A Methodology of Power Demand Prediction

Corina MARTINEAC, Oana ONET, Simona ARDELEAN,
Claudiu VERMEŞAN, Mihai HOPÎRTEAN, Tudor VESA
Technical University of Cluj-Napoca,FDEE Electrica Distribuţie Transilvania Nord, Utrecht University
str.C.Daicoviciu nr.15, RO-400020 Cluj-Napoca, str. I. Măcelaru nr 28 A, RO- 40038 Cluj-Napoca,
Heidelberglaan 2,3584CS Utrecht, THE NETHERLANDS
Corina.Martineac@eps.utcluj.ro, onetz_oana@yahoo.com, Simona.Ardelean@eps.utcluj.ro,
Claudiu.Vermesan@tnd.electrica.ro, M.I.Hopirtean@students.uu.nl, Tudor.Vesa@staff.utcluj.ro

***Abstract* — The end use consumers of energy conversion, transport and distribution determine market demand of energy. This should be balanced by the offer provided by producers in the energy system.**

Ensuring the balance between demand and supply in electricity is a complex process of dynamic nature and requires a strict balance of electrical power at each moment for ensuring a stable energy system, since the electricity can not be stored. This balancing of supply and demand of energy must be fulfilled both technically and economically. Recording activity and energy consumption parameters that depend on them is followed by a prediction of consumption and an analysis of the results to improve the final information quality.

The study presented in the paper focuses on predicting trends in energy consumption, holding for database real-time readings of energy consumption related to a calendar year, filed for a distribution operator. The chosen prediction mechanism is the simple linear regression method, since the only variable that was considered is the history of consumption. The presented prediction methodology uses the regression instrument of Excel running linear regression analysis using least squares method to find a line that corresponds to a set of observations.

It was intended that the indexes of quality of energy consumption trend is as good as they can be in order to increase the chances that forecast made to adequately reflect the future actual data and the maximum percentage error is as small as it can be.

***Index Terms* — load curves, power demand, prediction, statistical indicators, trend**

I. LOAD CURVES

A. Daily load curves. Problems and main indexes.

Usually, the daily load curves represent the active power demanded by the consumers.

The problems of daily load curves:

- Classifications, criteria: the building type, the kind of consumer, the kind of load, determination methods, (measuring, calculation, forecast);
- Main indexes: daily energy; maximal, minimal, medium power, flattening coefficient, duration of use of the daily peak power, time losses, power factor; mean square root, power variation coefficient, correlation coefficient;
- Load curves modeling: determinist; probabilistic (probabilistic hourly level, the normal distribution law is respected, indexes: average, irregularity, dispersion, standard deviation)
- Load curves forecast: classical methods (directly: static, dynamic: indirect); modern methods (heuristic; artificial intelligence: neuronal network, expert systems, decision tree).

II. POWER DEMAN PREDICTION

A. The need for power demand prediction

Predicting energy and power consumption aims to anticipate energy and power consumption calculations based on the analysis and interpretation of .diverse data set, in order finally to achieve an accurate match between estimated and actual consumption.

The activity of power demand prediction involves solving the following problems:

- Identifying the power demand causes in order of importance.
- Determination qualitative and quantitative shape of the law (correlation) between cause and effect.
- Using the correlation previously established for effective future power demand prediction.
- Checking in time predicted results.

Balancing energy supply with demand is made on two distinct levels:

- technical (at the design stage and then in the operational phase);
- economical.

B. Mathematical model of power demand prediction

Following the phase of collection, selection and processing the original data from the database, the mathematical model of power demand is established.

Power demand prediction highlights the existence of four main components which determine the energy curve shown in figure 1: the tendency or trend, T, (the main component that establishes the essential form of energy change), the cyclical component, C, (due to fluctuating and slow-acting causes), the seasonal component, S, (caused by certain parameters that have seasonal fluctuations) and the random component ε (due to accidental causes). By projecting the power demand, the variation of each component is estimated separately, achieving the final result by summing the results of the predicted components.

If cyclical components, seasonal and random are small compared to the trend, then influence on variation in power demand can be neglected and all is related just about trend prediction. This is most common in practice, being further developed in this paper.

