RESEARCH ARTICLE

Optimization of energy mix in a smart grid using knapsack problem and forecasts based on HMM.

* Hasnaa Bazine, Mustapha Mabrouki.
Industrial Engineering Laboratory, Faculty of Science and Technologies, Sultan Moulay Slimane University, Beni Mellal, Morocco.

Manuscript Info	*Abstract*
	The energy activities of mankind make it necessary to consider our resources, showing the proportion of renewable energy that allows us to consider a real sustainable development. Renewable resources are an extremely abundant deposit. However, they are mostly fluctuating. Network managers must therefore develop renewable generation forecasting tools to better manage the balance production / consumption and thus improve the integration of this production into networks. It is about how much of the maximum total power given to renewable energy compared to the demand. In this paper, we will optimize the energetic mix through modeling the total resource share of energy using the Knapsack problem, then we'll predict the fluctuating variables namely consumption and renewable origin production using the Hidden Markov Model.

Introduction:-
To optimize is to seek the best compromise between needs and resources. In our case, the resources are the various sources of energy, the need is energetic uses and the best compromise is obtained by minimizing the cost, ie using maximum renewable energy. This distribution of different sources of primary energy consumed in the production of different types of energy is called the energetic mix.

When the problem of data takes unique values known by decision makers, it is called deterministic optimization. Which is not the case for the power grid, where the unpredictability of production from renewable energies and sometimes that of demand also implies a stochastic optimization.

The smart power grid is therefore a typical example of a complex phenomenon given various factors (the heterogeneity of the actors, the material aspect and the divergent economic issues…). To study these complex systems, modeling method represents a perfect and an essential tool. In general, the biggest challenge of smart grids is to find the best energetic mix, that is to say an effective way to share the energetic resources while ensuring the balance between production and consumption. This work aims to model the total share of energy based on the Knapsack problem, then we'll predict the fluctuating variables namely consumption and renewable origin production using the Hidden Markov Model in order to optimize the energetic mix and ensure an optimal demand response. In our work, the objective is to maximize the supply or production without exceeding the requested amount. By the way, we will optimize the welfare function by maximizing the utility function, while taking into account the random nature of consumption and renewable origin production, which are governed by many factors. Consumption, prediction and modeling the sharing of energy resource will be studied using many accurate algorithms in this work.

Material and Methods:-

In this paper, we will consider three sources of supply that is renewable, supply one day in advance and the energy purchased for balancing in real time.

Our work will propose a methodology, based on the knapsack problem and Hidden Markov Model, which will enable us firstly to use the maximum renewable energy despite its randomness, and also to buy in advance one day enough energy to minimize the instantaneous balancing energy. To achieve these objectives, the forecast electricity demand and production from renewable energy sources is essential. Generally, optimization tools are based on algorithms as, fuzzy logic, neural network or another. Our approach apply the knapsack problem to model the total share of energy used in order to ensure an optimal demand response, and Hidden Markov Model to forecast electricity demand and production from renewable energy sources. In order to validate our model, we have chosen to focus on two different databases. The first one concerns historical consumption data of 2013 of the faculty of science and technology of Beni Mellal. This database, which consists of 365 values, represents the daily consumption of the institution for 2013. While the second one concerns the production of renewable energy for 2015 of twelve amorphous panels installed in the same faculty.

Result and Discussion:-

The total share of energy:-

We consider a set of users served by a utility or power company. Each user has a set of appliances. The energy management is to decide how much each user must be able to consume. The company decides how much capacity it needs to get a day in advance, and when renewable production is achieved, how many balancing energy must be purchased in order to meet global demand. Energy sources considered are:

- ❖ Renewable sources: Pr(t)
- ❖ Supply one day in advance: Pd(t)
- ❖ Instant balancing energy: Pb(t)

The sharing of energy can therefore be seen as a multi-dimensional knapsack problem in its stochastic version.
The Multidimensional Knapsack Problem (MKP) is a NP-hard problem which has many practical applications, such as processor allocation in distributed systems, cargo loading, or capital budgeting. The goal of the MKP is to find a subset of objects that maximizes the total profit while satisfying some resource constraints. [1]

User model:-

Each user has an appliance that operates with pi(t) probability, which reached the $U_i(q_i(t))$ utility when consuming $q_i(t)$.

