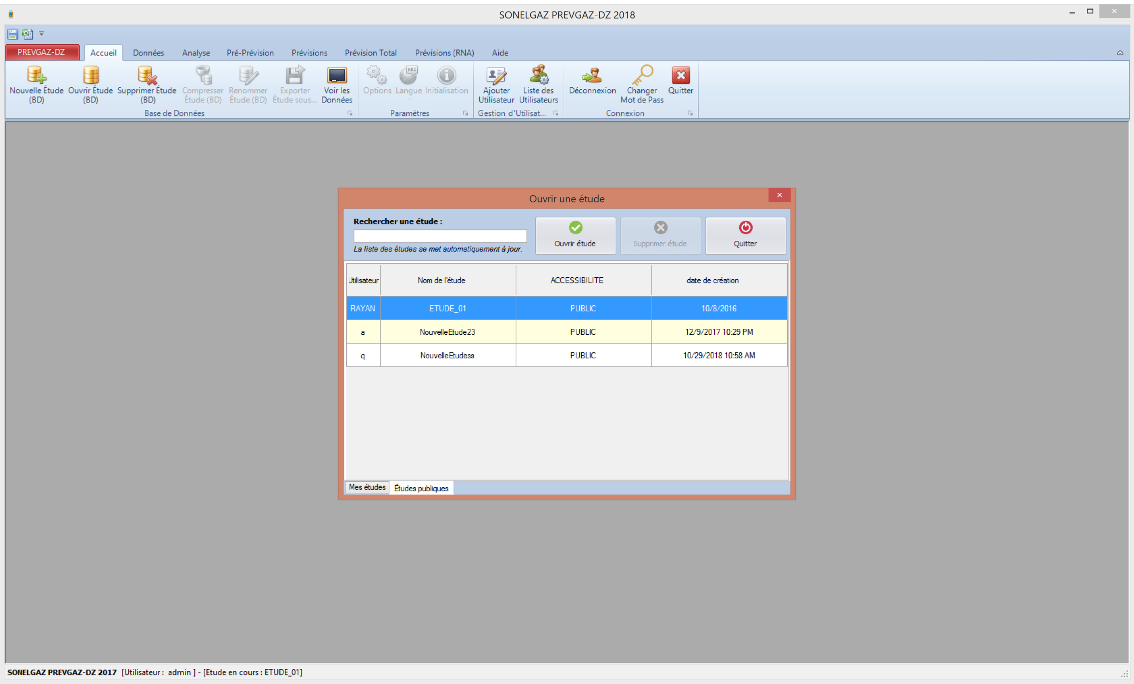
As presented in Figure. 6.6 This tab is shown at the first execution of the software where it constitutes of three sub-sections holding the following tasks:

1. database management: This part allows to create, open, delete, compress, modify and save database instance.

2. factory reset: Where it gives the option to resetting the whole software parameters to its initial state.

3. profiles management: In this sub-section, beside the user's connection

task (login and logout), the administrator will have the ability to add and remove any users on the database.



La liste des études se met automatiquement à jour.

Utilisateur	Nom de l'étude	ACCESSIBILITE	date de création
RAYAN	ETUDE_01	PUBLIC	10/8/2016
a	NouvelleEtude23	PUBLIC	12/9/2017 10:29 PM
q	NouvelleEtudess	PUBLIC	10/29/2018 10:58 AM

SONELGAZ PREVGAZ-DZ 2017 [Utilisateur : admin] - [Etude en cours : ETUDE_01]

Figure 6.6: Home tab.

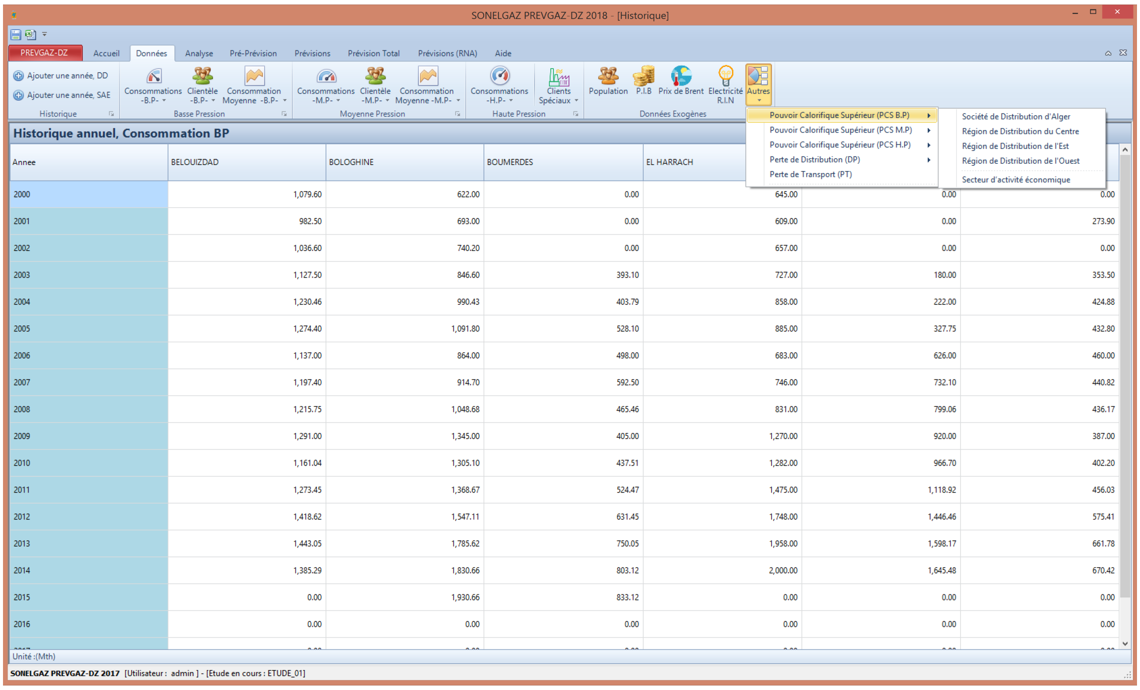
6.8.2 Data tab

This tab is customised to display the entire historical data (yearly natural gas consumption and and clientele size) and the essential exogenous variables data which are used as inputs fed into the prediction models. As shown in Figure. 6.7, this section is composed of 5 parts:

- Historical consumption: Allows to adjust the history data (Add, modify and delete functions of a year).
- Low pressure data menus: These menus shows the annual consumption and customers data for each distribution division or department in both residential sector and economical activities sector.
- Medium pressure data menus: Shows the same data as in the low pressure sector which was recorded in the medium pressure sector.

- High pressure data menu: As mentioned before in Section. 6.7, this menu shows only the global annual natural gas consumption and the total number of customers in the high pressure sector.

- Exogenous data menus: Show data corresponds to population, PIB, Brent oil price, RIN production and others. The others menu holds the following factors:
  - High calorific value (BP, MP and HP) for each distribution division and SAE department.
  - distribution loss.
  - transport loss.



