CONSUMERS LOAD PROFILE CLASSIFICATION CORELATED TO THE ELECTRIC ENERGY FORECAST

Ioan FELEA, Florin DAN, Simona DZIŢAC

University of Oradea E-mail: simona.dzitac@gmail.com

This paper focuses over short term load forecast (STLF). The method used for classification of consumers is to examine the statistical database containing the values of power measurements and analyse the correlation between them. Based on two types of forecasting and database analysis, namely the correlation, a consumer classification in consumer profiles will be done. Later we will highlight the best performing forecast method for a consumer, comparing forecast by a classical method (linear regression - LR) and forecast using dynamic method (artificial neural networks - ANN). Conclusions will be drawn on the basis of research results.

Key words: short term load forecast, artificial neural network forecast, linear regression forecast, power measurement correlation.

1. INTRODUCTION

A prerequisite for developing an accurate forecasting model is thorough understanding of the consumption patterns to be modelled [1-3]. This knowledge about the behaviour of the load is learned from experience with the use of consumer data and statistical analysis of consumption in the past. Electricity consumers operating in a similar economic environment and climate are usually similar consumers, behaviour and consumption forecast models developed for one consumer can easily be adapted for use at another consumer [4, 5]. Load supplied by a power distribution system (EE) has a dynamic development and reflects directly the activities and conditions in the environment [6].

2. DATABASE PRESENTATION

The database was accomplished by measurements at different consumer types, in the transformer station serving the users. The database was obtained from the average hourly power measurements in a year interval. Some of the measurements were provided by energetics engineers of industrial entities referred to, some was obtained by measurements performed by the University of Oradea with the professional equipment of the Research Centre "Energy Process Management." Due to confidentiality clauses of the electricity provider and its customers, in some cases, data are composed only of hourly average powers and a summary description of the consumers or of the consumer groups served.

2.1. Consumer C1 (residential consumer)

The consumer that we will analyse is represented by a neighbourhood of apartment blocks in a small town located in the west of the country, which is representative for residential users ($85\% \div 90\%$). Measurements begin on Tuesday – 01.01.2008 until Wednesday – 31.12.2008. Measurements were made in the transformer station serving the residential neighbourhood.[4]

Elements characterizing the load curve are shown in Figs. 1, 2, 3 and Table 1.



Fig. 1 – 3D graphical representation of the hourly load curve in weekly series, throughout the year, for consumer C1 (hourly P_{med} in kW).

1	able	j
-		

Statistical description of the database C1 (P_{med})

Function	Value[kW]
Minimum(min)	116,6
Maximum(max)	345,3
Median(median)	214,12
Average(mean)	217,8689
Quintiles(quantile)	q ₂ =185,485
	q ₃ =247,725
Dispersion(var)	1913,022
Standard deviation(sd)	43,7381



Fig. 2 – Box and whisker diagram for consumer C1 (hourly P_{med} in kW).



Fig. 3 – Graphical plot of the correlation between daily load curves from the same week and the week before.

2.2. Consumer C2 (residential consumer)

The consumer that we will analyse is represented by a neighbourhood of houses in a small town located in the west of the country, which is representative for residential users (80%). Measurements begin on Tuesday - 01.01.2008 until Wednesday - 31.12.2008. Measurements were made in the transformer station serving the residential neighbourhood [7].

Elements characterizing the load curve are shown in Figs. 4, 5, 6 and Table 2.



Fig. 4 – 3D graphical representation of the hourly load curve in weekly series, throughout the year, for consumer C2 (hourly P_{med} in kW).

Table	2
-------	---

Statistical description of the database C2 (P_{med})

Function	Value[kW]
Minimum(min)	58,66
Maximum(max)	198.588
Median(median)	118,404
Average(mean)	120,2795
Quintiles(quantile)	$q_2 = 101,556$
	$q_3 = 136,190$
Dispersion(var)	565,56
Standard deviation(sd)	23,78151



Fig. 5 – Box and whisker diagram for consumer C2 (hourly P_{med} in kW).



Fig. 6 – Graphical plot of the correlation between daily load curves from the same week and the week before.

