Paper accepted for presentation at 2009 IEEE Bucharest Power Tech Conference, June 28th - July 2nd, Bucharest, Romania

Long Term Multi-Scale Analysis of the Daily Peak Load Based on the Empirical Mode Decomposition

M.Ould Mohamed Mahmoud, F.Mhamdi, M.Jaïdane-Saïdane

***Abstract*—In this paper, an original technique to explore the long term load dynamics using a multi-scale analysis of the daily peak load based on the Empirical Mode Decomposition (EMD) is presented. The signal is decomposed into intrinsic oscillatory components called Intrinsic Mode Functions (IMFs). These modes are derived from the signal itself and not on a specific basis function. In this work, the EMD is used to extract and separate the suitable load component for long term forecast. Physical interpretations and statistical description of the modes are discussed. A comparison is made between load components extracted by the EMD approach and that of a classical multiple linear regression model. The load component predictability was investigated using the mutual information function. Real load data of the Tunisian power systems are used in this study.**

***Index Terms*—Load modeling, empirical mode decomposition, model regression, load analysis, predictability.**

I. Introduction

The load modeling and forecasting is an important aspect in the development of any model for electricity planning and for short term operation[1], [2], [5]. The load characteristics change as the factors affecting load (the day of the week, holiday periods, seasonal variations, the weather, socioeconomic trends.) change. Thus for mid and long term load forecasting it is of great interest to extract and isolate the seasonal effects on daily, weekly, and yearly time scales and trend from the underlying load series and treat them separately. For this we propose in this study to introduce an empirical analysis algorithm for extraction of the meaningful load component from the global consumption particularly to extract the trend load component.

The main motivation of the empirical modal decomposition (EMD) is to define decompositions of the load signals which do not use any predetermined bases (as the wavelet techniques, Fourier transform) which depend on the choice of a particular basis function. Furthermore, the EMD is an adaptive method which is entirely empirical and preserves the characteristics in the separate IMFS. It is particularly powerful for time series data that are nonstationary and non linear, explaining why it has been successfully applied in many engineering fields, e.g. [6], [8], [7].

The load data used in this study are represented by the daily peak load series (see Fig.1) recoded in 2000-2006.

This work was supported by the 2005/2007 VRR (Projet de Valorisation des Rsultats de Recherche du Ministère de la Recherche Scientifique de la Technologie et du Développement des Compétences) research project

M.Ould Mohamed Mahmoud, F.Mahamdi and M.Jaïdane-Saïdane are with the Signals and Systems Research Unit, Ecole Nationale d'Ingénieurs de Tunis (ENIT), Tunisia e-mail: (med.barrar@enit.rnu.tn, meriem.jaidane@planet.tn, mhamdifarouk@yahoo.fr).

978-1-4244-2235-7/09/$25.00 ©2009 IEEE

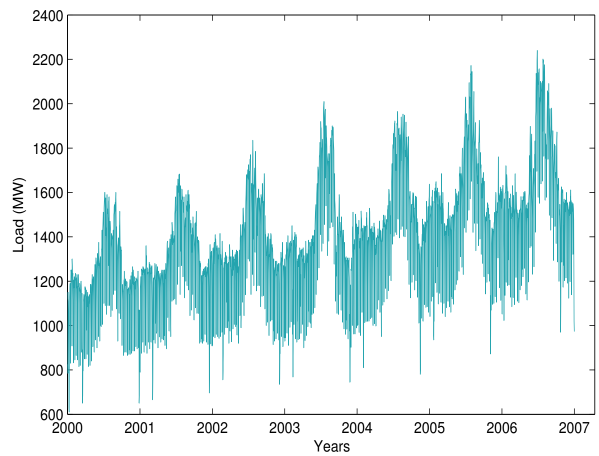


Fig. 1. The daily peak load evolution 2000-2006 from Tunisian power systems.

As shown in Fig. 1, the load data reflects as in others country [3] :

- seasonal patterns
- high, variability and fluctuations
- long-term trend

The purpose of this paper is to present the preliminary results relative to the application of the EMD technique to the Tunisian daily peak load data recorded in the period 2000 to 2006 (see Fig.1) which is inherently nonlinear and nonstationary time series. We propose after the IMFs construction a comparison between load components extracted by the EMD approach and those obtained by applying a classical multiple linear regression model. To compare the predictability of different load components, the mutual information function was used here.