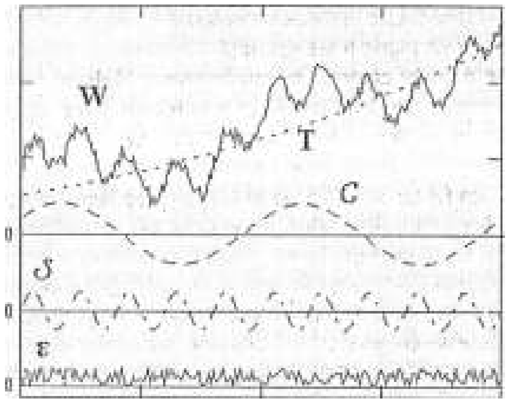

Figure 1. Components of mathematical model of power demand curve.

III. METHODS USED IN PREDICTION

A. Regression method

The method implements by means of analytical expressions, known as regression functions, show how the dependent variable y evolves in relation with changes of one or more independent variables X.

The general form of the regression function is:

$$Y_x = f(X_1, X_2, \ldots, X_k) + e \tag{1}$$

where "e" is the perturbing random variable, or error, representing the effect of all unspecified factors that are difficult to quantify or are insignificant.

The main types of regression models are:

- unifactorial regression;
- Simple curvilinear regression and correlation;
- Multiple regression and correlation can be expressed by a linear function or a curvilinear function.

B. Simple linear regression

It is a regression model in which the dependent variable (y) changes linearly under significant influence of a single independent variable (x).

Analysis and prediction model used will be the unifactorial linear regression:

$$Y_{x_i} = a + bx_i + e_i \tag{2}$$

where "a" and "b" are unknown parameters of the function, which are to be estimated

C. Simple linear regression and regression model validation

Estimation of parameters "a" and "b" of the linear regression equation is achieved by the method of least squares. This method is based on minimizing the sum of squared errors, meaning the minimization of the observed squared deviations values sum (Yi) from the theoretical values (Yx): (3)

Normal equation systems become:

$$\begin{cases} na + b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \\ a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i \end{cases} \tag{4}$$

By solving the system of equations the parameters a and b are obtained, as follows

$$a = \frac{\sum y_i \sum x_i^2 - \sum x_i \sum x_i y_i}{n\sum x_i^2 - \left(\sum x_i\right)^2} \tag{5}$$

$$a = \bar{y} - b\bar{x} \tag{6}$$

$$b = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - \left(\sum x_i\right)^2} \tag{7}$$

IV. STUDY CASE ON LOAD CURVES

A. Load curves corresponding to characteristic weeks

For this case study hourly readings of power demand for the years 2006, 2007 and 2008 are used. A first selection and processing of collected data is made, followed by viewing them in graphical form, trough the average load curves specific to a particular period.

The concepts of monthly typical week and annual typical week are introduced in this paper.

Monthly typical week presents load curves for each day of a typical week of a month. They are obtained by making the average consumption for 24 hours for all the days with the same name within a month. Thereby one can speak about a typical week of January, February, etc.

Annual typical week shows the typical average load curves for each day of a typical week of a year, mentioning that they are obtained by making the average consumption during the 24 hours for all 52 days with the same name during a year.

B. Monthly typical week - load curves

Graph of figure 2 presents day load curves characteristic of "Monday" in the monthly typical weeks. The presented graphs correspond to the days of 12 months related to the monthly typical week of January, February, etc. It is noticed that on Monday the power demand keep the same shape. Each graph is obtained by performing the power demand average for the days of "Monday" of the month.

Based on the same principle all weekdays for each month separately is represented graphically.

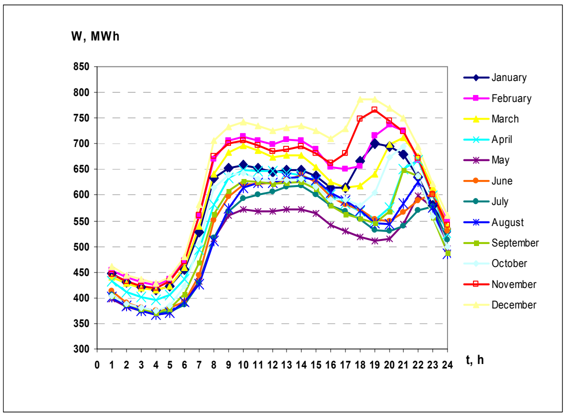

Figure 2. Monday load curves from the monthly typical week of the year 2006.

A monthly typical week covers average hourly energy consumption for the day Monday, Tuesday, Wednesday, etc. for the specific month. One can consider a monthly typical week for "January", "February", "March", "April", etc.