2.3. Consumer C3 (industrial consumer)

Industrial consumer that we will analyse named C3, is an industrial park from the west side of the country, and it consists of a foundry and metal processing specific industries branches. Measurements begin on Tuesday - 01.01.2008 until Wednesday - 31.12.2008. Measurements were made in the transformer station serving the industrial park [2]. Elements characterizing the load curve are shown in Figs. 7, 8, 9 and Table 3.



Fig. 7 – 3D graphical representation of the hourly load curve in weekly series, throughout the year, for consumer C3 (hourly P_{med} in kW).

Table 3

Statistical description of the database C3 (Pmed)

Function	Value[kW]
Minimum(min)	22,15
Maximum(max)	197.34
Median(median)	96,56
Average(mean)	101,3341
Quintiles(quantile)	q ₂ =81,90
	q ₃ =119,50
Dispersion(var)	653,7822
Standard deviation(sd)	25,56916



Fig. 8 – Box and whisker diagram for consumer C3 ($hourly P_{med} in kW$).



Fig. 9 – Graphical plot of the correlation between daily load curves from the same week and the week before.

2.4. Consumer C4 (commercial consumer)

The commercial consumer that we will analyse, Lotus Centre Oradea, is represented by a large commercial centre, which includes a supermarket, "Carrefour", a cinema, "Hollywood Multiplex" and a large number of commercial entities. The C4 consumer is located in Oradea. Measurements begin on Wednesday - 01.07.2009 until Saturday - 31.07.2010. Measurements were made in the transformer station serving the commercial centre [7]. Elements characterizing the load curve are shown in Figs. 10, 11, 12 and Table 4.



Fig. 10–3D graphical representation of the hourly load curve in weekly series, throughout the year, for consumer C4 (hourly P_{med} in kW).

Fig. 11 – Box and whisker diagram for consumer C4 (hourly P_{med} in kW).

*	(
Function	Value[kW]
Minimum(min)	250
Maximum(max)	2262
Median(median)	1052
Average(mean)	920,6476
Quintiles(quantile)	q2 =291
	q3 =1393
Dispersion(var)	363350,3
Standard deviation(sd)	602,7855



Fig. 12 Graphical plot of the correlation between daily load curves from the same week and the week before

2.5. Consumer C5 (industrial consumer)

This industrial consumer is a foundry from Oradea belonging to SC Turnătorie Iberica S.A. Recordings starts on Sunday – 01.02.2009 01:00 to Saturday – 01/31/2010 24:00. Consumer S.C. Turnătorie Iberica S.A. (C8) operates in three shifts, six days a week, with an exception of two weeks in which they worked in two shifts. Given the two periods of interruption of work for 2 weeks and the power audit results [8], the results obtained in this case are presented in Figs. 13, 14, 15 and Table 5.



Fig. 13 – 3D graphical representation of the hourly load curve in weekly series, throughout the year, for consumer C5 (hourly P_{med} in MW).

Table .	5
---------	---

Statistical description of the database C5 (P_{med})

Function	Value[MW]
Minimum(min)	0.052954
Maximum(max)	0.534616
Median(median)	0.306
Average(mean)	0.2620182
Quintiles(quantile)	q ₂ =0.112478
	$q_3 = 0.392000$
Dispersion(var)	0.02115484
Standard deviation(sd)	0.1454470



Fig. 14 – Box and whisker diagram for consumer C5 (hourly P_{med} in MW).



Fig. 15 – Graphical plot of the correlation between daily load curves from the same week and the week before.

3. DESCRIPTION OF THE SOFTWARE ENVIRONMEN AND MATHEMATICAL MODEL FOR THE IMPLEMENTATION OF ARTIFICIAL NEURAL NETWORK (ANN)

3.1. Software environment used

R-Statistics is a language and environment for statistical computing and graphics. It is a project developed by GNU and is similar to the S language and environment and which was developed at Bell Laboratories (formerly AT & T, now Lucent Technologies) by John Chambers and colleagues. R-Statistics

Table 4

Statistical description of the database C4 (P_{med})

offers a wide variety of statistical (linear and nonlinear modelling, classical statistical tests, time series analysis, classification, clustering, etc.) and graphical techniques, and is very extensible. R-Statistics Language is often the vehicle of choice for research in statistics and R-Statistics provides an open source solution for this.