$$\underline{q}(t) \le q_i(t) \le \overline{q}_i(t) \tag{1}$$

$$\sum_t q_i(t) \le Q_i$$

Company model:-

$$Pb(t) = D(t) - Pr(t) - Pd(t) \tag{2}$$

The users are numbered by the index i varying from 1 to n. The qi number represents the user number i consumption. The ability of our bag will be denoted Q which represents the maximum consumption.

$$D(t) = \sum_{\{i,\,x_i=1\}} q_i(t) = \sum_i \delta_i q_i(t) \tag{3}$$

δi = 1 with pi(t) and δi = 0 with 1 − pi(t) pi(t) is the probability that the user operates the appliance.

The objective function or utility function:

$$U_i(q_i) = \sum_i \delta_i U_i(q_i) \tag{4}$$

Our goal here is to minimize the amount Pb(t), which is usually expensive, which is why we must predict the amount D(t) and Pr(t) in order to have a sufficient quantity Pd(t).

The forecast of consumption:-
Electricity consumption is governed by many factors, made of complicated and large dynamic systems. The prediction of electricity consumption is of great economical interest for players on the global electricity market, since an accurate prediction of the consumption is needed to obtain the best prices on the day-ahead market and to avoid purchasing on the more expensive real-time market. The demand forecasting models can be classified depending on many criteria, Hernandez et al [2] classified demand forecasting models according to two different criteria: the forecasting horizon and the aim of the forecast. The most important forecasting horizons are weekly, daily and hourly, and the aims of the forecast can be based on the number of values to predict, where we find forecasts with only one value (next hour's load, next day's peak load, next days total load, etc.); and forecasts with multiples values, such as next hours, peak load plus another parameter (for example, aggregated load) or even next days hourly forecast- the so-called load profile.

A growing body of literature exists on the topic of energy consumption or demand forecasting. For example Binh.p et al [3] have developed an approach for demand forecasting by combining Hidden Markov Model and Bayesian method. Teixeira and Zaverucha [4] presents a new hybrid system that merges Fuzzy logic with dynamic Bayesian Network, called the Fuzzy Hidden Markov Predictor. Almeshaiei and soltan [5] presents a methodology based on decomposition and segmentation of the load time series, while Zhang and Wang [6] presents a Fuzzy Wavelet Neural Network approach for annual electricity consumption in high consumption city.

Many contributions consist of presenting a methodology, and showing its accuracy on a simulated dataset. The element that was the most forecasted is the electricity price where many models in the literature were presented by H.V. Haghi et al. [7], some are based on the game theory which takes the analysis of Nash equilibrium as a key point. The other simulation models, also called structural or fundamental models, try to consider not only production cost but also the agents behaviors impact on power market. There are also the statistical models known as black-box models that carry out price evolution analysis from a statistical point of view without examining the underlying physical details of the system. Our objective here will be to predict energy demand based on the history, our demarche is built on the hidden Markov model.

Hidden Markov Model:-
The hidden Markov Model has been applied to many forecasting problems, for example, a Hidden Markov Model approach for forecasting stock price for interrelated markets was presented by M. Hassan and B. Nath [8]. HMM's have been extensively used for pattern recognition and classification problems because of its proven suitability for modeling dynamic systems. L. Long et al. Used the HMM to forecast the bidding behavior of advertisers using the historical auction data of advertisers [9]. And S. Nootyaskool and W. Choengtong used the HMM to predict foreign exchange rate [10] ; the prediction model here, uses four factors, dollar index, interest rate, inflation rate and economic growth. The main idea of this work is a technique of encoding four factors into one observation sequence to train HMM.

