IEEE Transactions on Power Apparatus and Systems, Vol. PAS-101, No. 9 September 1982

FORECASTING OF HOURLY LOAD BY PATTERN RECOGNITION

-A DETERMINISTIC APPROACH-

Abdolhosein S. Dehdashti
Member, IEEE

James R. Tudor
Senior Member, IEEE

Michael C. Smith
Non-Member

Dept. of Electrical Engineering
University of Missouri-Columbia
Columbia, Missouri 65211

Abstract - An algorithm using pattern recognition techniques and current weather parameters to forecast hourly electric load with a lead time of one to three hours is proposed. The lead time can be increased to at most 24 hours with the use of forecasted weather parameters.

This method is intended for small area power systems.

The assumption that load patterns are a direct result of climatological changes (weather sensitive load) was made throughout this research. The methodology of this algorithm is explained step by step, the results of actual data testing are presented, and finally method characteristics along with accuracy improvements are discussed. The content of this paper should be of direct benefit to utility engineers for power system planning and operation.

1. INTRODUCTION

Pattern recognition techniques are generally applied in the study of variables whose total physical principles behind their variation is unknown, but certain kinds of measurements explain their behavior. Figure 1 shows a standard system of pattern recognition which will be adopted throughout this research.

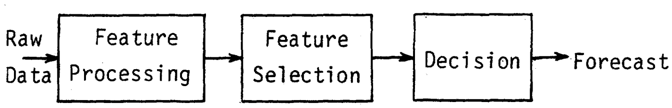


FIGURE 1

2. LOAD FORECASTING

Considerable research has been done on an accurate method of forecasting hourly load especially in the past two decades. Most of the methods developed either use recent load information or weather-load relationships, but not both. In this proposed model of forecasting hourly load, both of the above plus current weather information are employed. The result is an algorithm which is highly adaptive to both sudden weather and load changes.

82 WM 124-6 A paper recommended and approved by the IEEE Power System Engineering Committee of the IEEE Power Engineering Society for presentation at the IEEE PES 1982 Winter Meeting, New York, New York, January 31-February 5, 1982. Manuscript submitted August 31, 1981; made available for printing December 16, 1981.

Daily electrical load patterns in the same geographical area were assumed to have been repeated sometime in the past as a result of similar climatological conditions.

The basic 24-hour load pattern was divided into eight three-hour intervals (see figure 2). This division allows for various weather parameters to be considered for different times of the day and is likely to obtain monotonic changes of load patterns in every interval, thus easing interpolation in the decision section of figure 1. Additionally this division conforms with the official weather data from NOAA (National Oceanic and Atmospheric Administration) which is available at three-hour intervals.

Within each three-hour interval it is assumed that changes in weather pattern do not affect the load significantly.

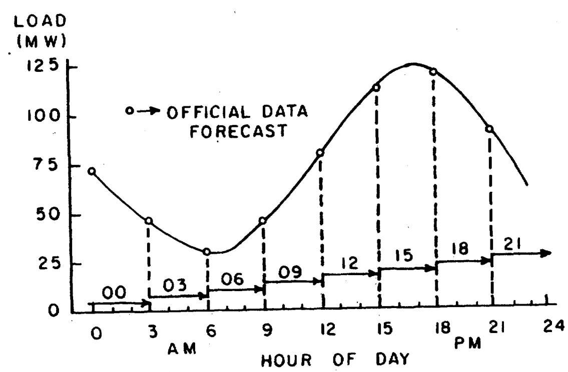


FIGURE 2

3. FEATURE PROCESSING

The purpose of feature processing is to extract, accumulate, and evaluate the relative significance of all the weather variables in the vicinity of an electrical power system.

Auto-correlation analysis was performed since it is a well-known fact that daily load patterns correlate closely with the previous days climatic condition. Factor analysis was used to determine the communality [4] of all the weather variables which will be used in feature selection.

Sixteen weather variables were selected at this stage (see Appendix). The choice of independent variables will of course depend on the season being analyzed. For example, humidity is significant in the summertime analyses, and in the same manner, windchill factor is significant in wintertime analyses.

0018-9510/82/0900-3290$00.75 © 1982 IEEE

4. FEATURE SELECTION

In this section various means were used to find the most efficient independent variables that explain similar load patterns on an electrical power system. Two types of studies were performed on weather variables from feature processing:

1. All Combination Regression
2. Stepwise Linear Regression analysis

In study 1, the multiple correlation coefficient squared (R^2) factors for all the combinations of weather variables were determined. Study 2 was performed to obtain total squared error (Mallow's Cp) and F statistics [5].