As shown power demand increases in the morning peak area and evening peak area as the site of these peaks should be provided adequate reserves of energy.

These load curves provide information on power demand aspect both for a period of 24 hours and depending on the season. Figure 2 emphases that in the winter months approaching consumption of 800 MWh and in the summer months power demand reduces by around 400 MWh.

Shape of load curves specific of working days are very different than shape of the load curves for Saturdays and Sundays or holidays over year, as shown in figure 3.

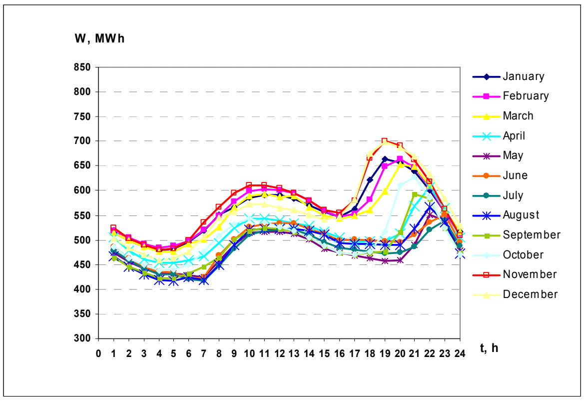

Figure 3. Sunday load curves from the monthly typical week of the year 2006.

Similarly corresponding graphs for average hourly change in power demand for all days of the monthly typical weeks for the years 2007 and 2008 are presented, figures 4 and 5 showing only specific curves for the days of Monday.

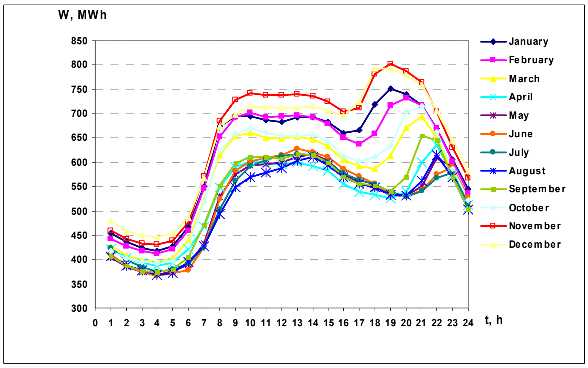

Figure 4. Monday load curves from the monthly typical week of the year 2007.

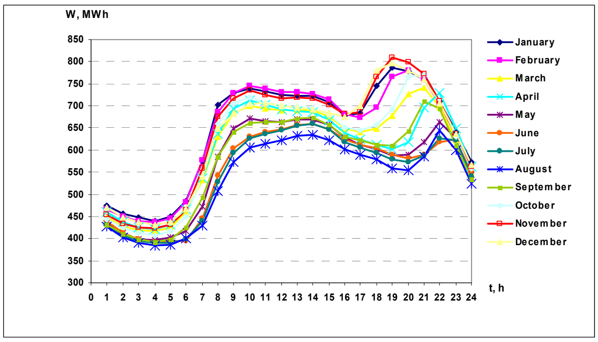

Figure 5. Monday load curves from the monthly typical week of the year 2008.

It may be noticed that for the same day, load curves have similar shape whether is part of the monthly typical week, or annual typical week.

Also different consumption peak loads day and night depending on season appear.

C. *Annual typical week – load curves*

Studying the variation of specific average power demand, during the seven days of annual typical week, other types of graphics are obtained.

Chart of figure 6 presents the load curves specific for *annual typical week* corresponding to the three studied years 2006, 2007 and 2008. Each load curve is obtained by performing the average energy consumption for all the days of "Monday" during that year. And it's obvious to notice that in any of the three years on "Monday" the variation of energy consumption has the same shape.

Same analysis was done for load curves of the days of "Tuesday", "Wednesday", "Thursday", etc. and these days together form an annual typical week.

This method of overlapping the profiles of energy consumption is a method used in practice quite frequently. It may also apply to a monthly typical week. With this overlap of load curves is observed that all Mondays have a similar allure, and it is assumed that the day "Monday" of the following year, 2009, power demand keeps the same shape. In fact it seeks the identity between hourly consumer behavior of year "n" and year "n-1", in our case in 2009 compared to the years 2006 ÷ 2008.