A Hidden Markov Model (HMM) is a first order Markov Chain, where each state is related to an observation of the data set via a conditional distribution. In general, the unknown elements are the states, the transition matrix and the initial probabilities of the chain, but their values can be calculated with the use of the conditional distributions and the given data set ; this fact makes the chain and its related stochastic process known as hidden. In general, a HMM is characterized by the following parameters. [11]

a. N : number of states in the model, to choose this number, in our case, we will use the Bayesian Information Criterion (BIC) ;
b. M : number of of distinct observation symbols per state
c. O=O1,O2,. . . ,OM the observation sequence ;
d. A : N*N transition matrix of the Hidden Markov Chain. By definition of Markov Chain, we have:

$$a_{i,j} = P\left(X_{t+1} = j \backslash X_t = i\right) \qquad i, j = 1, \ldots, N \qquad (5)$$

$$(6)$$

$$\sum_{m'}^{M} a\left(m, m'\right) = 1$$

Where Xt is the hidden process which describes the state visited at time t.

e.   b : a set of N emission probability density functions. Given the observation O at time t, we have:

$$bj(o) = p(Yt = o \backslash Xt = j) \qquad j = 1, \ldots, N \qquad (7)$$

$$\sum_{j=1}^{i} b(i, j) = 1$$

Where Yt is the observable process at time t.

f.   π initial state distribution {πi}. Where λ = (A, B, π) is the overall HMM model.

Bayesian Information Criterion:-

The Bayesian Information Criterion (BIC), also known as the Schwarz criterion, is a model selection criterion among a finite set of models for describing a dataset. The selected model is the one with the lowest BIC. This criterion is based, in part, on the likelihood function The Bayesian Information Criterion is derived from Akaike information criterion proposed by Gideon Schwarz in 1978. But unlike the Akaike information criterion, the penalty for the BIC depends on the size of the sample and not only the number of parameters. The Bayesian Information Criterion is given by the following equation:

$$BIC = -2 \ln \hat{L} + k \ln(n) \qquad (8)$$

$$\text{With } \hat{L} = p\left(x \mid \hat{\theta}, \lambda\right)$$

Where:
- ❖ x= observed data;
- ❖ θ= the parameters of the model;
- ❖ n=the number of data points in x, the number of observations;
- ❖ K= the number of free parameters to be estimated;
- ❖ $\hat{L}$ = the maximized value of the likelihood function of the model λ, to calculate it we will use the EM algorithm;
- ❖ $\hat{\theta}$ = the parameter values that maximize the likelihood function ;

Expectation Maximization (EM) algorithm:-

The EM algorithm is an efficient iterative procedure to compute the Maximum Likelihood (ML) estimate in the presence of missing or hidden data. In ML estimation, we wish to estimate the model parameter(s) for which the observed data are the most likely. Each iteration of the EM algorithm consists of two processes: The E-step, and the M-step. In the expectation, or E-step, the missing data are estimated given the observed data and current estimate of the model parameters. This is achieved using the conditional expectation, explaining the choice of terminology. In the M-step, the likelihood function is maximized under the assumption that the missing data are known. The estimates of the missing data from the E-step are used in lieu of the actual missing data. Convergence is assured since the algorithm is guaranteed to increase the likelihood at each iteration.

Considering a sample x = (x1, . . . , xn) of individuals in a law f(xi , θ) set by θ, we seek to determine the θ maximizing the log-likelihood given by:

$$L(x, \theta) = \sum_{i=1}^{n} \log f(x_i, \theta) \qquad (9)$$

Modeling:-

➔ Let S be all outputs or possible observations with S = {S1, S2,…, St − 1} observations of demand.

➔ the request of the next day is considered as the target parameter and the unknown sought to predict, from the previously sequence.

➔ Let E be the set of possible states E = {E1, E2, . .., Em}

➔ The transition probability distribution matrix between states A = {aij, ≤ i, j ≤ N} with aij = P(the state j at t-1/ the state i at t).

➔ The probability distribution matrix of observation symbols B = {bij, 1 ≤ i ≤ N} with bij = P(observation j at t/ state i at t).

➔ The probability distribution of the initial state matrix I = {Ii , 1 ≤ i ≤ N}, Ii = P{q = Si} which defines the probability that the hidden state I is the initial state. The actual load curve of consumption of the Faculty of Science and Technology of Beni Mellal is presented in Fig 1.
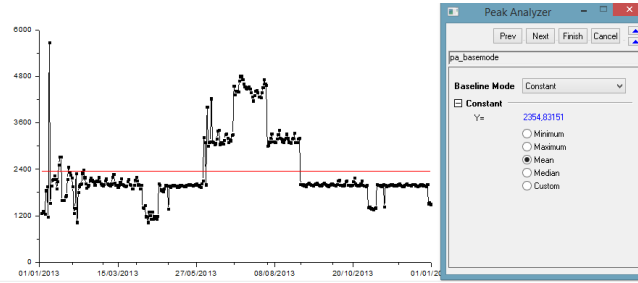
Figure 1:- Real consumption for 2013

Descriptive statistics:-
For all of the observed values, the mean and the standard deviation are calculated.