II. Daily peak load decomposition derived from EMD

A. Empirical Mode Decomposition Technique

Empirical Mode Decomposition (EMD) has recently been introduced by Huang et al. [4], as an important alternative to traditional methods for analyzing time series such as wavelet methods, Fourier methods.

The concept of the EMD is to decompose data p_t into so-called intrinsic mode functions (IMF) [4] :

$$p_t = \sum_{k=1}^{K} IMF_k(t) + r(t) \quad (1)$$

- K is the number of IMFs
- IMF is defined as a function that satisfies the following two properties : 1. The mean value of the upper envelope defined by the local maxima and the lower envelope defined by the local minima must be zero; 2. The number of zero-crossings and the number of extrema are equal or differ at most by one.
- $r(t)$ denotes the final residue which can be considered as the trend component [15].

The algorithm for the extraction of IMFs from real load data is called sifting and it consists of the following steps [4], [6]:

i Initialize the residue $r_0(t) = p_t$ (p_t represent the daily peak load), set $g_0(t) = r_{k-1}(t)$ and $i = 1$; and index of IMF $k = 1$;

ii Construct the lower (minima of load data) $Imin_{i-1}$ and the upper (maxima of load data) $Imax_{i-1}$ envelopes of the signal by the cubic spline method.

iii Calculate the mean values m_{i-1} by averaging the upper envelope and the lower envelope :
$m_{i-1} = [Imax_{i-1} + Imin_{i-1}]/2$

iv Subtract the mean from the original signal
$g_i = g_{i-1} - m_{i-1}$ and $i = i+1$. and repeat steps (ii)-(iv) until g_i being an IMF. If so, the kth IMF is given by $IMF_k = gi$

v update residue $r_k(t) = r_{k-1}(n) - IMF_k$. This residual component is treated as a new data and subjected to the same process described above to calculate the next IMF_{k+1}.

vi Repeat the steps above until the final residual component becomes a monotonic function. Fig.2 illustrates the sifting process

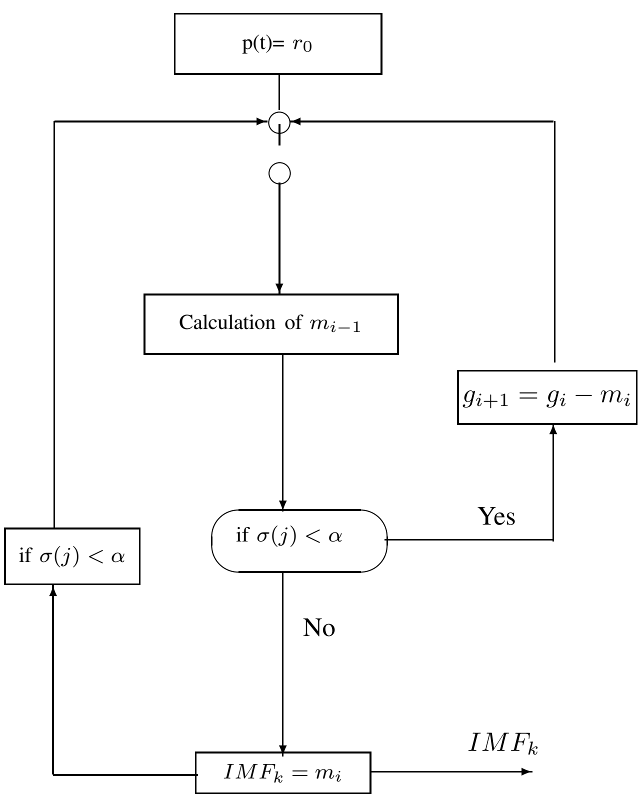


Fig. 2. *Sifting process to compute EMD components for the load data* $p(t)$

At the end of EMD the log load series p_t is represented as:

$$p_t = \sum_{k=1}^{K} IMF_k(t) + r(t) \tag{2}$$

The advantage of this method is the fact that the IMFs or (mode decomposition) which are generated by this method are strongly related to the load data $p(t)$.