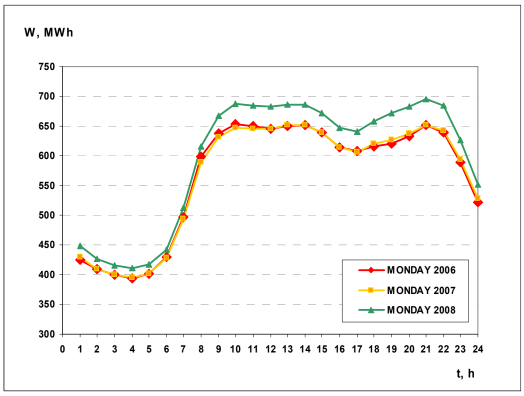

Figure 6. Monday load curves from the annually typical week.

V. APPLICATION OF A SIMPLRE LINEAR REGRESSION

A. *Predicting the trend in power demand*

This study is based on the trend prognosis of the power demand. The simple linear regression method has been chosen as a prediction instrument because the only variable taken into account is the historical consumption.

The historical consumption is based on the 52 weeks of the year, a special attention paying to the (holydays) feast-days: Christmas, Easter, January the 1st, the 1st of May and December the 1st. Due to the fact that in these days the power demand is reduced, they are considered as days with low consumption, corresponding to the days of Sunday.

For the following prediction the Excel's simple linear regression method is used, based on least-squares method for finding a function that corresponds to an observed data base. A Regression analysis tool, included by request in Microsoft Excel Analysis Tool Pack application is used.

In this case the dependent variable is the energy, W, and the independent variable is the time, t.

The prediction is made for two random weeks of 2006 and one week of 2007, as well as 2008. For a good approximation of the trend it is recommended to respect the condition: $k \geq 5$ where k is the number of data considered for the previous period before forecasting.

The value $k = 20$ is chosen and the aim is to predict the value of day 21. Twenty consecutive values corresponding to the power demand for the days of "Monday" are chosen to predict the power demand for the 21st day of "Monday".

The same way has been applied to predict the 21st day of Tuesday, Wednesday, etc, thus obtaining a prediction for the entire week. This week corresponds to the third week of May.

Data from table 1 represent the results achieved through prognosis instrument regression of Excel for Monday. It makes a comparison between forecast (using time intervals - 24 hours of the day) and effective consumption values. These are followed by error, and statistical indicators. These tables include also minimum values, maximum and average of the errors absolute and percentage. Similar tables for each day of the week have been obtained. Figure 7 represents the power demand prediction for the whole week (the 21st week of the year 2006) in according with these tables.

TABLE I. POWER DEMAND FOR THE 21ST DAY OF MONDAY FOR THE YEAR 2006.

	Tide	Energy forecast	Real demand??	Absolut error	Percentage error	Mean absolut deviation	Mean percentage error
Mo 21	1	425.09	394.48	-30.61	-7.76%	30.61	7.76%
	2	404.18	380.01	-24.17	-6.36%	24.17	6.36%
	3	396.08	371.24	-24.83	-6.69%	24.83	6.69%
	4	404.53	361.03	-43.51	-12.05%	43.51	12.05%
	5	394.09	357.47	-36.62	-10.24%	36.62	10.24%
	6	405.67	369.14	-36.54	-9.90%	36.54	9.90%
	7	448.30	425.32	-22.98	-5.40%	22.98	5.40%
	8	494.47	526.52	32.05	6.09%	32.05	6.09%
	9	533.27	574.03	40.76	7.10%	40.76	7.10%
	10	548.33	594.67	46.35	7.79%	46.35	7.79%
	11	564.43	586.83	22.41	3.82%	22.41	3.82%
	12	565.95	604.58	38.63	6.39%	38.63	6.39%
	13	565.56	605.97	40.40	6.67%	40.40	6.67%
	14	567.54	603.19	35.65	5.91%	35.65	5.91%
	15	549.88	594.11	44.23	7.44%	44.23	7.44%
	16	524.53	571.43	46.90	8.21%	46.90	8.21%
	17	525.37	564.73	39.36	6.97%	39.36	6.97%
	18	515.55	551.65	36.11	6.55%	36.11	6.55%
	19	502.57	546.50	43.93	8.04%	43.93	8.04%
	20	506.38	561.59	55.20	9.83%	55.20	9.83%
	21	531.40	573.87	42.48	7.40%	42.48	7.40%
	22	577.40	602.60	25.20	4.18%	25.20	4.18%
	23	553.07	592.21	39.14	6.61%	39.14	6.61%
	24	482.40	513.48	31.08	6.05%	31.08	6.05%
						36.63	**7.23%**
MIN				**-43.51**	**-12.05%**		
MAX				**55.20**	**9.83%**		
MED				**18.36**	**2.36%**		

Figure 7 presents the variation of power demand prediction for the 21st weeks of the year 2006.

