NONLINEAR DATA MODELING METHODS
FOR MULTIDIMENSIONAL SIGNAL ANALYSIS

Mehmet FİDAN

Ph. D. Dissertation

Graduate School of Science
Electrical and Electronics
Engineering Program

April-2015

JÜRİ VE ENSTİTÜ ONAYI

Mehmet Fidan'ın **"Nonlinear Data Modeling Methods for Multidimensional Signal Analysis"** başlıklı **Elektrik-Elektronik Mühendisliği** Anabilim Dalındaki, Doktora Tezi 27.03.2015 tarihinde, aşağıdaki jüri tarafından Anadolu Üniversitesi Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliğinin ilgili maddeleri uyarınca değerlendirilerek kabul edilmiştir.

		Adı Soyadı	İmza
Üye (Tez Danışmanı) :		**Prof. Dr. ÖMER NEZİH GEREK**	**……..**
Üye	**:**	**Prof. Dr. VAKIF CAFER**	**……..**
Üye	**:**	**Doç. Dr. RİFAT EDİZKAN**	**……..**
Üye	**:**	**Doç. Dr. FATİH ONUR HOCAOĞLU**	**……..**
Üye	**:**	**Doç. Dr. HAKAN ÇEVİKALP**	**……..**

Anadolu Üniversitesi Fen Bilimleri Enstitüsü Yönetim Kurulu'nun …….. tarih ve …….. sayılı kararıyla onaylanmıştır.

Enstitü Müdürü

ABSTRACT

Ph.D. Dissertation

Nonlinear Data Modeling Methods for Multidimensional Signal Analysis

Mehmet FİDAN

Anadolu University
Graduate School of Sciences
Electrical and Electronics Engineering Program

Supervisor: Prof. Dr. Ömer Nezih GEREK
2015, 80 Pages

In this dissertation, various novel stochastic and deterministic nonlinear data models are proposed for the analysis of the discrete signals that are defined in multidimensional spaces. The purpose of choosing the multidimensional space is to analyze the data according to multiparameters at the same time. The reason of analysing nonlinear methods is their compatibility with chaotic and nonlinear behaviour of the data which are measured for natural events and in this thesis natural events are taken as case study. The comparison of stochastic and deterministic methods gives the opportunity to choose the most suitable model for the data of handled problem. As stochastic models, one and multidimensional versions of Mycielski method and different versions of Markov Chain Models are proposed. As deterministic models, multidimensional polynoms, multidimensional splines, multidimensional Empirical Mode Decomposition and Wavelets are chosen. In addition a Markovian error tuning model is designed as an infrastructure to test these models, which is inspired from time varying and time invariant versions of the Hidden Markov Model. These comparative works try to reveals the phenomenon underlies the natural events as wind speed, solar radiation, temperature which are taken as case study in this work.

Keywords: Data Modeling, Time Varying Models, Time Invariant Models, Mycielski, Markov Chains, Hidden Markov Models

ÖZET

Doktora Tezi

Çok Boyutlu İşaretlerin Analizi için Doğrusal Olmayan Veri Modelleme Yöntemleri

Mehmet FİDAN

Anadolu Üniversitesi
Fen Bilimleri Enstitüsü
Elektrik-Elektronik Mühendisliği Anabilim Dalı

Danışman: Prof. Dr. Ömer Nezih GEREK
2015, 80 Sayfa

Bu tezde, çok boyutlu uzaylarda tanımlanmış ayrık işaretler için çeşitli yenilikçi olasılıksal ve belirlenimci veri modelleme yöntemleri önerilmiştir. Çok boyutlu uzayın seçilmesinin amacı, ilgili veriyi aynı anda birden fazla parametreye bağlı olarak inceleyebilmektir. Doğrusal olmayan modellerin incelenmesinin nedeni, bu modellerin gerçek hayat için ölçülmüş verilerin karmaşık ve doğrusal olmayan davranışıyla uyumudur ve bu tezde de bu tür gerçek olay davranışları örnek olay incelemesi olarak ele alınmıştır. Olasılıksal modeller olarak Mycielski yönteminin tek ve çok boyutlu varyasyonları ve Markov zincir modelinin çeşitli sürümleri önerilmiştir. Belirlenimci yöntemler olarak çok boyutlu polinomlar, çok boyutlu kobra eğrileri, çok boyutlu görgül kip ayrışımı ve dalgacıklar incelenmiştir. Ele alınan yöntemler, Saklı Markov Modelin zaman değişimli ve zaman değişimsiz varyasyonlarından esinlenerek geliştirilmiş bir hata düzeltme modelinin tahminci bileşeninde kullanılmıştır. Bu sayede farklı olasılıksal ve belirlenimci yöntemlerin geliştirilen hibrit modele katkısı karşılaştırmalı olarak ortaya konmuştur. Bu karşılaştırmalı çalışma, tez için örnek olay incelemesi olarak seçilen rüzgar hızı, güneş ışıması ve sıcaklık gibi olayların altında yatan olguyu açığa çıkarmaya çalışmaktadır.

Anahtar Kelimeler: Veri Modelleme, Zaman Değişimli Modeller, Zaman Değişimsiz Modeller, Mycielski, Markov Zincirleri, Saklı Markov Modeller

ACKNOWLEDGEMENTS

First and foremost I thank my supervisor Prof. Dr. Ömer Nezih GEREK for providing to study and complete my Ph.D. dissertation. I am extremely grateful for our many discussions which helped guide me in the right direction, his motivating attitude, his availability as a supervisor and his many insightful suggestions.

And most importantly I would like to thank you my family for their encouragement and support and for keeping me honest.

Mehmet FİDAN

April 2015

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

PDF	: Probability density function
HMM	: Hidden Markov Model
DTMC	: Discrete-Time Markov Chain
RMSE	: Root mean square error
CA	: Approximation component in wavelet decomposition
CH	: Horizontal component in wavelet decomposition
CV	: Vertical component in wavelet decomposition
CD	: Diagonal component in wavelet decomposition
Myc	: Mycielski algorithm
LP	: Linear prediction
AR	: Autoregressive Model
MA	: Moving Average
ARMA	: Autoregressive Moving Average Model
TIMTIP	: Time-Invariant Markov Chains with the Predictor Error Defined by Time-Invariant PDF
TIMTVP	: Time-Invariant Markov Chains with the Predictor Error Defined by Time-Varying PDF
PTIMTIP	: Piecewise Time-Invariant Markov Chains with the Predictor Error Defined by Time-Invariant PDF
PTIMTVP	: Piecewise Time-Invariant Markov Chains with the Predictor Error Defined by Time-Varying PDF

1. INTRODUCTION

Discrete data sets, which are collected from natural or human related events with a specific sampling rate, can be inspected in either linear or nonlinear perspective. Least-squares regression (Moser, 1996a), maximum likelihood estimation (Moser, 1996b), autoregressive model(Hayes, 1996) and autoregressive moving average model(Hayes, 1996) can be mentioned as most common linear models. Since the dependence of multi-parameters cause nonlinear behavior in time or frequency domain (which can not be expressed explicitly with linear models) on these data sets, nonlinear perspective is more convenient for obtaining beneficial information from data of real life problems.

In the previous works nonlinear models like Markov Chains (Sahin and Sen, 2001; Shamshad et al., 2005; Hocaoglu et al., 2008; Hayes and Djokic, 2013) and Hidden Markov Models (Hocaoğlu et al., 2010) were used for prediction and modeling of wind speed data. In addition Wavelet based models were used for prediction of Solar radiation (Cao and Cao, 2005; Cao and Cao, 2006; Mellit et al., 2006). Wavelet methods are also used for modeling some energy related data (Ece and Gerek, 2004; Gerek and Ece, 2004). Empirical Mode Decomposition is one of the nonlinear methods which was used for data analysis and modeling purpose (Huang et al., 1998; Huang et al., 2003; Lei et al., 2013; Jiang et al., 2014). Mycielski based Wind Speed data analysis can also be mentioned for nonlinear data analysis (Hocaoglu et al., 2009; Fidan et al., 2012). Spline method is one of the methods that was used for solar radiation data (Childs et al., 1984; Genç et al., 2002). In addition Fourier series based solar radiation analysis is one of the way in the literature (Fidan et al., 2009; Hocaoğlu et al., 2009b; Fidan et al., 2014)

In addition the dependence of multi-parameters can cause a requirement of multidimensional analysis. Moreover a simultaneous analysis for multi-periods (for example a simultaneous hourly, daily, monthly and yearly analysis) can be necessary for a sophisticated comprehension about the phenomenon underlies the case study. Due to the multi-parameters and multi-periods cases,

the multidimensional analysis becomes crucial for an advanced data research. 2d linear filters can be examples for multidimensional data analysis methods (Marzetta, 1980; Hocaoğlu et al., 2008). In addition multidimensional Wavelet analysis were used for modeling electrical power data(Ece and Gerek, 2004) and sound and vibration data taken from electrical motors (Germen et al., 2014). Multifractal analysis is one of the multidimensional data analysis method that should be mentioned (Lopes and Betrouni, 2009; Pal et al., 2014). Bidimensional Empirical Mode Decomposition method is one of the multidimensional data analysis method that was used for modeling and analysis of various type of data (Nunes et al., 2005; Fauchereau et al., 2008; WU et al., 2009; Islam et al., 2012).

In this thesis a Markov transition supported error tuning model is proposed. This model has a structure which permits to be combined with different kinds of one and multi-dimensional nonlinear data models. This proposed model is inspired by Hidden Markov Model(Rabiner and Juang, 1986) which is explained in Section 1.1. The model has variations according to the time-varying and time invariant cases. These variations are explained in detail in Chapter 4.

Through the proposed error tuning model, behavior of the data are explained by the combination of a nonlinear model and Markovian model. Basic properties of nonlinear models used in this thesis are explained in Section 1.2 and each single nonlinear method are explained in Chapter 2 in detail. The types of Markovian Models, which are used in the variations of the error tuning model, are handled in Chapter 3.

In Chapter 5, multidimensional representation of one dimensional data is explained. The purpose and advantages of multidimensional representation are stated and the methodology of multidimensional representation are expressed by the steps in Chapter 5.

The case studies of the thesis are selected as wind speed prediction, solar radiation modeling, prediction of energy demand and motor fault diagnosis and these case studies are explained and illustrated with comparative modeling results in Chapter 6.

The different types of data handled in the case studies give the opportunity to make a brief conclusion about the proposed models. These comparative results are also helpful to obtain some opinions about the convienience of hybrid model with selected nonlinear models. This brief conclusion and the opinions are stated in Chapter 7. The possible future works are also mentioned in Chapter 7.

1.1. Markov Transition Supported Error Tuning Model

Even if the selected model is stochastic or deterministic for the data of real life problem, the selected model probably will not fit the data perfectly. The reason of imperfect fitting is the data itself which consists both deterministic and stochastic components. In addition, although the selected problem consists purely stochastic process, the process may not be a stationary process, which shows time varying statistical characteristics. To achieve these problems, an error tuning model is proposed which is inspired by Hidden Markov Model (HMM). HMM is a stochastic model which is developed for unobservable stochastic processes, which can only be observed by the relation of another stochastic process that is observable (Rabiner and Juang, 1986). Hidden Markov models are used in various areas including speech recognition(Ferreiros et al., 1995; Russell et al., 1996; Gales, 1998; Takiguchi et al., 2001), handwriting recognition(Bengio et al., 1995; Jianying et al., 1996; Schlapbach and Bunke, 2004), gesture recognition(Hyeon-Kyu and Kim, 1999; Deng and Tsui, 2000; Yang et al., 2007) and applications in bioinformatics(Delorenzi and Speed, 2002; Kall et al., 2005; Shah et al., 2006; Xu et al., 2008). Moreover there are new generalization methods for hidden Markov models, which are pairwise Markov models and triplet Markov models, that permit inspecting complex data structures (Pieczynski, 2010; Boudaren et al., 2014) and the modeling of non-stationary data(Duanchao et al., 2012; Li, 2012)

In the original HMM, there are two processes. First one is hidden process and second one is observable process. The hidden process is measured by means of observable variable. In the proposed error tuning model, a predictor model is used as observable process.

Predictor Model: This process is defined by selected nonlinear prediction model (predictor). The selected prediction model obtains its output by using finite or infinite history. However this is not the ultimate prediction. This prediction is fitted on the center of a Gaussian (Normal) distribution(Dixon and Massey, 1969) with the root mean square error of the previous predictions as variance of the distribution and by the sampling of continuous distribution, finite number of candidate values for the prediction with their possibilities are obtained. This process is determined in detail in Section 1.1.1 and the selected prediction models (predictors) Chapter 2.

Markov Process: According to the nature of hidden process, there is no observation for the next sample. However there are data for the counts of the transitions between previous and next samples according to the history data. These data are used to construct Markov transition matrix which is used in discrete time Markov chains (DTMC). DTMC is explained in detail in Section 1.1.2 and its variations are explained in Chapter 4.

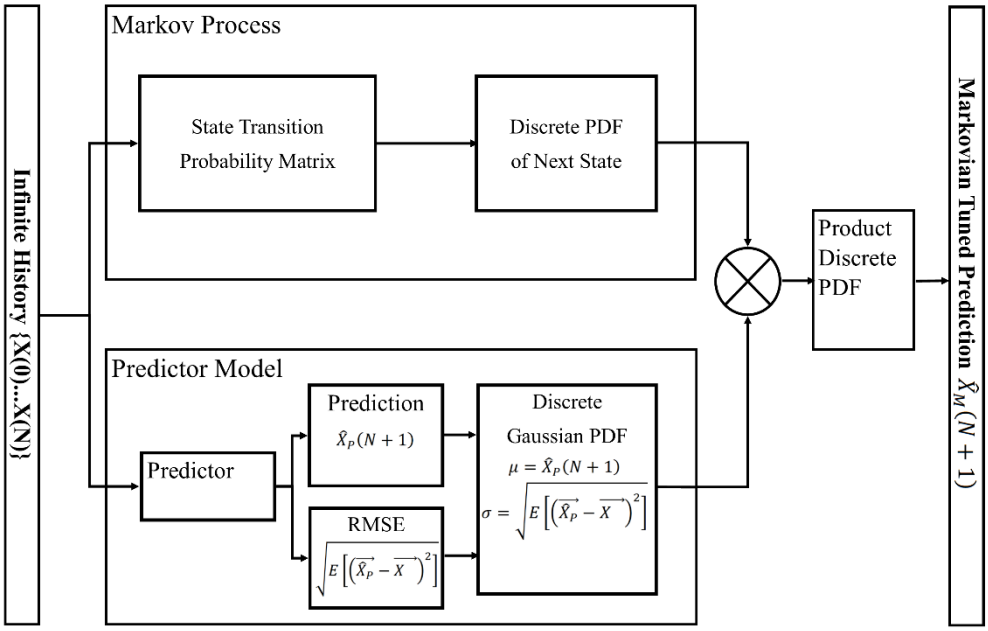

Figure 1.1. General structure of Markov transition supported error tuning model

1.1.1. Predictor

In this thesis, predictor is an operator which estimates future values of a discrete-time signal with a specific function of previous samples. This function can use whole previous samples (infinite history) or a specific portion of previous samples (finite history).

The predictor that is used in the hybrid model will produce a prediction for the next data sample. However this prediction will not be the ultimate prediction of the hybrid model. Depending to the prediction value and the prediction errors of the previous predictions, a Gaussian distribution will be produced. The midpoint of the distribution will be the prediction value. The standard deviation of the distribution will be changed depending to the RMSE of the prediction model. This RMSE is taken as time-varying and time-invariant in different hybrid models which is explained in Chapter 4 in details. The Gaussian distribution can be explained with Equation (1.1) and illustrated as Figure 1.2 (Dixon and Massey, 1969):

$$f(x,\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \tag{1.1}$$

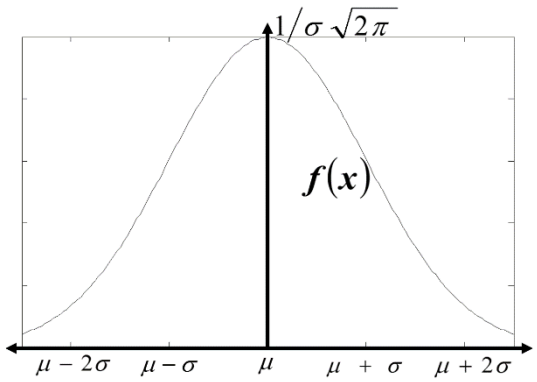

Figure 1.2. Graph of the prediction error centered Gaussian distribution

1.1.2. Discrete Time Markov Chain for State Transition Probabilities

A stochastic process $\{X_n, n \geq 0\}$ on state space $\boldsymbol{S}$ is said to be a discrete-time Markov chain (DTMC) if Equation (1.2) is satisfied for all states S_i and S_j in $\boldsymbol{S}$.

$$P\left(X_{n+1} = S_j \middle| X_n = S_i, X_{n-1}, \cdots, X_0\right) = P\left(X_{n+1} = S_j \middle| X_n = S_i\right) \quad (1.2)$$

A DTMC $\{X_n, n \geq 0\}$ is said to be time invariant if Equation (1.3) is satisfied.

$$P\left(X_{n+1} = S_j \middle| X_n = S_i\right) = P\left(X_1 = S_j \middle| X_0 = S_i\right) \quad (1.3)$$

If the transition probability from S_i to S_j is taken as Equation (1.4), then the state transition probability matrix A is obtained as Equation (1.5) for N number of states.

$$a_{ij} = P\left(X_{n+1} = S_j \middle| X_n = S_i\right) \quad (1.4)$$

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1N} \\ \vdots & \ddots & \vdots \\ a_{N1} & \cdots & a_{NN} \end{bmatrix} \quad (1.5)$$

In order to obtain the state transition probabilities with the two samples of the discrete data with m number of sample distance from each other, the state transition probability matrix is obtained as Equation (1.6).

$$A^m = \begin{bmatrix} a_{11} & \cdots & a_{1N} \\ \vdots & \ddots & \vdots \\ a_{N1} & \cdots & a_{NN} \end{bmatrix}^m \quad (1.6)$$

According to distributions obtained from predictor and DTMC the decision of the ultimate prediction is obtained as shown in Figure 1.3.