B. Relevant parameters of the EMD algorithm

The EMD decomposition is sensitive to the choice of the method of interpolation and the choice of the stopping criterion

- Interoplation procedure to find lower and upper envelopes
To find local minimum and maximum points, a sliding window is moved through the time series. The minimum and maximum values within the window are picked out as the local minimum and maximum points. The interpolation to produce two envelopes from local maxima and local minima points can be done in different ways [4]. We here use the cubic splines interpolation method which has the advantage to be simple and sufficient [4].
- The stopping criterion for the sifting algorithm
In order to accomplish the second IMF condition, several criterions has been defined [10], [4].
In this paper we have used the criterion proposed in [10], which consists in comparing the amplitude of the mean of the upper and lower envelopes with the amplitude of the corresponding IMF. This criterion is based on two thresholds (θ_1 and θ_2) and a tolerance parameter (α).

$$\sigma(j) = |\frac{m_{i-1}(j)}{Imax_{i-1}(j) - Imin_{i-1}(j)}| \tag{3}$$

- $\sigma(j) \leq \alpha$, $(\forall j)$
- $\theta_1 \leq j\sigma(j) \leq \theta_2$

In this work we have used the default values proposed in [10] $\alpha = 0.05$, $\theta_1 = 0.05$ and $\theta_2 = 0.5$.

C. Daily peak load IMFs components

We have applied the empirical mode decomposition method described by the above algorithm to daily peak load from 2000 to 2006. Fig.3 shows the complete decomposition.

For physical interpretation of the obtained components, we calculate the mean period defined as the number of data samples divided by the total number of maxima [14] :

$$Mean\ \ period = \frac{data\ \ length}{number\ \ of\ \ local\ \ maximum} \tag{4}$$

the percentage of variance of each component defined as the ratio of variance between a specified IMF and the original signal $(\sigma^2_{IMFk(t)}/\sigma^2_{p(t)})$.
Table I shows the parameters values of the mean period and the variance provides the idea about the amount of information contained by each IMF components.

Based on the preliminary results of EMD decomposition of the daily peak load it possible to notice the following :

- IMF1 to IMF3 are the high frequency components and represents the very short term fluctuations demand related probably to difference between the working day and

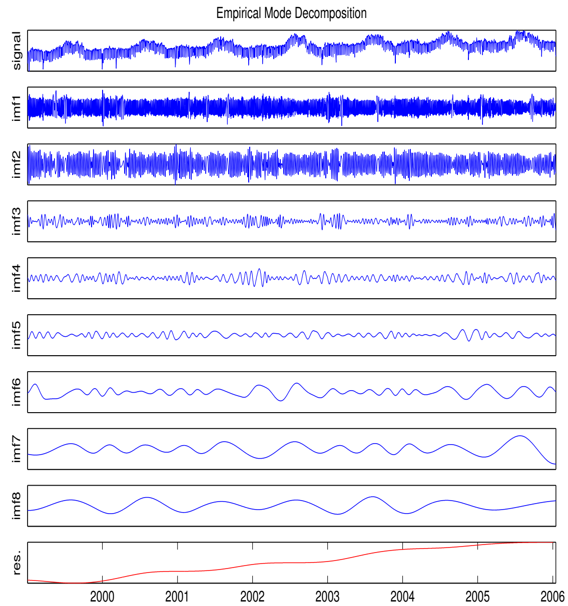


Fig. 3. EMD applied to Tunisian peak load series (2000-2006) to find the IMF (1-8) components and the final residue (signal trend).

TABLE I
AVERAGE PERIOD, IN DAYS, AND % OF VARIANCE OF THE LOAD COMPONENT (IMFS) OBTAINED WITH EMD ALGORITHM

imf	Mean period (days)	% of variance
imf1	3.49	0.35
imf2	7.00	0.38
imf3	14.20	0.13
imf4	24.60	0.13
imf5	50.13	0.10
imf6	116.22	0.18
imf7	232.45	0.25
imf8	426.16	0.23

weekend load data. The corresponding period is on the order of 3 days and one week and two weeks.

- IMF4 to IMF5 their period is between 1-2 months for these IMFs the percentage of variance is very small indicating that such IMFs are not significant.
- IMF6 has approximately 3 months periodicity oscillations, it allow to capture mid-term effects described by seasonal variations. As shown in Fig.2 The IMF7 has very regular oscillations on yearly time scales. So the peaks load in these fluctuations corresponds with the summer dates marked by a high temperature, and the troughs with winter where the temperature decrease.
- Trend : It is represented as the residue component in EMD, it could represent the major trend of load demand in the long term which may be related to economic growth in Tunisia.