Historique annuel, Consommation BP

Annee	BELOUIZDAD	BOLOGHINE	BOUMERDES	EL HARRACH		
2000	1,079.60	622.00	0.00	645.00	0.00	0.00
2001	982.50	693.00	0.00	609.00	0.00	273.90
2002	1,036.60	740.20	0.00	657.00	0.00	0.00
2003	1,127.50	846.60	393.10	727.00	180.00	353.50
2004	1,230.46	990.43	403.79	858.00	222.00	424.88
2005	1,274.40	1,091.80	528.10	885.00	327.75	432.80
2006	1,137.00	864.00	498.00	683.00	626.00	460.00
2007	1,197.40	914.70	592.50	746.00	732.10	440.82
2008	1,215.75	1,048.68	465.46	831.00	799.06	436.17
2009	1,291.00	1,345.00	405.00	1,270.00	920.00	387.00
2010	1,161.04	1,305.10	437.51	1,282.00	966.70	402.20
2011	1,273.45	1,368.67	524.47	1,475.00	1,118.92	456.03
2012	1,418.62	1,547.11	631.45	1,748.00	1,446.46	575.41
2013	1,443.05	1,785.62	750.05	1,958.00	1,598.17	661.78
2014	1,385.29	1,830.66	803.12	2,000.00	1,645.48	670.42
2015	0.00	1,930.66	833.12	0.00	0.00	0.00
2016	0.00	0.00	0.00	0.00	0.00	0.00

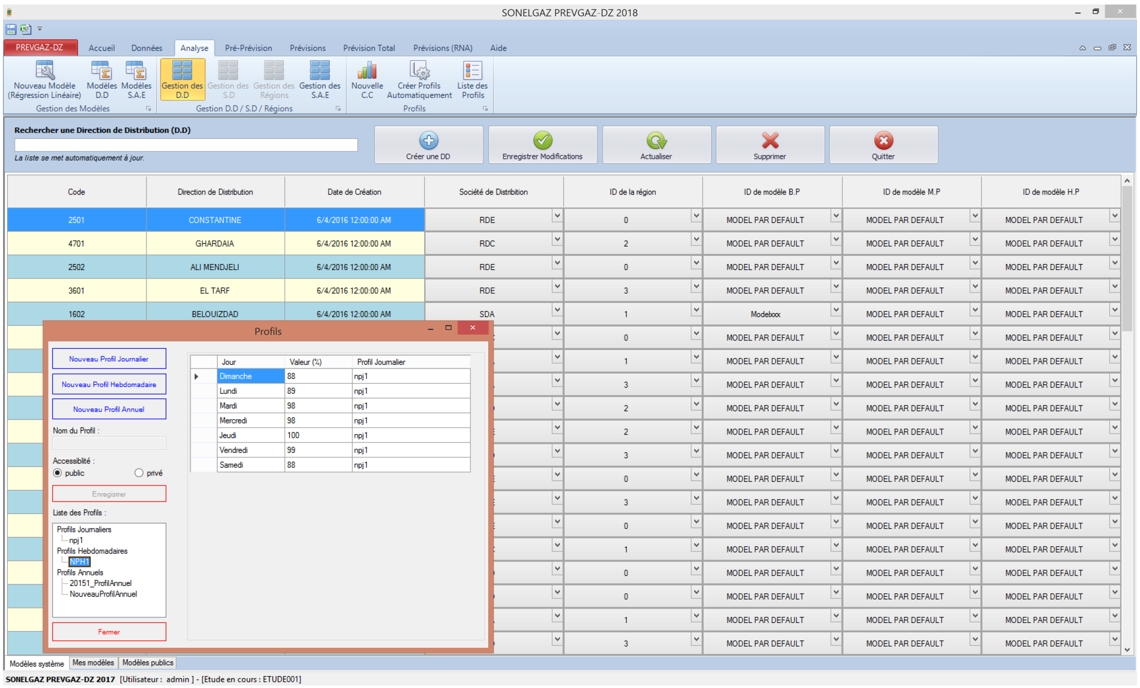
Unité :(Mth)

SONELGAZ PREVGAZ-DZ 2017 [Utilisateur : admin] - [Etude en cours : ETUDE_01]

Figure 6.7: Data tab.

6.8.3 Analyse tab

This tab is comprises three big parts as shown in Figure. 6.8 focusing on: MLR models management, distribution offices assignment to the distribution divisions as well as their assignment to one of the principal regions (north, south or high plateau) and consumption profiles management.



Rechercher une Direction de Distribution (D.D)

La liste se met automatiquement à jour.

Code	Direction de Distribution	Date de Création	Société de Distribution	ID de la région	ID de modèle B.P	ID de modèle M.P	ID de modèle H.P
2501	CONSTANTINE	6/4/2016 12:00:00 AM	RDE	0	MODEL PAR DEFAULT	MODEL PAR DEFAULT	MODEL PAR DEFAULT
4701	GHARDAIA	6/4/2016 12:00:00 AM	RDC	2	MODEL PAR DEFAULT	MODEL PAR DEFAULT	MODEL PAR DEFAULT
2502	ALI MENDJELI	6/4/2016 12:00:00 AM	RDE	0	MODEL PAR DEFAULT	MODEL PAR DEFAULT	MODEL PAR DEFAULT
3601	EL TARF	6/4/2016 12:00:00 AM	RDE	3	MODEL PAR DEFAULT	MODEL PAR DEFAULT	MODEL PAR DEFAULT
1602	BELOUIZDAD	6/4/2016 12:00:00 AM	SDA	1	Modelxxx	MODEL PAR DEFAULT	MODEL PAR DEFAULT
				0	MODEL PAR DEFAULT	MODEL PAR DEFAULT	MODEL PAR DEFAULT
				1	MODEL PAR DEFAULT	MODEL PAR DEFAULT	MODEL PAR DEFAULT
				3	MODEL PAR DEFAULT	MODEL PAR DEFAULT	MODEL PAR DEFAULT
				2	MODEL PAR DEFAULT	MODEL PAR DEFAULT	MODEL PAR DEFAULT
				2	MODEL PAR DEFAULT	MODEL PAR DEFAULT	MODEL PAR DEFAULT
				3	MODEL PAR DEFAULT	MODEL PAR DEFAULT	MODEL PAR DEFAULT
				0	MODEL PAR DEFAULT	MODEL PAR DEFAULT	MODEL PAR DEFAULT
				3	MODEL PAR DEFAULT	MODEL PAR DEFAULT	MODEL PAR DEFAULT
				0	MODEL PAR DEFAULT	MODEL PAR DEFAULT	MODEL PAR DEFAULT
				1	MODEL PAR DEFAULT	MODEL PAR DEFAULT	MODEL PAR DEFAULT
				0	MODEL PAR DEFAULT	MODEL PAR DEFAULT	MODEL PAR DEFAULT
				0	MODEL PAR DEFAULT	MODEL PAR DEFAULT	MODEL PAR DEFAULT
				1	MODEL PAR DEFAULT	MODEL PAR DEFAULT	MODEL PAR DEFAULT
				3	MODEL PAR DEFAULT	MODEL PAR DEFAULT	MODEL PAR DEFAULT

Nom du Profil :

Accessiblité :
● public ○ privé

Liste des Profils :
- Profils Journaliers
  - npj1
- Profils Hebdomadaires
  - NPH1
- Profils Annuels
  - 20151_ProfilAnnuel
  - NouveauProfilAnnuel

Jour	Valeur (%)	Profil Journalier
Dimanche	88	npj1
Lundi	89	npj1
Mardi	98	npj1
Mercredi	98	npj1
Jeudi	100	npj1
Vendredi	99	npj1
Samedi	88	npj1

SONELGAZ PREVGAZ-DZ 2017 [Utilisateur : admin] - [Etude en cours : ETUDE001]

Figure 6.8: Analyse tab.