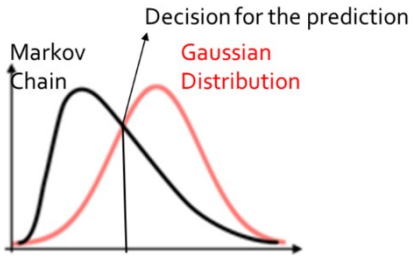

Figure 1.3. Ultimate decision of the Markov transition supported error tuning model

1.2. Overview to Nonlinear Data Models

The superposition property is the necessary condition of the linear models. However, the data taken from a real-life event does not show the superposition mostly. Therefore alternative models are needed which are nonlinear. These nonlinear models can be separated as deterministic and stochastic nonlinear models inside. The deterministic models are used for the systems which does not consist randomness. On the other hand, the stochastic models are defined by random variables.

1.2.1. Stochastic Models

An estimation process which is depending to the probability distributions of the potential candidates for the prediction can be called stochastic model (Pinsky and Karlin, 2010). These probability distributions or probabilistic definitions are obtained from the observations for the statistical characteristics of the historical data with suitable time series techniques. In order to get more reliable predictions, longer historical data should be observed with more repeats. These models are the methods of measuring dynamic correlations of the series of the random events.

A stochastic process in discrete state space, which is modelled by a stochastic model, is represented by the set of random variables $\{X_n\}$ where the $n \in \mathbb{Z}$ denotes time. A modellable random variable X_n is expected to be

dependent to past random variables $\{X_{n-i}\}$ where $i \in \mathbb{Z}^+$. These dependencies can be expressed by Equation (1.2) and Equation (1.3) for a standard Markovian model.

In this thesis Markov and Mycielski based methods are held as stochastic models, explained in Chapter 2 with details.

1.2.2. Deterministic Models

Mathematical models, which are expressed by means of well-defined relationships among states and events and do not include any component for random variation, are called deterministic models (Heinz, 2011). In such models, a given input will always cause the same output. On the other hand, stochastic models use ranges of values for variables in the form of probability distributions.

In this thesis, polynomials, wavelets, splines and empirical mode decomposition are chosen as deterministic models.

2. SELECTED METHODS FOR PREDICTOR

2.1.Mycielski Based Methods

Mycielski algorirthm is designed for 1d data, which has the advantage using infinite history rule (Jacquet et al., 2002). Mycielski algorithm makes its prediction by finding the longest pattern that is repeated at the end of the infinite history. Infinite history means the whole past samples of the data. Depending on its strong prediction capability, Mycieslki algorithm can be used in different applications. Some of these applications can be listed as wind speed prediction (Hocaoğlu et al., 2009a), synthetic wind speed data generation (Fidan et al., 2012), pseudo-random number generator(Ehrenfeucht and Mycielski, 1992) and cyphering by using pseudo-random number generator(Fidan and Gerek, 2008). In this thesis, different bidimensional variations of the Mycielski Algorithms are developed. These developed bidimensional versions of the Mycielski algorithm will be expected to obtain better prediction values than the original one dimensional Mycielski algorithm when they are applied on bidimensional real-life data or images. The reason of this expectation is the 2d pattern detection ability of bidimensional versions which does not exist on the original version. These novel algorithms are applied for the prediction of hourly energy demand in a city, prediction of wind energy and also predictive coding and compression of 2d images and the comparative of the algorithms are illustrated in Chapter 6.

2.1.1. 1d-Mycielski Algorithm

The Mycielski algorithm performs a prediction using the total exact history of the data samples. The basic idea of the algorithm is to search for the longest suffix string at the end of the data sequence which had been repeated at least once in the history of the sequence. The search starts with a short (length = 1) template size and continues increasing the template length as long as matches are found in the history. When the longest repeating sequence is determined, the

value of the sample right after the longest repeating template is assigned as the prediction value. The rule estimates that if this pattern had appeared like this in the past, then it is supposed to behave the same now. This predictor can be generalized with the expression in Equation (2.1).

$$\hat{x}[n+1] = f_{n+1}(x[1], \cdots, x[n]) \tag{2.1}$$

The function f (.) performs an iterative algorithm that starts from the shortest data segment at the end (i.e. length one sample: $x[n]$), then one by one increases the data segment length to the left side as $(x[n-1], x[n]), (x[n-2], x[n-1], x[n])$, etc. Meanwhile, the segments are searched from the end point to the start point by sliding over the samples. Several matches could be found during the algorithm run. At a point of a ‘‘no-match”, a probably longer segment will not be encountered anywhere in the past sequence. At that point, the prediction is made as the next sample value of the latest encountered (1 shorter) matching string. Naturally, the algorithm searches through the whole data sequence repeatedly for each prediction step, and it has high computational requirements. The overall scheme can be analytically expressed as Equation (2.2).

$$\begin{aligned} m = arg \max_{L} &\begin{Bmatrix} x[k] = x[n], x[k-1] = x[n-1] \\ , \cdots, x[k-L+1] = x[k-L+1] \end{Bmatrix} \\ f_{n+1} &= \hat{x}[n+1] = x[m] \end{aligned} \tag{2.2}$$

The above predictor can be re-described in words as an attempt to estimate the next sample in the currently ongoing random process as the most probable sample that had occurred in the history of the data sequence. The most probable is taken as the longest repeating chain of data samples.

2.1.2. 2d-Mycielski Algorithm

First step of the developing two dimensional version of the Mycielski algorithm is identifying the shape of the searched pattern and how it will be widened for the next search step. The widening routine is shown for original Mycielski and 2d-Mycielski in Figure 2.1.

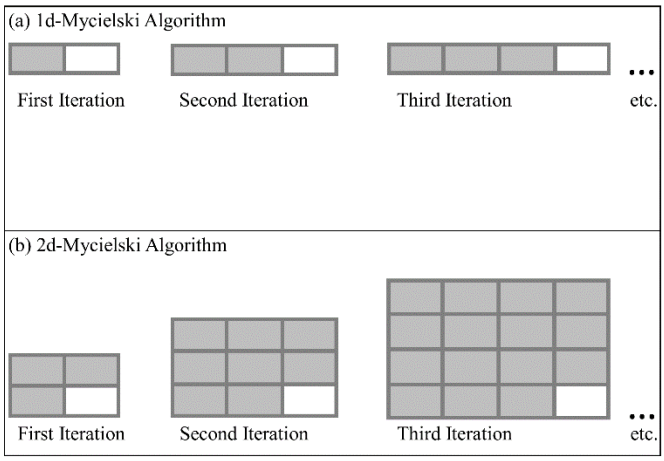

Figure 2.1. Widening iterations of (a) 1d, and (b) 2d-Mycielski algorithms

Second step is identifying the history which is the region that the searched pattern can be repeated. In this thesis, the limits of the history is identified with the index values of the sample that will be predicted. In other words the history is the full region which stands left upper side of the sample. However, the searched pattern is excluded from that region to avoid the intersection of searched pattern and probable repeated version of it in the past. Depending to these explanations the infinite history where the searched pattern is excluded can be illustrated as Figure 2.2.

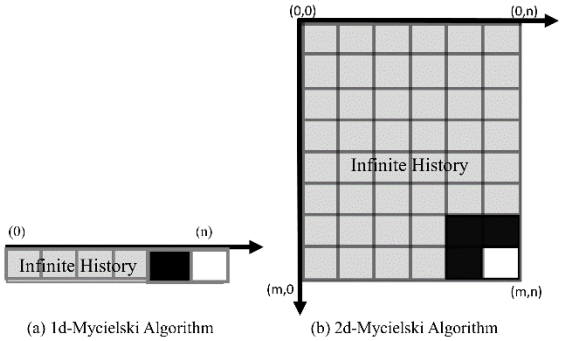

Figure 2.2. Infinite history rules of (a)1d, and (b)2d-Mycielski algorithms

Any 2d predictor depending to infinite history can be expressed as Equation (2.3).

$$\hat{x}[n+1,m+1] = f\begin{pmatrix} x[1,1], x[1,2], x[2,1], x[2,2], \\ \cdots, x[n+1,m], x[n,m+1] \end{pmatrix} \tag{2.3}$$

Due to the eqution above, the bidimensional Mycielski algorithm can be generalized as Equation (2.4).

$$(i,j) = \arg\max_{\{k,l\}} L \coloneqq \left\{ \{k,l\} \middle| \begin{matrix} x[k+1,l] = x[n+1,m], \\ x[k,l+1] = x[n,m+1], x[k,l] = x[n,m], \cdots, \\ x[k-L+1,l-L+1] = x[n-L+1,m-L+1] \end{matrix} \right\},$$
$$f_{n+1,m+1} = \hat{x}[n+1,m+1] = x[i,j] \tag{2.4}$$

Depending to (2.4), the prediction value is the sample at the right lower corner of the found largest repeated pattern. If there are more than one repeat for the largest searched pattern, the closest one should be taken as default repeat. In other words the scanning procedure should be than in a direction from closer sample to the farrest sample.

There is only one single choice for the scanning direction of orginal Mycielski Algorithm. Since the distance between two points in one dimension is

exactly the same for all possible d_p metrics that use L^P norm, the direction from closer point to farest point is single. In other words, the variation of p in the d_p metric does not cause any change the measurement of the distance. According to these explanations, the single scanning direction which can be expressed by the simplest d_1 Manhattan metric and the other complicated metrics, which is from right to left side of the searched pattern is the single choice for orginal Mycielski algorithm.

Nevertheless for the two dimensions, the measured distance will be changed when the p in the d_p norm is changed. Therefore there are infinite number of choices for the scanning direction. In this thesis, the scanning directions defined by d_1(Manhattan), d_2(Euclidean) and d_∞(Chebyshev) metrics are chosen(Shannon, 2007). In addition the performances of these three algorithms are compared.

L^P or p norm are used to define the spaces with finite number of dimensions. d_P metrics are the metrics which define the distance between two points in the space with L^P norm. These d_P metrics are also named as Minkowski distance. For the real valued R^n vector space, the d_P metric between two points can be defined as in Equation (2.5).

$$d_p(x,y) = (\sum_{i=1}^{n} |x_i - y_i|^P)^{1/P} \tag{2.5}$$

In Manhattan metric the p value in the equation above is taken as 1. For the two dimensional real valued R^2 space, the distance between any two points is expressed with Manhattan metric as in Equation (2.6).

$$d_1(x,y) = |x_1 - y_1| + |x_2 - y_2| \tag{2.6}$$

According to Manhattan metric, the unit circle can be shown as in the Figure 2.3.

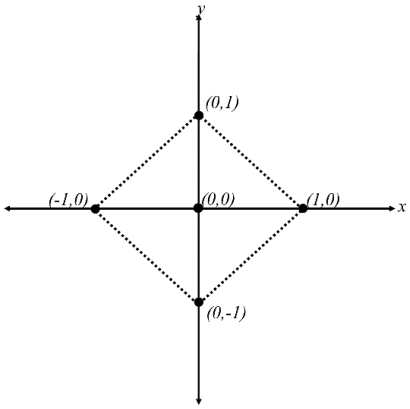

Figure 2.3. The Unit Circle defined by Manhattan metric

The scanning direction has a similar shape with unit circle of Manhattan metric which can be shown as in the following Figure 2.4.

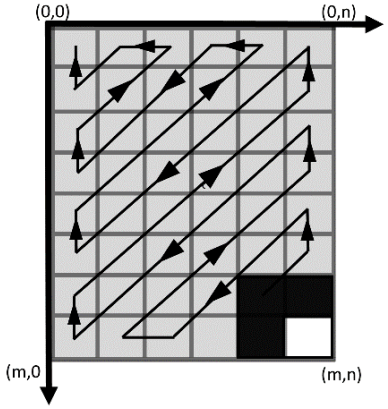

Figure 2.4. Scanning direction defined by Manhattan distance

In Euclidean metric the p value in Equation (2.5) is taken as 2. For the two dimensional real valued R^2 space, the distance between any two points is expressed with Manhattan metric as in Equation (2.7).

$$d_2(x,y) = (|x_1 - y_1|^2 + |x_2 - y_2|^2)^{1/2} \tag{2.7}$$

According to Euclidean metric, the unit circle can be shown as in the Figure 2.5

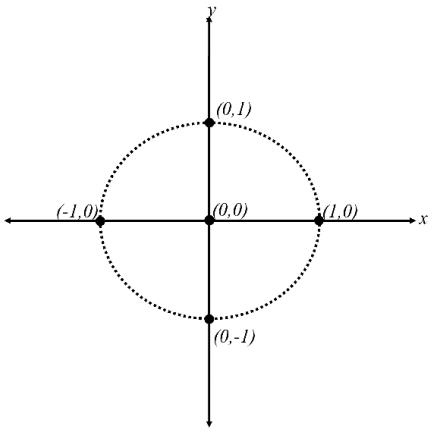

Figure 2.5. The unit circle defined by Euclidean metric

The scanning direction has a similar shape with unit circle of Euclidean metric which can be shown as in the Figure 2.6.

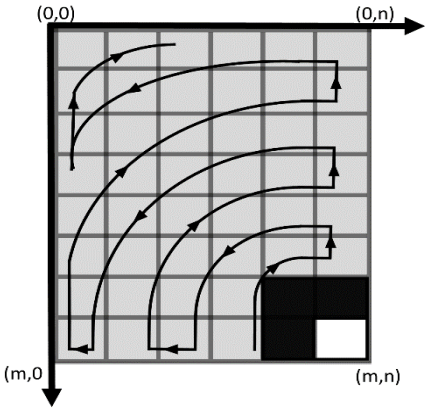

Figure 2.6. Scanning direction defined by Euclidean distance

To define a perfect circle scanning order, the index values of samples which are defined Cartesian coordinates should be converted to polar coordinates with Equation (2.8).

$$\overline{v_{(m+1,n+1)(k+1,l+1)}} = (m-k, n-l) = (r, \theta)$$
$$r = \sqrt{[(m-k)^2 + (n-l)^2]}, \theta = \arctan\left(\frac{n-l}{m-k}\right)$$
(2.8)

For obtaining a perfect circle shaped scanning direction, samples of the infinite history should be ordered depending to their polar coordinates. This

ordering scheme has two steps. In first step samples should be ordered according to radius *r* component. In the second step, the samples which have the same radius should be ordered according to angle θ component. Therefore scanning scheme will be started from smallest radius to largest radius, and in the group of samples which have same radius scanning scheme will be from smallest angle to largest in the first circle, then will be swapped to the order from largest to smallest angle for second circle, and swapped to the order from smallest to largest angle and swapping will be continued for the following circles.

Euclidean based scanning direction is the exact method to find the closest repeated pattern. However due to the computation of polar coordinates this is the most complex search scheme which increases the complexity of search algorithm exponentially.

The Chebyshev metric is calculated by taking limit of (2.5), when the *p* goes to infinity. For the two dimensional real valued R^2 space, the distance between any two points is expressed with Chebyshev metric as in Equation (2.9).

$$d_\infty(x,y) = \lim_{p\to\infty}(|x_1 - y_1|^P + |x_2 - y_2|^P)^{1/P} = max\{|x_1 - y_1|, |x_2 - y_2|\} \tag{2.9}$$

According to Chebyshev metric, the unit circle can be shown as in Figure 2.7.

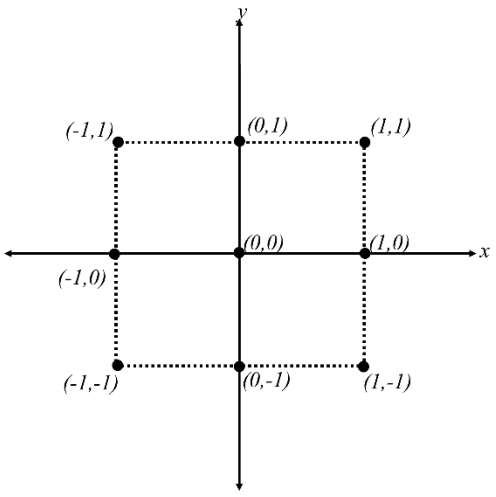

Figure 2.7. The unit circle defined by Chebyshev metric

The scanning direction has a similar shape with unit circle of Chebyshev metric which can be shown as in Figure 2.8.

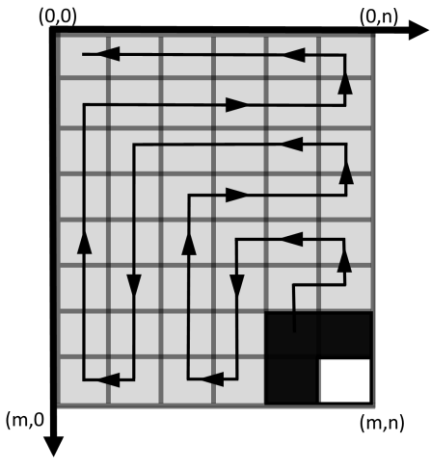

Figure 2.8. Scanning direction defined by Chebyshev distance

In the Chebyshev based scanning scheme, the scanning scheme is going to be similar to original one dimensional Mycielski algorithm, while the distance is becoming larger. For the purpose of catching the repeated patterns in the infinite history, a similary test should be applied between searched pattern and the candidate for the repeat in the history. In the original Mycielski Algorithm this similarity test is applied with using Hamming distance (Forney Jr, 1966). For the binary valued two different series the Hamming distance can be calculated as Equation (2.10).

$$d_{Hamming} = (c_1, c_2) = \sum_{i=1}^{n} [c_1(i) \oplus c_2(i)] \tag{2.10}$$

If the Hamming distance is 0, then the compared series will be assumed as the same. On the other case, they will be assumed different. Hamming distance is the simplest distance that can be applied to binary valued series. However, in the real valued series the algorithm will probably fail to catch similar patterns with Hamming distance, because of impossibility of repeating a real value. In the previous work, a tolerated distance is suggested to use instead

of Hamming distance(Hocaoglu et al., 2009). This tolerated distance is also used in this thesis also, which can be defined as Equation (2.11).

$$f_{tolerated}(x_1, x_2) = \begin{cases} 0, |x_1 - x_2| < \dfrac{max\{|x_1|, |x_2|\}}{10} \\ 1, |x_1 - x_2| \geq \dfrac{max\{|x_1|, |x_2|\}}{10} \end{cases}$$

$$d_{tolerated}(c_1, c_2) = \sum_i \sum_j f_{tolerated}\big(c_1(i,j), c_2(i,j)\big) \qquad (2.11)$$

As in the Hamming distance, if the tolerated distance is 0, then the compared series will be assumed as the same. On the other case, they will be assumed different.This tolerated distance allows max 10% RMSE between searched pattern and the found repeated pattern.