III. EMD AND CLASSICAL MULTIPLE LINEAR REGRESSION MODELING

In this section we analyze the relationships between the IMFs peak load components of the Empirical Mode Decomposition and that extracted by the use of the Least Square modeling estimation. There are various models [13], [16], [17], [18], [19] for modeling and forecasting the load consumption applied in many countries.

In our case, Tunisian daily peak load demand (p_t) from 2000 to 2006 was used. This series presents the following characteristics :

- linear trend that related to the long run economic growth.
- weekly seasonal component
- annual seasonal component related to the industrial activities and temperature fluctuations.

To consider the seasonal variations independent seasonal dummy variables that capture the day of the week (w_{it}), holiday (h_t) and month (m_{jt}) effects are included in the model. The classical multiple linear model can be represented as folow :

$$\ln(p_t) = \sum_{i=2}^{7} \alpha_i w_{it} + \gamma h_t + \sum_{j=2}^{12} \beta_j m_{jt} + c + at + \varepsilon_t \quad (5)$$

$$= \sum_{i=1}^{2} SC_i(t) + T(t) + \varepsilon_t \quad (6)$$

where

- $SC_1(t) = \sum_{i=2}^{7} \alpha_i w_{it}$: represents the weekly seasonal load component.
- $SC_2(t) = \sum_{j=2}^{12} \beta_j m_{jt}$: annual seasonal component.
- $T = c + at$: trend load component.
- a, c, α_i, $i = 2..7$,β_j, $j = 2..12$ are the regression coefficients to be estimated.

The model was identified using sample data from 2000 to 2006. Linear regression has been performed using Eviews. Based on the Auto Correlation Function (ACF) and the Partial Auto-Correlation Function (PACF) of the residue, a first-order autoregressive process in the error term (eq. 4).

$$(1 - \phi_1 L)\epsilon_t = \xi_t \quad (7)$$

where L is the backward shift operator which is a special notation used to simplify the representation of lag values.

The parameter estimation is given at the table below : The

TABLE II
SUMMARY OF PARAMETERS ESTIMATION

par.	est. (10^{-3})	p-value	par.	est. (10^{-3})	p-value
$\widehat{c}$	6979	0	$\widehat{\beta}_5$	-15.68	0.02
$\widehat{a}$	0.13	0	$\widehat{\beta}_6$	15.25	0
$\widehat{\alpha}_2$	9.66	0	$\widehat{\beta}_7$	49.48	0
$\widehat{\alpha}_3$	9.64	0	$\widehat{\beta}_8$	46.22	0
$\widehat{\alpha}_4$	10.30	0	$\widehat{\beta}_9$	19.26	0.01
$\widehat{\alpha}_5$	4.07	0.03	$\widehat{\beta}_{10}$	0.60	0.04
$\widehat{\alpha}_6$	-48.53	0	$\widehat{\beta}_{11}$	-25.41	0.02
$\widehat{\alpha}_7$	-61.65	0	$\widehat{\beta}_{12}$	-12.40	0.04
$\widehat{\beta}_2$	-23.40	0	$\widehat{\gamma}$	-71.56	0
$\widehat{\beta}_3$	-37.92	0	$\widehat{\phi}_1$	88.30	0
$\widehat{\beta}_4$	-35.10	0			

$R^2 = 95.5\%$

determination coefficient (eq. 6) is calculated to determine the robustness of the model.

$$R^2 = \frac{\sum_{t=1}^{2557}(ln(\hat{p_t}) - ln(\bar{\hat{p_t}}))^2}{\sum_{t=1}^{2557}(ln(p_t) - ln(\bar{p_t}))^2} \quad (8)$$

where t indicates the day in the sample from 2000 to 2006. The results in table II illustrate the parameters estimation of the model. As expected, all the parameters are significant which improve that daily peak load time series contains a daily and annual seasonal components. It is of great interest to extract this seasonality and remove it from the underlying time series. In the section below we will investigate if the EMD, which can fully capture the local fluctuation of data, can be considered as an efficient method for seasonal extraction in complex signal. To evaluate its accuracy we have calculated the correlation coefficients between seasonal components and IMFs.