The result of F statistics will be used for weighing purposes in the Decision section of Figure 1.

5. DECISION

5.1 Cluster Analysis

Variables selected through feature selection were used to perform cluster analysis to separate the different dates into classes with similar climatic parameters. Hierarchical cluster analysis by means of Euclidean distance was used to separate different classes. Variation within a class (cluster) represents the unknown nature of load.

The optimum number of classes would be one with different classes as dense and as far as possible [10]. This is called natural cluster (see Discussion). In general, further division of a certain class as long as total number of classes exceeds the natural cluster number will result in satisfactory results (see Figure 3).

Weights can be given to different selected variables to enhance or reduce their relative importance and also to normalize different variables.

The number of classes could also be dependent on the available library size and number of elements needed in each class for the operation to be performed.

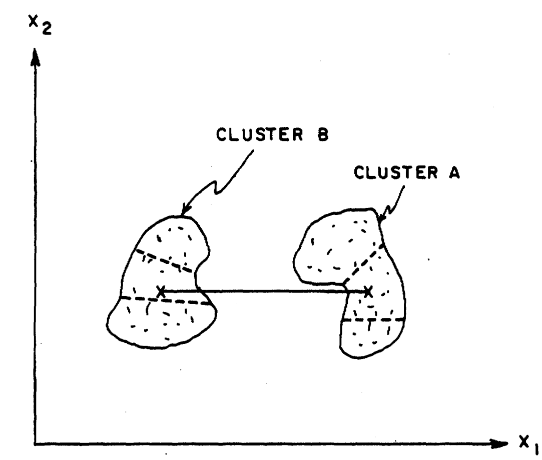


FIGURE 3

5.2 Growth Elimination

In order to use load information dating several years back, load growth will be eliminated by the concept of percent spread [8]. Load values will be converted to percent spread values while preserving the shape.

$$\text{Percent spread} = S_t = \frac{L_t - L_{min}}{L_{max} - L_{min}} \times 100 \tag{1}$$

where

L_t = load at time t (MW)

L_{min} = Minimum load for that season or year (MW)

L_{max} = Maximum load for that season or year (MW)

In performing the conversion each load value will be converted to a value between zero and 100. Zero corresponds to the minimum load for the period of study, and 100 corresponds to the maximum load of the same period.

Results show that there is no need to adjust for very small growths.

5.3 Interpolation

Cluster analysis will identify the best dates for load forecasting of a particular day. Depending on the lead time either current weather information (lead time 1-3 hrs.), or forecasted weather information (lead time up to 24 hrs.) will be used in the cluster analysis to identify the above dates.

It can be proven that if the ratios of minimum to maximum load for two particular days are equal, the corresponding piecewise shapes will only be different by a scalar.

Percent spread factor will convert all the load to a common base by preserving the shape which will satisfy the above condition.

Interpolation will be repeated 8 times for the three-hour intervals to obtain a base load curve. We will have:

$$\frac{F'(t)}{F(t)} = K \pm E \tag{2}$$

where:

$F'(t)$ = Actual load curve at time t

$F(t)$ = Base load curve at time t

K = Constant (scalar)

E = Error

so, the forecast for time (t+1) will be

$$F''(t+1) = KF(t+1) \pm E \tag{3}$$

where Constant K will be changed adaptively (1-3 hours) as new load information is received.

All the results will be in percent spread which will be converted back to MW by a reverse process.

Figure 4 shows a complete block diagram of the proposed method.

6. DATA LIBRARY

A different data library should be built for each season under consideration. A larger time span of data should be included for transitional seasons to account for all possible variations of climatic conditions.

In general, the larger the data library, the more accurate the forecast will be. Note that redundant climatic conditions can be discarded since they serve no useful purpose. Actual data testing later showed that more recent information yields a better forecast. This is due to the fact that there exists a gradual change in load pattern that is caused by demographic and economic changes of every environment.

Special occasions such as football games with unusually large audiences could be treated by adding another variable into feature classification.

Our actual library contained five years of data (weekends and legal holidays excluded) and resulted in satisfactory results. The actual size of the library could also be dependent on diversity of load.