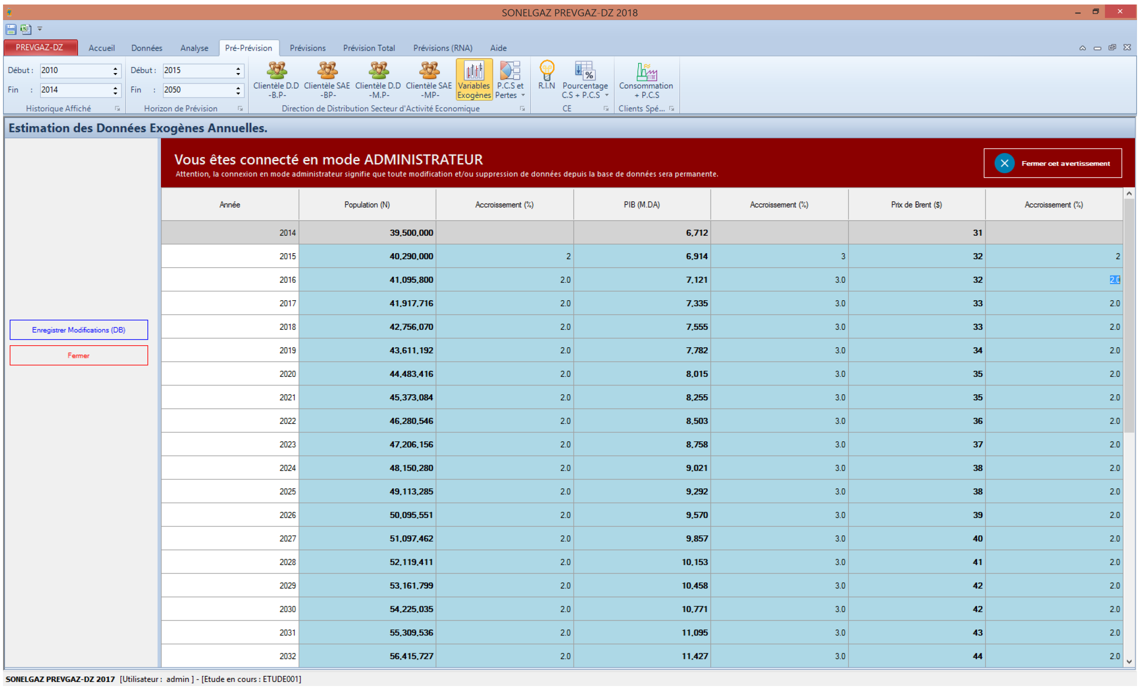
Table. 6.7 summarises a brief description of the functionalities included in this tab.

Table 6.7: Descriptive table of the analyse tab's functionalities

Function		**Action**
Models management	New MLR model	Create new MLR model
	DD models	DD MLR models management
	SAE models	SAE MLR models management
DD assignment	DD management	Assign each distribution office to its division and region, add new offices or divide an office into new two ones.
	SAE management	Select MLR models for each SAE department.
Consumption profiles management	New load curve	New load curve's creation
	Auto-profile creation	Auto-profile creation according to a given data.
	Profiles list	Creating and customising consumption profiles

6.8.4 Pre-forecasting tab

This page is mainly designed to be the post process to occur the forecasting. As presented in Figure. 6.9, this page allows to setting up the prediction parameters as well as the future values of the exogenous variables.



Historique Affiché
Début : 2010
Fin : 2014

Horizon de Prévision
Début : 2015
Fin : 2050

Estimation des Données Exogènes Annuelles.

Vous êtes connecté en mode ADMINISTRATEUR
Attention, la connexion en mode administrateur signifie que toute modification et/ou suppression de données depuis la base de données sera permanente.

Année	Population (N)	Accroissement (%)	PIB (M.DA)	Accroissement (%)	Prix de Brent ($)	Accroissement (%)
2014	39,500,000		6,712		31	
2015	40,290,000	2	6,914	3	32	2
2016	41,095,800	2.0	7,121	3.0	32	2.0
2017	41,917,716	2.0	7,335	3.0	33	2.0
2018	42,756,070	2.0	7,555	3.0	33	2.0
2019	43,611,192	2.0	7,782	3.0	34	2.0
2020	44,483,416	2.0	8,015	3.0	35	2.0
2021	45,373,084	2.0	8,255	3.0	35	2.0
2022	46,280,546	2.0	8,503	3.0	36	2.0
2023	47,206,156	2.0	8,758	3.0	37	2.0
2024	48,150,280	2.0	9,021	3.0	38	2.0
2025	49,113,285	2.0	9,292	3.0	38	2.0
2026	50,095,551	2.0	9,570	3.0	39	2.0
2027	51,097,462	2.0	9,857	3.0	40	2.0
2028	52,119,411	2.0	10,153	3.0	41	2.0
2029	53,161,799	2.0	10,458	3.0	42	2.0
2030	54,225,035	2.0	10,771	3.0	42	2.0
2031	55,309,536	2.0	11,095	3.0	43	2.0
2032	56,415,727	2.0	11,427	3.0	44	2.0

SONELGAZ PREVGAZ-DZ 2017 [Utilisateur : admin] - [Etude en cours : ETUDE001]

Figure 6.9: Pre-forecasting tab.

In summary, in order to predict the natural gas consumption, the SONELGAZ company's experts will have to submit the following parameters:

- Exogenous variables concerning both residential (number of clients) and economical sectors(exogenous data).
- Exogenous variables concerning Electric Central's consumption forecasting.
- special clients consumption estimations.

6.8.5 Forecasting tab

As the main section in the software, this part focuses on calculating the future natural gas consumption in the medium and long term using the MLR models. These calculation are based on estimated data that had been saved in 'Pre-forecasting tab'.

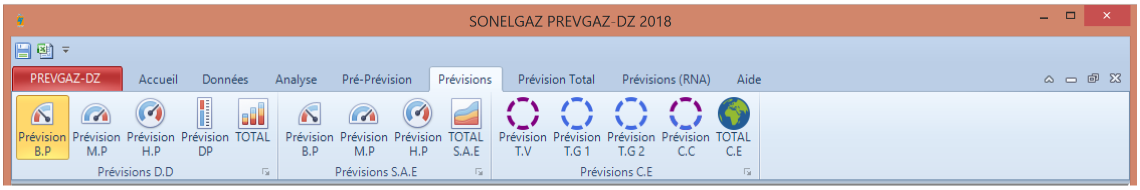


Figure 6.10: Forecasting tab menu.

