Mathematical models for Algerian long medium and short load forecasting: prevelec-dz and prevos-dz

Khadir M. T. [1], Boughrira A.[2]

[1]LabGED, Computer Science department,
Badji Mokhtar University, P.O. Box 12, 23005, Annaba, Algeria
[2]Centre Régional de Conduite Annaba S.TOS/DCS/CRCAN, 23000, Alegria.

Article Info

Article history:

Received: 20/01/2025
Revised: 28/03/2025
Accepted: 30/04/2025

Keywords:

Forecasting, Mathematical Modeling, Artificial Intelligence, Machine learning

ABSTRACT

Electricity demand plays a critical role in the development of growing economies. To meet daily electricity needs and support demand, utility companies and operators must utilize accurate and efficient forecasting tools based on strong linear and nonlinear mathematical models. This paper will present briefly two tools developed over the course of a 10 years collaboration period between the academic sector represented by the LabGED laboratory and the Algerian national electricity company (SONELGAZ) namely: PREVOS-DZ and PREVELC-DZ respectively for short/medium and long term forecasting. Short term load forecasting aid in both daily production planning where long-term forecasting helps development strategies in terms of planning and construction of production means. The proposed software includes multiple regression methods as well as Artificial Neural Networks (ANNs) as basis for the large number of forecasting models in the software back end part. Forecasting capabilities extend beyond national demand to include regional and special case requirements such as special days, Ramadhan etc. PREVOS-DZ and PREVELEC-DZ are Implemented in C#, ensuring seamless compatibility with Excel spreadsheets, widely used by system operators. The predictive performance of forecasting models was validated using 2018 data, for short term forecasting. Part of the results are presented in this paper.

Corresponding Author: *khadir@labged.net*

1. INTRODUCTION

Electrical grids are intricate, interconnected systems that continue to grow in complexity. Their size and structure are closely tied to the level of development in a given region, the number of customers served, and economic demands. This makes managing them a challenging task, encompassing energy generation, transmission, and distribution. This paper presents two valuable tools for forecasting short medium and long term electrical Algerian load.

Different forecasting scenarios are tailored to specific management needs, categorized based on forecast horizons: short-term (1 hour to 1 week ahead, STLF), medium-term (several weeks to a few years, MTLF), and long-term (more than a year ahead). These forecasts serve distinct purposes: short-term forecasts support hourly electricity production planning, medium-term forecasts aid in managing production units and scheduling maintenance, while long-term forecasts inform the design and construction of future production infrastructure to meet anticipated demand.

I order to develop reliable STLF models, various paradigms and methodologies have been proposed in the literature, generally categorized into two main approaches: On one side, traditional methods include regression models, classical time-series approaches, and Autoregressive Moving Average (ARMA) models. For example, Sargunaraj et al. [1] introduced a two-stage approach using Kalman filtering [2]. In their method, mean load predictions are first generated, followed by real-time error adjustments using the Kalman filter. Kyriakides and Polycarpou [3] reviewed and compared these traditional approaches, providing valuable insights and conclusions.

Artificial Intelligence (AI) and machine learning techniques have recently demonstrated significant efficiency, thanks to advancements in computational power. Artificial Neural Networks (ANNs) [4] have been widely tested and successfully applied in various experiments and practical scenarios [5]. Similarly, other AI paradigms, such as Support Vector Machines (SVMs) [6], have been employed effectively. For instance, in [7], SVMs were used to interpolate load and temperature data. Fuzzy logic techniques have also been applied to handle abnormal situations in load demand [8].

Additionally, numerous hybrid approaches combining AI techniques have been implemented to enhance model efficiency [9] [10] [11]. Recently, Lusis et al. [12] explored the impact of calendar effects, data granularity, and training set length on the accuracy of day-ahead load forecasting for residential customers. Furthermore, deep

©UBMA – 2025

learning architectures have emerged as a promising approach, enabling the integration of influential input data by first extracting optimal features and then performing forecasting [13]. However, this method is currently not feasible for the Algerian market due to a lack of sufficient data.