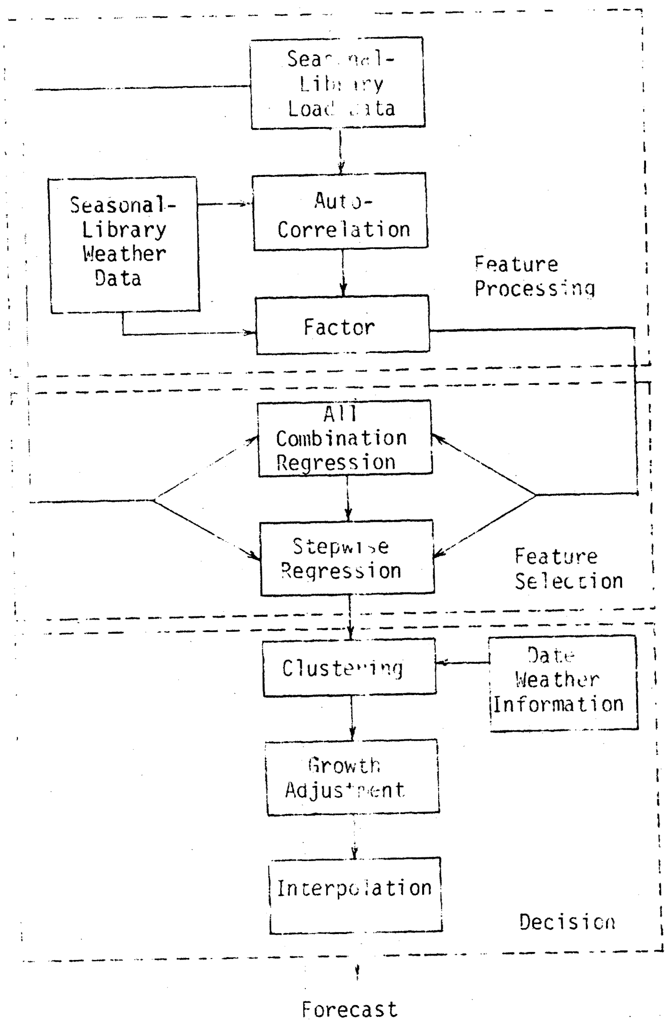


FIGURE 4

7. DATA TESTING

Load information from the city of Columbia, Missouri was used to test actual performance of the proposed method.

Columbia is a typical college town with summer 1980 peak load of 123 MW. The summer of 1980 was unusually hot and was therefore used to test this method.

Columbia has a modest load structure. In 1980, 31% was residential, 26% commercial, 26% industrial, 11% university, and 6% other.

8. RESULTS

Table I shows the selected weather variables for each of the eight intervals.

Table I
Best Four Variables by the Order of Importance

PERIOD / VARIABLES	00 1-3 AM	03 4-6 AM	06 7-9 AM	09 10-12 AM	12 1-3 PM	15 4-6 PM	18 7-9 PM	21 10-12 PM
ONE	X_1	H_3	X_1	X_1	X_4	V_1	X_1	X_1
TWO	H_3	X_1	H_3	H_3	X_2	X_2	V_3	V_3
THREE	J	F	V_1	V_3	H_3	U	X_3	U
FOUR	X_5	X_4	J	X_3	V_3	H_3	H_3	X_2

Table II shows the results of regression.

Table II
Total Multiple Correlation Coefficient (R) Estimates

INTERVAL	00	03	06	09	12	15	18	21
R	.925	.920	.881	.926	.941	.935	.934	.921

Table III summarizes the result of clustering.

Table III
Optimal Number of Clusters

INTERVAL	00	03	06	09	12	15	18	21
NUMBER OF CLUSTERS	68	71	56	57	76	71	75	74

9. SAMPLE

Figure 5 is the actual and forecasted (lead time one hour) load curves for July 30, 1980 which was an exceptionally hot day. The maximum temperature for the day was 111^{O}F which was about 15^{O}F higher than normal. Maximum error of 4% with an average error of 2% was computed by comparing forecasted and actual load curves.

The above accuracy is acceptable since rounded load values were used (for example, 50MW:49.51-50.49MW), and also due to the fact that metering errors are included.

10. ACCURACY IMPROVEMENTS

The result of our study showed an increased error for particular day's forecast. This was caused by using information from days that the University was not in session to forecast load for days that the university was in session. This diversity problem was overcome by restricting the same period information and forecasting.