Table I:- The mean and the standard deviation for consumption data.

Observations	Obs. with missing data	Obs. without missing data	Minimum	Maximum	Mean	Standard deviation
365	0	365	905,588	5677,932	2347,629	861,790

To represent the distribution of data, we built a histogram, as we see in Fig 2, and for each of its intervals we calculated the frequency and density in Table 2.

Table II:- Frequency and density for consumption intervals.

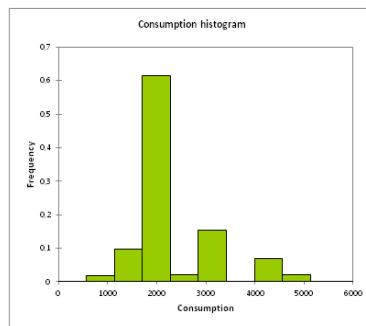| Lower bound	Upper bound	Effective	Frequency	Density
0	570	0	0,000	0,000000
570	1140	7	0,019	0,000034
1140	1710	36	0,099	0,000173
1710	2280	224	0,614	0,001077
2280	2850	8	0,022	0,000038
2850	3420	56	0,153	0,000269
3420	3990	0	0,000	0,000000
3990	4560	25	0,068	0,000120
4560	5130	8	0,022	0,000038
5130	5700	1	0,003	0,000005

Figure 2:- Consumption histogram for 2013

According to the histogram in Fig 2, it is observed that most of the values are between 1710 Kwh and 2280 Kwh. Initially, we consider a three-dimensional Gaussian distribution as the observation probability density function:

$$bj(S) = \sum C_{jm} \phi \left[S, \mu_{jm} U_{jm} \right] \quad , 1 \le j \le N \tag{10}$$

With:
- ❖ S the observation vector modeled.
- ❖ C_{jm} mixture coeff. for the m -th mixture in state j
- ❖ μ_{jm} mean vector for the m-th mixture component state j
- ❖ U_{jm} la covariance matrix for the m-th mixture component in state j

❖   Φ Gaussian density, where $\sum_{m=1}^{M} C_{jm} = 1$

Each observation is generated by the probability density function associated with one state to no time.
The probability of transition from state m to state m' m' is equal to:

$$a(m, m') = p(S_t = m' / S_{t-1} = m) \tag{11}$$

With $\sum_{m'=1}^{M} a(m, m') = 1$

The likelihood that the model is in state m at the initial time is equal to:

$$I(m) = P(S_0 = m) \tag{12}$$

With $\sum_{m=1}^{M} I(m) = 1$

The likelihood that the model emits when the observation j in state i:

$$b(i, j) = P(y_t = j / s_t = i) \tag{13}$$

With $\sum_{j=1}^{N} b(i, j) = 1$

The weighted average of the Gaussian mixture will be the prediction of observation of demand the following day:

$$S_{t+x} = \sum_{j=1}^{N} Ax(i, j) E(bij) \tag{14}$$

With:
❖   E(bij) : the expectation of the probability density function
❖   Ax(i, j) : distribution probability matrix of state transition from state i to j for x steps.

It is important to select the number of states. Increasing the number of states give place to a model with not only higher approximation capabilities but also with higher risk of over-fitting and computational effort. Moreover, models with too many states are also more different to interpret. In this experiment, different architectures have been adjusted in order to select the size of the model. After using the Bayesian Information Criterion we obtain the following results.

Table III:- The results for the models of BIC criterion for the case of consumption.