2.1.3. Mycielski Algorithm for 2d-Pattern Search In A Spiral Plane

In the previous section the 2d-Mycielski Model is constructed. In this section data will be converted into 2d spaces from 1d space. However the continuity property of 1d will not be lost. For conservation of continuity property, representation in archimedian spiral coordinate system is proposed (Curtis, 1960; Benedetto and Wu, 2000). This spiral coordinate system satisfies to search 2d patterns in 1d order. This model is constructed for real-life events like wind speed for catching 2d patterns (hour and day are the two dimensions) without loosing continuity of sequential days.

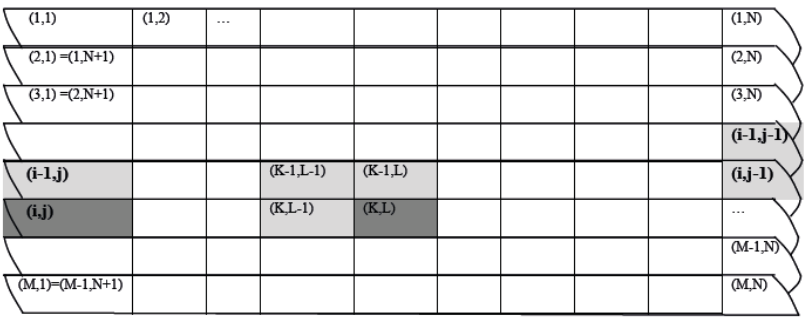

Figure 2.9. A symbolic illustration of spiral coordinate system

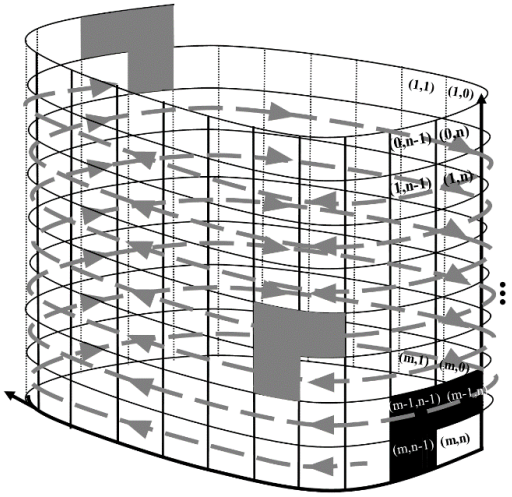

Figure 2.10. A symbolic illustration of scanning direction in the spiral coordinate system

2.1.4. Mycielski Decomposition

This method is inspired by the idea of iterative combinations of residual prediction. Predictions are started for longest period (least sampling frequency). For each step prediction errors are calculated. In the next step the prediction period is shortened and new predictions are made on residual data (prediction error of first iteration). This iterative prediction steps are continued until the smallest prediction period is reached. A four-level Mycielski decomposition can be illustrated as Figure 2.11. As a result of application of N-level Mycieslki decomposition, N+1 number of components are constructed.

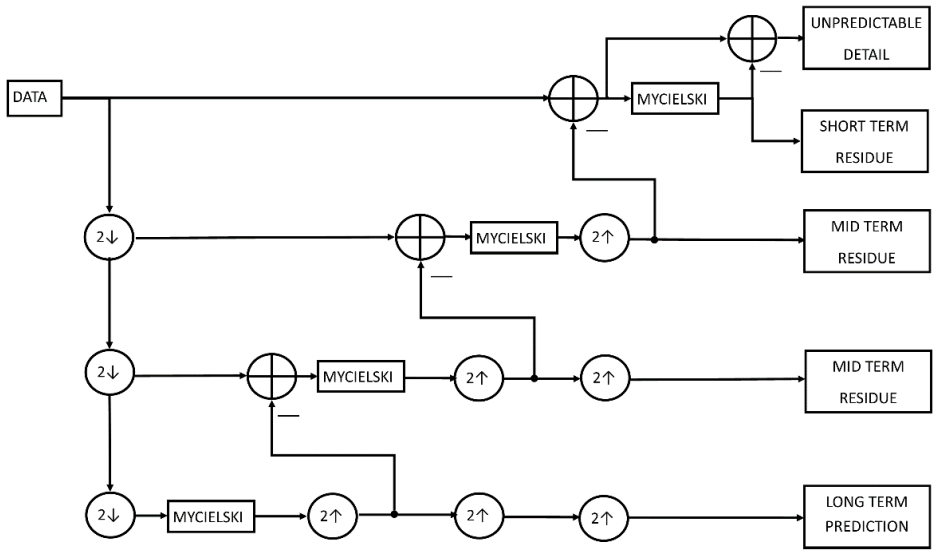

Figure 2.11. Four-level Mycielski decomposition

By means of Mycielski decomposition, the predictability of the data for long term and short term can be observed. In addition, the statistics of the unpredictable component give some opinions about randomness and nonlinearity of the inspected data. The constructed decomposition model is a suitable model for energy demand data, which have monotonicly increasing components and fluctuating components as seen from EMD analysis in Section.

2.2.Wavelet Based Methods

In this section, data is modelled by using Wavelet decomposition. The wavelet decomposition seperates the signal into its low and high energy components and gives the opportunity to inspect energy component and detail component individually. This model is chosen for modeling and prediction of solar radiation.

The energy component of solar radiation can be simple modelled by 2d-cosine function which is depending to hour and day variables. The other component of the signal is modelled as following.

In this application the main purpose is construct a one year approximation from applying inverse Wavelet transform to the mathematical models of energy and detail componenets of the solar radiation data.

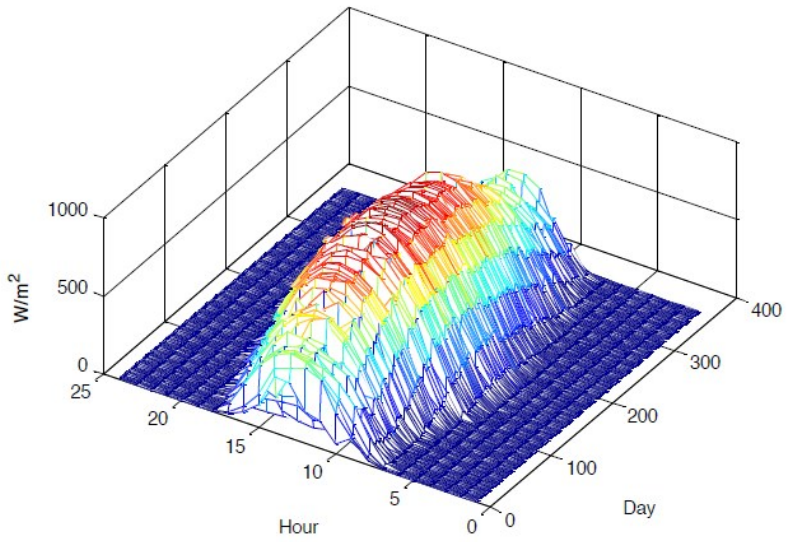

Figure 2.12. 2d representation of one year solar radiation data

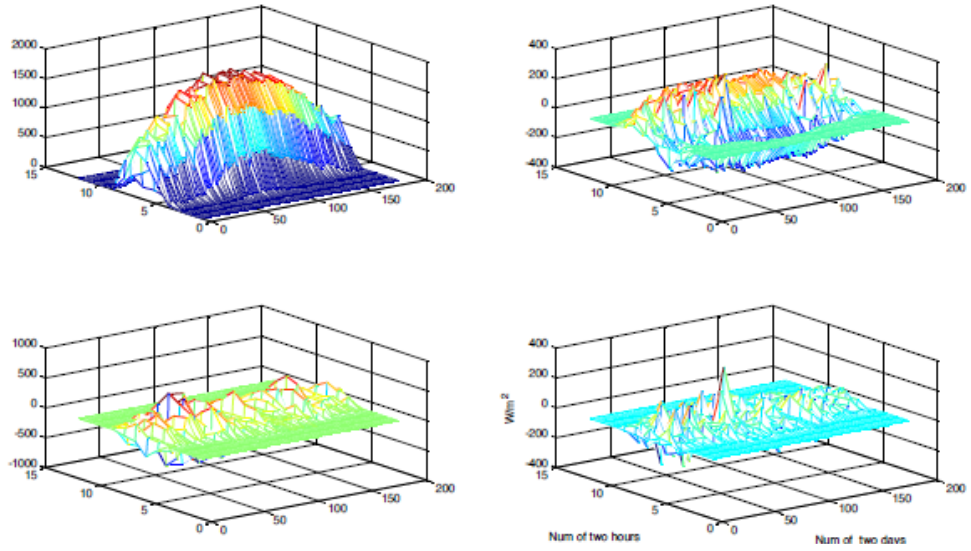

Figure 2.13. The components of the solar radiation data after the single layered Haar Wavelet transform CA, CH, CV, CD

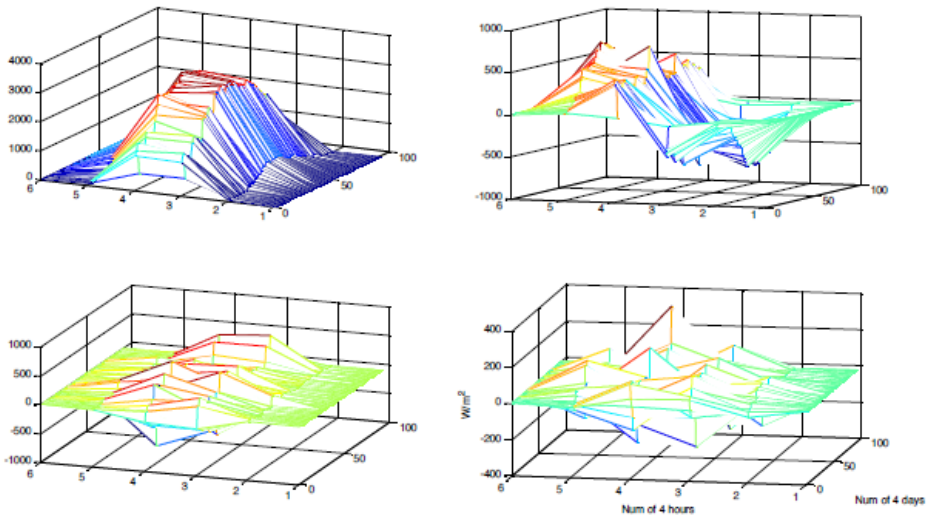

Figure 2.14. The components of the solar radiation data after the application of second layer of the Haar Wavelet transform CAA, CAH, CAV, CAD

CAA and CAH component in the figure above are defined with polynomials in 5th degrees whose coefficients are also polynomials in 5^{th} degrees which can be shown as below.

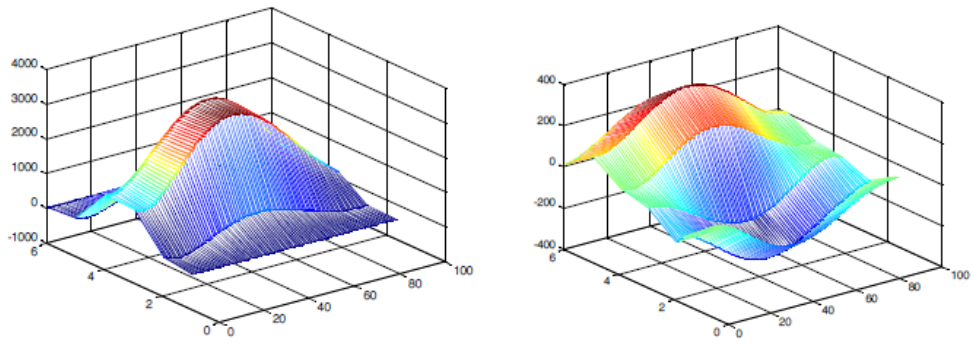

Figure 2.15. CAA and CAH constructed by polynomials with polynomial valued varying coefficients

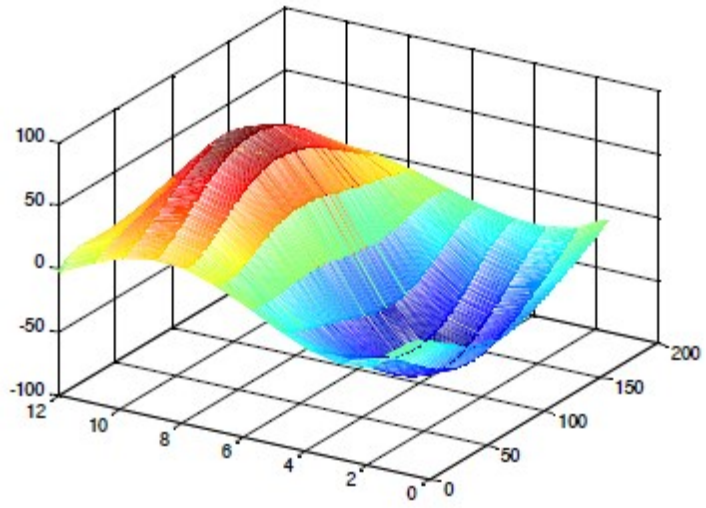

Figure 2.16. CH constructed by polynomials with polynomial valued varying coefficients

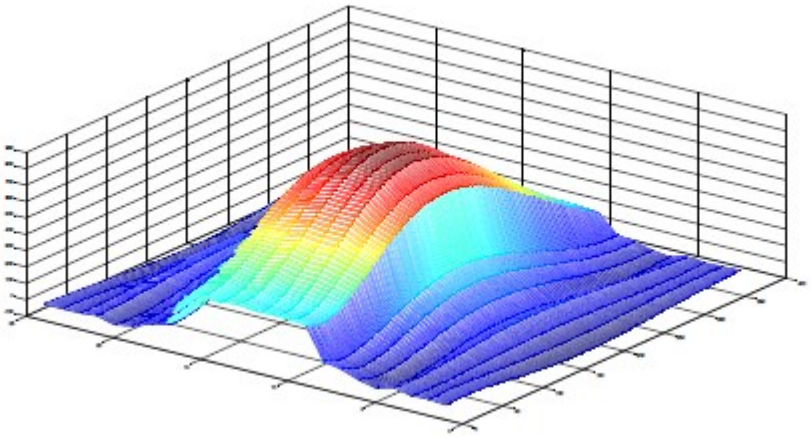

Figure 2.17. General solar radiation model constructed by obtained CAA, CAH and CH components

2.3. Linear Predictors

Linear prediction is a mathematical operation where future values of a discrete-time signal are estimated as a linear function of previous samples. In digital signal processing, linear prediction is often called linear predictive coding (LPC) and can thus be viewed as a subset of filter theory. In system analysis (a

subfield of mathematics), linear prediction can be viewed as a part of mathematical modeling or optimization.

A linear system is such that produces its output as a linear combination of its current and previous inputs and its previous outputs (Parsons, 1987). It can be described as time-invariant if the system parameters do not change with time. Mathematically, linear time-invariant (LTI) systems can be represented by Equation (2.12) (Makhoul, 1975).

$$y(n) = \sum_{j=0}^{q} b_j x(n-j) - \sum_{k=1}^{p} a_k y(n-k) \tag{2.12}$$

This is the general difference equation for any linear system, with output signal y and input signal *x*, and scalars b_j and a_k, where *j* changes from 1 to *q* and k changes from 1 to p where the *p* and *q* identify the order of the system.

2.3.1. 1d Linear Predictor

In this thesis all-pole model and least-squares model are taken as 1d linear predictor. The all-pole model, which is also mentioned as autoregressive model and explained with details in the research article of Weruaga (Weruaga, 2007), can be expressed with Equation (2.13).

$$x[n] = (\sum_{k=1}^{P} a_k x[n-k]) + e[n] \tag{2.13}$$

In the equation *P* denotes the model order and a_k are the autoregressive coefficients. The most popular method for estimation of autoregressive coefficients is the Yule-Walker method(Yule, 1927; Walker, 1931).

If the *P* is taken as 3 then (2.13) can be written as Equation (2.14).

$$x[n] = a_1 x[n-1] + a_2 x[n-2] + a_3 x[n-3] + e[n] \tag{2.14}$$

The coefficients a_1, a_2 and a_3 have the relation defined in Equation (2.15) with the correlation values R_{XX} .

$$\begin{bmatrix} R_{XX}(0) & R_{XX}(-1) & R_{XX}(-2) \\ R_{XX}(1) & R_{XX}(0) & R_{XX}(-1) \\ R_{XX}(2) & R_{XX}(1) & R_{XX}(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} R_{XX}(-1) \\ R_{XX}(-2) \\ R_{XX}(-3) \end{bmatrix} \quad (2.15)$$

The relation in Equation (2.15) can be represented as Equation (2.16).

$$\boldsymbol{R} \cdot \vec{\boldsymbol{a}} = \vec{\boldsymbol{r}} \quad (2.16)$$

Therefore the a_k coefficients can be obtained by Equation (2.17).

$$\boldsymbol{R^{-1}} \cdot \vec{\boldsymbol{r}} = \vec{\boldsymbol{a}} \quad (2.17)$$

2.3.2. 2d Linear Predictor

Due to predictive image coding literature, it is known that a 2d-matrix can be efficiently modeled by linear predictive filters(Maragos et al., 1984). An other work shows that solar radiation data can be modelled by 2d linear predictor(Hocaoğlu et al., 2008).

The structure of the 2d linear predictor is similar to 1d linear predictor. However the difference is the identification of the neighbourhood of the unknown value which is used for prediction. In order to obtain a simple 2d linear predictor with 3 coefficients, the neighbourhood is defined as in Figure 2.18.