One of the strengths R-Statistics is the ease with which generate quality graphics, including mathematical symbols and formulas, where necessary [10].

R-Statistics is available as Free Software under the terms of the Free Software Foundation GNU General Public License in source code form. It compiles and runs on a wide variety of UNIX platforms and similar systems (including FreeBSD and Linux), Windows and MacOS.

The term "environment" is intended to characterize as a fully planned and coherent system, rather than a basic build tools very specific and inflexible, as is frequently the case with other software for data analysis.

R-Statistics is designed around a programming language and allows users to add additional functionality by defining new functions. Much of the system itself is written in programming, making it easy for users to follow the algorithmic choices made. For computationally intensive tasks, the codes C, C + + and Fortran can be added to that link and used at runtime. Advanced users can write C code to manipulate R-Statistics objects directly [11].

3.2. The mathematical model used

The model used for correlation is:

$$r = \operatorname{Correl}(x, y) = \frac{\sum (x - \overline{x})(y - \overline{y})}{\sqrt{\sum (x - \overline{x})^2 (y - \overline{y})^2}},$$
(1)

where: x and y are the averages of the two data base to correlate.

The mathematical model used is an application of artificial neural network multilevel (feed forward) by the method of gradient descent (back propagation) by minimizing an error function which is not a euclidian type [12-16].

Multilevel feed forward networks are trained by supervised methods, which involve the use of training instances of the form: (X^p, t^p) , where: $X^p = (X^p_1, X^p_2, ..., X^p_N)$ is the input vector for the training p; $t^p = (t^p_1, t^p_2, ..., t^p_M)$ is the vector of desired outputs for p; N is the number of input units of the network; M is the number of output units.

Considering F(X) the function associated to the processing of the problem, according to the input *X*, then:

$$t^p = F(X^p) \,. \tag{2}$$

The output obtained by processing the input data using neural network is denoted by:

$$O^{p} = (O_{1}^{p}, O_{2}^{p}, ..., O_{M}^{p}).$$
 (3)

 O^p can be considered as the result of processing of the input, X^p , by using the function $F_w(w;X^p)$, based network implemented as an approximation of F(X). Therefore:

$$O^p = F_w(w; X^p). \tag{4}$$

Recorded error at the processing in the network of the input vector X^p , the measured error in a unit of output U_j and denoted by e_j^p is expressed as the difference between desired and actual output achieved, namely:

$$e_i^p = t_i^p - O_i^p \quad . \tag{5}$$

The E^p error, recorded at the processing trough the network of the input vector X^p and set the whole neural network is obtained by combining the error e_i^p , based on a relationship of the form:

$$E^{p} = \sum_{j=1}^{M} f(e_{j}^{p}).$$
 (6)

For error calculation we will use the error E^p and the Zero Based Log-Sigmoid Function:

$$f(x) = \frac{e^{a+bx}}{1+e^{a+bx}}.$$
 (7)

The mathematical model used for classic forecast, linear regression is:

$$a + bx$$
, (8)

$$a = \overline{y} - b\overline{x} , \qquad (9)$$

$$b = \frac{\sum (x - \overline{x})(y - \overline{y})}{\sum (x - \overline{x})^2},$$
(10)

where: x and y are the averages of the known data base

To identify the performance level of the application we will use the mean absolute percentage of errors (MAPE) and root mean squared error (RMSE).

MAPE =
$$100 \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right|,$$
 (11)

RMSE =
$$\sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}$$
, (12)

where: *n* – total number of the forecast values; *y* – real value; \hat{y} – forecast value.

Performance of the forecasting process is presented in tables and graphs, exemplifying the forecast results for each consumer analysed. To analyse the results, we specify that, according to ANRE [4, 5] the maximum legal allowable communicated forecast error, equals 25%. In this paper we start from the premise that an acceptable forecast would be below 5%.