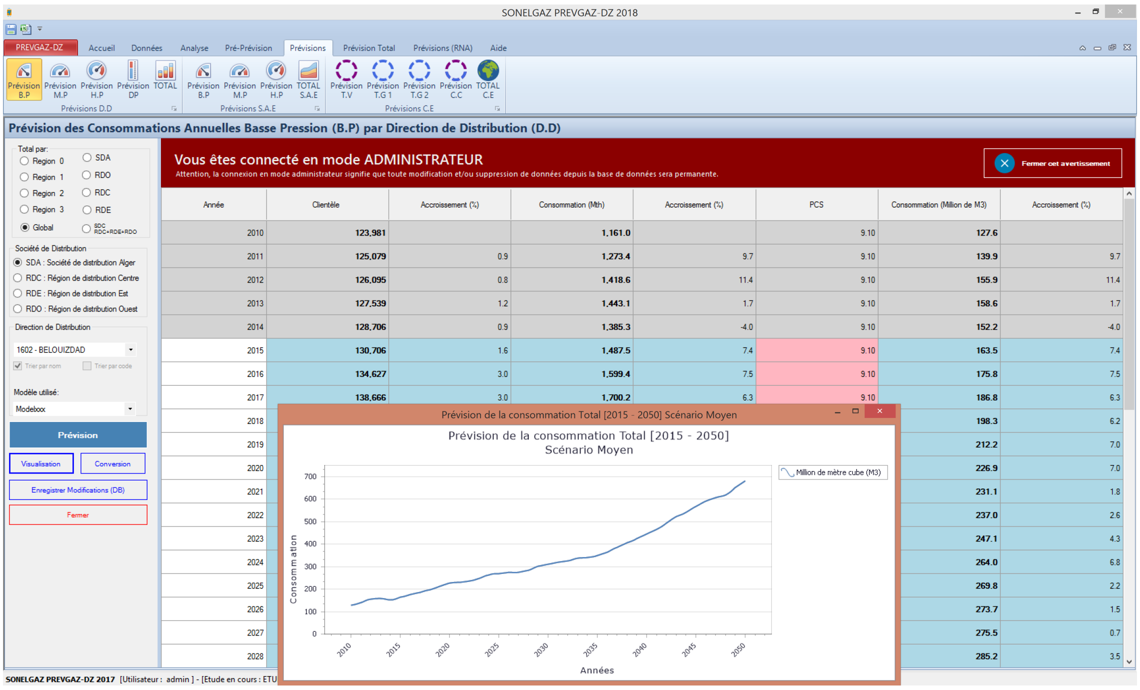
In addition to the multiple forecasting bases that are shown in Figure. 6.10, the following Table. 6.8 summarises all the functionalities found in this tab.

Table 6.8: Descriptive table of the forecasting tab's functionalities

Function		**Action**
Residential sector forecast	LP forecasting	Forecasting the LP consumption for the residential sector.
	MP forecasting	Forecasting the MP consumption for the residential sector.
	HP forecasting	Forecasting the HP consumption for the residential sector.
	DP forecasting	Totalising the LP and MP gas consumption forecast to obtain the DP forecast.
	Total	Totalising the Annual natural gas consumption for the residential sector with all losses been calculated.
SAE forecast	LP forecasting	Forecasting the LP consumption for the economical sector.
	MP forecasting	Forecasting the MP consumption for the economical sector.
	HP forecasting	Forecasting the HP consumption for the economical sector.
	Total SAE	Totalizing the Annual natural gas consumption for the economical sector with all losses been calculated.
CE forecast	ST forecasting	Forecasting the steam turbines consumption.
	GT.1 forecasting	Forecasting the low power gas turbines consumption (GT¡100MW).
	GT.2 forecasting	Forecasting the high power gas turbines consumption (GT¿100MW).
	CCT forecasting	Forecasting the combined cycle gas turbines consumption.
	Total CE	Totalizing the Annual consumed natural gas for electrical power generation.

6.8.5.1 Residential sector natural gas consumption forecasting

For only low pressure and medium pressure sectors, as been shown in Figure. the user can select one of 4 distribution divisions and then forecast each one of its distribution offices separately. In the other hand, high pressure natural gas consumption forecasting is achieved at the national level only.



Prévision des Consommations Annuelles Basse Pression (B.P) par Direction de Distribution (D.D)

Total par:
○ Region 0 ○ SDA
○ Region 1 ○ RDO
○ Region 2 ○ RDC
○ Region 3 ○ RDE
● Global ○ SDC RDC+RDE+RDO

Société de Distribution
● SDA : Société de distribution Alger
○ RDC : Région de distribution Centre
○ RDE : Région de distribution Est
○ RDO : Région de distribution Ouest

Direction de Distribution
1602 - BELOUIZDAD
☑ Trier par nom ☐ Trier par code

Modèle utilisé:
Modeboox

Vous êtes connecté en mode ADMINISTRATEUR
Attention, la connexion en mode administrateur signifie que toute modification et/ou suppression de données depuis la base de données sera permanente.

Année	Clientèle	Accroissement (%)	Consommation (Mth)	Accroissement (%)	PCS	Consommation (Million de M3)	Accroissement (%)
2010	123,981		1,161.0		9.10	127.6	
2011	125,079	0.9	1,273.4	9.7	9.10	139.9	9.7
2012	126,095	0.8	1,418.6	11.4	9.10	155.9	11.4
2013	127,539	1.2	1,443.1	1.7	9.10	158.6	1.7
2014	128,706	0.9	1,385.3	-4.0	9.10	152.2	-4.0
2015	130,706	1.6	1,487.5	7.4	9.10	163.5	7.4
2016	134,627	3.0	1,599.4	7.5	9.10	175.8	7.5
2017	138,666	3.0	1,700.2	6.3	9.10	186.8	6.3
2018						198.3	6.2
2019						212.2	7.0
2020						226.9	7.0
2021						231.1	1.8
2022						237.0	2.6
2023						247.1	4.3
2024						264.0	6.8
2025						269.8	2.2
2026						273.7	1.5
2027						275.5	0.7
2028						285.2	3.5

SONELGAZ PREVGAZ-DZ 2017 [Utilisateur : admin] - [Etude en cours : ETU

Figure 6.11: Belouizdad low pressure natural gas consumption forecasting.

In response to the company's needs, once the natural gas consumption is forecasted for each distribution office in the three pressure levels, the company's experts will be able to have two sums, the first is the DP (explained in Table. 6.8) by considering a loss factor that represents amount of natural gas being lost during the distribution. The second sum is the final national consumption for all three pressure sectors, to calculate of this latest another loss factor is considered which represents the amount of natural gas being lost in transport.

6.8.5.2 Economical sector natural gas consumption forecasting

In the economical activity sector of the market, SONELGAZ is mainly interested by five sections namely:

- Residential.
- Agricultural.
- Two industrial subsections (energy-intensive industries and small and medium industries).

- Two service sector subsections (non-market services and the rest of the section)
- The rest of sections.

The global consumption forecast of the economic activities sector is simply given by summing the forecast of all the last mentioned sections, and this is the case for LP, MP and HP sectors. Furthermore, the only sections included in the high pressure sector are the energy-intensive industries and the small and medium industries subsections which represents the industrial sector.