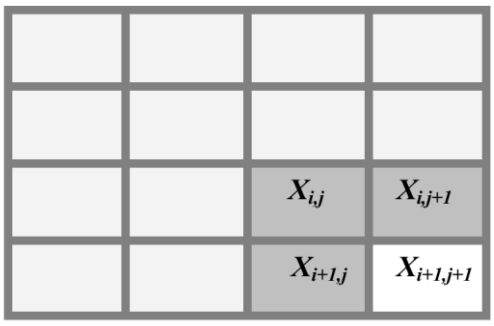

Figure 2.18. Neighbourhood of unknown value $X_{i+1,j+1}$ for 2d linear prediction

According to Figure 2.18, 2d linear predictor defined as Equation (2.18).

$$X_{i+1,j+1} = a_1 X_{i,j} + a_2 X_{i,j+1} + a_3 X_{i+1,j} + e_{i+1,j+1} \tag{2.18}$$

The coefficients a_1, a_2 and a_3 have the relation defined in Equation (2.19) with the correlation values R_{XX} .

$$\begin{bmatrix} R_{XX}(0,0) & R_{XX}(-1,-1) & R_{XX}(-1,0) \\ R_{XX}(1,1) & R_{XX}(0,0) & R_{XX}(0,1) \\ R_{XX}(1,0) & R_{XX}(0,-1) & R_{XX}(0,0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} R_{XX}(-1,-1) \\ R_{XX}(-1,0) \\ R_{XX}(0,-1) \end{bmatrix} \tag{2.19}$$

The relation in (2.19) can be represented as Equation (2.20).

$$\boldsymbol{R} \cdot \vec{\boldsymbol{a}} = \vec{\boldsymbol{r}} \tag{2.20}$$

Therefore the a_k coefficients can be obtained by Equation (2.21).

$$\boldsymbol{R^{-1}} \cdot \vec{\boldsymbol{r}} = \vec{\boldsymbol{a}} \tag{2.21}$$

2.4. Splines

In mathematics, a spline is a numeric function that is piecewise-defined by polynomial functions, and which possesses a sufficiently high degree of smoothness at the places where the polynomial pieces connect (which are known as knots)(Judd, 1998; Chen, 2009)

In interpolating problems, spline interpolation is often preferred to polynomial interpolation because it yields similar results to interpolating with higher degree polynomials while avoiding instability due to Runge's phenomenon. In computer graphics, parametric curves whose coordinates are given by splines are popular because of the simplicity of their construction, their ease and accuracy of evaluation, and their capacity to approximate complex shapes through curve fitting and interactive curve design.

The most commonly used splines are cubic spline, cubic B-spline and cubic Bézier spline.

B-splines were investigated as early as the nineteenth century by Nikolai Lobachevsky. The term "B-spline" was coined by Isaac Jacob Schoenberg and is short for basis spline. A spline function is a piecewise polynomial function of degree k in a variable x. The places where the pieces meet are known as knots. The number of knots must be equal to, or greater than k+1. Thus the spline function has limited support. The key property of spline functions is that they are continuous at the knots. Some derivatives of the spline function may also be continuous, depending on whether the knots are distinct or not. A fundamental theorem states that every spline function of a given degree, smoothness, and domain partition, can be uniquely represented as a linear combination of B-splines of that same degree and smoothness, and over that same partition.

2.5. Empirical Mode Decomposition

EMD is a method of breaking down a signal without leaving the time domain (Huang et al., 1998). It can be compared to other analysis methods like Fourier Transforms and wavelet decomposition. The process is useful for analyzing natural signals, which are most often non-linear and non-stationary. This parts from the assumptions of the methods we have thus far learned (namely that the systems in question be LTI, at least in approximation).

EMD filters out functions which form a complete and nearly orthogonal basis for the original signal. Completeness is based on the method of the EMD; the way it is decomposed implies completeness. The functions, known as Intrinsic Mode Functions (IMFs), are therefore sufficient to describe the signal, even though they are not necessarily orthogonal.

The fact that the functions into which a signal is decomposed are all in the time-domain and of the same length as the original signal allows for varying frequency in time to be preserved. Obtaining IMFs from real world signals is important because natural processes often have multiple causes, and each of these causes may happen at specific time intervals. This type of data is evident

in an EMD analysis, but quite hidden in the Fourier domain or in wavelet coefficients.

2.5.1. 1d Empirical Mode Decomposition

1d Empirical Mode Decomposition is method which express a signal in terms of its component Intrinsic Mode Functions (IMF). An IMF is a function which has only one extreme between zero crossings, and has a mean value of zero. The process of obtaining IMF is called as sifting process. In addition there is a locality parameter in EMD, which controls number of iteration and number of IMFs of EMD. This parameter also identifies the calculation time and the effectiveness of the EMD. According to locality parameter the sifting parameter is repeated *k* times.

In the sifting process, firstly upper and lower envelopes of a signal $x(t)$ are calculated. Upper envelope is obtained from cubic-spline interpolation of local maxima of $x(t)$. As similar lower envelope is obtained from cubic-spline interpolation of local minima of $x(t)$. The mean of upper and lower envelopes gives the local mean function $\mathrm{m_1(t)}$.

The first component $\mathrm{h_1(t)}$is calculated as Equation (2.22).

$$\mathrm{h_1(t) = x(t) - m_1(t)} \tag{2.22}$$

In the second sifting step, $\mathrm{h_1(t)}$ is handled as the original signal and the local mean function $\mathrm{m_{11}(t)}$ is obtained from $\mathrm{h_1(t)}$ with the same way of obtaining $\mathrm{m_1(t)}$
from $x(t)$. $\mathrm{h_{11}(t)}$ is calculated as Equation (2.23).

$$\mathrm{h_{11}(t) = h_1(t) - m_{11}(t)} \tag{2.23}$$

This sifting procedure is repeated *k* times, until $\mathrm{h_{1k}(t)}$ which is an IMF is obtained as Equation (2.24).

$$h_{1k}(t) = h_{1k-1}(t) - m_{1k}(t) \tag{2.24}$$

In the inverse order $h_{1k}(t) = c_1(t)$ is accepted as first IMF component, which has highest frequency (shortest period) in all IMFs. Therefore the last IMF is $h_{11}(t) = c_k(t)$ which has the lowest frequency. In addition there is residual component, which can be thought as noise of the data. This residual component does not give any information about the frequency of the harmonics of the data. According to these obtained IMFs $x(t)$ can be expressed as Equation (2.25).

$$x(t) = c_1(t) + c_2(t) + \cdots + c_k(t) + r_k(t) \tag{2.25}$$

2.5.2. 2d Empirical Mode Decomposition

The ability to effectively classify and segment images based on textural features is of key importance in scene analysis, medical image analysis, remote sensing and many other application areas. Feature extraction is the first stage of image texture analysis. To extract the 2d-IMF during the sifting process, we have used morphological reconstruction to detect the image extrema and RBF to compute the surface interpolation. A 2d-IMF is a zero-mean 2d AM–FM component. The image AM–FM decomposition is partially unsupervised feature based segmentation algorithm, whereas the EMD is fully unsupervised. The 2d sifting process is started by finding the maxima and the minima of the 2d data by morphological reconstruction based on geodesic operators and the upper and lower 2d envelopes are constructed by radial bias functions. After this choices for construction envelopes, the similar sifting process is applied on the data two obtain 2d-IMFs.

2.6. Polynomial Methods

Polynomials are the most popular deterministic methods for explanation of the real life data. The following proposed methods are constructed for 2d data. Each two methods are constructed by the least squares method based polynomial regression of the local extrema (Gergonne, 1974). The polynomial regression with *m*th degree polynomial is defined as Equation (2.26), where the x_i's are coordinates of the extrema and y_is are the values of extrema.

$$y_i = [\sum_{k=0}^{m} a_k (x_i)^k] + \varepsilon_i \tag{2.26}$$

Matrix notation of *n* number of extrema with *m*th degree polynomial is shown in Equation (2.27).

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 & {x_1}^2 & \dots & {x_1}^m \\ 1 & x_2 & {x_2}^2 & \dots & {x_2}^m \\ 1 & x_3 & {x_3}^2 & \dots & {x_3}^m \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & {x_n}^2 & \dots & {x_n}^m \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_n \end{bmatrix} \tag{2.27}$$

Equation (2.27) can be represented as Equation (2.28).

$$\vec{y} = X\vec{a} + \vec{\varepsilon} \tag{2.28}$$

According to Equation (2.28) a_k coefficients can be estimated as Equation (2.29).

$$\vec{a} = (X^T X)^{-1} X^T \vec{y} \tag{2.29}$$

2.6.1. Multiplication of polynomials under the assumption of seperability

In this method firstly local extremum points are found. After that step, projections on x axis of these extremum points are defined with a single polynomial fuction $f(x)$ and projections of these extremums on y axis are defined with single polynomial function $g(y)$. According to select appropriate $f(x)$ and $g(y)$ functions least square method is applied to the projections. After that the function that express whole data $S(x,y)$ is found from the multiplication of these two polynomial functions as in Equation (2.30).

$$S(x,y) = \frac{f(x)\, g(y)}{\sqrt{|f(x)||g(y)|}} \tag{2.30}$$

2.6.2. Polynomials with polynomial valued varying coefficients

In this method, the local extrema are detected in all of the rows(x axis) of the 2d data. Then these extrema are fitted to Nth degree polynomials by least squares method. The a_k coefficients of each rows construct kth column. Afterward local extrema of the each constructed column are found. These local extrema are fitted to Mth degree polynomials by least squares method. Obtained polynomial model is represented by Equation (2.31).

$$\begin{aligned} S(x,y) &= \textstyle\sum_{i=0}^{N} g_i(y)\, x^i \\ g_i(y) &= \textstyle\sum_{j=1}^{M} a_j y^j \end{aligned} \tag{2.31}$$

2.7. Fourier Series Method

The model illustration is carried out over the data acquired in Izmir region of Turkey between years 2004 and 2005. The harmonic analysis enables us to construct the daily solar radiation data model in a unified and compact form. The idea is inspired from the fact that, the Discrete Fourier Transform of

hourly solar variation in a day can be easily modeled by few simple harmonic components superimposed by a hard-to-predict random noise. Therefore, the approach deals with the general behavior of the data using Fourier analysis.

The literature survey about hourly solar radiation of a specific region yields the existence of various mathematical models. The classical model for hourly solar radiation was presented by S. N. Kaplanis (Kaplanis, 2006), which is also shown in Equation (2.32).

$$I(h, n_j) = a(n_j) + b(n_j) \cdot \cos\left(\frac{2\pi h}{24}\right) \tag{2.32}$$

The classical model has one DC and one cosine component for expressing the hourly behavior of solar radiation in a day. The coefficients $a(n_j)$ and $b(n_j)$ vary according to the day, n_j. This model is altered in another work as in Equation (2.33) (Kaplanis and Kaplani, 2007).

$$I(h, n_j) = A(n_j) + B(n_j) \cdot \frac{e^{-\mu(n_j)x(h)} \cos\left(\frac{2\pi h}{24}\right)}{e^{-\mu(n_j)x(h=12)}} \tag{2.33}$$

This model has an additional exponential term that depends on $\mu(n_j)$, which is the solar beam attenuation coefficient that is modeled using extra-terrestrial radiation and daily global solar radiation.

In this thesis, the proposed model is inspired by the above classical models. The model in Equation (2.32) considers only one harmonic term, and Equation (2.33) contains nonlinearities. In that aspect, our proposal is an extension of Equation (2.32) to multiple harmonics, avoiding nonlinearities. Naturally, the incorporation of harmonics indicates the utilization of FS coefficients. It must be noted that some previous researches about solar radiation modeling also include FS theory (Dorvlo, 2000; Rahoma and Hassan, 2007; Fidan et al., 2009; Hocaoğlu et al., 2009b). However, in those works, the Fourier series expansion was used for modeling daily solar radiation (instead of hourly solar radiation data).

As indicated above, the classical model shown in Equation (2.32) can be thought as a Fourier series with one cosine harmonic. Here, the classical model is expanded using other cosine and sine harmonics as in Equation (2.34).

$$I(h,n_j) = a_{n_j,0} + \sum_{i=1}^{N}\left[a_{n_j,i}\cdot\cos\left(\frac{2\pi ih}{24}\right) + b_{n_j,i}\cdot\sin\left(\frac{2\pi ih}{24}\right)\right] \tag{2.34}$$

The main goal of this expansion is to improve the accuracy of the model. If one-day data is modeled with Equation (2.34), one-year data can be notated in a matrix, which is constructed from FS coefficients, as shown in Equation (2.35).

$$\begin{bmatrix} a_{1,0} & a_{1,1} & \cdots & a_{1,N} & b_{1,1} & \cdots & b_{1,N} \\ \vdots & & & & & & \\ a_{n_j,0} & a_{n_j,1} & \cdots & a_{n_j,0} & b_{n_j,1} & \cdots & b_{n_j,N} \\ \vdots & & & & & & \\ a_{365,0} & a_{365,1} & \cdots & a_{365,N} & b_{365,1} & \cdots & b_{365,N} \end{bmatrix} \tag{2.35}$$

Given the hourly solar radiation for a day as $I(h,n_j)$, where h stands for the hour of day n_j, the FS coefficients, $a_{n_j,0}$, $a_{n_j,i}$ and $b_{n_j,i}$ can be calculated as shown in Equations (2.36), (2.37) and (2.38), respectively.

$$a_{n_j,0} = \frac{\sum_{h=-11}^{12} I(h,n_j)}{24} \tag{2.36}$$

$$a_{n_j,i} = \frac{\sum_{h=-11}^{12}\left[I(h,n_j)\cos\left(\frac{2\pi ih}{24}\right)\right]}{\sum_{h=-11}^{12}\left[\cos\left(\frac{2\pi ih}{24}\right)\right]^2} \tag{2.37}$$

$$b_{n_j,i} = \frac{\sum_{h=-11}^{12}\left[I(h,n_j)\sin\left(\frac{2\pi ih}{24}\right)\right]}{\sum_{h=-11}^{12}\left[\sin\left(\frac{2\pi ih}{24}\right)\right]^2} \tag{2.38}$$

These parameters will be used in the NN model as input/output data. The recovery of the hourly information is obtained by the inverse Fourier transform.

Starting with the Fourier model of the hourly solar radiation in Equation (2.34) , the coefficient, $a_{n_j,0}$, corresponds to the mean (average) of the values

for day n_j. Experimentally, it was observed that the relation between $a_{n_j,1}$ and $a_{n_j,0}$ fits to a simple linear model in the form of $a_{n_j,1}=m\,a_{n_j,0}+n$. Consequently, a coarse inter-coefficient model is obtained as:

$$\begin{aligned} a_{n_j,0} &= I(n_j), \\ a_{n_j,1} &= mI(n_j) + n \end{aligned} \tag{2.39}$$

In Kaplanis et al, $H(n_j)$, which is the daily solar radiation at surface, was defined as:

$$H(n_j) = c_1 + c_2 \cos\left(\frac{2\pi n_j}{364} + c_3\right) \tag{2.40}$$

Experimentally, the equivalence of $I(n_j)$ to the above expression is confirmed, thus we have:

$$I(n_j) = c_1 + c_2 \cos\left(\frac{2\pi n_j}{364} + c_3\right) \tag{2.41}$$

as well. According to the dependence of $a_{n_j,1}$ and $a_{n_j,0}$, we have:

$$\begin{aligned} a_{n_j,1} &= m \cdot \left[c_1 + c_2 \cos\left(\frac{2\pi n_j}{364} + c_3\right)\right] + n, \\ &= d_1 + d_2 \cos\left(\frac{2\pi n_j}{364} + d_3\right) \text{ where } d_3 \cong c_3 \end{aligned} \tag{2.42}$$

The coefficients c_1, c_2 and c_3, can be calculated as in Equations (2.43-2.45).

$$c_1 = \frac{\sum_{n_j=-181}^{182} I(n_j)}{364} \tag{2.43}$$

$$c_2 = \max\left(\frac{\sum_{n_j=-181}^{182} I(n_j)\cos\left(2\pi(n_j+k)\right)}{\sum_{n_j=-181}^{182}\left[\cos\left(2\pi(n_j+k)\right)\right]^2}\right), \quad k = -181,-180,\cdots,181,182. \tag{2.44}$$

$$c_3 = \frac{2\pi}{364}\arg\max_{M}\left\{\left(M = \frac{\sum_{n_j=-181}^{182} I\left(n_j\right)\cos\left(2\pi\left(n_j+k\right)\right)}{\sum_{n_j=-181}^{182}\left(\cos\left(2\pi\left(n_j+k\right)\right)\right)^2}\right), \quad k=-181,-180,\ldots 181,182. \right\} \tag{2.45}$$

3. DISCRETE TIME MARKOV CHAINS

3.1. Time-Invariant Markov Chains

The time-invariant Markov Chain model is the model explained in Section 1.1.2, which assumes that state transition probabilities will not be changed over time. According to calculate heathy transition probabilities, there should be enough number observations about the signal which could be take a long period. Therefore there should enough history data to calculate sufficient transition matrix.

3.2. Piecewise Time-Invariant Markov Chains

In some kind of data, the statistical characteristics can show seasonal changes. However, these characteristics does not change inside that seasons. For example, the wind speed shows different state transitions in winter and summer seasons. However the same seasons can have the same statistical charateristics. Therefore transition matrix can show piecewise time invariant property. This model is thought for the signals which do not have a stable statistical characteristics. Therefore state transition matrix should be updated over time. In this kind of Markov Chain models, the size of sliding window should be enough to calculate meaningful transition probabilities. This model also shows randomness of the signal. This model was mentioned in a previous research(Hosseini et al., 2012). The transition probability in piecwise-time invariant case can be shown as Equation (3.1).

$$P\left(X_{n+1}(T_k) = S_j \middle| X_n(T_k) = S_i, X_{n-1}(T_k), \cdots, X_0(T_k)\right) = P\left(X_{n+1}(T_k) = S_j \middle| X_n(T_k) = S_i\right) \quad (3.1)$$

T_k is the specific time period. *k* denotes the transition probaility is determined for which time period. For instance if the wind speed model is handled, T_1 shows

the winter period and T_2 shows the summer period. Therefore transition probabilities show variations in different time periods. However it is invariant in every single time periods. According to this property, A DTMC $\{X_n, n \geq 0\}$ is said to be piecewise time-invariant if Equation (3.2) is satisfied.

$$P\big(X_{n+1}(T_k) = S_j \big| X_n(T_k) = S_i\big) = P\big(X_1(T_k) = S_j \big| X_0(T_k) = S_i\big) \quad (3.2)$$

If the transition probability from S_i to S_j is taken as Equation (3.3), then the state transition probability matrix A is obtained as Equation (3.4) for N number of states.

$$a_{ij}(T_k) = P\big(X_{n+1}(T_k) = S_j \big| X_n(T_k) = S_i\big) \quad (3.3)$$

$$A(T_k) = \begin{bmatrix} a_{11}(T_k) & \cdots & a_{1N}(T_k) \\ \vdots & \ddots & \vdots \\ a_{N1}(T_k) & \cdots & a_{NN}(T_k) \end{bmatrix} \quad (3.4)$$

4. VARIATIONS OF ERROR TUNING MODEL

4.1. Time-Invariant Markov Chains with the Predictor Error Defined by Time-Invariant PDF (TIMTIP)

If the signal shows a stable correlation with the chosen predictor model, predictor error which will be used for Gaussian distribution can be calculated as constant. The 3d representation of Gaussian distribution shown in Figure 4.1 illustrates that, however the center of the Gaussian distributions changes due to time, the standard deviation and the maximum state probability do not change. In addition if the signal itself is also statistically stable, a time invariant model can be used for the calculation of state transition matrix. This hybrid model is expected to more useful for stochastic signals, which are tried to be predicted by stochastic prediction models.