4. RESULTS

Forecast parameters of ANN used in each training / forecast are:

- Limit of the learning cycles (epochs) -10000; Minimum weighting value delta 0,0001;
- Initial weight -0.3; Learning rate -0.3; Momentum -0.6; Neurons in hidden layer -0

- Activation function - logistic sigmoid with 0 based (unipolar sigmoid)

4.1. Forecast for consumer C1

The forecast is highlighted by hourly load curve presented in Fig. 16 and specific elements of characterization, given in Table 6 for ANN forecast(Epochs=10000, Weighting delta =0,0042), and Fig. 17 and Table 7 for linear regression.



Fig. 16 - One week ahead hourly forecast for C1 using ANN.



Fig. 17 - One week ahead hourly forecast for C1 using LR.

One week ahead forecast accuracy for C1 using ANN		One week ahead forecast accuracy for C1 using LR			
	MAPE[%]	RMSE[kW]		MAPE[%]	RMSE[kW]
24.11.2008-Monday	4.55682971	12.61629313	24.11.2008-Monday	1.52113832	5.569226
25.11.2008-Tuesday	4.288897015	12.44614283	25.11.2008-Tuesday	2.196557248	6.640759
26.11.2008-Wednesday	3.870885164	11.32889093	26.11.2008-Wednesday	2.359317288	9.447599
27.11.2008-Thursday	4.52945905	11.70518844	27.11.2008-Thursday	0.97142627	3.19834
28.11.2008-Friday	3.050005917	8.632927989	28.11.2008-Friday	1.255686621	4.525399
29.11.2008-Saturday	3.611697555	10.16733637	29.11.2008-Saturday	2.408969195	7.094272
30.11.2008-Sunday	4.077218056	11.85733583	30.11.2008-Sunday	3.096172985	8.209386
	MAPE[%]	RMSE[kW]		MAPE[%]	RMSE[kW]
	3.997856067	11.25058793		1.972752561	6.684785

4.2. Forecast for consumer C2

The forecast is highlighted by hourly load curve presented in Fig. 18 and specific elements of characterization, given in Table 8 for ANN forecast(Epochs=969, Weighting delta =0,0001), and Fig. 19 and Table 9 for linear regression.



Table 6

Fig. 18 - One week ahead hourly forecast for C2 using ANN.

Table 8



Fig. 19 - One week ahead hourly forecast for C2 using LR.

Table 9

	MAPE[%]	RMSE[kW]
4.11.2008-Monday	4.380227466	7.322981
5.11.2008-Tuesday	2.420037496	3.65774
6.11.2008-Wednesday	3.723682762	5.449896
7.11.2008-Thursday	5.548457278	8.181593
3.11.2008-Friday	2.508608415	3.918072
9.11.2008-Saturday	4.204878811	6.851871
0.11.2008-Sunday	5.139414125	7.5532
	MAPE[%]	RMSE[kW]
	3.989329479	6.133622

One week ahead forecast accuracy for C2 using LR

	MAPE[%]	RMSE[kW]
24.11.2008-Monday	2.417028168	4.109297
25.11.2008-Tuesday	2.138612073	3.389452
26.11.2008-Wednesday	2.011919607	3.529789
27.11.2008-Thursday	2.087474302	3.783835
28.11.2008-Friday	2.183351077	3.892344
29.11.2008-Saturday	3.440507444	5.339462
30.11.2008-Sunday	3.299443094	4.609076
	MAPE[%]	RMSE[kW]
	2.511190824	4.141347

4.3. Forecast for consumer C3

The forecast is highlighted by hourly load curve presented in Fig. 20 and specific elements of characterization, given in Table 10 for ANN forecast(Epochs=192, Weighting delta =0,0001), and Fig. 21 and Table 11 for linear regression.



Fig. 20 - One week ahead hourly forecast for C3 using ANN.



Fig. 21 – One week ahead hourly forecast for C3 using LR.