The paper is organized as follows: a non-exhaustive description of both used approaches for instance: Multiple Linear regression (MLR) and Artificial Neural Networks (ANN) are presented in Section 2 and 3. In the fourth section, the framework of PREVOS-DZ and PREVELEC-DZ are presented with experiments and results shown in Section five. Finally, a conclusion on both technical and cooperation aspects is drawn

2 MULTIPLE LINEAR REGRESSION

2.1 MLR Principles

Regression analysis is a linear method used to identify relationships within a set of input- output data, establishing causality between variables. In the context of electrical load, causality must be determined between exogenous inputs such as temperature, humidity, past electrical load, and the future predicted load. The development of an appropriate model depends entirely on the collection of accurate and relevant data from the key variables of interest. Regression is then applied to estimate the quantitative influence of causal variables on the dependent variable they affect. Historically, regression techniques have been pivotal in the field of economic statistics, commonly referred to as "Econometrics."

Multiple linear regression is specifically employed in PREVOS-DZ and PREVELEC-DZ models to examine the impact of multiple independent variables on a single dependent variable. A multiple linear regression model with more than one explanatory variable is expressed mathematically as shown in equation (1) :

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p \quad (1)$$

In this context, y represents the output variable, βi are the regression parameters ($i = 1, 2,..., p$), xi are the input variables
($i = 1, 2, ... p$), and ε denotes the random error term. Exponential regression is particularly important when the trend of a time series exhibits exponential behavior. This is often observed in load growth patterns in developing countries, including Algeria. It is also applicable when an exponential curve provides the best fit for data that does not change linearly over time [14]. Exponential regression models are primarily utilized for long-term predictions in the presence of a clear trend and are expressed as follows :

$$y = \beta_0 \times \beta_1^{x_1} \times \beta_2^{x_2} \times ... \times \beta_q^{x_q} \quad (2)$$

Where βi are parameters for the n independent variables.

2.2 MLR models for Load forecasting

In Multiple Linear Regression (MLR) models, the inputs consist of lagged load values de- rived from autocorrelation analysis, along with nonlinearly related exogenous variables in this case, hourly temperature estimates.

The electric load L at hour t is modeled using several selected past load values in addition to temperature estimates. These lagged values include L_{t-1}, L_{t-2}, L_{t-3}, L_{t-24}, and L_{t-168}, which represent, respectively, the loads from the previous three hours, the same hour on the previous day, and the same hour on the previous week. Additionally, the day type context is encoded using one-hot encoding for each day of the week (DoW) and a separate input for special days (ToD). This results in a total of 11 input variables, as described in equation (**??**).

$$L_t = a_1 L_{t-1} + a_2 L_{t-2} + a_3 L_{t-3} + a_4 L_{t-24} + a_5 L_{t-168}$$
$$+ b_1 T_{t-1} + b_2 T_{t-2} + b_3 T_{t-3} + b_4 T_{t-24} + DoW + ToD \quad (3)$$

For long term MLR load forecasting the same strategy is followed, however with different exogenous variables. Indeed, for long term forecasting, temperatures are no longer a valuable external factor, whereas the LOF (Lodging occupation Factor) derived from population and constructed and finalized lodging units. Note that the most important number of constructed flats or lodging unit are state construction within a large national construction program. This gives the following MLR model:

$$L_t = a_1 L_{t-1} + a_2 L_{t-2} + a_3 L_{t-3} + ... + a_n L_{t-n} + b_1 LOF_{t+1} + b_2 LOF_{t+2} + ... + b_3 LOF_{t+m} \quad (4)$$

Here t is expressed in years. Other external variables may be integrated such as *GDP* and other economic factors.

©UBMA – 2025

3 ARTIFICIAL NEURAL NETWORKS

Artificial Neural Network (ANN) is a data processing approach inspired by biological nervous systems, mirroring the way the brain processes information. McCulloch and Pitts [15] pro- posed the first mathematical and computational model of biological neurons in 1943, introducing formal functions for neural computations.

The characteristics of ANN information processing make it a powerful computational technique capable of learning from examples and generalizing to cases not presented during training. ANNs are composed of a set of neurons, where each neuron operates autonomously and under- goes two phases:

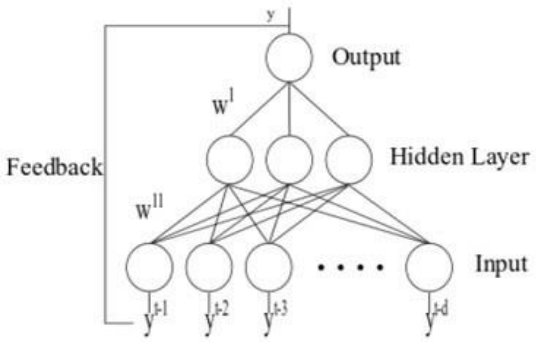


FIGURE 1 – Feed Forward MPC Neural Network.