$$\rho_{T,P} = \frac{cov(T,P)}{\sqrt{cov(T,T)cov(P,P)}} \quad (9)$$

where cov is the covariance matrix defined as

$$cov(X,Y) = E[(X - E(X))(Y - E(Y))] \quad (10)$$

We consider

TABLE III
CORRELATION COEFFICIENTS BETWEEN SEASONAL COMPONENTS AND IMFS

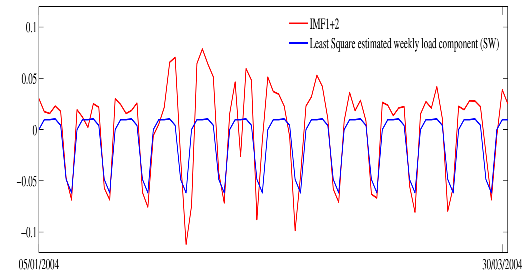
IMF	SW^{LS}	SA^{LS}
IMF1	0.42	0.00
IMF2	0.60	0.00
IMF3	0.01	-0.01
IMF4	0.00	-0.03
IMF5	0.01	0.02
IMF6	0.00	0.55
IMF7	0.00	0.50
IMF8	0.00	0.47

- $\widehat{SW}^{LS} = \sum_{i=2}^{7}\widehat{\alpha}_i w_{it}$: Least Square estimated weekly seasonal load component.
- $\widehat{SA}^{LS} = \sum_{j=2}^{12}\widehat{\beta}_j m_{jt}$: Least Square estimated annual seasonal load component.

Table III presents the computation of the correlations between the estimated weekly, annual seasonal variation and each IMF. As illustrated in the table III, we notice that there are significant correlations between the Least Square estimated daily and annual seasonal components and some IMFs. In particular, we notice the significant correlation between the estimated weekly seasonal load component and the IMF1, 2. However, it is possible to aggregate some IMFs to find better results (see table below)

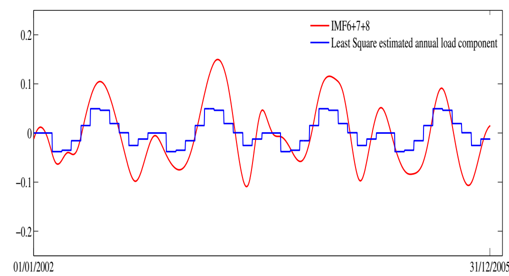
TABLE IV
CORRELATION BETWEEN SEASONAL COMPONENTS AND AGGREGATED IMFS

aggregated IMFs	SW^{LS}	SA^{LS}
IMF1+2	0.72	
IMF6+7+8		0.85



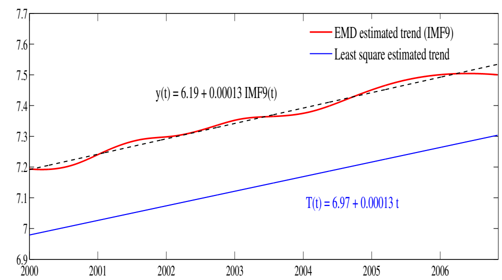
a.

IMFs 1-2 make up the weekly seasonal component



b.

IMFs 6-8 make up the annual seasonal component



c.

The EMD and Least Square estimated trends have the same slope

Fig. 4. Reconstructed peak load components by aggregation of IMFs

As shown in Fig.3, we can see that

- weekly seasonal component can be reconstructed by aggregation of the first and second IMFs (Fig.4.a). This is confirmed by the raised value of correlation coefficient equal to 0.72.
- the sum of IMF 6, 7 and 8, is very close with annual seasonal component estimated by the Least Square estimated linear regression model (Fig.4.b).
- the trend components have the same slope in the two case (Fig.4.c).

The results shown that the EMD gives an interesting load decomposition into IMF components with physical meaning related to time series characterizes. These are very interesting results since, unlike multiple linear regression model, this method is a non-model based approach to seasonal adjustment of time series. This allows method to capture the changes in the structure of the seasonal variation that can occur.

In the following section we propose to analyze the load component predictability. This predictability problem of daily peak load was the object of a previous study [11] based on measure of Lyapunov exponent. We propose now the use of mutual information function[9] combined with EMD load decomposition.

IV. MULTI SCALE PREDICTABILTY OF THE DAILY PEAK LOAD SERIES

Due to the different physical nature of information containing in each IMF, it is important to know how these components contribute to the original signal, and to measure their predictability degree. This can be done in measuring the lack of information by the mutual information function ($I_{yy}(\tau)$) which is a measure of the information that one can obtain from the past.