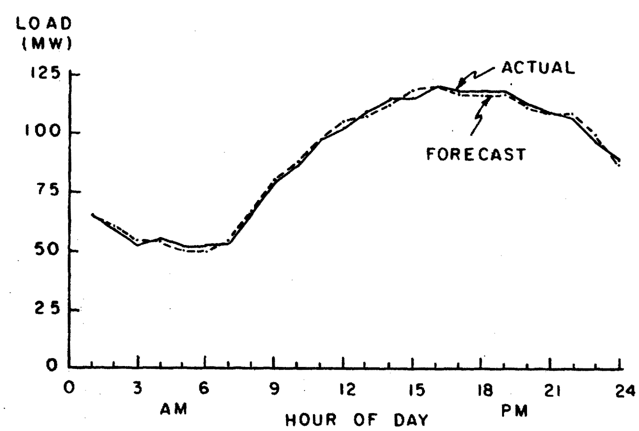


FIGURE 5

11. DISCUSSION

The most sensitive part of this forecasting method is feature selection. Several complications arose during the course of development of this algorithm. For example, F statistics tend to lose their significance with collinear weather parameters. Some of the collinearity is removed by factor analysis which tends to group variables with high communality. If collinearity exist among weather variables, the problem was overcome by weighing down such a variables in the clustering procedure. Independent variables with high communality values are desirable in this approach since general attributes are the objective in the clustering procedure.

While the high number of variables used in this research effects the number of degrees of freedom available, each was found essential in order to obtain the necessary separation within and between clusters (in effect number of clusters present was increased).

At several points, compromises were made in order to choose the most efficient weather parameters good for a three-hour interval. These compromises were necessitated due to the fact that for any interval, the three corresponding hourly loads patterns were used, and at times not all the three results were totally consistent.

The reader should be cautioned that F statistics, multiple correlation, etc. are estimates and should not be used literally.

All the analyses used in the proposed method are linear. Linearity of the selected weather variables should be checked for accurate results.

The maximum distance between two observations of a particular cluster is the diameter of that cluster. Assume that X is the number of distances within clusters (all less than maximum diameter), and similarly, Y is the total number of distances less than the maximum diameter. If the ratio of X to Y is close to unity and corresponds to a number of clusters less than total number of observations, this will indicate a natural cluster (natural grouping).

Optimum number of clusters is an indication of number of load patterns necessary to forecast the load for any interval.

Clusters with single observations could be considered "Maverick" as long as they are adequately separated from closeby clusters.

12. CONCLUSIONS

A new linear, deterministic, and adaptive method of forecasting hourly load using a weather sensitive model is presented.

1. This method is intended for small area power systems due to limitation of geographical area, and applicability of weather parameters. In a large power system diversity of loads distort the weather sensitive pattern of load.
2. The fact that system uses a seasonal library treats the load as a seasonal effect which accomodates seasonal changes of load and is desirable.
3. Since the most recent pattern in each cluster is used for forecasting purposes, any kind of recent load management procedures such as pricing will automatically be accomodated.
4. The use of eight three-hour intervals with current weather parameters as input makes the algorithm highly adaptive to sudden weather changes such as passing of cold fronts in the winter time.
5. The fact that most recent patterns in the clusters are used for forecasting purposes, allows discarding of outdated repeated climatic conditions and helps to keep a moderate library size.
6. For even the hottest day, average accuracy of 2% (lead time one hour) was obtained. Considering the error present in the input of our data base, this accuracy is acceptable and comparable to many other existing methods.
7. If a lead time of 24 hours is used, forecasted weather parameters have to be employed in the clustering procedure, and accuracy of the forecast is expected to be affected to some degree.
8. Forecasting for weekends and legal holidays can be approached in exactly the same way. In this case load patterns for only weekends and legal holidays should be used in the seasonal library of weather and load data.
9. Method works best for geographical areas that have uniform weather patterns with well defined seasons.

13. APPENDIX

The following weather variables from NOAA were used in the study:

1. V_1 = Maximum temperature for the day
2. V_2 = Minimum temperature for the day
3. V_3 = Average temperature for the day
4. W = Average dewpoint for the day
5. U = Minutes of sunshine for the day
6. Z = Average windspeed for the day
7. H_1 = Maximum temperature for the previous day
8. H_2 = Minimum temperature for the previous day
9. H_3 = Average temperature for the previous day
10. F = Average dewpoint for the previous day
11. J = Minutes of sunshine for the previous day
12. X_1 = Air temperature at the period
13. X_2 = Wet bulb temperature at the period
14. X_3 = Dewpoint at the period
15. X_4 = Humidity at the period
16. X_5 = Windspeed at the period

REFERENCES

[1] P. C. Gupta and K. Yamada, "Adaptive Short Term Load Forecasting of Hourly Load Using Weather Information.", IEEE Transaction on Power Apparatus and Systems, Vol. PAS-91, pp. 2085-2094, 1972.

[2] P. P. Mathewman and N. Nicholson, "Techniques for Load Prediction in the Electricity-Supply Industry.", Proceedings of IEE, Vol. 155, No. 10, pp. 1451-1457, 1968.