6.8.5.3 Power plan consumption forecasting

The achieved forecast in this part is about the natural gas that being consumed by three types of power plans for electricity generation: steam turbines, low-power gas turbines (GT<100MW), high-power gas turbines (GT>100MW) and combined cycle gas turbines. Basically, the power plan consumption forecasting is achieved based on a prior prediction of the electrical production (RIN), therefore, the consumption forecast does not require any models to be designed. Besides using the RIN factor as input variable, a specific estimated consumption value for each turbine (th/GWh) and the gross heating value are also added to the equation in order to calculate the natural gas forecast.

6.8.6 Forecasting with MLPs tab

This section focuses on forecasting the residential natural gas consumption using the artificial neural networks models. Similarly to sub-section. 6.8.5.1, this part does use MLPs to model the annual natural gas consumption for each distribution office on LP and MP sectors, and also to models the national HP sector consumption.

6.8.6.1 Total tab

This part actually manages the obtained natural gas consumption forecast for the residential sector, the economic activities sector and the electricity generation sector. Alongside with Figure. 6.12, Table. 6.9 shows three different strategies to calculate the total forecast.

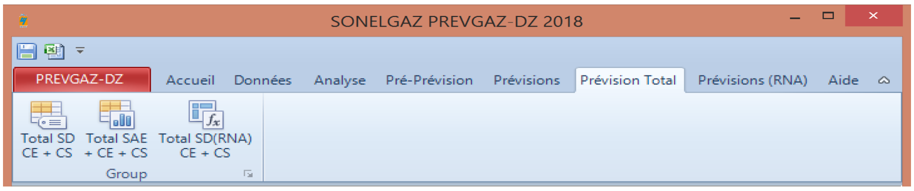


Figure 6.12: Total tab menu.

Table 6.9: Descriptive table of the analyse tab's functionalities

Function		**Action**
Total forecast	Total SD + CE + CS	Totalize the national consumption forecast based on the MLR results for the residential sector.
	Total SAE + CE + CS	Totalize the national consumption forecast based on the economic activities sector.
	Total SD (ANN) + CE + CS	Totalize the national consumption forecast based on the MLP results for the residential sector.

6.9 Conclusion

In summary, Artificial Neural Networks models for yearly basis estimation of natural gas consumption studied and presented, and which trained with the Levenberg-Marquardt learning algorithm. Despite the limited quantity of data between 2000 and 2014, the multiple MLPs method achieved an encouraging and very acceptable results compared to both, the method that uses a single MLP model and the Linear Regression models, especially for generalization on test data. Finally, the multiple MLPs method can reliably and accurately be used for forecasting natural gas consumption and is therefore an excellent tool in helping the SONELGAZ company's experts to make the best decisions.

Chapter 7

Conclusions

In this thesis, we have investigated a batch of artificial and statistical techniques to see how can they model the Algerian natural gas consumption on the short and long term bases. Concerning the short term modeling, the thesis begins by studying the basic two-stage forecasting approach, the study is tended to examine and analyse an hourly natural gas consumption data using multiple benchmark clustering techniques beside developing GP models to achieve the consumption prediction.

Beginning with the use of K-means as a parametric centroid based method, HDBSCAN as a non-parametric density based method and MOHGP time series based method. According to the elbow method that looks at the total within-cluster sum of squares to validate the number of clusters, 3 clusters is the best choice to be used for K-means. Being non-parametric classification methods, the auto-generated numbers of clusters for HDBSCAN and MOHGP are 5 and 6 respectively. Furthermore, another adopted consumption profiles grouping was a result of combining the 3 obtained clusters by K-means with other 2 clusters, these latest were recognized by HDBSCAN representing the month of Ramadan and special period of national and religious holidays. These different natural gas consumption profile segmentations led to creating distinct amount and types of AR-GP models, where they fit according to the shape of the resulted clusters.

Classifying load curves into a huge amount of groups or adopting many different models does not necessarily improve the forecasting results, even in the case of using powerful clustering techniques. Contrarily, properly segmented and classified clusters can enhance the overall quality of the developed models considerably.

In the same horizon, this thesis proposed a new forecasting approach based on RNNs method, particularly LSTM models. The novelty of the proposed approach consists in resolving the two-stage approach's deficiency by developing a MLP to estimate the next day consumption profile and thus creates a role of a forecasting monitor. The second part of the proposed approach is then focuses on building a comprehensive Long Short Term Memory (LSTM) recurrent models according to load behavior.

In order to assess the FM-MLP's estimation quality, the dataset was randomly split into training, validation and test sets with 60%, 20%5 and 20% respectively. The FM-MLP shows a powerful ability by achieving a consumption profiles classification with zero miss-classed profile during training and validation process. Regarding test set that contains 167 load profiles, the FM-MLP had a single miss-classed profile with results a classification accuracy score of 98.43%.

During the LSTM models creation phase, based on results obtained in Chapter.4 where the selected number of clusters is 5, different RNN topologies and features vectors were tested and employed in order to reach the best fit of natural gas consumption data. Based on the results in the Chapter. 5 we come to the following conclusions:

In a 24 hours forecasting, historical consumption represented in the lagged loads (C_{t-1}),..., (C_{t-12}) have the most important effect on initiating an interesting natural gas consumption modeling.

Weather condition data also considered as a key factor that improves the models prediction. Due to the lack of hourly temperature data provided by the meteorological services, a special technique for the purpose of estimating daily temperature profiles is used, these hourly points are generated based on the minimum and maximum temperature of the day.

In addition to historical and exogenous data, another type of input is used holding calendar information. The use of this kind of attributes especially day and hour indicators can properly adjust the prediction by helping the LSTM models capture the intraday periodic pattern and distinguishing between working days and holidays. Furthermore, a monthly indicator variable may allow to catch the long term change by recognizing months and seasons.

Due to less computational time and resources to find the most appropriate model configuration, Random search technique was used to perform 30 experiments with a combination of different values of learning rate and hidden layer neurons count.

After comparing the results with four alternative benchmark approaches: MLP, LSTM, SARIMAX and MLP models, especially with the 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.

7.1 Future work

Regarding the short term forecasting which represents the main subject of this thesis, the following work will continue by studying and exploring a bigger dataset whether by extending the natural gas consumption data or by integrating some auxiliary exogenous information sources (possibly correlated) such as humidity, wind speed, pollution levels and even trying to utilise the temperature values in other forms. In most cases, adding new factors can improve the prediction accuracy.

Despite the good results obtained by our novel proposed approach presented in Chapter. 5, we believe that more possible improvements deserve to be exploited for the respect of future works such as:

- In the same direction of adopting hybrid approaches, a variety of multi-model architectures can be also be investigated if an ideal schema with right models are chosen.
- Using other types and structures of RNN might be interesting and could probably surpass LSTM model in terms of modeling accuracy.
- Since DNN has become the 'talk of the town' between the majority of researchers with machine learning backgrounds, reviewing a bigger dataset which is recorded for a period of years would allow to practically apply this kind of models.

Bibliography

Akpinar, M., Adak, M. & Yumusak, N. (2017), 'Day-ahead natural gas demand forecasting using optimized ABC-based neural network with sliding window technique: The case study of regional basis in turkey', *Energies* **10**(6), 781.
URL: *https://doi.org/10.3390/en10060781*

Akpinar, M. & Yumusak, N. (2013), Forecasting household natural gas consumption with arima model: A case study of removing cycle, *in* 'Proceedings of 2013 7th International Conference on Application of Information and Communication Technologies', pp. 1–6.