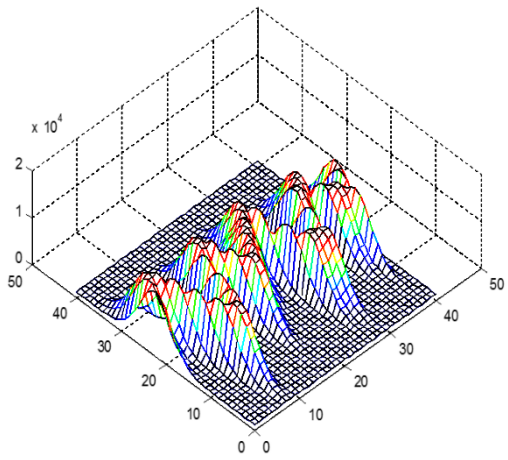

Figure 4.1. Example of Gaussian distribution fitting with time invariant RMSE of predictor

4.2. Time-Invariant Markov Chains with the Predictor Error Defined by Time-Varying PDF (TIMTVP)

If the performance of the predictor is changing over time, the prediction error, which will be used for Gaussian distribution, should be thought as time varying. The 3d representation of Gaussian distribution shown in Figure 4.2 illustrates that, both the center of the Gaussian distributions, the standard deviation and the maximum state probability changes due to time.

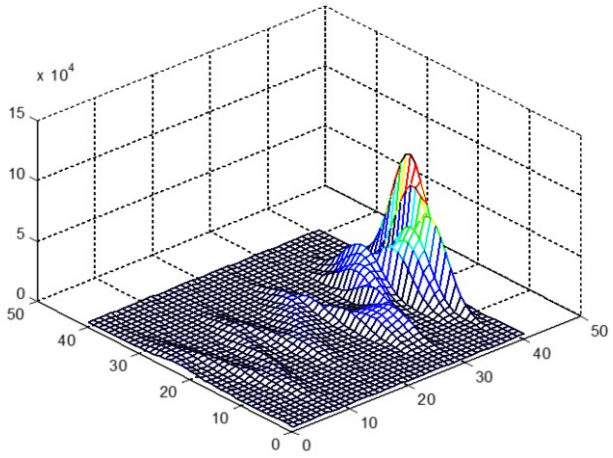

Figure 4.2. Example of Gaussian distribution fitting with time varying RMSE of predictor

On the other hand if the signal itself is statistically stable a time invariant model can be used for the calculation of state transition matrix. This hybrid model is expected to more useful for stochastic signals, which are tried to be predicted by deterministic prediction models.

4.3. Piecewise Time-Invariant Markov Chains with the Predictor Error Defined by Time-Invariant PDF (PTIMTIP)

If the signal shows a stable correlation with the chosen predictor model, predictor error which will be used for Gaussian distribution can be calculated as constant. On the other hand if the signal itself is statistically unstable, a time varying model can be used for the calculation of state transition matrix. This hybrid model is expected to more useful for deterministic signals, which are predicted by deterministic prediction models.

4.4.Piecewise Time-Invariant Markov Chains with the Predictor Error Defined by Time-Varying PDF (PTIMTVP)

If the performance of the predictor is changing over time, the prediction error, which will be used for Gaussian distribution, should be thought as time varying. On the other hand if the signal itself is statistically unstable, a time varying model can be used for the calculation of state transition matrix. This hybrid model can be useful for deterministic signals that are tried to be predicted by stochastic prediction models.

5. INCREASING NUMBER OF DIMENSIONS FOR MULTIDIMENSIONAL ANALYSIS

In some one dimensional signals, the behaviour of the signal can changes with a meaningful period. According to illustrate this kind of changes, multidimensional analysis is a more suitable way than one dimensional analysis. However to increase the number of dimensions, a meaningful period should be used. For example, for solar radiation if we take hour as a first dimension, day is a good choice for second dimension. Since the solar mnovements have the similarities in the same hours of different days, day is the best choice for second dimension. However choosing the second dimension can not be as trivial as solar radiation data. For example, for financial data this period can be changed. To choose the best candidate for second dimension, two diferent ways can be applied. Firstly signal will be low pass filterred for noise deletion than the poer spectral density will be analyzed. The maximum value of power spectral density shows the frequence of strongest harmonic of the signal which gives the most suitable period for second dimension. In second way signal will also be low pass filterred. Then the autocorrelation of low-pass filtered signal will be analyzed. The second larger local maxima of the autocorrelation will give us the best candidate for the second dimension. The autocorrelation is calculated by Pearson product-moment correlation coefficient which can be defined as Equation (5.1) (Rodgers and Nicewander, 1988).

$$r_{\chi_i\chi_j} = \frac{\sum_{n=0}^{N-1}[(\chi_i(n)-\overline{\chi_i})(\chi_j(n)-\overline{\chi_j})]}{\sqrt{[\sum_{n=0}^{N-1}(\chi_i(n)-\overline{\chi_i})^2]\left[\sum_{n=0}^{N-1}(\chi_j(n)-\overline{\chi_j})^2\right]}} \tag{5.1}$$

For instance the 1d-autocorrelation of the 4 years long hourly wind speed data of İzmir is plotted as Figure 5.1.

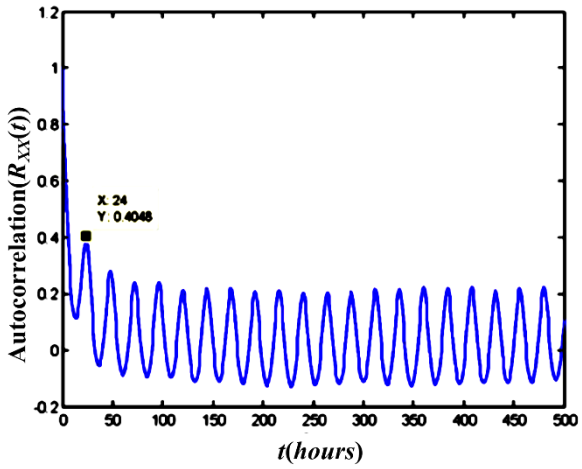

Figure 5.1. One dimensional autocorrelation of 4 years long hourly wind speed data of İzmir

According to Figure 5.1. the maximum correlation after $R_{xx}(0)$ is $R_{xx}(24)$. That means 24 hours is most suitable period for first dimension. Therefore the second dimension shows the day of the year. According to this selections wind speed can be represented as Figure 5.2.

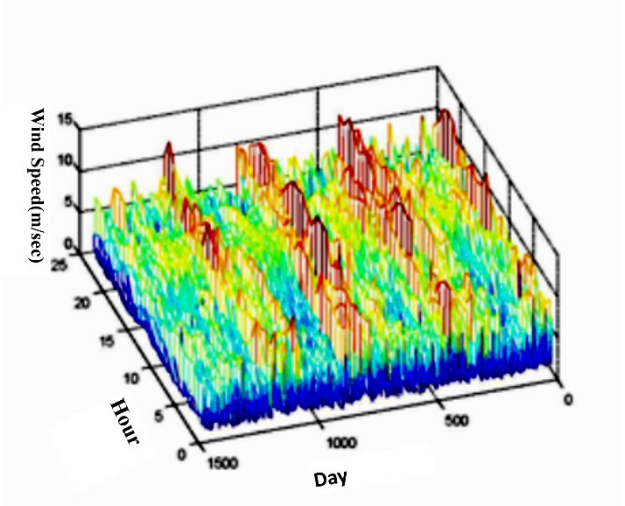

Figure 5.2. Two dimensional representation of 4 years long hourly wind speed data of İzmir

In other example, motor sound data taken by a microphone which is specified in Section 6.4 is given as Figure 5.3. The autocorrelation and 2d image representation according to the period found in autocorrelation are shown as Figure 5.4 and Figure 5.5 respectively.

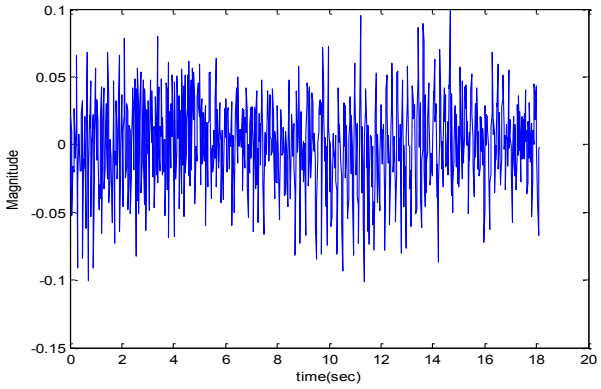

Figure 5.3. Sound data taken by a microphone

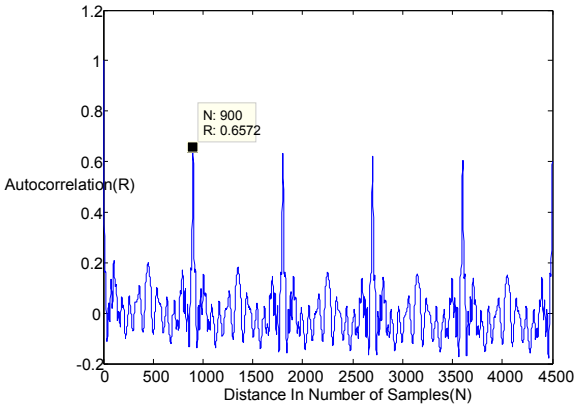

Figure 5.4. 1d autocorrelation of the sound data

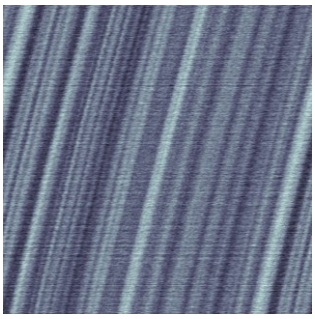

Figure 5.5. 2d image representation of sound data recorded from first microphone over Motor1

6. CASE STUDIES AND COMPARATIVE RESULTS

6.1.Wind Speed Prediction

Wind speed data modeling and prediction remains an important subject for energy planning. Specifically, the prediction of general statistics and the distribution of the wind data are of vital importance for determining the wind regime of a region on earth. Generally, a Weibull distribution and its parameters are used for wind regime studies because it gives a fair (but coarse) information regarding the overall wind potential of a site (Ulgen and Hepbasli, 2002; Ettoumi et al., 2003; Celik, 2004). More specific modeling and time variation analysis of wind speed data was examined in relatively fewer works. Because of the transitional behavior of wind speed, Markov models became popular in that area. As examples, Sahin and Sen have modeled the wind speed data measured from the Marmara region of Turkey using first order Markov chains (Sahin and Sen, 2001). In one of the previous studies, first order Markov chain models were used for synthetic generation of hourly wind speed time series in the Corsica region (Torre et al., 2001). Shamshad et al. have generated hourly wind speed data using first and second order Markov chains and compared the first and second order Markov chains using wind speed data measured from two different regions in Malaysia (Shamshad et al., 2005). In their study, it was concluded that the wind speed behavior slightly improves by increasing the Markov model order. Recently, Hocaoglu et al. also modeled the wind speed data using Markov chains. It is presented in that study that increasing the state size of Markov has important effects for the quality of generated data from a Markov process (Hocaoglu et al., 2008).

This study presents a different and novel approach to the analysis of time variations in wind speed data by means of short term prediction. The prediction approach utilizes the Mycielski algorithm, which is occasionally used for communication applications. The underlying flavor of Markov modeling consists of constructing a value transition probability table using a training set of recorded wind speed samples. Although the idea of the Mycielski algorithm

also depends on learning from past samples, unlike the Markov approach of building transition probabilities, it considers the past data samples as a whole during the prediction.

In this case study, in order to show the efficiency of the Mycielski approach in wind speed data modeling, the data belonging to different geographical regions of Turkey (Kayseri, Izmir and Antalya) are distinctively selected. The data were recorded in the year 2005, and were in the ranges of 0–11.5, 0.1–13.6 and 0–15.3 for the Kayseri, Izmir and Antalya regions, respectively. Since the model requires integer values (so that exact comparisons can be made), the data is first converted to wind speed states basically by rounding to the nearest integers. These integerized values will be notated as ‘‘rounded measurements”.

After the wind speed values are converted to wind states by rounding, the Mycielski prediction is applied throughout the available data. As suggested by the Mycielski prediction method, the wind states data sequence is examined by looking for the longest template ending at the end of the sequence which had appeared in the history of the wind state sequence. The prediction procedure is applied for each wind state. The main motive in applying the Mycielski algorithm to wind speed data was the relatively stationary behavior of wind data. As a matter of fact, the prediction methods that depend on Markov models also rely on this assumption. The stationary behavior can be illustrated by the following example. Let m indicate a wind state within a time interval, t. It is expected that, in the next time interval, the wind speed state will most probably be within the same state (m), or near to this state (m+1 or m-1). Large deviations from the state value of m are rare. Another assumption of the Mycielski predictor is the ‘‘repeated” behavior of the data. This behavior is found to be reasonable considering the short term cyclic pattern corresponding to day-night transitions, and longer term patterns corresponding to seasonal weather variations. Therefore, it is assumed that there should be some sequences (long or short) that repeat themselves in the history of the data. This idea was translated into forms of transition probabilities in Markov models. Here, the idea is tested in its absolute repeating structure.

The wind speed behavior consists both deterministic and probabilistic processes. Therefore combination of a stochastic and deterministic models is assumed to fit the wind speed data. In the previous studies Markov and Mycielski based methods are used for prediction of wind speed which are both stochastic models. In this novel approach deterministic part of the wind is attempted to be more distinguishable. The wind speed data for 4 hours with substraction of daily means can be represented as Figure 6.1 in 2d and oly for winter seasons it can be represented as Figure 6.2.

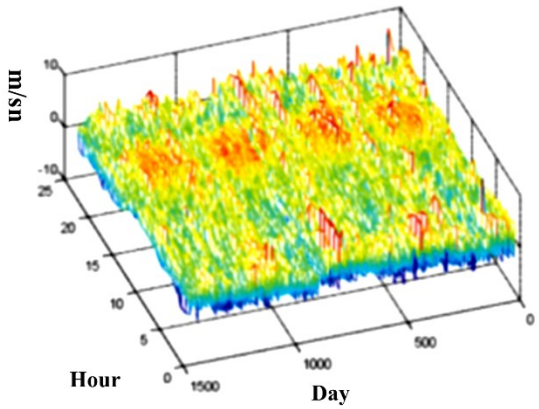

Figure 6.1. 4 year wind speed data with substraction of means of days

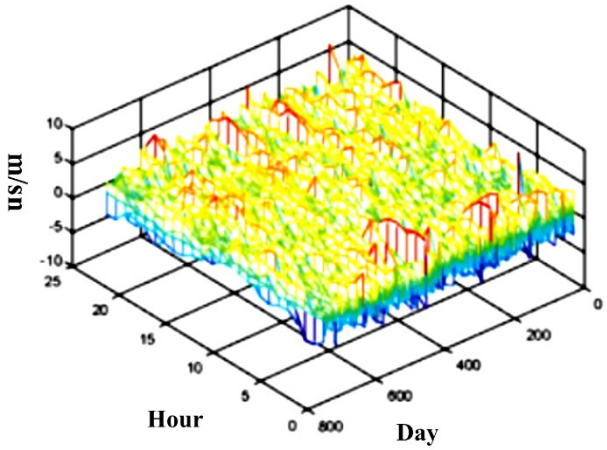

Figure 6.2. 4 year wind speed data with substraction of daily means for only winter seasons

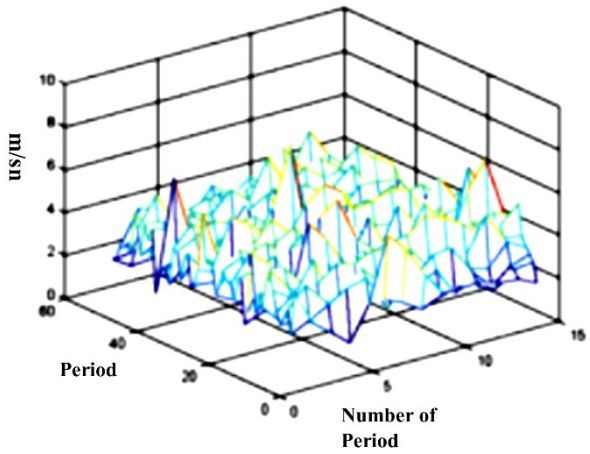

Figure 6.3. 2d representation of 4 year long daily wind speed data with arbitrary selected period

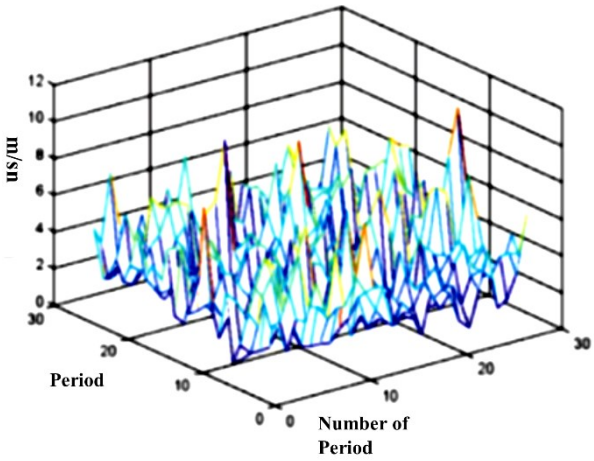

Figure 6.4. 2d representation of 4 year long daily wind speed data for winter seasons with arbitrary selected period

As a result of the illustrations above, daily means of the wind speed shows stochastic behavior. However hourly residues (especially for seasonal illustrations) have a deterministic behavior. That shows a hybrid stochastic and deterministic approach is more suitable for wind speed.