1. The computation of the weighted sum of the inputs, given by the following equation:

$$a = \sum w_i \times e_i \qquad (4)$$

where w_i and e_i represent the weight and input of neuron i, respectively.

2. A transfer function (or activation function) that calculates the neuron's state based on the previously computed sum.

The primary advantages of ANNs stem from the scalability of connections between neu- rons, which are determined through a learning process. Once trained, ANNs can solve nonlinear problems and approximate any nonlinear function.

Various network topologies have been proposed in the literature based on connections bet- ween neurons, including Feed Forward [16], Recurrent Neural Network [17], and Self-Organizing Map (SOM) [18]. In this study, a Non-Linear Auto Regressive Neural Network with Exogenous Inputs (NARX) Multi-Layer Perceptron (MLP) is utilized (see Fig. 1). This model is a recurrent neural network inspired by the linear Auto Regressive with Exogenous Inputs (ARX) model. The NARX model consists of an input layer, a hidden layer, and a single-neuron output layer, similar to a Feed Forward neural network, except it includes a feedback loop from the output layer to the input layer (illustrated in **??**). The output is defined by the following equation:

$$\hat{y}_t = \sum_{i=1}^{d} W_i y_{t-1} + \sum_{j=1}^{m} B_i T_{t-1} + WoD + ToD \qquad (5)$$

where W_i and B_j are the weights to be estimated, d is the autoregressive delay and m the exogenous input delay on the exogenous variable T being here the temperature, WoD and ToD the two hot encoded inputs described earlier.

In ANN, instead of using a standard linear regression, the parameters are estimated using a learning algorithm e.g. the Retro propagation algorithm, the Levenberg Marquardt algorithm, or resilient propagation, etc. The added feature of the recurrent model, is the inclusion of past model outputs within the input vector. This ensures multi-step ahead predictions and allows the use of more nonlinear model types such as Nonlinear Box Jenkins.

©UBMA – 2025

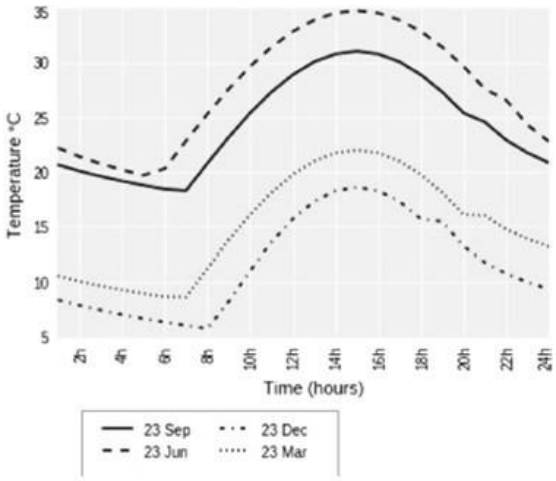


FIGURE 2 – Temperature profile estimation results

4. AVAILLABLE DATA

4.1 Exogenous Inputs (Temperature)

For the purpose of this work, daily hourly temperature loads are estimated based on the minimum ($T_{\min}$) and maximum ($T_{\max}$) daily temperatures provided by SONELGAZ.
The construction of daily temperature loads is based on an idealized temperature curve using Linvill's equations [**?**], which assume that the maximum daily temperature occurs 2 hours after solar noon and that the temperature cycle after sunrise follows a sine curve:

$$T_{ih} = T_{\min_i} + (T_{\max_i} - T_{\min_i}) \sin\left(\frac{\pi t_a}{DL_i}\right) \tag{6}$$

where Tih is the temperature at hour h, $T_{\max i}$ and $T_{\min i}$ are the maximum and minimum temperatures for day i, DL_i is the number of hours from sunrise to sunset, and t_a is the time in hours ($t_a \geq 0$) counted from sunrise. After sunset, the temperature is assumed to follow a logarithmic decline:

$$T_h = T_s - (T_s - T_{\min_{i+1}}) \ln(t_b) \tag{7}$$

where T_S is the sunset temperature determined from Eq. (**??**), and $T_{\min i+1}$ is the minimum temperature of the following day ($i + 1$), which is replaced by the minimum temperature of the same day for hours before sunrise. The time t_b represents the time in hours after sunset plus 1 hour ($t_b \geq 1$). Linvill's equations require input parameters such as sunset time, sunrise time, and day length, which are computed using the procedures described in [**?**, **?**] based on geographic latitude.