The mutual information function is defined as

$$I_{yy}(\tau) = \sum_{i}^{N} pr(y(i), y(i+\tau)) log \frac{pr(y(i), y(i+\tau))}{pr(y(i)).pr(y(i+\tau))} \quad (11)$$

wher N the sample size, pr is a measure of probability, $pr(x, y)$ is the joint probability.

$I_{yy}(\tau)$ represents the amount of information that one can know about signal values that are separated by the time lag τ. The mutual information function shows a minimum when the delayed signal is strongly statistically independent from the original signal. On the computation of $I_{yy}(\tau)$ see [9], [12].

The computed mutual information function for load components IMFs are shown in Fig.5. We notice that mutual function

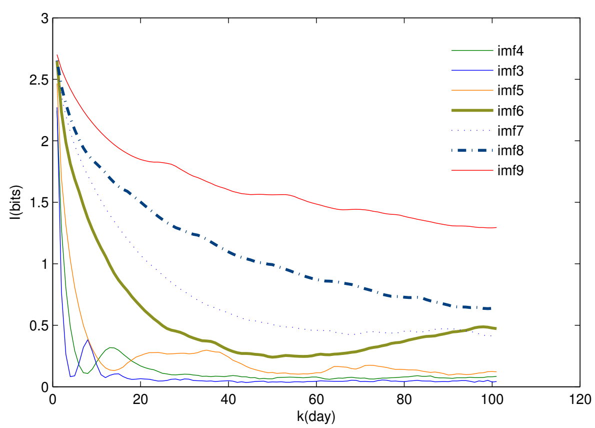


Fig. 5. Measure of load component predictability : The mutual information function computed for the obtained IMFs of the daily peak load 2000-2006.

decreases quickly for the higher frequency modes (IMF1 to IMF4). However, we notice that the mutual information decreases slowly in the last constituents (IMFS 5, 6, 7, 8 and 9). This decreasing is connected to long-term memory effect. Thus the IMF6 to IMF9 bring enough information for load dynamic characterization. For this reason the IMFs that seems to be more suitable for a long-term forecast. are the IMFS 5, 6, 7, 8 and 9.

V. CONCLUSION

The EMD method provide a new decomposition for long term load demand analysis. Especially, the IMFs components produced by the EMD method are usually physical and allow to capture long-run seasonality, short-run effects and trend effect. A comparative study between the EMD decomposition and a classical parametric method based on regression analysis was performed.

The mutual information function,which on can characterize the predictability, was used here to measure the load component predictability. In particular we concluded that there are some load components that are more adapted for average and long term load forecast and others less predictable for these horizons. These preliminary results show that the EMD seems to be a powerful approach, suitable for nonstationary load analysis.

REFERENCES

[1] A.R. Osareh, J. Pan, S. Rahman *"An efficient approach to identify and integrate demand-side management on electric utility generation planning"*, Electric Power Systems Research 36 (1996) 3- I1.

[2] B.H. Baily, J.R. Doty, R. Perez, R. Stewart, *"Evaluation of a demand side management photovoltaic system"*, IEEE Transactions on Energy Conversion 8 (4) (1993).

[3] Shamsuddin Ahmed *"Seasonal models of peak electric load demand", Technological Forecasting & Social Change"* 72 (2005) 609622.

[4] Huang, N.E., Shen, Z., Long, S.R., Wu, M.C., Shih, H.H., Zheng, Q., Yen, N., Tung, C.C., and Liu, H.H., *"The empirical mode decomposition and the Hilbert spectrum for nonlinear and nonstationary time series analysis"*, Royal Society London, 1998, p 903-995.

[5] Narcis Nabona, Csar Gil, and Jens Albrecht, *"Long-Term Thermal Power Planning at vew Energie Using a Multi-Interval Bloom and Gallant Method"*, IEEE Transactions on power Systems, VOL. 16, NO. 1, february 2001.

[6] Flandrin, P., Rilling, G., and Goncalves, P., *"Empirical Mode Decomposition as a Filter Bank"*, IEEE Signal Processing Letters, 2004, p 112 114.

[7] Nilanjan.S, Siddharth.S, and Paulo F. Ribeiro, *"An Improved HilbertHuang Method for Analysis of Time-Varying Waveforms in Power Quality"*, IEEE Transaction on Power systems, Vol. 22, NO. 4, Novemeber 2007.