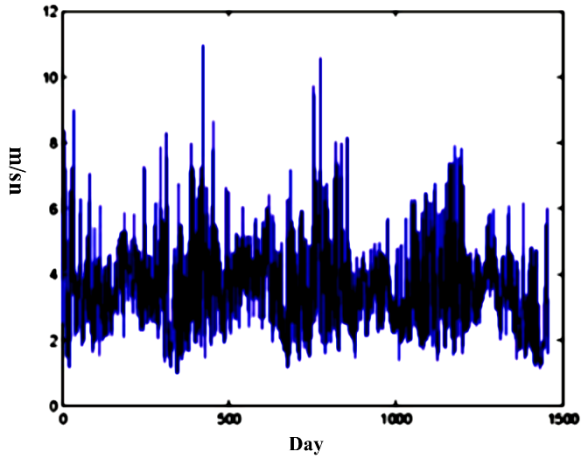

Figure 6.5. One dimensional representation of 4 year long daily wind speed data for winter seasons

The daily wind speed data which is shown in Figure 6.5 has the autocorrelation shown in Figure 6.6.

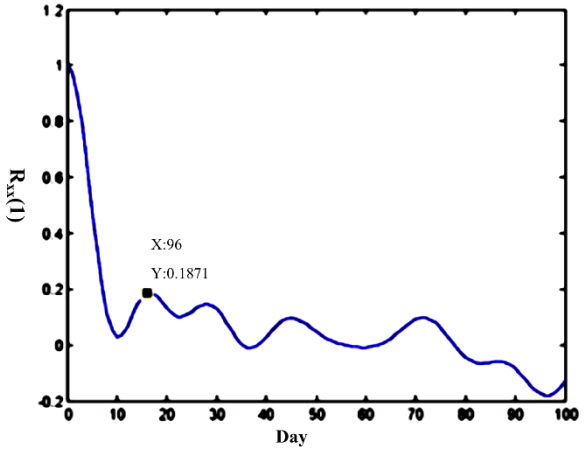

Figure 6.6. Autocorrelation of daily wind speed data

In order to found period in Figure 6.6 the daily wind speed data is represented in two dimensions as Figure 6.7 with 96 day long time period selection in one dimension.

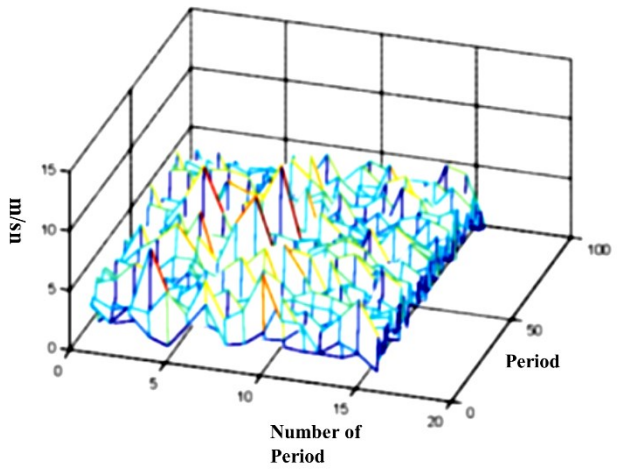

Figure 6.7. 2d representation of daily wind speed data with the selected period of 96 days long

The daily data is supposed to be modelled by Markovian , 1d-Mycielski, spiral Mycielski (for 2d representation) or another stochastic model.

Table 6.1. Comparative RMSE (m/sec) results of Mycielski based prediction methods and their prediction error tuned versions

Wind Speed Locations	1d-Myc	Spherical Myc	TIMTIP of Spherical Myc	PTIMTIP Of Spherical Myc
Kayseri	1.09	0.98	1.02	0.94
Antalya	1.37	1.28	1.20	1.12
İzmir	1.52	1.48	1.52	1.30

Table 6.2. Comparative RMSE (m/sec) results of linear prediction(LP) based prediction methods and their prediction error tuned versions

Wind Speed Locations	1d-LP	2d-LP	TIMTIP of 2d-LP	PTIMTIP Of 2d-LP
Kayseri	1.02	0.99	0.96	0.89
Antalya	1.25	1.20	1.18	1.15
İzmir	1.36	1.28	1.30	1.27

6.2. Solar Radiation Prediction

Forecasting and accurate modeling of solar radiation in a particular region remains to be an important engineering problem. Efficiency of solar energy generators, solar heat systems, or even the architectural design applications depends on the robust solar radiation model. In a recent study the importance of monthly, daily and hourly solar radiation for the sizing of solar systems is analyzed (Khare and Rangnekar, 2014). In another study the crucial role of the solar radiation model on a solar chimney power plant is explained (Guo et al., 2014). The radiation data is naturally stochastic due to atmospheric effects. The eminent randomness makes it difficult to forecast the solar radiation in any hour of the day accurately. Consequently, many researchers deal with the solar radiation forecasting or modeling issue from a wide scale of mathematical and practical approaches. The approaches range from the correlation behavior (spectral density) of the data, to more adaptive methods, including neural networks (NNs). Typical studies are summarized below. Several studies are carried out using Markov models (Aguiar et al., 1988; Maafi and Adane, 1989).

Kaplanis and Kaplani developed a stochastic simulation model for PV sizing (Kaplani and Kaplanis, 2012). A different stochastic model based on hidden Markov model is developed in a previous study (Hocaoğlu, 2011). A model is constructed using a novel visualization method by Hocaoglu et al. (Hocaoğlu et al., 2008). In that study, the sequential solar radiation data are rendered in a 2-dimensional (2d) matrix to utilize image-processing methods such as optimal coefficient linear prediction filters with Neural Networks. Almorox et al. employed air temperature data to predict daily solar radiations in Madrid (Almorox et al., 2011). There are some more examples for utilization of NNs in both daily and hourly solar radiation data prediction models (Mellit et al., 2005; Mellit et al., 2006; Moghaddam and Seifi, 2011). In another recent study principal component analysis is applied to model the variability of solar radiation data (Zarzo and Marti, 2011). Yang et al. use Support Vector Machines for short-term solar radiation prediction(Xiyun et al., 2013). An alternative approach for the solar radiation data analysis is the harmonic (Fourier) analysis. In such a study, harmonic analysis is performed over the daily solar radiation data belong to different regions of Turkey (Fidan et al., 2009; Hocaoğlu et al., 2009b). A similar study is also presented for Seeb region (Dorvlo, 2000). Harmonic analysis is performed for seven different regions in Oman in another study (Dorvlo and Ampratwum, 2000).

In this thesis, Izmir, Turkey is selected as the sample region. Using the methodology in Section 2.7 over the hourly solar radiation values in 2004, the $a_{n_j,0}$ and $a_{n_j,1}$ coefficients were obtained as:

$$a_{n_j,0} = 209.6827 + 120.6825\cos\left(\frac{2\pi n_j}{364} - 0.0173\right) \tag{6.1}$$

$$a_{n_j,1} = 317.4188 + 157.3008\cos\left(\frac{2\pi n_j}{364} + 0.0173\right) \tag{6.2}$$

The coefficients, c_3 and d_3 were naturally found the same, with a value difference within a small error limit of 0.0346 radians.

The daily calculated $a_{n_j,0}$ values and the corresponding model values (derived from Equation (6.1) and (6.2)) for the years 2004 and 2005 are shown in Figure 6.8.

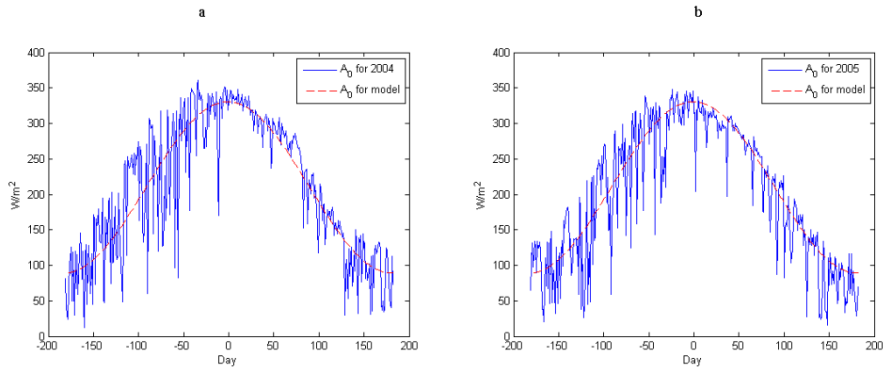

Figure 6.8. Actual and model values of a_0 in year (a) 2004 and (b) 2005

Figures 6.8 indicates that there is a reasonable resemblance between the model and the actual values, which verifies the efficiency of the proposed model. The sample-wise error between actual and model values for $a_{n_j,0}$ in year 2004 and 2005 is shown in Figure 6.9.

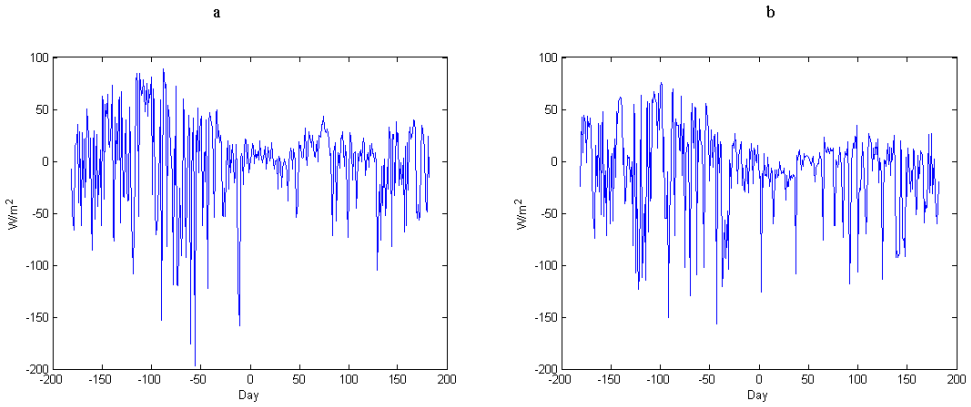

Figure 6.9. Sample difference between actual and model values of a_0 in year (a) 2004 and (b) 2005

Here, the average error between the $a_{n_j,0}$ model and $a_{n_j,0}$ calculated from the 2004 radiation data is -7.8082×10^{-16}, the root mean square error (RMSE) is 43.61. Comparing this error to the root mean energy of $a_{n_j,0}$ (for the same year,

2004), which is 230.545; the model is visibly capable of accurately estimating the radiation pattern. Similarly, the average error between actual and model generated $a_{n_j,0}$ values for year 2005 is 8.5609, and the RMSE is 42.67, compared to the root mean energy of $a_{n_j,0}$, which is 222.627. Achievement of fair model accuracy for two separate years indicates that the proposed model for the $a_{n_j,0}$ coefficient is pretty robust.

The same comparison suit is carried out over the $a_{n_j,1}$ values for 2004 and 2005. The model-versus-actual value comparisons are shown in Figure 6.10 and Figure 6.11.

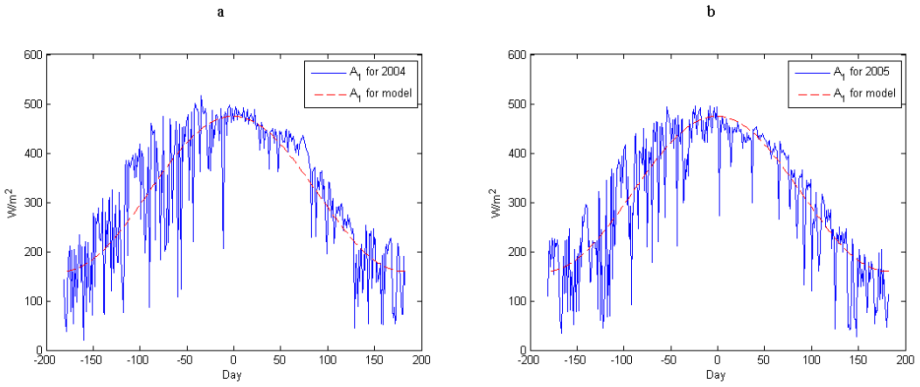

Figure 6.10. Actual and model values of a_1 in year (a) 2004 and (b) 2005

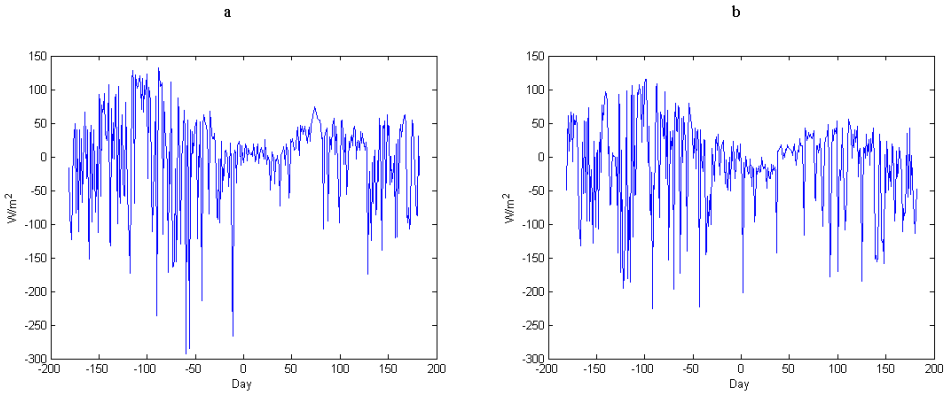

Figure 6.11. Sample difference between actual and model values of a_1 in year (a) 2004 and (b) 2005

The quantitative analysis of the model generated and actual values of the $a_{n_j,1}$ coefficient yield that, similar to the previous coefficient, the average value between the value and the model output is negligible, whereas the RMSE value causes an energy decrease by a factor over 5 (69.21 versus 343.380 for year 2004 and 66.88 versus 334.036 for year 2005).

6.3. Energy Demand Prediction

In this thesis daily energy demand data is attempted to be modelled. The energy demand prediction is a crucial problem for identifying future energy needs of cities and countries. According to this prediction, the size of necessary power plants are determining. The 4 year hourly power consumption of whole Turkey can be shown as Figure 6.12. The daily mean of the 4 year consumption data is shown in Figure 6.13.

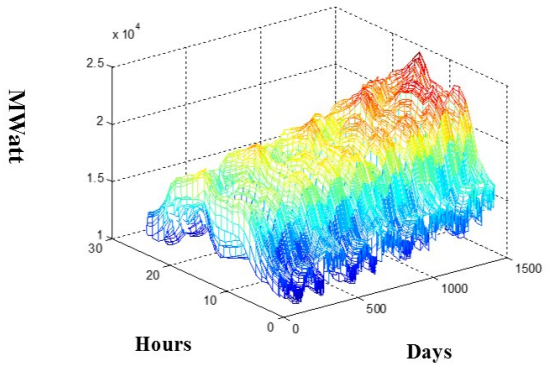

Figure 6.12. The four years long hourly power consumption of Turkey

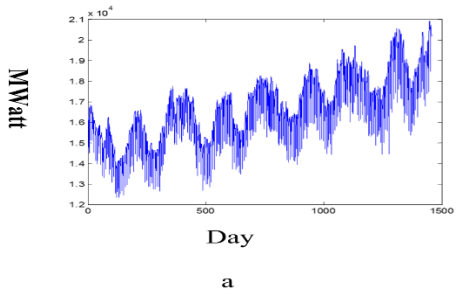

Figure 6.13. Daily mean of the four years long hourly data

As seen in Figure 6.13, The moving average of the energy demand shows and monotonic incremental behavior because of increasing population and consuming. If the EMD is applied to this daily mean data, the 3 modes can be plotted as Figure 6.14.

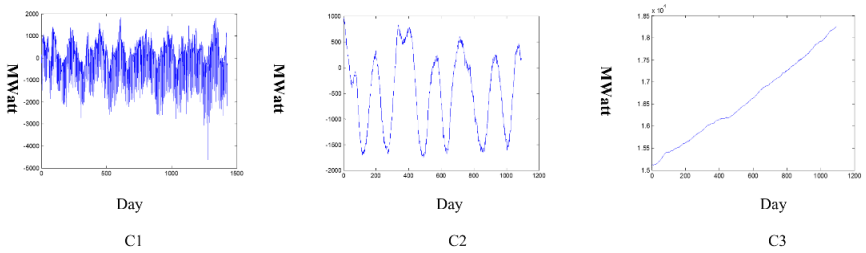

Figure 6.14. Three modes of daily meaned energy consumption data after EMD

The component C3 is monotonicly increasing; therefore it can be modelled by a polynomial or another deterministic model. In addition for C1 and C2, Mycielski algorithm can be applied because of the repeated behavior. The experiment shows that the energy demand behavior can also be inspected by Mycielski decomposition which gives an opportunity for displaying long term and short term behavior at the same time.

6.4.Motor Fault Diagnosis

Due to their simple construction, cost effective pricing and easy maintenance, the squirrel cage induction motors are the most preferable electrical motors in industry. In order not the interrupt the industrial processes

with unexpected failures, which cause downtime and repair costs, preventive maintenance strategies are essential for electric motors. Early diagnostics of incipient faults in induction motors are important to ensure safe operation and help to recognize and fix the problems with low costs and time.

Significant amount of research have been focused on the methods for the early detection of the mechanical and electrical faults in induction motors (Nandi et al., 2005). Among all the methods in literature, motor current signature analysis (MCSA) is one the most popular one, which provides an effective way to detect incipient faults. MCSA is mainly focuses on the analysis of the current data that supplied from the ac network to the induction motor with time-frequency analysis techniques like Fast Fourier Transform (FFT), Short Time Fourier Transform (STFT), Wavelet Transform or Wavelet Packet Transform (Benbouzid, 2000). However there is a bottleneck to apply this technique to induction motors in their working environment since in most cases obtaining data is a cumbersome process because additional circuitry like isolators or data acquisition cards and interface should be added between the supply and the test motors. Also it may not be possible to detach load from motor and run motor under no load condition. In order to get rid of disadvantages of current based techniques like MCSA, the acoustic and vibrational methods are getting popular since those approaches do not need detaching the motor from their working environment. Fault diagnosis methods based on vibration analysis has been applied for many years using the methods such like Fourier and Wavelet Transforms and showed significant success especially on detecting faulty bearing detection (Ocak and Loparo, 2004).