Model/Nb of states	2	3	4
V	-5292,144	-5272,838	-5159,139
E	-5992,501	-5852,630	-5864,430

The model V means here that we work with variable variance, while the model E means that the variances are equal for all states.
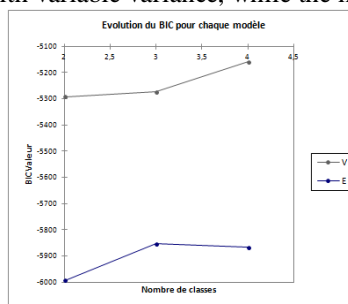

Figure 3:- The evolution of BIC criterion

According to BIC, the best mixture model is the V (variable variance) with 4 components. The EM algorithm has converged in 68 iterations. Therefore, the algorithm should be started with a maximum number of class over 4 which is the optimal number of states.

The transition path of the hidden states during all the period is shown in Fig 4.
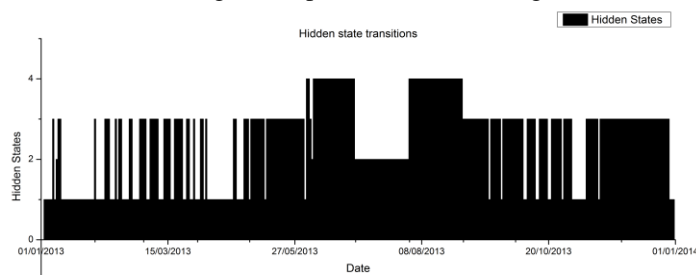

Figure 4:- Transition of hidden states.

By applying the set of equations of the model on our historical data the following prediction curve in Fig 5 is obtained.
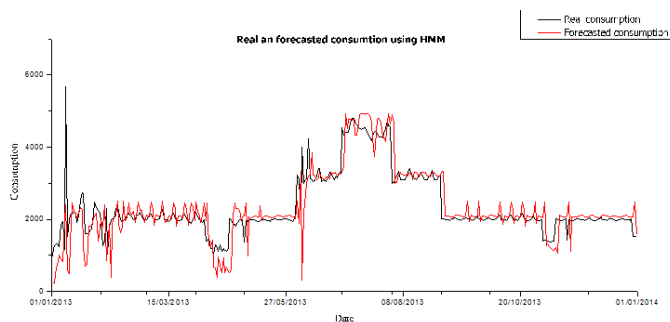

Figure 5:- Forecasted and real consumption using HMM for 2013

We notice that the consumption curve planned perfectly follows the real curve, except in a few points. We see that these points are characterized by excessive consumption. We therefore deduce that the model is correct given the chosen parameters; however, it is necessary to develop learning algorithms that will allow adjusting model parameters automatically for a better match between the actual and the expected. The parameter set of probabilities λ = {A, B, Π} is trained to fit the model at the training stage. The parameters are updated so as to achieve the best adaptation to the specific model.

The forecast of renewable production:-
In order to predict the renewable production we will use the same method as the case consumption. We will use the Hidden Markov Model to forecast the daily renewable production.

We will use the production of renewable energy of twelve amorphous panels, NXPower-AF model form of two parallel rows of six panels each, as a case of study.

Descriptive statistics:-
For all of the observed values, the mean and the standard deviation are calculated.

Table IV:- The mean and the standard deviation for production data.

Observations	Obs. with missing data	Obs. without missing data	Minimum	Maximum	Mean	Standard deviation
273	0	273	1,666	12,574	9,193	2,671

For renewable production data also we built a histogram, as we see in Fig 6, and for each of its intervals we calculated the frequency and density in Table 5.

Table V:- Frequency and density for production intervals.

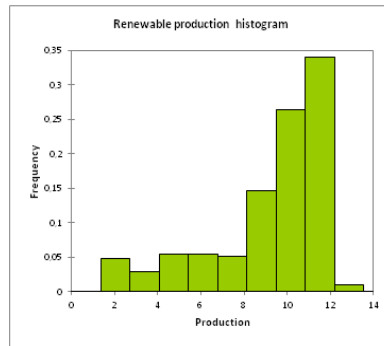| Lower bound	Upper bound	Effective	Frequency	Density
0	570	0	0,000	0,000000
570	1140	7	0,019	0,000034
1140	1710	36	0,099	0,000173
1710	2280	224	0,614	0,001077
2280	2850	8	0,022	0,000038
2850	3420	56	0,153	0,000269
3420	3990	0	0,000	0,000000
3990	4560	25	0,068	0,000120
4560	5130	8	0,022	0,000038
5130	5700	1	0,003	0,000005

Figure 6:- Production histogram.