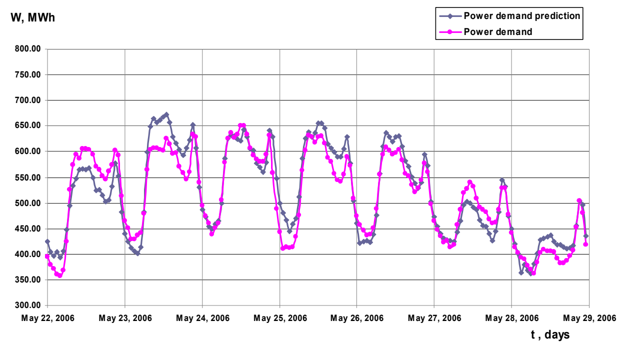

Figure 7. Power demand prediction for the 21st week of the year 2006.

It is pointed out that the expected values present a certain difference with real consumption, ones which can lead to some losses, especially economical, for the distribution operator. If consumption are underestimated then cost are incurred through cover damages caused by non-energy, the problems associated with improper sizing of the transport capacity and an overload of existing equipments. If consumption are overvalued, then unjustified funds are allocated to additional investments linked to increase the production capacity, transmission and increasing stocks of energy carriers.

Similarly prediction is done for the 11th week of 2006, corresponding to the third week of March. The prediction graph is shown in figure 8.

Unlike power demand prediction in May, it can be seen that for a smaller number of consumption readings, the accuracy is much higher.

Therefore it is recommended that any months in which the prognosis is made, to take into account a small number of readings consumption k <20.

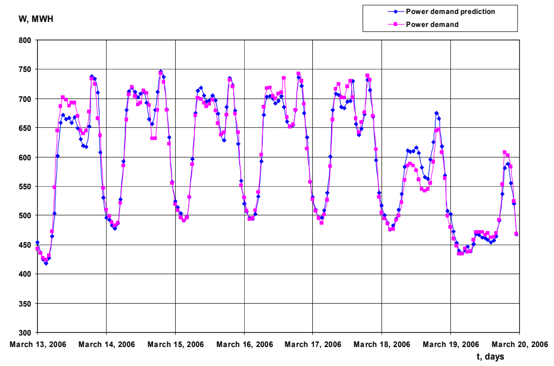

Figure 8. Power demand prediction for the 11th week of the year 2006.

Prediction made on the basis of prior reading for power demand both of the 11th week of the years 2007 and 2008 are shown in figures 9 and 10.

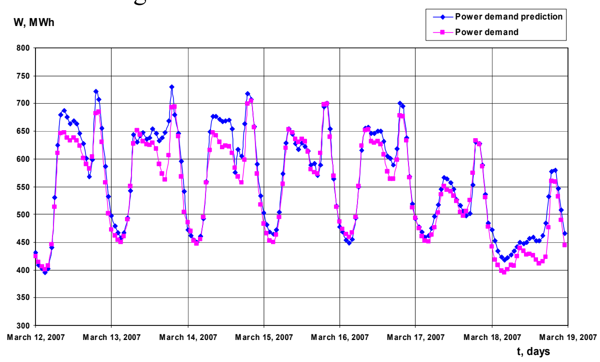

Figure 9. Power demand prediction for the 11th week of the year 2007.

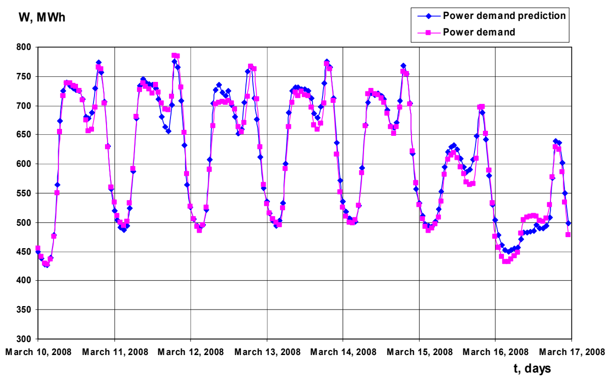

Figure 10. Power demand prediction for the 11th week of the year 2008.