Alamaniotis, M., Chatzidakis, S. & Tsoukalas, L. (2014), Monthly load forecasting using kernel based gaussian process regression, *in* 'Proceedings of MedPower 2014', Institution of Engineering and Technology.

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

Azadeh, A., Asadzadeh, S. & Ghanbari, A. (2010), 'An adaptive network-based fuzzy inference system for short-term natural gas demand estimation: Uncertain and complex environments', *Energy Policy* **38**(3), 1529–1536.
URL: *https://doi.org/10.1016/j.enpol.2009.11.036*

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

Baldacci, L., Golfarelli, M., Lombardi, D. & Sami, F. (2016), 'Natural gas consumption forecasting for anomaly detection', *Expert Systems with Applications* **62**, 190–201.
URL: *https://doi.org/10.1016/j.eswa.2016.06.013*

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

Behrouznia, A., Saberi, M., Azadeh, A., Asadzadeh, S. M. & Pazhoheshfar, P. (2010), An adaptive network based fuzzy inference system-fuzzy data envelopment analysis for gas consumption forecasting and analysis: The case of south america, *in* '2010 International Conference on Intelligent and Advanced Systems', IEEE.
URL: *https://doi.org/10.1109/icias.2010.5716160*

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

Bianchi, F. M., Maiorino, E., Kampffmeyer, M. C., Rizzi, A. & Jenssen, R. (2017), *Recurrent Neural Networks for Short-Term Load Forecasting*, Springer International Publishing.
URL: *https://doi.org/10.1007/978-3-319-70338-1*

Box, G. (2013), Box and jenkins: Time series analysis, forecasting and control, *in* 'A Very British Affair', Palgrave Macmillan UK, pp. 161–215.
URL: *https://doi.org/10.1057/9781137291264_6*

Box, G. E. P. & Jenkins, G. M. (1994), *Time Series Analysis: Forecasting and Control*, 3rd edn, Prentice Hall PTR, Upper Saddle River, NJ, USA.

Box, G. & Jenkins, G. (1976), *Time series analysis: forecasting and control*, Holden-Day series in time series analysis and digital processing, Holden-Day.
URL: *https://books.google.dz/books?id=1WVHAAAAMAAJ*

Bozkurt, O. O., Biricik, G. & Taysi, Z. C. (2017), 'Dataset for "artificial neural network and sarima based models for power load forecasting in turkish

electricity market"'.
URL: *https://zenodo.org/record/375589*

Breiman, L., Friedman, J., Stone, C. & Olshen, R. (1984), *Classification and Regression Trees*, The Wadsworth and Brooks-Cole statistics-probability series, Taylor & Francis.
URL: *https://books.google.dz/books?id=JwQx-WOmSyQC*

Calinski, T. & Harabasz, J. (1974), 'A dendrite method for cluster analysis', *Communications in Statistics - Theory and Methods* **3**(1), 1–27.
URL: *https://doi.org/10.1080/03610927408827101*

Campello, R. J. G. B., Moulavi, D. & Sander, J. (2013), Density-based clustering based on hierarchical density estimates, *in* 'Proceedings of Knowledge Discovery and Data Mining (PAKDD'13)', Springer Berlin Heidelberg, p. 160–172.

Campello, R. J. G. B., Moulavi, D., Zimek, A. & Sander, J. (2015), 'Hierarchical density estimates for data clustering, visualization, and outlier detection', *ACM Transactions on Knowledge Discovery from Data* **10**(1), 1–51.

Chen, N., Qian, Z., Meng, X. & Nabney, I. T. (2013), Short-term wind power fore- casting using gaussian processes, *in* 'Proceedings of International Joint Conference on Artificial Intelligence (IJCAI'13)', Beijing, China.

Chen, Y., Chua, W. S. & Koch, T. (2018), 'Forecasting day-ahead high-resolution natural-gas demand and supply in germany', *Applied Energy* **228**, 1091–1110.
URL: *https://doi.org/10.1016/j.apenergy.2018.06.137*

Chou, J.-S. & Tran, D.-S. (2018), 'Forecasting energy consumption time series using machine learning techniques based on usage patterns of residential householders', *Energy* **165**, 709–726.
URL: *https://doi.org/10.1016/j.energy.2018.09.144*

Clark, L., Lou, D., Michelle, D., T., G. & C., G. (2018), 'Day-ahead load forecasting using support vector regression machines', *International Journal of Advanced Computer Science and Applications* **9**(3).
URL: *https://doi.org/10.14569/ijacsa.2018.090305*

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

Demirel, O. F., Zaim, S., Caliskan, A. & Ozuyar, P. G. (2012), Forecasting natural gas consumption in istanbul using neural networks and multivariate time series methods.

Du, P., Wang, J., Yang, W. & Niu, T. (2018), 'Multi-step ahead forecasting in electrical power system using a hybrid forecasting system', *Renewable Energy* **122**, 533–550.
URL: *https://doi.org/10.1016/j.renene.2018.01.113*

Elamin, N. & Fukushige, M. (2018), 'Modeling and forecasting hourly electricity demand by SARIMAX with interactions', *Energy* .
URL: *https://doi.org/10.1016/j.energy.2018.09.157*

Ervural, B. C., Beyca, O. F. & Zaim, S. (2016), 'Model estimation of ARMA using genetic algorithms: A case study of forecasting natural gas consumption', *Procedia - Social and Behavioral Sciences* **235**, 537–545.
URL: *https://doi.org/10.1016/j.sbspro.2016.11.066*

Eshel, G. (n.d.), 'The yule walker equations for the ar coefficients'.

Fagiani, M., Squartini, S., Gabrielli, L., Spinsante, S. & Piazza, F. (2015), 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, pp. 185–194.
URL: *https://doi.org/10.1007/978-3-319-18164-6_18*

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

Farfar, K. E. & Khadir, M. T. (2018), '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*

Farfar, K. E., Khadir, M. T. & Laib, O. (2015), Comparison of serial and parallel approaches using artificial neural networks for algerian short term load forecasting, *in* 'Third International Conference on Advances in Computing, Electronics and Electrical Technology - CEET 2015', Institute of Research Engineers and Doctors.
URL: *https://doi.org/10.15224/978-1-63248-056-9-23*

Fayaz, M. & Kim, D. (2018), 'A prediction methodology of energy consumption based on deep extreme learning machine and comparative analysis in residential buildings', *Electronics* **7**(10), 222.
URL: *https://doi.org/10.3390/electronics7100222*

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

Franco, A. & Fantozzi, F. (2015), 'Analysis and clustering of natural gas consumption data for thermal energy use forecasting', *Journal of Physics: Conference Series* **655**, 1–10.