Table 7

Table 10 One week ahead forecast accuracy for C3 using ANN

Table 11

One week ahead forecast accuracy for C3 using LR

	5	e		5	e
	MAPE[%]	RMSE[kW]		MAPE[%]	RMSE[kW]
24.11.2008-Monday	7.503933232	8.042164	24.11.2008-Monday	3.349937243	3.781569
25.11.2008-Tuesday	8.434014921	9.986919	25.11.2008-Tuesday	6.997833166	7.896136
26.11.2008-Wednesday	5.294470956	6.509274	26.11.2008-Wednesday	3.825449717	5.074666
27.11.2008-Thursday	8.515690096	9.874305	27.11.2008-Thursday	3.20886107	4.246588
28.11.2008-Friday	5.953928064	7.535833	28.11.2008-Friday	4.121178237	5.482632
29.11.2008-Saturday	4.793994534	5.525859	29.11.2008-Saturday	3.503595503	4.596913
30.11.2008-Sunday	6.379335167	6.985652	30.11.2008-Sunday	5.843209987	6.096536
	MAPE[%]	RMSE[kW]		MAPE[%]	RMSE[kW]
	6.696480996	7.780001		4.407152132	5.46147

4.4. Forecast for consumer C4

The forecast is highlighted by hourly load curve presented in Fig. 22 and specific elements of characterization, given in Table 12 for ANN forecast(Epochs=10000, Weighting delta =0,0067), and Fig. 23 and Table 13 for linear regression.



Fig. 22 - One week ahead hourly forecast for C4 using ANN.

 Table 12

 One week ahead forecast accuracy for C4 using ANN



Fig. 23 - One week ahead hourly forecast for C4 using LR.

Table 13

One week ahead forecast accuracy for C4 using LR

	MAPE[%]	RMSE[kW]
21.06.2010-Monday	17.84275528	215.1867
22.06.2010-Tuesday	8.882101952	142.789
23.06.2010-Wednesday	18.48234401	283.9598
24.06.2010-Thursday	13.164723	131.1198
25.06.2010-Friday	9.492240273	156.6805
26.06.2010-Saturday	12.56749897	205.1811
27.06.2010-Sunday	14.31308146	225.5948
	MAPE[%]	RMSE[kW]
	13 53496356	194 3588

	MAPE[%]	RMSE[kW]
21.06.2010-Monday	6.022212111	64.58014
22.06.2010-Tuesday	8.891695738	115.782
23.06.2010-Wednesday	9.552367008	140.5267
24.06.2010-Thursday	9.985557988	102.1905
25.06.2010-Friday	12.17938133	152.8444
26.06.2010-Saturday	7.29327607	91.52135
27.06.2010-Sunday	9.764512242	111.7653
	MAPE[%]	RMSE[kW]
	9.098428927	114.6397

4.5. Forecast for consumer C5

The forecast is highlighted by hourly load curve presented in Fig. 24 and specific elements of characterization, given in Table 14 for ANN forecast(Epochs=10000, Weighting delta =0,0414), and Fig. 25 and Table 15 for linear regression.



Fig. 24 - One week ahead hourly forecast for C8 using ANN.



Fig. 25 - One week ahead hourly forecast for C8 using LR.

8

Table 14

One week ahead forecast accuracy for C8 using ANN

 Table 15

 One week ahead forecast accuracy for C8 using LR

	5	e			e
	MAPE[%]	RMSE[kW]		MAPE[%]	RMSE[kW]
24.01.2010-Monday	20.27999671	0.056191	24.01.2010-Monday	9.96497843	49.04666
25.01.2010-Tuesday	9.833870146	0.047871	25.01.2010-Tuesday	6.78337481	32.97864
26.01.2010-Wednesday	13.5136903	0.061088	26.01.2010-Wednesday	4.700193049	20.83411
27.01.2010-Thursday	12.04850749	0.053663	27.01.2010-Thursday	12.22401199	39.3909
28.01.2010-Friday	10.04786445	0.04593	28.01.2010-Friday	142.9779979	112.1321
29.01.2010-Saturday	34.22835576	0.05948	29.01.2010-Saturday	25.05841631	34.24369
30.01.2010-Sunday	22.60971056	0.022043	30.01.2010-Sunday	14.52347297	16.8041
	MAPE[%]	RMSE[kW]		MAPE[%]	RMSE[kW]
	17.50885649	0.049466		30.89034936	52.79

5. CONCLUSION

Forecast consumption is the main element of analysis of this research, forecasting being the factors influencing the development or modification of decisions at various stages of management of the electricity supply service [17].