An example of daily temperature profiles generated using this method is shown in Fig. 2. Seasonal variations are clearly noticeable when comparing minimum and maximum temperatures. Minimum temperatures consistently occur at sunrise due to two mechanisms:

1.The Earth's surface stops absorbing solar radiation after sunset, ceasing to heat up.

2. Conduction and convection transport heat energy upward into the atmosphere, replacing warm surface air with cooler air from above due to atmospheric mixing [**?**].

The results obtained demonstrate a better approximation of daily temperature loads compared to linear and cubic interpolations. The model effectively captures the rise and fall of temperatures as minimum and maximum values are reached. Moreover, this estimation of temperature profiles improves model performance by providing a more accurate representation of exogenous inputs [**?**].

4.1.1 Load data

The load data used in this study represents the Algerian quarterly and hourly electric load over a four-year period (2013-2017). Data from earlier periods were excluded to avoid negatively affecting the model's learning process and the search for optimal weights and biases. This decision was based on significant changes in load behavior starting in 2013, driven by the wides-pread adoption of air conditioners. Consequently, data from 2013 onwards were chosen as they more accurately reflect the current load patterns.

To validate and evaluate the performance improvements achieved by the Artificial Neural Network (ANN) (see Section III), a traditional time-series analysis technique, specifically multiple linear regression (see Section II), is used as a benchmark.

©UBMA – 2025

The data are divided into three subsets: 60% for the training set to create the model, 20% for the validation set to fine-tune the model and optimize its parameters, and the optimization of hyper-parameters to avoid over fitting and the last data set for testing (20%),

5 RESULTS AND DISCUSSION

5.1 Developed models for short, mid and long term forecasting

A total of 216 models were developed for the PREVOS-DZ tool. These models were collaboratively validated by the authors from both academic and industrial institutions. They address a wide range of prediction requirements and are classified based on the following criteria:

Model Type: Differentiating by day types, special periods, and specific days. Region Models: National-level or regional-level predictions. Seasonal and Monthly Models: Adjusted for seasonal variations and, in some cases, monthly trends. Prediction Horizons: Covering quarterly, hourly, and weekly forecasting needs. Tables 1 provide an overview of the developed models, categorized by type for Artificial Neural Network (ANN) models. The same types of models are also developed for LRM types.

As for the long term, not less than 105 models where implemented National, regional and economic sectors previsions for all three level of voltages namely : Low Voltage (LV), Medium Voltage (MV) and High Voltage (HV) as shown in table 2.

5.2 Model accuracy for different models

In order to assess model quality, Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) are used to obtain comprehensive and objective evaluation for the performance of the proposed approach. Overall results in terms of MAPE, are presented in Table **??** and graphical evaluation for a 24hour forecast is given for random days in Figures 4, showing the prediction using a quarterly overall national model for a summer day (2nd June 2016). It can be seen that the daily seasonality is well respected and that the ANN model performs better than the MLR one.

ANN model still outperforms the MLR one in most cases as SLTM is highly nonlinear. Although it is the case for most model types, and for the average evaluation, for some specific days the MLR model might outperforms the ANN one.

Table 3 summarizes the efficiency evaluation for ANN short term load forecasting models.

It is clearly seen that the MAPE error falls whithin the margin error rate of ±2%.

TABLE 1 – Overview of Regional Load Models

Type	**National**	**Annaba**	**Setif**	**Oran**	**Alger**	**Sud**	**Adrar**
	15m **15m** **1H**	**1H** **1H**	**15m** **15m**	**1H** **1H**	**1H** **15m**	**1H** **1H**	**15m** **1H** **15m**
Days (Sat, WD, WE)	3 3	3 3	3 3	3 3	3 3	3 3	3 3
Season	4 4	4 4	4 4	4 4	4 2	4 2	4 2
Month	12 0	0 0	0 0	0 0	0 0	0 0	0 0
Ramadan	1 1	1 1	1 1	1 1	1 1	1 1	1 1
Global	1 1	1 1	1 1	1 1	1 1	1 1	1 1
Total/Region	21 9	9 7	9 7	9 7	9 7	9 7	9 7
Total RLM : 112							