B. *Statistical indicators*

Time series contains n observations made for each prediction and there are n prediction errors; in this case n = 24 (hours).

The statistical indicators are:

ME - mean error:

$$ME = \frac{1}{n} \cdot \sum_{k=1}^{n} (\hat{y}_k - y_k) \tag{8}$$

MAD - Mean absolute deviation:

$$MAD = \frac{1}{n} \cdot \sum_{k=1}^{n} |\hat{y}_k - y_k| \tag{9}$$

MAPE – mean percentage error:

$$MAPE = \frac{1}{n} \cdot \sum_{k=1}^{n} \left| \frac{\hat{y}_k - y_k}{y_k} \right| \cdot 100 \tag{10}$$

MSE – mean squared error:

$$MSE = \frac{1}{n} \cdot \sum_{k=1}^{n} (y_k - \hat{y}_k)^2 = \hat{\sigma}^2 \qquad (11)$$

The prediction study is accompanied by the statistical indicators calculation to evaluate its performance. Among the statistical indicators that describe the accuracy of a mathematical model some indicators have been selected in order to observe the numerical accuracy of the forecast.

Mean error is an estimated value that is intended to be as close to zero as possible. If the mean error of the prognosis differs appreciably from zero then there is a threshold in the indicated prognosis (a continuous disturbing component). It may be an indication that the cycle of time has been modified in a way which was not noticed by the used method.

The two indicators MAD and MSE are measuring the errors of the prognosis. The errors are wanted to be as small as possible.

If the errors are uniformly distributed, the standard deviations of the prognosis errors have a connection with MAD thru:

$$\hat{\sigma} = \sqrt{\frac{\pi}{2}} \cdot MAD \cong 1{,}25 MAD \qquad (12)$$

this constitutes a verification of the obtained values for the indicators and a view over the prognosis performance made in the case study.

Tables 2 ÷ 5 collate statistical indicators calculated for energy forecasts made in this study. Calculated percentage errors are below the threshold of 10%

TABLE II. STATISTICAL INDICATORS FOR 21ST WEEK OF 2006

DAY	ME	MAD	MAPE	MSE	Sqrt(π/2)*MAD	σ=Sqrt(MSE)
MONDAY	18.36	36.63	7.23%	1412.16	45.79	37.58
TUESDAY	-19.50	35.04	6.30%	1508.92	43.80	38.84
WEDNESDAY	-1.90	13.10	2.38%	443.71	16.37	21.06
THURSDAY	-31.58	32.05	6.41%	1334.79	40.06	36.53
FRIDAY	-6.90	17.37	3.24%	386.75	21.71	19.67
SATURDAY	12.30	17.80	3.64%	457.93	22.25	21.40
SUNDAY	-12.18	17.31	4.35%	399.96	21.64	20.00

TABLE III. STATISTICAL INDICATORS FOR 11TH WEEK OF 2006

DAY	ME	MAD	MAPE	MSE	Sqrt(π/2)*MAD	σ=Sqrt(MSE)
MONDAY	15.15	20.92	3.30%	611.82	26.14	24.73
TUESDAY	-5.36	11.43	1.80%	255.48	14.29	15.98
WEDNESDAY	-5.36	8.05	1.25%	98.85	10.06	9.94
THURSDAY	7.38	9.25	1.39%	185.65	11.57	13.63
FRIDAY	3.73	12.54	6.12%	231.32	15.68	15.21
SATURDAY	-18.98	18.98	3.36%	529.76	23.72	23.02
SUNDAY	4.10	8.77	1.73%	133.20	10.97	11.54

TABLE IV. STATISTICAL INDICATORS FOR 11st WEEK OF 2007

DAY	ME	MAD	MAPE	MSE	Sqrt(π/2)*MAD	σ=Sqrt(MSE)
MONDAY	-16.71	20.61	3.46%	571.19	25.76	23.90
TUESDAY	-20.87	24.47	4.26%	1029.53	30.59	32.09
WEDNESDAY	-21.52	24.07	4.03%	966.39	30.09	31.09
THURSDAY	-3.45	9.67	1.77%	130.43	12.08	11.42
FRIDAY	-8.96	13.56	6.17%	275.47	16.95	16.60
SATURDAY	-5.28	10.09	1.97%	146.62	12.61	12.11
SUNDAY	-27.54	27.54	6.36%	917.95	34.42	30.30