Vibration signal can be obtained via a contact device such as accelerometers, however in some specials cases it may be a little difficult. For example the surface of the test motor may be irregular and the device can't be located properly on the test motors. Also the surface may be greasy or humid due to the hard operating conditions like high temperature or humidity (Lu et al., 2012). On the other hand, sound based fault diagnosis of motors offers a great advantage that solves all these problems is being contactless. Only external microphones located around the operating motor are enough to record the

information. Unlike vibration based analysis, there is very limited literature on fault diagnosis of induction motors which based on techniques of sound analysis. Also some of the works analyze the sound data which is recorded in an echo-free silent environment, which seems impractical for the real life applications, especially in industrial processes (Benko et al., 2004).

In this case study, sound data is collected from six different induction motors, first five of them having specific incipient mechanical or electrical faults, via five microphones surrounding the test rig. Sound data can be analyzed like current or vibration data and can possibly contain many fault related information for diagnosis. Therefore this part of the thesis aims to extract the useful information which can classify the healthy mode from the faulty modes, and distinguish the fault types between the faulty modes. This work has two main strategies for extracting necessary features from sound data that used to express motor fault type. In the first strategy, features obtained directly by calculating cross correlation coefficients of the sound data recorded by the microphone pairs. Ten different correlation coefficients are calculated for each motor for every trial, in other words, all possible pair combinations of five different microphones are used to extract features. In the second strategy additional features are extracted by using 2d wavelet decomposition of the grayscale images which are obtained by converting one dimensional sound data which are recorded by each separate microphone to 2d grayscale images. 2d representation is more convenient for feature extraction because there are many image processing tools in literature. The 2d representation of the sound data has many advantages over the regular one dimensional data such that useful data can be extracted that we couldn't have from regular data (Do and Chong, 2011). These two dimensional images are expected to show different type of textures and in past works for texture analysis, basically Wavelet based methods were used (Chang and Kuo, 1993; Laine and Fan, 1993; Unser, 1995; Fukuda and Hirosawa, 1999; Sebe and Lew, 2000; Arivazhagan and Ganesan, 2003; Hsieh et al., 2003; Ece and Gerek, 2004; Gerek and Ece, 2004; Semler et al., 2005; Dettori and Semler, 2007; Younus and Yang, 2012). These wavelet based texture analysis methods are applied on different kind of research areas including

classification of tomography images (Semler et al., 2005; Dettori and Semler, 2007), finger-print classification (Hsieh et al., 2003), analysis of SAR images (Fukuda and Hirosawa, 1999), power quality analysis from 2d represented power quality event data (Ece and Gerek, 2004; Gerek and Ece, 2004) and fault analysis of rotating machinery by using data obtained from infrared thermography (Younus and Yang, 2012).

6.4.1. Fault Types and Experimental Setup

In this work, six induction motors are used for obtaining the acoustic data. Motors are driven directly from AC network. A 3-phase, 25 kVA, Δ-Y connected isolation transformer is located between laboratory setup and the network and the motors are supplied with the output of this transformer. Test motors are 3-phase and 2-pole squirrel cage induction motors rated at 2.2 kW and 380-V line to line.

Some commonly encountered mechanical and electrical faults are created synthetically on the five of these motors. The first of these identical motors is left healthy as being a reference to others. In two of these motors, broken rotor bar fault realized by drilling holes to 3 and 5 rotor bars respectively over 18 bars. This type of faults cause a rise in magnitude at adjacent side band frequency components at the twice of the slip frequency sidebands located symmetrically at around the main frequency which the stator coils are supplied by. However, monitoring the increase in magnitude at the predicted sideband frequencies alone may mislead the classification because different types of faults may give rise to same frequencies. Also when the slip is small, these sidebands come close to the main frequency and it becomes very difficult to see them by inspecting the spectral information due the dominancy of the fundamental component.

Bearing deficiencies are another commonly encountered mechanical fault type in induction motors. Due to improper lubrication, corrosion and contamination, the bearing surfaces and balls inside the inner and outer races

may lose their perfection. Those types of faults cause abnormal noise and vibration during operation. Even the operation of the motor is not affected seriously; these incipient faults may deteriorate fast. Only considering current as fault diagnostic criteria it is not possible to track the acoustical traces of faults due to the noise and vibration, and may seriously leads ineffective classification. In the test bed, the bearings of two test motors are replaced with defective ball bearings of the motors that had been used for plenty many hours in industry. One of these two used bearings has problem with alignment and the other bearing has ball defect complication. Last type of fault, shorted stator winding problem, is generated to another motor with pealing the insulation of two adjacent coils for a few millimeters and soldering them together. The Table 6.3 shows the fault types of the setup.

Table 6.3. Table of faulty motors in experimental setup

Motor #	**Fault Type**
M1	Healthy Motor
M2	Bearing Fault (misalignment)
M3	Broken Rotor Bars (5 over 18)
M4	Bearing Fault (ball defect)
M5	Broken Rotor Bars (3 over 18)
M6	Short Circuit in Stator Winding

Figure 6.15. Laboratory setup

The resistor values on load banks are adjusted to six different values such that the motors are driven 3.6, 4.1, 4.7, 4.9, 5 and 5.4 Amperes stator current. Actually, these levels represent the conditions that motors run under different loads. First, motors start to operate and left for a few minutes before loading and then, resistor values changed consecutively to the levels given above. At each current level approximately 30 seconds of sound data is collected via a full transparent analog amplifier and through 5 microphones which are located around the test rig. One of these microphones is located approximately more than a half meter above the center of the test rig. Other remaining 4 microphones are set to construct the edges of a square which surrounds the test rig and located at a little lower height compared to the microphone at the center. The locations of the microphones are very important when constructing the cross correlation between them. This settlement of microphone array gives as a virtual hemisphere which covers all the experimental setup which can be seen in the Figure 6.16.

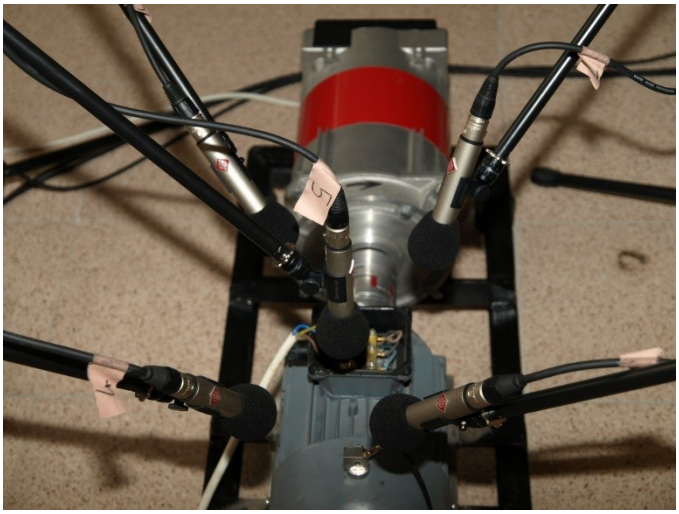

Figure 6.16. The settlement of the microphones over the test motor

The sound data is collected from this microphone array. The microphones are connected to an analog amplifier and approximately 20 to 30 seconds of sound data is collected for six motors under 6 different loading conditions.

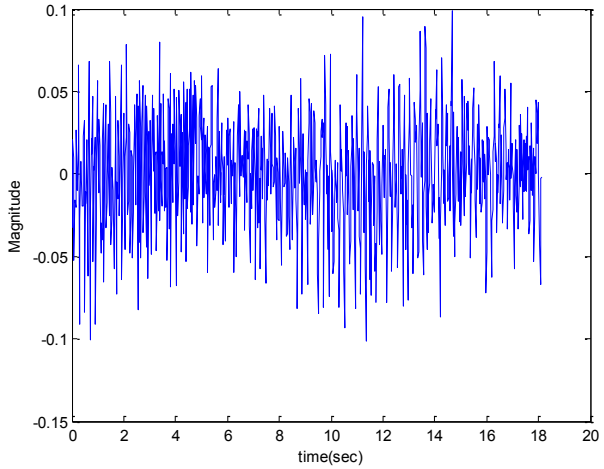

Figure 6.17. Sound data recorded by first microphone for Motor1

The sound data is digitized with a sampling frequency of 44.1 kHz which gives us digitized files of length approximately 1 million samples for each microphone. These procedures are repeated 3 times for each test motor by dissembling the motor from the test bad and reattaching it in order to obtain different types of working conditions. It was necessary because in every trail, the operator may place the motor slightly different from the previous trail and these small discrepancies create different artifacts in sound data. Since the test rig has mechanical parts like screws and wrenches, which fixes the motor to the metal test rig assembly, it is important to find features which are independent from operator related differences.

$$r_{\chi_i\chi_j} = \frac{\sum_{n=0}^{N-1}[(\chi_i(n)-\overline{\chi_i})(\chi_j(n)-\overline{\chi_j})]}{\sqrt{[\sum_{n=0}^{N-1}(\chi_i(n)-\overline{\chi_i})^2]\left[\sum_{n=0}^{N-1}(\chi_j(n)-\overline{\chi_j})^2\right]}} \tag{6.3}$$

In the Equation (6.3) χ_i and χ_j denote array data taken from different microphones (channels) where *i* and *j* denote different microphones. Here N is the number of samples in these arrays which is taken as 800000 in this work. Since 5 channels are used, there are $\binom{5}{2}$ in other words 10 coefficients to calculate. Therefore first 10 features of the feature set are obtained from these cross correlation coefficients.

Second strategy is an indirect method which uses 2d grayscale images obtained from sound data for feature extraction instead of using the data itself. While converting the one dimensional sound data to grayscale image, the amplitude of each sound data samples are normalized between 0 and 255, which is the pixel intensity range for grayscale. After normalization process, square shaped grayscale images are obtained by using this normalized data. 2d image representation is new in this area and has great potential in detecting power quality related events (Ece and Gerek, 2004; Gerek and Ece, 2004).

When driven directly from ac network, induction machines revolve very close but a little slower than the synchronous speed even under no load. It is impossible to reach synchronous speed unless a frequency invertor is located between the motor and the network. Also under loading conditions, the revolution speed of the motor decrease slowly. Because of these reasons, determination of the exact real rotor frequency is a cumbersome process and vital importance for constructing 2d images from sound data. The widths of the images are determined due to the length of a complete period of the power signal. Since the sampling frequency is 44.1 kHz, and motor fundamental frequency is 50 Hz, for each complete cycle, 882 samples are necessary. However in practice the fundamental frequency comprises fluctuations and in general less than 50 cycles, the 882 samples includes more samples than one cycle data. In these situations, in order to construct the 2d grayscale image with non-overlapping period segments, determination of the sample size of a complete period is crucial. Here the autocorrelation can solve this problematic issue. In order to determine the exact frequency, 1d autocorrelation sequence of each trial is analyzed. This autocorrelation sequence for motor1 under 3.6A is represented as in Figure 6.18.

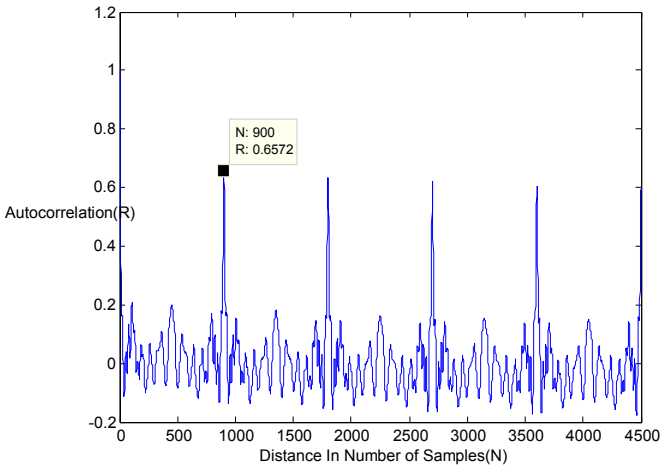

Figure 6.18. 1d Autocorrelation of Motor1 under 3.6 A

The values under other loading conditions are also calculated in the same manner and they are rounded to the nearest integer because the autocorrelation sequence gives non integer values. Rounded results for sample sizes of a single period for each loading conditions are given in tabular form in Table 6.4.

Table 6.4. Sample size of a period of the noise data collected from motors run under different loads

	3.6 A	**4.1 A**	**4.7 A**	**4.9 A**	**5.0 A**	**5.4 A**
M1	900	906	912	914	916	919
M2	905	910	914	917	918	921
M3	902	908	915	917	919	923
M4	901	907	908	916	919	923
M5	903	910	912	919	922	926
M6	901	906	913	914	915	920

First element of normalized data is assigned as first pixel value and assigned to the top-left corner of the image. The succeeding element is assigned to the right of first pixel and this procedure continues till the number of element that we determined by autocorrelation. After the last pixel assigned the first row, we continue these same procedure for second row and repeat the same procedure until we reach a square image whose width is the value calculated by

autocorrelation. The ultimate image representation of the data recorded by the first microphone over Motor1 can be shown as Figure 6.19.

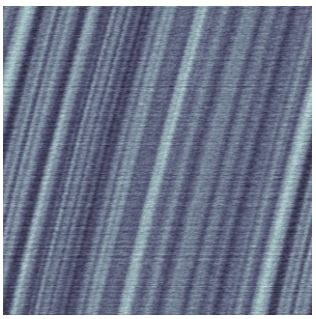

Figure 6.19. 2d Image Representation of Sound Data Recorded from first microphone over Motor1

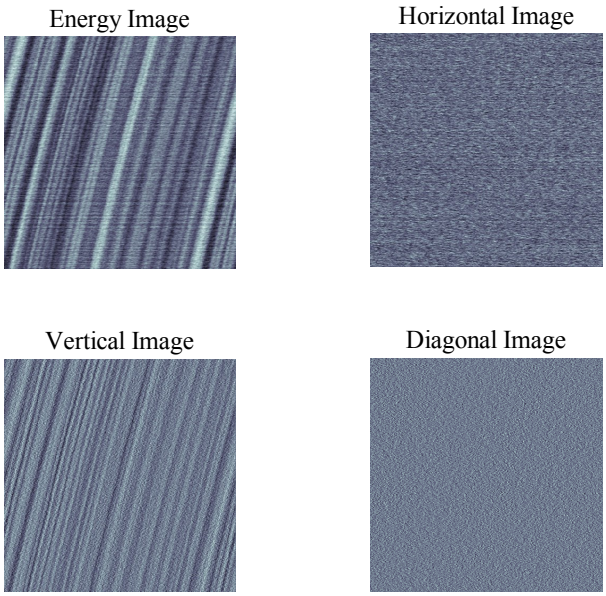

Figure 6.20. Wavelet decomposition of the 2d Image Representation given in Figure 6.16.

Images are obtained for each data for the recordings every separate five microphones for 6 test motors under 6 different loads for 3 different trials. Therefore, image database consist of 540 grayscale images, whose widths are ranging from 900 to 926 according to their autocorrelation value. In order to extract third group of features, single level 2d wavelet transformation is applied to these images. In 2d wavelet transform first, each row of the image is filtered by low-pass and high-pass filters. Then, outputs of both low-pass and high-pass filters are downsampled by 2 in order to obtain middle images L and H

respectively. L is the low-pass filtered and downsampled version and H is the high-pass filtered and downsampled versions of the original image on the vertical direction. Then, each column of these new images, L and H, are again filtered on the horizontal direction by low-pass and high-pass filters. Outputs of every filter are again downsampled by 2 in order to obtain four sub-band (LL, LH, HL, HH) images. In here, LL is the low-pass filtered version of the original image on vertical and horizontal directions and called as approximation image. Approximation image is also called the energy image. HL, named as vertical image, is vertically high-pass and horizontally low-pass filtered, LH, named as horizontal image, is vertically low-pass and horizontally high-pass filtered, and HH, named as diagonal image, is both horizontally and vertically high-pass filtered and downsampled versions of the original image. The wavelet decomposition of the image shown in Figure 6.19 can be demonstrated as in Figure 6.20.

By combining these 4 sub-band images, an image that has equal samples to the original can be obtained. When carefully inspecting these images, different textures which may be indicators of features related to certain faults can be seen even by eye as seen in Figure 6.21. The Figure 6.21 consists of diagonal (detail) images of the 6 different motor types.

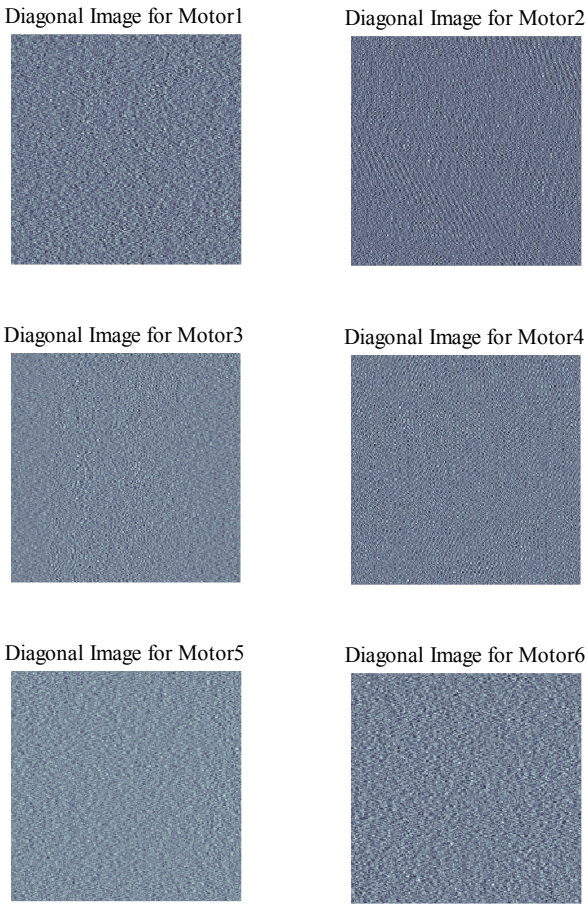

Figure 6.21. Diagonal image for different motor types

After one level 2d wavelet transform, 6 different features are extracted from these 4 sub-band images. First and second features are the root mean square energy of the vertical and diagonal images. Remaining features are related to the correlation. The row correlation of the horizontal image and the column correlations of the approximation, vertical, diagonal and energy images are used. In this selection, column correlation is more important because rows correspond to periods, but in columns it is possible to detect different type of patterns according to the type of the fault. In this work, the correlations between two adjacent rows or columns are calculated over the sub-images and the mean of these calculated values are taken as feature.