For each of the three types of consumers analysed, each of the forecasting methods used are suitable: static method – linear regression and dynamic method – based on artificial neural networks. There is an interdependent relationship between the correlations performed on the database and the optimal type of forecast for a particular type of consumer, as follows: Residential consumers are characterized by a correlation coefficient between the values of the previous week and this week that are nearly one. For this type of consumer, there is a much better accuracy when using traditional forecasting methods. Industrial and commercial consumers have a correlation coefficient which varies greatly, meaning very weak correlation between this week values and the week before analysis. For this type of consumer applying the ANN based forecast results in much better accuracy.

REFERENCES

- 1. WILLIS M.L., SCHAUER A.E., NORTHCOTE J.E.D., VISMOR T.D., Forecasting distribution system loads using curve shape clustering, IEEE Trans. Power Apparatus Syst., **102**, *4*, pp. 893–901, 1983.
- 2. JIGORIA-OPREA DAN, LUŞTREA BUCUR, et all., Human machine interface for daily load short term forecasting using recursive artificial neural network, Proceedings of the **9th** WSEAS, Genoa, Italy, 2009.
- 3. DUMBRAVĂ VIRGIL, et al., *Case study regarding the load profile utilisation*, Energetica Magazine, **56**, *4*, 2008
- 4. ANRE 73 decree from 6 august 2009, regarding the approval of the electric energy forecast procedure, M. Of. 591, 2009.
- 5. ANRE Electric energy forecast procedure from 6 august 2009, M. Of. 591, 2009.
- 6. GOUDE Y., *Tracking the best predictor with a detection based algorithm*, In Proceedings of the Joint Statistical Meetings (JSP), 2008b.
- 7. FELEA I., et al., Documentation for the electric power balance at S.C. Lotus Centre S.A., Oradea University, 2010.
- 8. FELEA I., DAN F., et al., Documentation for the electric power balance at S.C. Zahärul Oradea S.A., Oradea University, 2009.
- 9. FELEA I., DAN F., Summary of the energy audit performed on the machines that develop aluminium parts by injection, Journal of Sustainable Energy, I, 4, pp. 53–60, 2010.
- 10. THEUBL STEFAN, ZEILEIS ACHIM, Collaborative Software Development Using R-Forge The R Journal, 1, pp. 9–14, 2009.
- 11. MURRELL PAUL, Drawing Diagrams with R, The R Journal, Invited paper on "The Future of R", 1/1, pp. 14–21, 2009.
- 12. CONSTANTA BODEA, Artificial Intelligence Lecture (ASE Bucuresti) www.ia.ase.ro/IA/Curs20.pdf view at 01/03/2011.
- 13. KILYENI ŞT., et al., Daily load forecasting using recursive Artificial Neural Network vs. classic forecasting approaches Proceedings of the 5th International Symposium on Applied Computational Intelligence and Informatics (SACI), Timişoara, 2009 pp.487–490.
- 14. SHANTI SWARUP, K., SATISH, B., Integrated ANN approach to forecast load, IEE Computer Applications in Power, pp. 46–51, 2002.
- DZIŢAC SIMONA, FELEA I, DZIŢAC I, VESSELENYI T., An Application of Neuro-Fuzzy Modelling to Prediction of Some Incidence in an Electrical Energy Distribution Center, International Journal of Computers, Communications & Control, III, Supplementary Issue (3S), pp. 287–292, 2008.
- 16. GAVRILAS M., Artificial intelligence and her appliance in energetics, "Gh. Asachi" Publishing House, 2002.
- 17. FILIP F., LEIVISKÄ K., Large-Scale Complex Systems, in Springer Handbook of Automation (S. Y. Nof, Ed.), Springer, Dordrecht, pp. 619-638.

Received September 26, 2011