TABLE 2 – Overview of Network and Voltage Level Previsions models

Electrical Network	**Voltage Level Previsions**	**Sectorial Previsions**
RIN	LV : 5 (ESA), 5 (DC) MV: 6 (ESA), 5 (DC) HV : 5 (DC)	5 (DC) 6 (ESA)
RIS-PIAT	2	
Scenarios	3 3 3	3 3
Sub-Total	21 36	33 15
Total number of RLM networks	105	

©UBMA – 2025

TABLE 3 – Performance of ANN Models for 1 Hour and 24 Hours Ahead Predictions

ANN Model	1 Hour Ahead		24 Hours Ahead	
	MAE (MWh)	MAPE (%)	MAE (MWh)	MAPE (%)
Global Model (1h)	126.18	1.58%	466.74	6.52%
Seasonal Models (1h)	118.08	1.52%	241.10	3.14%
Day Type Models (1h)	126.20	1.62%	210.39	2.61%
Monthly Models (1h)	122.18	1.58%	203.52	2.59%
Ramadan Model (1h)	122.81	1.49%	273.92	3.21%
Global Model (15mn)	109.16	1.43%	424.77	6.36%
Seasonal Models (15mn)	116.63	1.53%	288.01	3.69%
Day Type Models (15mn)	136.59	1.78%	472.76	6.88%

TABLE 4 – Comparison of Real and Predicted Voltage Levels for 2011

Voltage	Real 2011	Predicted 2011
Low Voltage	0.98215	0.99765
Mid Voltage	0.88591	0.89658
High Voltage	0.91952	0.92188

Concerning Long term Load forecasting the results of the RLM model for the national Case is given in Table 4 and displayed in the prediction figure taken from the PREVELEC-DZ software**??**.

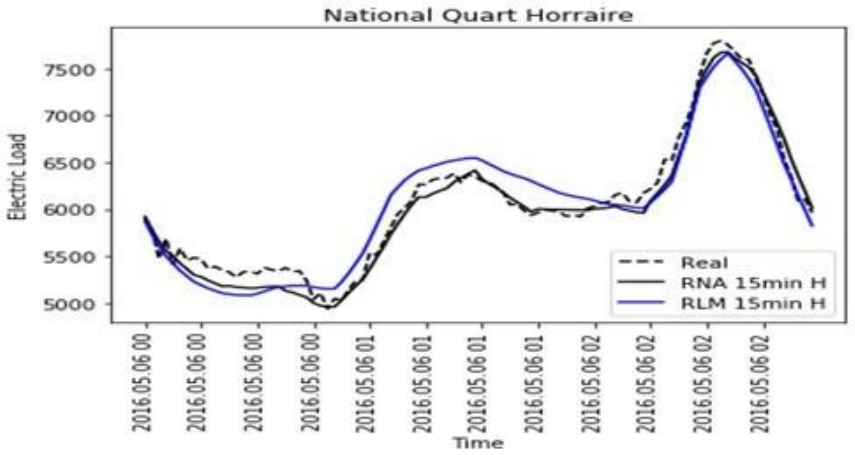


FIGURE 3 – National ANN global STLF model performances for (2nd June 2016)