Friedman, J. H. (1991), 'Multivariate adaptive regression splines', *The Annals of Statistics* **19**(1), 1–67.
URL: *https://doi.org/10.1214/aos/1176347963*

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

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

Gers, F. A., Schraudolph, N. N. & Schmidhuber, J. (2002), 'Learning precise timing with LSTM recurrent networks', *jmlr* **3**, 115–143.
URL: *http://nic.schraudolph.org/pubs/GerSchSch02.pdf*

Ghadimi, N., Akbarimajd, A., Shayeghi, H. & Abedinia, O. (2017), 'A new prediction model based on multi-block forecast engine in smart grid', *Journal*

of Ambient Intelligence and Humanized Computing **9**(6), 1873–1888.
URL: *https://doi.org/10.1007/s12652-017-0648-4*

Gooijer, J. G. D. & Kumar, K. (1992), 'Some recent developments in nonlinear time series modelling, testing, and forecasting', *International Journal of Forecasting* **8**(2), 135–156.
URL: *https://doi.org/10.1016/0169-2070(92)90115-p*

Graves, A., Fernández, S. & Schmidhuber, J. (2005), 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, pp. 799–804.

Hebb, D. O. (1962), *The organization of behavior: a neuropsychological theory*, Science Editions.

Hensman, J., Lawrence, N. D. & Rattray, M. (2013), 'Hierarchical bayesian modelling of gene expression time series across irregularly sampled replicates and clusters', *BMC Bioinformatics* **14**(1), 252.

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

Hochreiter, S., Bengio, Y. & Frasconi, P. (2001), 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.

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

Hong, T., Gui, M., Baran, M. E. & Willis, H. L. (2010), Modeling and forecasting hourly electric load by multiple linear regression with interactions, *in* 'IEEE PES General Meeting', IEEE.
URL: *https://doi.org/10.1109/pes.2010.5589959*

Hong, W.-C. (2013), *Intelligent Energy Demand Forecasting*, Vol. 10.

Hooshmand, R.-A., Amooshahi, H. & Parastegari, M. (2013), 'A hybrid intelligent algorithm based short-term load forecasting approach', *International Journal of Electrical Power & Energy Systems* **45**(1), 313–324.
URL: *https://doi.org/10.1016/j.ijepes.2012.09.002*

Hornik, K., Stinchcombe, M. & White, H. (1989), 'Multilayer feedforward networks are universal approximators', *Neural networks* **2**(5), 359–366.

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

Jang, J.-S. (1993), 'ANFIS: adaptive-network-based fuzzy inference system', *IEEE Transactions on Systems, Man, and Cybernetics* **23**(3), 665–685.
URL: *https://doi.org/10.1109/21.256541*

Jang, J.-S. R. et al. (1991), Fuzzy modeling using generalized neural networks and kalman filter algorithm., *in* 'AAAI', Vol. 91, pp. 762–767.

Jin, M., Zhou, X., Zhang, Z. M. & Tentzeris, M. M. (2012), 'Short-term power load forecasting using grey correlation contest modeling', *Expert Systems with Applications* **39**(1), 773–779.
URL: *https://doi.org/10.1016/j.eswa.2011.07.072*

Jovanović, R. Ž., Sretenović, A. A. & Živković, B. D. (2015), 'Ensemble of various neural networks for prediction of heating energy consumption', *Energy and Buildings* **94**, 189–199.
URL: *https://doi.org/10.1016/j.enbuild.2015.02.052*

Ju-Long, D. (1982), 'Control problems of grey systems', *Systems & Control Letters* **1**(5), 288–294.
URL: *https://doi.org/10.1016/s0167-6911(82)80025-x*

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

Kermanshahi, B. & Iwamiya, H. (2002), 'Up to year 2020 load forecasting using neural nets', *International Journal of Electrical Power & Energy Systems* **24**(9), 789–797.
URL: *https://doi.org/10.1016/s0142-0615(01)00086-2*

Kim, M., Choi, W., Jeon, Y. & Liu, L. (2019), 'A hybrid neural network model for power demand forecasting', *Energies* **12**(5), 931.
URL: *https://doi.org/10.3390/en12050931*

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

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

Laib, O., Khadir, M. T. & Mihaylova, L. (2018), 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)', pp. 55–61.

Laib, O., Khadir, M. T. & Mihaylova, L. (2019), 'Toward efficient energy systems based on natural gas consumption prediction with LSTM recurrent neural networks', *Energy* **177**, 530–542.
URL: *https://doi.org/10.1016/j.energy.2019.04.075*

Lauret, P., David, M. & Calogine, D. (2012), 'Nonlinear models for short-time load forecasting', *Energy Procedia* **14**, 1404–1409.
URL: *https://doi.org/10.1016/j.egypro.2011.12.1109*

Lecun, Y., Bottou, L., Bengio, Y. & Haffner, P. (1998), 'Gradient-based learning applied to document recognition', *Proceedings of the IEEE* **86**(11), 2278–2324.
URL: *https://doi.org/10.1109/5.726791*

Lee, K., Cha, Y. & Ku, C. (n.d.), 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*

Li, G.-D., Wang, C.-H., Masuda, S. & Nagai, M. (2011), 'A research on short term load forecasting problem applying improved grey dynamic model', *International Journal of Electrical Power & Energy Systems* **33**(4), 809–816.
URL: *https://doi.org/10.1016/j.ijepes.2010.11.005*

Liu, H., Liu, D., Zheng, G., Liang, Y. & Ni, Y. (2004), Research on natural gas load forecasting based on support vector regression, *in* 'Fifth World Congress on Intelligent Control and Automation (IEEE Cat. No.04EX788)', IEEE.
URL: *https://doi.org/10.1109/wcica.2004.1343263*

Liu, N., Babushkin, V. & Afshari, A. (2014), 'Short-term forecasting of temperature driven electricity load using time series and neural network model', *Journal of Clean Energy Technologies* **2**(4), 327–331.
URL: *https://doi.org/10.7763/jocet.2014.v2.149*

Lourenço, J. M. & Santos, P. J. (2012), 'Short-term load forecasting using a gaussian process model: The influence of a derivative term in the input regressor', *Int. Dec. Tech.* **6**(4), 273–281.
URL: *http://dx.doi.org/10.3233/IDT-2012-0143*

Lourenco, J. & Santos, P. (2010), Short term load forecasting using gaussian process models, Technical Report 5, Systems Engineering and Computers, INESC - Coimbra.

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

Ma, Y. & Li, Y. (2010), 'Analysis of the supply-demand status of china's natural gas to 2020', *Petroleum Science* **7**(1), 132–135.
URL: *https://doi.org/10.1007/s12182-010-0017-9*

Mamlook, R., Badran, O. & Abdulhadi, E. (2009), 'A fuzzy inference model for short-term load forecasting', *Energy Policy* **37**(4), 1239–1248.
URL: *https://doi.org/10.1016/j.enpol.2008.10.051*

Maritz, J., Lubbe, F. & Lagrange, L. (2018), 'A practical guide to gaussian process regression for energy measurement and verification within the bayesian framework', *Energies* **11**(4), 935.

Mbamalu, G. & El-Hawary, M. (1993), 'Load forecasting via suboptimal seasonal autoregressive models and iteratively reweighted least squares estimation', *IEEE Transactions on Power Systems* **8**(1), 343–348.

McCulloch, W. S. & Pitts, W. (1943), 'A logical calculus of the ideas immanent in nervous activity', *The Bulletin of Mathematical Biophysics* **5**(4), 115–133.
URL: *https://doi.org/10.1007/bf02478259*

Minsky, M. & Papert, S. A. (2017), *Perceptrons: An introduction to computational geometry*, MIT press.
ΩÖzmen et al.