According to the histogram in Fig 6, it is observed that most of the values are between 9,5018 Kw and 10,850 Kw. After using the Bayesian Information Criterion we obtain the following results.

Table VI:- The results for the models of BIC criterion for the case of production.

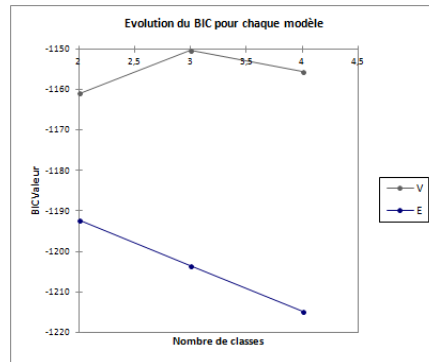| Model/Nb of states	2	3	4
V	-1160,935	-1150,384	-1155,536
E	-1192,375	-1203,586	-1214,798

Figure 7:- The evolution of BIC criterion.

According to BIC, the best mixture model is the V (Variable variance) with 3 components. The EM algorithm has converged in 287 iterations. Therefore, the algorithm should be started with a maximum number of classes over 3 which is the optimal number of states.

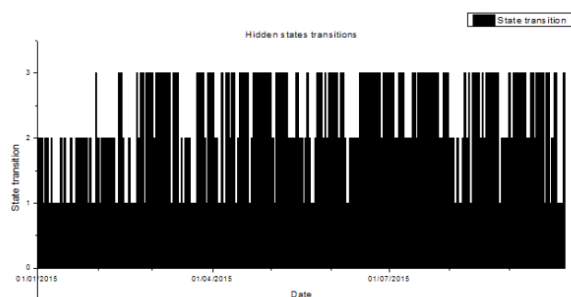The transition path of the hidden states during all the period is shown in Fig 8.

Figure 8:- Transition of hidden states

By applying the set of equations of the model on our historical data the prediction curve in Fig 9 is obtained.
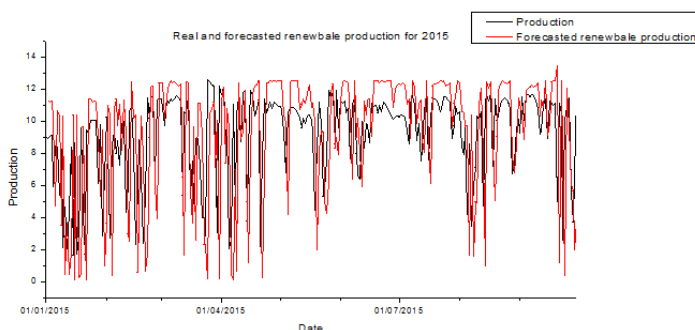

Figure 9:- Forecasted and real renewable production using HMM for 2015

We note that the production curve is under the same shape as that of the actual production. However, the two curves do not always have the same amplitude. This problem must be solved by adapting the model parameters, ie using learning algorithms.

Out-of-sample test:-
When testing a model on historical data, it is beneficial to reserve a time period of historical data for testing purposes. The initial historical data on which the model is tested and optimized is referred to as the in-sample data. The data set that has been reserved is known as out-of-sample data. This setup is an important part of the evaluation process because it provides a way to test the idea on data that has not been a component in the optimization model. As a result, the idea will not have been influenced in any way by the out-of-sample data and we will be able to determine how well the system might perform on new data

We reserved seven months of 2014 for this purpose, and after having applied the model to the data we obtained the curve in Fig 10.
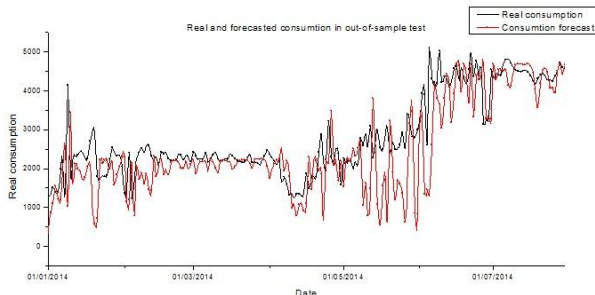

Figure 10:- Forecasted and real consumption in out-of-sample test

We note that the consumption curve provided following the real consumption in the out-of-sample test also.