[8] Ooninckx, P.J., and Hermand, J.P., *"Empirical mode decomposition of ocean acoustic data with constraint on the frequency range"*, Proceedings of the Seventh European Conference on Underwater Acoustics, Delft, 2004.

[9] H.P.Bernhard, G.Kubin, *"A fast mutual information calculation algorithm"*, Elsevier Science B.V., Amsterdam, 1994.

[10] G. Rilling, P. Flandrin, and P. Gonalvs, *"On Empirical Mode Decomposition and its Algorithms"*,IEEE-Eurasip Workshop on Nonlinear Signal and Image Processing NSIP-03, Grado (I) 2003.

[11] Ould Mohamed M.Mohamed, M.Jaïdane, J.Souissi, *"Varaiability of predictabilty of the daily peak load using Lyapunov exponent approach: case of Tunisian Power System"* IEEE Power Engineering Society, PowerTech07, Suizerlan 2-5 july 2007.

[12] A.M.Fraser, H.L.Swinney, *"Independent coordinates for strange attractors from mutual information"*, Physical Review A, vol.33, no.2, pp. 1134-1140, Feb.1986

[13] Ching-Lai Hor, Simon J. Watson, Shanti Majithia *"Daily load forecasting and maximum demand estimation using ARIMA and GARCH"*, 9th Probabilistic Methods applied to Power System Conference (PMAPS), June 2006.

[14] Haiyong Zhang Qiang Gai *"Research on Properties of Empirical Mode Decomposition Method"*, Proceedings of the 6th World Congress on Intelligent Control and Automation, June 21 - 23, 2006, Dalian, China

[15] Christopher D. Blakely *" Besov Spaces and Empirical Mode Decomposition for Seasonal Adjustment in Nonstationary Time Series"* Washington Statistical Society Seminars August, 2006.

[16] Hanjie Chen, Yijun Du, John N. Jiang *"Weather sensitive short-term load forecasting using knowledge-based ARX Models"*, IEEE PES General Meeting, 12-16 June 2005, 190-196 Vol. 1.

[17] S.Mirasgedis,Y.Sarafidis,E.Georgopoulou,D.P.Lalas, M.Moschovits, F.Karagiannis, D.Papakonstantinou *"Models for Mid-Term Electricity Demand Forecasting Incorporating Weather Influences"*, Energy 31 (2006) 208-227.

[18] Patrick E. McSharry, Sonja Bouwman, Gabriël Bloemhof *" Probabilistic Forecasts of the Magnitude and Timing of Peak Electricity Demand"* IEEE Transaction on Power Systems, Vol. 20, No. 2, May 2005

[19] N.S.Sisworahardjo, A.A.El-Keib,J.Choi, J.Valenzuela, R.Brooks, I. El-Agtal *"A Stochastic Load Model for an Electricity Market"*, Electric Power Systems Research Electr. vol 76(6-7):500-508.

M.Ould Mohamed Mahmoud He obtained the Diplôme d'ingènieur and MS degree in electrical engineering from Ecole Nationale d'ingènieurs de Tunis (ENIT) in 2004. He is currently pursuing the Ph.D degree. He cooperates with the Direction of Studies and Planning (DEP) in Tunisian Company of Electricity and Gas, His research interests include load modeling, Load predictability, probabilistic methods and its applications in power systems and Multiscale analysis.

Farouk Mhamdi Received the Engineer degree in Statistic and Data Analysis from the Ecole Supérieure de Statistiques et de l'Analyse de l'Information de tunis (ESSAIT), Tunisia, in 2005, and the Master's degree in econometric from the Ecole Polytechnique de Tunis, Tunisia, 2006. He is currently working towards the Ph.D. degree in Applied Mathematics. He is a Member of the Unit Signaux et Systèmes (U2S), ENIT.

Merieme Jaïdan Received the M.S. degree in electrical engineering from the Ecole Nationale d'Ingnieurs de Tunis (ENIT), Tunisia, in 1980. From 1980 to 1987, she worked as a Research Engineer at the Laboratoire des Signaux et Systmes, CNRS/Ecole Suprieure d'Electricitè, France. She received the Doctorat d'Etat degree in 1987. Since 1987, she was with the ENIT, where she is currently a Full Professor at Communications and Information Technologies Department. She is a Member of the Unit Signaux et Systmes, ENIT.