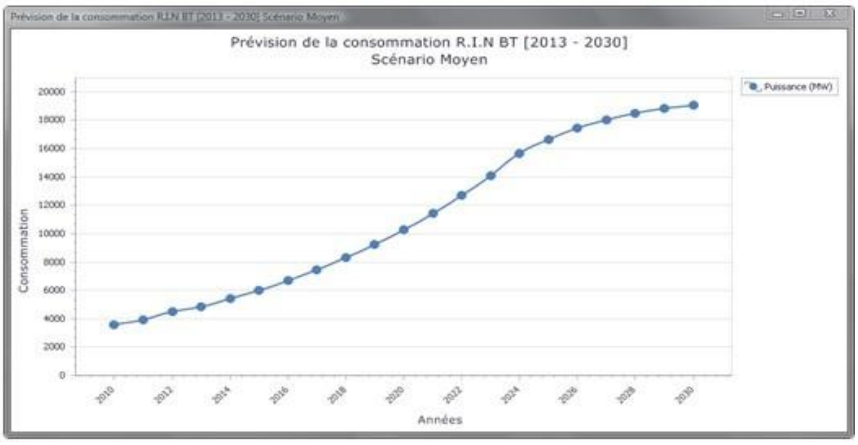


FIGURE 4 – National RLM global LTLF model performances

©UBMA – 2025

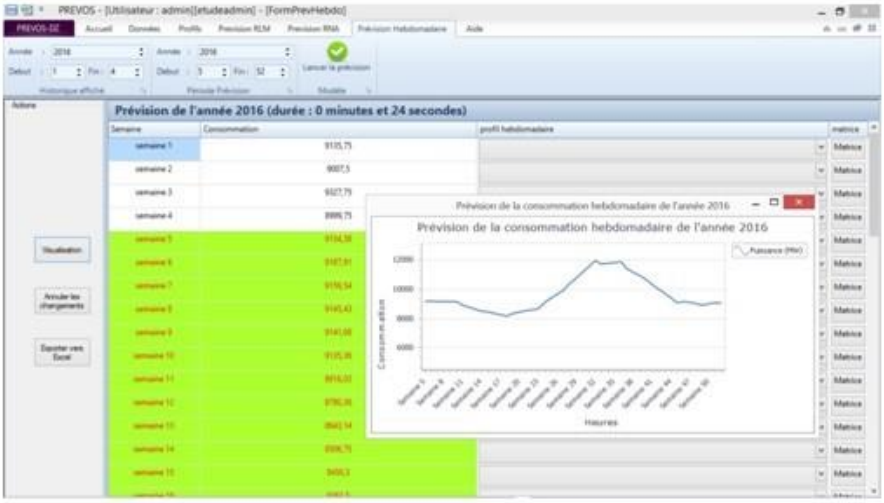


FIGURE 5 – PREVOS-DZ GUI

5.3 PREVELEC-DZ and PREVOS-DZ Graphical user interface (GUI) examples

All the developed models Tables 2 and 1, as well as other functionalities and data are re- grouped to constitute the prevision software's PREVOS-DZ and PREVELEC-DZ. Fig. **??** shows the main windows performing a SLTF prevision based on previous data and the corresponding results [19, 20].
Fig. **??** shows the GUI for PREVELEC-DZ, and an example for LTLF showing the tabs for different scenario's and voltage levels (LV (BT), MV (MT) and HV (HT)).

6 CONCLUSIONS

Artificial Neural Networks (ANNs), particularly the Multi-Layer Perceptron (MLP), have demonstrated their effectiveness as powerful tools for modeling and forecasting time series, as evidenced in this study on predicting Algerian electrical load. ANNs consistently outperformed traditional Multiple Linear Regression (MLR) models in most scenarios. However, certain special days exhibited significant prediction inaccuracies, with Mean Absolute Percentage Errors (MAPE) exceeding 9%. These days correspond to unique load profiles that deviate considerably from the average load patterns for their respective periods, day types, or regions.

In conclusion, the development of a comprehensive tool that integrates accurate prediction models for various regions, seasons, and day types, alongside midterm weekly prediction models, profile generation tools, and MLR model creation capabilities, represents a significant asset for our partner, the System Operator (OS). This tool not only improves the accuracy of load predictions compared to previous methods but also enhances decision-making processes. Furthermore, the strong collaboration with a university laboratory located near the OS provides a foundation for ongoing development, enabling the introduction of innovative solutions to address yet unresolved challenges.