Reliability of predictions:-
The forecasting error analysis depends on their extent. This is to compare the prediction to an observed statistical series. This calculation, seemingly simple, in fact raises two problems: what achievements should I choose? What statistical tools should be used to quantify the errors? For example if we take the case of consumption, we calculate the Mean Absolute Percentage Error (MAPE). This error expresses the accuracy as a percentage. The equation to calculate the MAPE error is:

$$\sum \frac{|y_t - y'_t|}{n} \times 100 \qquad \text{With} \quad (y_t \neq 0) \tag{17}$$

The forecast mean absolute percentage error in our case is 11.09%. The curve of MAE error is presented in Fig 11.
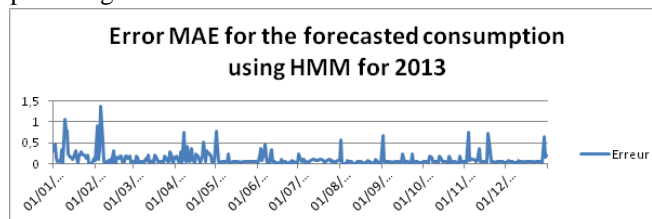

Figure 11:- Error MAE for the forecasted consumption using HMM for 2013

Conclusion and perspectives:-
In this paper, we have modeled the total share of energy using the Knapsack problem, then we predicted the fluctuating variables namely consumption and renewable origin production using the Hidden Markov Model in order to ensure an optimal demand response. In our model we took into consideration only the historical data of consumption and renewable production. Although it is a fairly pertinent parameter, it is still insufficient to model the problem with precision. So as expected, our model provides curves that perfectly follow the real curves, with amplitude default in some points. This problem can be due to the absence of other factors affecting consumption or electric production. In the following works, we aim to realize a detailed study including relatively all possible factors. The next work will examines the relationship between electricity demand or renewable production and many other related factors.

References:-
1. **I. Alaya, C. Solnon and K. Ghédira**, "Ant algorithm for the multidimensional knapsack problem", International Conference on Bioinspired Optimization Methods and their Applications (BIOMA), 2004.
2. **L. Hernandez, C. Baladron, J. Aguiar, B. Carro, A. Sanchez-Esguevillas, J. Lloret and J. Massana**, "A Survey on Electric Power Demand Forecasting: Future Trends in Smart Grids, Microgrids and Smart Buildings", IEEE Communications Surveys & Tutorials, vol. 16, no. 3, pp. 1460-1495, 2014.
3. **P. Binh, T. Loan and T. Khoa,** "Application of Hidden Markov and Bayes for Demand Forecasting", 2006 IEEE Power India Conference.
4. **M. Teixeira and G. Zaverucha**, "Fuzzy hidden Markov predictor in electric load forecasting", 2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No.04CH37541).
5. **E. Almeshaiei and H. Soltan,** "A methodology for Electric Power Load Forecasting",Alexandria Engineering Journal, vol. 50, no. 2, pp. 137-144, 2011.
6. **P. Zhang and H. Wang,** "Fuzzy Wavelet Neural Networks for City Electric Energy Consumption Forecasting", Energy Procedia, vol. 17, pp. 1332-1338, 2012.
7. **Haghi, H. V. and S. M. Tafreshi** (2007). An overview and verification of electricity price forecasting models. Power Engineering Conference, 2007. IPEC 2007. International, IEEE.
8. **M. Hassan and B. Nath,** "Stock market forecasting using hidden Markov model: a new approach", 5th International Conference on Intelligent Systems Design and Applications (ISDA'05), 2005.
9. **Lili Long, Hongbin Dong, Yue Pan, Li Huangfu, Naikang Gou and Xingmei Wang,** "Forecasting model for bidding behavior of advertisers based on HMM", 2015 International Joint Conference on Neural Networks (IJCNN), 2015.
10. **S. Nootyaskool and W. Choengtong,** "Hidden Markov Models predict foreign exchange rate", 2014 14th International Symposium on Communications and Information Technologies (ISCIT), 2014.
11. **S. Rebagliati, E. Sasso and S. Soraggi,** "Market forecasting using Hidden Markov Model", arXiv, 2015.