TABLE V . STATISTICAL INDICATORS FOR 11th WEEKS OF 2008

DAY	ME	MAD	MAPE	MSE	Sqrt(π/2)*MAD	σ=Sqrt(MSE)
MONDAY	-4.91	7.65	1.19%	135.60	9.56	11.64
TUESDAY	9.59	13.17	2.00%	244.44	16.46	15.63
WEDNESDAY	-4.29	16.08	2.35%	477.98	20.10	21.86
THURSDAY	-11.62	12.42	1.91%	236.07	15.53	15.36
FRIDAY	-2.61	5.51	6.18%	46.78	6.89	6.84
SATURDAY	-10.15	12.89	2.23%	265.01	16.11	16.28
SUNDAY	-1.27	16.68	3.39%	320.50	20.85	17.90

Comparing tables 2 and 3 shows a higher precision for a smaller number ("k") of readings.

In tables 3÷5, the values of the last two columns emphasizes that the condition (13) is fulfilled.

$$\sqrt{MSE} = \sqrt{\frac{\pi}{2}} \cdot MAD \qquad (13)$$

The power demand prediction study was made for the power demand corresponding to the first month of the year. It is recommended to be extended for each month of the year.

VI. CONCLUSIONS

The paper is meant to develop both theoretical and practical issues of power demand prediction which is current and widely discussed.

- Importance of an accurate prediction is necessary because of their technical and economical implications. As the distribution operator is forced to trade energy, due to the differences to power demand evaluation, the own technological consumption can be reduced or increased as it estimates. In this context power demand forecast is very important in electricity billing.
- As noted in the current study, using simple linear regression gives good results for some months but it is necessary to make an analysis for all months.
- The purpose of the work was to obtain a maximum percentage error of up to 10%, so this performance prognosis indicator is as small. Thus, it was intended that the indexes of power demand trend be good enough to increase the chances that the prediction reflects adequately the future data.
- Serious errors in the top portion of predicted load curves indicate that some other parameters than history must be taken into account, which could influence the power demand such as temperature, demographic factors, which lead to use the multiple linear regression method.
- The power demand prediction is dynamic and it must be a permanent, continuous activity and to base any study of energy development.

REFERENCES

[1] Albert, Hermina., Mihăilescu A. Pierderi de putere şi energie în reţelele electrice. Editura Tehnică, Bucureşti, 1997, ISBN 973-31-1071-X

[2] Baciu , A., Baciu , C. I. Energia electrică şi mediul înconjurător. Editura Tehnică, Bucureşti, 1982.

[3] Bott , E. Microsoft office XP. Editura Teora, Bucureşti, 2002 .ISBN-973-20-0089-9

[4] Carabogdan I. Gh., Mihăileanu C. Bilanţuri energetice. Editura Tehnică, Bucureşti, 1986

[5] Chindriş , M., Cziker , A. C., Miron, Anca, Tomoiagă, B. Managementul energiei electrice Aplicaţii. Casa Cărţii de Ştiinţă, Cluj-Napoca, 2009, ISBN 978-973-133-492-9.

[6] Darie, S., Vădan, I. Producerea, Transportul şi Distribuţia Energiei Electrice. Instalaţii pentru transportul şi distribuţia energiei electrice. U.T. PRES, Cluj-Napoca, 2003, ISBN 973-662-036-0.

[7] Luştrea , B. Prognoza consumului de energie . Editura Agir, Bucureşti, 2001 ISBN 973-8130-34-4.

[8] Montgomery, D.C., Jennings, C.L., Kulahci, M. Introduction to time series analysis and forecasting. Wiley & Sons Inc. Publication, New Jersey, 2008, ISBN 978-0-471-65397-4.

[9] Popescu, Th., Demetriu, S. Practica modelării şi predicţiei seriilor de timp. Metodologia Box-Jenkins. Editura Tehnică, Bucureşti, 1991, ISBN 973-31-0249-0.

[10] Preitl, Ş., Precup, R.E. Introducere în conducerea fuzzy a proceselor. Editura Tehnică, Bucureşti, 1997.

[11] Studiu ANRE realizat de KEMA: 1. Feasibility of Load Profiling; 2. Assessing the feasibility of load profiling in Romania, 2006

[12] Tertişco, M., Stoica, P., şi alţii. Modelarea şi predicţia seriilor de timp, Editura Academiei, Bucureşti, 1985.