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

Panapakidis, I. P. & Dagoumas, A. S. (2017), 'Day-ahead natural gas demand forecasting based on the combination of wavelet transform and ANFIS/genetic algorithm/neural network model', *Energy* **118**, 231–245.
URL: *https://doi.org/10.1016/j.energy.2016.12.033*

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

Plaut, D. C., Nowlan, S. J. & Hinton, G. E. (1986), 'Experiments on learning by back propagation.'.

Potočnik, P., Soldo, B., Šimunović, G., Šarić, T., Jeromen, A. & Govekar, E. (2014), 'Comparison of static and adaptive models for short-term residential

natural gas forecasting in croatia', *Applied Energy* **129**, 94–103.
URL: *https://doi.org/10.1016/j.apenergy.2014.04.102*

Potočnik, P. & Govekar, E. (2010), Practical results of forecasting for the natural gas market, *in* 'Natural Gas', Sciyo, New York, chapter 16, pp. 371–392.

Potočnik, P., Govekar, E. & Grabec, I. (2008), Building forecasting applications for natural gas market, *in* N. David & T. Michel, eds, 'Natural gas research progress', Nova Science Publishers, New York, pp. 505–530.

Prakash, A., Xu, S., Rajagopal, R. & Noh, H. (2018), 'Robust building energy load forecasting using physically-based kernel models', *Energies* **11**(4), 862.
URL: *https://doi.org/10.3390/en11040862*

Raghavendra, N. S. & Deka, P. C. (2016), *Multistep Ahead Groundwater Level Time-Series Forecasting Using Gaussian Process Regression and ANFIS*, Springer India, New Delhi, pp. 289–302.

Rezaie-Balf, M., Maleki, N., Kim, S., Ashrafian, A., Babaie-Miri, F., Kim, N. W., Chung, I.-M. & Alaghmand, S. (2019), 'Forecasting daily solar radiation using CEEMDAN decomposition-based MARS model trained by crow search algorithm', *Energies* **12**(8), 1416.
URL: *https://doi.org/10.3390/en12081416*

Rosenblatt, F. (1957), *The Perceptron, a Perceiving and Recognizing Automaton Project Para*, Report: Cornell Aeronautical Laboratory, Cornell Aeronautical Laboratory.
URL: *https://books.google.dz/books?id=P_XGPgAACAAJ*

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

Ruiz-Abellón, M., Gabaldón, A. & Guillamón, A. (2018), 'Load forecasting for a campus university using ensemble methods based on regression trees', *Energies* **11**(8), 2038.
URL: *https://doi.org/10.3390/en11082038*

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

Rumelhart, D. E., Hinton, G. E. & Williams, R. J. (1985), Learning internal representations by error propagation, Technical report, California Univ San Diego La Jolla Inst for Cognitive Science.

Rumelhart, D. E., Hinton, G. E. & Williams, R. J. (1986), Parallel distributed processing: Explorations in the microstructure of cognition, vol. 1, MIT Press, Cambridge, MA, USA, chapter Learning Internal Representations by Error Propagation, pp. 318–362.
URL: *http://dl.acm.org/citation.cfm?id=104279.104293*

Rumelhart, D. E., Hinton, G. E., Williams, R. J. et al. (1988), 'Learning representations by back-propagating errors', *Cognitive modeling* **5**(3), 1.

Saber, A. Y. & Alam, A. K. M. R. (2017), Short term load forecasting using multiple linear regression for big data, *in* '2017 IEEE Symposium Series on Computational Intelligence (SSCI)', IEEE.
URL: *https://doi.org/10.1109/ssci.2017.8285261*

Schraudolph, N. N. (2002), 'Fast curvature matrix-vector products for second-order gradient descent', *Neural Computation* **14**(7), 1723–1738.
URL: *https://doi.org/10.1162/08997660260028683*

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

Sharma, S. (1995), *Applied Multivariate Techniques*, Wiley.
URL: *https://books.google.co.uk/books?id=6iURRAAACAAJ*

Shchetinin, E. Y. (2018), Cluster-based energy consumption forecasting in smart grids, *in* 'Developments in Language Theory', Springer International Publishing, pp. 445–456.
URL: *https://doi.org/10.1007/978-3-319-99447-5_38*

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

Suykens, J., Lemmerling, P., Favoreel, W., de Moor, B., Crepel, M. & Briol, P. (1996), 'Modelling the belgian gas consumption using neural networks', *Neural Processing Letters* **4**(3), 157–166.
URL: *https://doi.org/10.1007/bf00426024*

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

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

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

Taylor, J. W. (2010), 'Triple seasonal methods for short-term electricity demand forecasting', *European Journal of Operational Research* **204**(1), 139–152.
URL: *https://doi.org/10.1016/j.ejor.2009.10.003*

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

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

Tonković, Z., Zekič-Sušac, M. & Somolanji, M. (2009), 'Predicting natural gas consumption by neural networks', *Tehnički vjesnik* **16**(3), 51–61.

Vapnik, V. N. (2000), *The Nature of Statistical Learning Theory*, Springer New York.
URL: *https://doi.org/10.1007/978-1-4757-3264-1*

Wang, Z.-X., Li, Q. & Pei, L.-L. (2018), 'A seasonal GM(1, 1) model for forecasting the electricity consumption of the primary economic sectors', *Energy* **154**, 522–534.
URL: *https://doi.org/10.1016/j.energy.2018.04.155*

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

Yang, Y., Li, S., Li, W. & Qu, M. (2018), 'Power load probability density forecasting using gaussian process quantile regression', *Applied Energy* **213**, 499–509.
URL: *https://doi.org/10.1016/j.apenergy.2017.11.035*

Yao, A. W., Chi, S. & Chen, J. (2003), 'An improved grey-based approach for electricity demand forecasting', *Electric Power Systems Research* **67**(3), 217–224.
URL: *https://doi.org/10.1016/s0378-7796(03)00112-3*

Yu, F. & Xu, X. (2014), 'A short-term load forecasting model of natural gas based on optimized genetic algorithm and improved BP neural network', *Applied Energy* **134**, 102–113.
URL: *https://doi.org/10.1016/j.apenergy.2014.07.104*

Zareipour, H., Bhattacharya, K. & Canizares, C. (2006), Forecasting the hourly ontario energy price by multivariate adaptive regression splines, *in* '2006 IEEE Power Engineering Society General Meeting', IEEE.
URL: *https://doi.org/10.1109/pes.2006.1709474*

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

Zhou, Q., Wang, S., Xu, X. & Xiao, F. (2008), 'A grey-box model of next-day building thermal load prediction for energy-efficient control', *International Journal of Energy Research* **32**(15), 1418–1431.
URL: *https://doi.org/10.1002/er.1458*

Zhuang, J., Chen, Y., Shi, X. & Wei, D. (2015), 'Building cooling load prediction based on time series method and neural networks', *International Journal of Grid and Distributed Computing* **8**(4), 105–114.
URL: *https://doi.org/10.14257/ijgdc.2015.8.4.10*