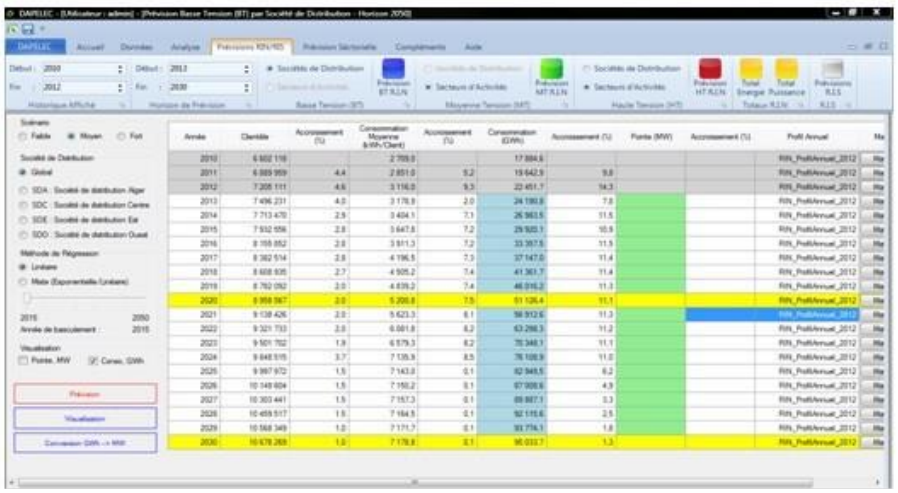


FIGURE 6 – PREVELEC-DZ GUI

©UBMA – 2025

REFERENCES

[1] S. Sargunaraj, DS. Gupta and S. Devi Short-term load forecasting for demand side managment. IEE Proc Gener Transm Distrib 144 :68-74 1997.

[2] P. Zarchan, H. Musoff. Fundamentals of Kalman filtering: a practical approach. American Institute of Aeronautics and Astronautics, Virginia, 2005.

[3] E. Kyriakides and M. Polycarpou, Short term electric load forecasting: a tutorial. In: Trends in neural computation. Springer, Berlin, pp 391-418, 2007.

[4] S.O., Haykin, Neural networks and learning machines. Pearson, New York, 2009. 16 :44- 55, 2001.

[5] H.S., Hippert, C.E., Pedreira and R.C., Souza, Neural networks for short-term load forecasting: a review and evaluation. IEEE Trans Power Syst 16 :44-55, 2001.

[6] V.N. Vapnik, The nature of statistical learning theory. Springer, New York, 1995.

[7] L. Yuancheng, F. Tingjian and Y. Erkeng, Study of support vector machines for short-term load forecasting. Proc Chin Soc Electr Eng 23 :55-59, 2003.

[8] R. Mamlook, O. Badran and E. Abdulhadi, A fuzzy inference model for short-term load forecasting. Energy Policy 37 :1239-1248, 2009.

[9] S. Fan and L. Chen, Short-term load forecasting based on an adaptive hybrid method. IEEE Trans Power Syst 21 :392-401, 2006.

[10] N. Amjady and F. Keynia, Short-term load forecasting of power systems by combination of wavelet transform and neuroevolutionary algorithm. Energy 34 :46-57, 2009.

[11] A. El Desouky and M., Elkateb, Hybrid adaptive techniques for electric-load forecast using ANN and ARIMA. IEE Proc Gener Transm Distrib 147 :213-217, 2000.

[12] P. Lusis, K.R., Khalilpour, L. Andrew and A. Liebman., Short-term residential load forecasting: Impact of calendar effects and forecast granularity, Applied Energy, 205(1), 654-669, 2017.

[13] W. He, Deep neural network based load forecast, Comput. Model. New Technology 18 (3), 258-262, 2014.

[14] Q. Zhang and T. Liu, A fuzzy rules and wavelet neural network method for mid-long term electric load forecasting, Second International Conference on Computer and Network Technology, Bangkok, pp. 442-446, Apr. 2010.

[15] W. S. McCulloch and W. Pitts, A logical calculus of the ideas immanent in nervous activity, The bulletin of mathematical biophysics, Vol. 5, No. 4, pp. 115-133, Dec. 1943.

[16] Z. Tang and P. A. Fishwick, Feedforward neural nets as models for time series forecasting, ORSA Journal on Computing, Vol. 5, No. 4, pp. 374-385, Nov. 1993.

[17] G. Dorffner, Neural networks for time series processing, Neural Network World, Vol. 6, No. 4, pp. 447-468, 1996.

[18] T. Kohonen, The self-organizing map, Neurocomputing, Vol. 21, No. 1, pp. 1-6, Nov. 1998.

[19] Khadir, M. T., Farah, N., Farfar, K., Bendaoud, N., Ameyoud, A., & Tenzer, N. (2019, February). PREVOS-DZ: A Short-Mid Term Algerian Electric Load Forecasting Software. In 2019 Algerian Large Electrical Network Conference (CAGRE) Algiers, (pp. 1-6). IEEE.

[20] Khadir, M. T., Chouireb, L., Attouchi, N., Farah, N., Nezzar, M.R., & Boulaaress, R. (2015). PREVELEC-DZ Mid and Long Term Algerian Electric Load Forecasting Soft- ware. In 2015 Algerian Large Electrical Network Conference (CAGRE), Algiers.

©UBMA – 2025