[3] K. Srinivasan and R. Pronovost, "Short Term Load Forecasting Using Multiple Correlation Models.", IEEE Transaction on Power Apparatus and Systems, Vol. 94, No. 5, pp. 1854-1858, Sept./Oct., 1975.

[4] Robert C. Tryon and Daniel Baily, "Cluster Analysis", McGraw-Hill, 1970.

[5] Daniel Cuthbert and Fred S. Wood, "Fitting Equations to Data", John Wiley and Sons, 1980.

[6] G.T. Heinemann, D.A. Norman, and E.C. Plant, "The Relationship Between Summer Weather and Summer Loads - A Regression Analysis", IEEE Transaction on Power Apparatus and Systems, Vol. PAS-85, No. 11, pp. 1144-1154 Nov, 1966.

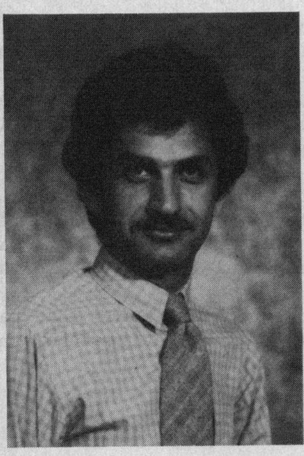
[7] J.T. Tou and R.C. Gonzalez, "Pattern Recognition Principles", Addison-Wesley, 1974.

[8] Edward C. Wolf, "A Dynamic Method for Evaluating Union Electric's Hourly Demand", EPRI Symposium on Load Forecasting, April, 1981.

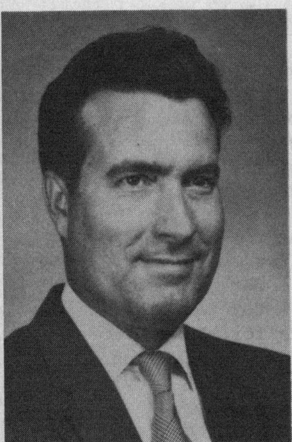
[9] B. Everitt, "Cluster Analysis", John Wiley and Sons, 1974.

[10] W. Meisel, "Computer Oriented Approaches to Pat-Recognition", Academic Press, 1972.

[11] SAS 79, SAS User's Guide, Statistical Analysis Systems, Cary, NC, 1979.



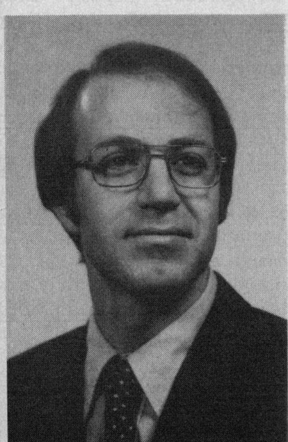
Abdolhosein S. Dehdashti was born in Iran on Jan. 1, 1956. He received his B.S.E.E. and M.S.E.E. in the area of power systems from the University of Missouri in 1978 and 1979 respectively, and is currently a PhD candidate in the same institution. His main areas of interest are power system planning and operation. Mr. Dehdashti is a member of IEEE.



James R. Tudor was born on March 26, 1922 at Fort Smith, Arkansas. He received the B.S.E.E. and the M.S.E.E. from the University of Missouri-Columbia in 1948 and 1950 respectively. He received the Ph.D. degree from Illinois Institute of Technology in 1960. His work experience includes approximately three years with Union Electric Company in St. Louis and the remainder as Asst. Professor, Assoc. Professor, and Professor of Electrical Engineering at the University of Missouri-Columbia.

From 1969 to 1974 he was Missouri Electric Utilities Professor of Power Systems. His technical specialties are in the power systems area with primary emphasis on computer applications.

His society memberships include: Eta Kappa Nu, Tau Beta Pi, Sigma Xi, MSPE and NSPE. He is a registered professional engineer in Missouri and a senior member of IEEE.



Michael C. Smith was born on May 31, 1949 in Winston - Salem, North Carolina. He received the B.S.I.E. and M.S.I.E. from the University of Tennesseee-Knoxville in 1971 and 1974 respectively. He received the Ph.D. from the University of Missouri-Columbia in 1977. His work experience includes assignments with the Buckeye Cellulose Corporation in Memphis and St. Mary's Medical Center in Knoxville, Tennessee. He has held academic appointments at Oregon State University and his present position as Assistant Professor of Industrial Engineering at the University of Missouri-Columbia. His technical interests include systems modeling and systems design applications in military, industrial and service industry environments. Dr. Smith is a member of Alpha Pi Mu and a senior member of A.I.I.E.