In the vertical and diagonal images there is strong information in their texture. These images have the distinguishable information mostly in their columns instead of rows. The main cause of this situation in the vertical image is the type of filter that is applied at the Wavelet decomposition to obtain this

vertical component. In addition, rows are approximately one period of the sound. Therefore there are no distinguishable differences between rows. According to this periodicity situation, the column information is also more dominant than the row information in the diagonal image. However the energy image is lack of texture information, the same situation of the diagonal image is also valid for the energy image. Therefore the column correlation is also used for energy image. The texture differences which are dependent to fault type can be identified by the relationship between neighbor columns of the images. For the calculation of column correlation Equation (6.4) is used, which is the mean of the cross-correlation coefficients of neighbor columns of the images.

$$r_{column} = \frac{\sum_{j=0}^{N-2} r_{I(\cdots,j)I(\cdots,j+1)}}{N-1} = \frac{\sum_{j=0}^{N-2}\left(\frac{\sum_{i=0}^{M-1}\left[\left(I(i,j)-\overline{I(\cdots,j)}\right)\left(I(i,j+1)-\overline{I(\cdots,j+1)}\right)\right]}{\sqrt{\left[\sum_{i=0}^{M-1}\left(I(i,j)-\overline{I(\cdots,j)}\right)^2\right]\left[\sum_{n=0}^{M-1}\left(I(i,j+1)-\overline{I(\cdots,j+1)}\right)^2\right]}}\right)}{N-1} \quad (6.4)$$

N:Number of Columns; *M*:Number of Rows

In spite of the dominance of the column information in texture, rows of the horizontal image also have some distinguishable information because of changes in the duration of one period of the sound. These periodicity changes cause some inconsistent shifting effects which can only be analyzed with relationship between neighbor columns of the horizontal image. For the row correlation Equation (6.5) is used, which is the mean of the cross-correlation coefficients of neighbor rows of the image.

$$r_{row} = \frac{\sum_{i=0}^{M-2} r_{I(i,\cdots)I(i+1,\cdots)}}{M-1} = \frac{\sum_{j=0}^{M-2}\left(\frac{\sum_{i=0}^{N-1}\left[\left(I(i,j)-\overline{I(i,\cdots)}\right)\left(I(i+1,j)-\overline{I(i+1,\cdots)}\right)\right]}{\sqrt{\left[\sum_{i=0}^{N-1}\left(I(i,j)-\overline{I(i,\cdots)}\right)^2\right]\left[\sum_{n=0}^{N-1}\left(I(i+1,j)-\overline{I(i+1,\cdots)}\right)^2\right]}}\right)}{M-1} \quad (6.5)$$

N:Number of Columns; *M*:Number of Rows

7. CONCLUSIONS

In this thesis, several stochastic and deterministic models are handled for generalization of real life data in one and multidimensional approach. According to express both deterministic and stochastic components of a unique real life problem as wind speed prediction, am novel error tuning model is improved. This novel model does not only combined stochastic and deterministic models, but also give the opportunity of inspecting in time varying case.

Time varying and piecewise-time invariant researches on the data give some alternative solutions for modeling non-stationary data, which is also the contribution of the proposed error tuning model.

In addition wavelet decomposition of solar radiation shows that high frequency and low frequency components of the solar radiation can be modelled in various models.

Fourier series approach shows also a similar result, which is naturally expected. A similar result is also obtained by empirical mode decomposition of the energy consumption data.

According to experiments of period estimation on motor sound and wind speed data, estimation of exact period can be said crucial for multidimensional nonlinear analysis. If the true period is estimated for the data, the multidimensional representation of the data according to this period gives a meaningful information.

The other contribution of the thesis is exhibition of the advantages of multidimensional versions of one dimensional methods like Mycielski and linear prediction methods. Multidimensional approach proves the information of the multidimensional patterns in the data. In addition using cylindrical coordinate system gives both advantage of continuity of one dimensional approach and advantage of multidimensional approach.

Finally, according to the results of the thesis it can be pointed that dividing a complex nonlinear real life problem into its atoms or its components and using a different model for each component depending to its own

characteristics, can be a more efficient strategy than using a single model for the whole problem.

REFERENCES

Aguiar, R. J., Collarespereira, M. and Conde, J. P. (1988), "Simple Procedure for Generating Sequences of Daily Radiation Values Using a Library of Markov Transition Matrices," *Solar Energy*, **40**(3), 269-279.

Almorox, J., Hontoria, C. and Benito, M. (2011), "Models for obtaining daily global solar radiation with measured air temperature data in Madrid (Spain)," *Applied Energy*, **88**(5), 1703-1709.

Arivazhagan, S. and Ganesan, L. (2003), "Texture classification using wavelet transform," *Pattern Recognition Letters*, **24**(9-10), 1513-1521.

Benbouzid, M. E. (2000), "A review of induction motors signature analysis as a medium for faults detection," *IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS*, **47**(5), 984-993.

Benedetto, J. J. and Wu, H. C. (2000), "Non-uniform sampling and spiral MRI reconstruction," *Wavelet Applications in Signal and Image Processing Viii Pts 1 and 2*, **4119**, 130-141.

Bengio, Y., LeCun, Y., Nohl, C. and Burges, C. (1995), "LeRec: a NN/HMM hybrid for on-line handwriting recognition," *Neural Comput*, **7**(6), 1289-1303.

Benko, U., Petrovcic, J., Juricic, D., Tavcar, J., Rejec, J. and Stefanovska, A. (2004), "Fault diagnosis of a vacuum cleaner motor by means of sound analysis," *Journal of Sound and Vibration*, **276**(3-5), 781-806.

Boudaren, M. E., Monfrini, E., Pieczynski, W. and Aissani, A. (2014), "Phasic Triplet Markov Chains," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **36**(11), 2310-2316.

Cao, J. C. and Cao, S. H. (2006), "Study of forecasting solar irradiance using neural networks with preprocessing sample data by wavelet analysis," *Energy*, **31**(15), 3435-3445.

Cao, S. H. and Cao, J. C. (2005), "Forecast of solar irradiance using recurrent neural networks combined with wavelet analysis," *Applied Thermal Engineering*, **25**(2-3), 161-172.

Celik, A. N. (2004), "A statistical analysis of wind power density based on the Weibull and Rayleigh models at the southern region of Turkey," *Renewable Energy*, **29**(4), 593-604.

Chang, T. and Kuo, C. J. (1993), "Texture analysis and classification with tree-structured wavelet transform," *IEEE Trans Image Process*, **2**(4), 429-441.

Chen, W. K. (2009), "*Feedback, Nonlinear, and Distributed Circuits*," Taylor & Francis.

Childs, R. E., Chuah, D. G. S., Lee, S. L. and Tan, K. C. (1984), "Analysis of Solar-Radiation Data Using Cubic-Splines," *Solar Energy*, **32**(5), 643-653.

Curtis, W. (1960), "Spiral antennas," *Antennas and Propagation, IRE Transactions on*, **8**(3), 298-306.

Delorenzi, M. and Speed, T. (2002), "An HMM model for coiled-coil domains and a comparison with PSSM-based predictions," *Bioinformatics*, **18**(4), 617-625.

Deng, J. W. and Tsui, H. T. (2000), "An HMM-based approach for gesture segmentation and recognition," *Proceedings of the Pattern Recognition, 2000. Proceedings. 15th International Conference on*, 679-682 vol.673.

Dettori, L. and Semler, L. (2007), "A comparison of wavelet, ridgelet, and curvelet-based texture classification algorithms in computed tomography," *Comput Biol Med*, **37**(4), 486-498.

Dixon, W. J. and Massey, F. J. (1969), "*Introduction to statistical analysis*," McGraw-Hill New York.

Do, V. and Chong, U. P. (2011), "Signal Model-Based Fault Detection and Diagnosis for Induction Motors Using Features of Vibration Signal in Two-Dimension Domain," *Strojniski Vestnik-Journal of Mechanical Engineering*, **57**(9), 655-666.

Dorvlo, A. S. S. (2000), "Fourier analysis of meteorological data for Seeb," *Energy Conversion and Management*, **41**(12), 1283-1291.

Dorvlo, A. S. S. and Ampratwum, D. B. (2000), "Harmonic analysis of global irradiation," *Renewable Energy*, **20**(4), 435-443.

Duanchao, C., Jianshe, K., Jianmin, Z. and Xinghui, Z. (2012), "Fault diagnosis of gearbox based on EEMD and HMM," *Proceedings of the Prognostics and System Health Management (PHM), 2012 IEEE Conference on*, 1-9.

Ece, D. G. and Gerek, O. N. (2004), "Power quality event detection using joint 2-D-wavelet subspaces," *IEEE Transactions on Instrumentation and Measurement*, **53**(4), 1040-1046.

Ehrenfeucht, A. and Mycielski, J. (1992), "A Pseudorandom Sequence--How Random Is It?," *American Mathematical Monthly*, 373-375.

Ettoumi, F. Y., Sauvageot, H. and Adane, A. E. H. (2003), "Statistical bivariate modelling of wind using first-order Markov chain and Weibull distribution," *Renewable Energy*, **28**(11), 1787-1802.

Fauchereau, N., Sinclair, S. and Pegram, G. (2008), "2-D Empirical Mode Decomposition on the sphere, application to the spatial scales of surface temperature variations," *Hydrology and Earth System Sciences Discussions*, **5**, 405-435.

Ferreiros, J., de Córdoba, R., Savoji, M. H. and Pardo, J. M. (1995), "Continuous Speech HMM Training System: Applications to Speech Recognition and Phonetic Label Alignment," *Speech Recognition and Coding*, A. R. Ayuso and J. L. Soler, Springer Berlin Heidelberg, **147,** 68-71.

Fidan, M. and Gerek, O. (2008), "Randomness analysis of Antimycielski number generator," *Proceedings of the Signal Processing, Communication and Applications Conference, 2008. SIU 2008. IEEE 16th*, IEEE, 1-4.

Fidan, M., Hocaoglu, F. O. and Gerek, O. N. (2009), "Effects of Temperature and Pressure Information in a Hybrid (Fourier Series/Neural Networks) Solar Radiation Model," *Proceedings of the Innovative Computing, Information and Control (ICICIC), 2009 Fourth International Conference on*, IEEE, 667-670.

Fidan, M., Hocaoglu, F. O. and Gerek, O. N. (2012), "Improved synthetic wind speed generation using modified Mycielski approach," *International Journal of Energy Research*, **36**(13), 1226-1237.

Fidan, M., Hocaoğlu, F. O. and Gerek, Ö. N. (2014) "Harmonic analysis based hourly solar radiation forecasting model." IET Renewable Power Generation.

Forney Jr, G. (1966), "Generalized minimum distance decoding," *Information Theory, IEEE Transactions on*, **12**(2), 125-131.

Fukuda, S. and Hirosawa, H. (1999), "A wavelet-based texture feature set applied to classification of multifrequency polarimetric SAR images," *IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING*, **37**(5), 2282-2286.

Gales, M. J. F. (1998), "Maximum likelihood linear transformations for HMM-based speech recognition," *Computer Speech and Language*, **12**(2), 75-98.

Genç, A., Kinaci, İ., Oturanç, G., Kurnaz, A., Bilir, Ş. and Özbalta, N. (2002), "Statistical Analysis of Solar Radiation Data Using Cubic Spline Functions," *Energy Sources*, **24**(12), 1131-1138.

Gerek, O. N. and Ece, D. G. (2004), "2-D analysis and compression of power-quality event data," *IEEE Transactions on Power Delivery*, **19**(2), 791-798.

Gergonne, J. D. (1974), "The application of the method of least squares to the interpolation of sequences," *Historia Mathematica*, **1**(4), 439-447.

Germen, E., Basaran, M. and Fidan, M. (2014), "Sound based induction motor fault diagnosis using Kohonen self-organizing map," *Mechanical Systems and Signal Processing*, **46**(1), 45-58.

Guo, P. H., Li, J. Y. and Wang, Y. (2014), "Numerical simulations of solar chimney power plant with radiation model," *Renewable Energy*, **62**, 24-30.

Hayes, B. and Djokic, S. (2013), "Advanced Markovian Wind Energy Models for Smart Grid Applications," *Proceedings of the 4th IEEE PES*

Innovative Smart Grid Technologies Europe (ISGT Europe), Copenhagen.

Hayes, M. H. (1996), "*Statistical digital signal processing and modeling*," John Wiley & Sons.

Heinz, S. (2011), "*Mathematical Modeling*," Springer.

Hocaoglu, F., Fidan, M. and Gerek, O. (2009), "Mycielski approach for synthetic wind speed data generation," *Proceedings of the Signal Processing and Communications Applications Conference, 2009. SIU 2009. IEEE 17th*, IEEE, 836-839.

Hocaoglu, F. O., Gerek, O. N. and Kurban, M. (2008), "The Effect of Markov Chain State Size for Synthetic Wind Speed Generation," *Proceedings of the The 10th International Conference on Probabilistic Methods Applied to Power Systems (PMAPS'08)*, Rincon, Puerto Rico, 1-4.

Hocaoğlu, F. O. (2011), "Stochastic approach for daily solar radiation modeling," *Solar Energy*, **85**(2), 278-287.

Hocaoğlu, F. O., Fidan, M. and Gerek, Ö. N. (2009a), "Mycielski approach for wind speed prediction," *Energy Conversion and Management*, **50**(6), 1436-1443.

Hocaoğlu, F. O., Fidan, M. and Gerek, Ö. N. (2009b), "A Novel Fourier Based Solar Radiation Model," *ICIC Express Letters*, **3**(4), 1101-1106.

Hocaoğlu, F. O., Gerek, Ö. N. and Kurban, M. (2008), "Hourly solar radiation forecasting using optimal coefficient 2-D linear filters and feed-forward neural networks," *Solar Energy*, **82**(8), 714-726.

Hocaoğlu, F. O., Gerek, Ö. N. and Kurban, M. (2010), "A novel wind speed modeling approach using atmospheric pressure observations and hidden Markov models," *Journal of Wind Engineering and Industrial Aerodynamics*, **98**(8), 472-481.

Hosseini, R., Le, N. and Zidek, J. (2012). Time-varying Markov models for dichotomized temperature series., THE UNIVERSITY OF BRITISH COLUMBIA, DEPARTMENT OF STATISTICS.

Hsieh, C. T., Lai, E. and Wang, Y. C. (2003), "An effective algorithm for fingerprint image enhancement based on wavelet transform," *Pattern Recognition*, **36**(2), 303-312.

Huang, N. E., Shen, Z., Long, S. R., Wu, M. L. C., Shih, H. H., Zheng, Q. N., Yen, N. C., Tung, C. C. and Liu, H. H. (1998), "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings of the Royal Society a-Mathematical Physical and Engineering Sciences*, **454**(1971), 903-995.

Huang, N. E., Wu, M. L. C., Long, S. R., Shen, S. S. P., Qu, W. D., Gloersen, P. and Fan, K. L. (2003), "A confidence limit for the empirical mode decomposition and Hilbert spectral analysis," *Proceedings of the Royal Society a-Mathematical Physical and Engineering Sciences*, **459**(2037), 2317-2345.

Hyeon-Kyu, L. and Kim, J. H. (1999), "An HMM-based threshold model approach for gesture recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **21**(10), 961-973.

Islam, M. R., Rashed-Al-Mahfuz, M., Ahmad, S. and Molla, M. K. I. (2012), "Multiband Prediction Model for Financial Time Series with Multivariate Empirical Mode Decomposition," *Discrete Dynamics in Nature and Society*, **2012**, 1-21.

Jacquet, P., Szpankowski, W. and Apostol, I. (2002), "A universal predictor based on pattern matching," *Ieee Transactions on Information Theory*, **48**(6), 1462-1472.

Jiang, X. S., Zhang, L. and Chen, X. Q. (2014), "Short-term forecasting of high-speed rail demand: A hybrid approach combining ensemble empirical mode decomposition and gray support vector machine with real-world applications in China," *Transportation Research Part C-Emerging Technologies*, **44**(0), 110-127.

Jianying, H., Brown, M. K. and Turin, W. (1996), "HMM based online handwriting recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **18**(10), 1039-1045.

Judd, K. L. (1998), "*Numerical Methods in Economics*," MIT Press.

Kall, L., Krogh, A. and Sonnhammer, E. L. (2005), "An HMM posterior decoder for sequence feature prediction that includes homology information," *Bioinformatics*, **21 Suppl 1**(suppl 1), i251-257.

Kaplani, E. and Kaplanis, S. (2012), "A stochastic simulation model for reliable PV system sizing providing for solar radiation fluctuations," *Applied Energy*, **97**, 970-981.

Kaplanis, S. and Kaplani, E. (2007), "A model to predict expected mean and stochastic hourly global solar radiation< i> I</i>(< i> h</i>;< i> nj</i>) values," *Renewable Energy*, **32**(8), 1414-1425.

Kaplanis, S. N. (2006), "New methodologies to estimate the hourly global solar radiation; Comparisons with existing models," *Renewable Energy*, **31**(6), 781-790.

Khare, A. and Rangnekar, S. (2014), "Optimal sizing of a grid integrated solar photovoltaic system," *IET Renewable Power Generation*, **8**(1), 67-75.

Laine, A. and Fan, J. (1993), "Texture Classification by Wavelet Packet Signatures," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **15**(11), 1186-1191.

Lei, Y. G., Lin, J., He, Z. J. and Zuo, M. J. (2013), "A review on empirical mode decomposition in fault diagnosis of rotating machinery," *Mechanical Systems and Signal Processing*, **35**(1-2), 108-126.

Li, Q. (2012), "Non-Stationary Pattern Recognition," *Speaker Authentication*, Springer Berlin Heidelberg**,** 61-73.

Lopes, R. and Betrouni, N. (2009), "Fractal and multifractal analysis: a review," *Med Image Anal*, **13**(4), 634-649.

Lu, W. B., Jiang, W. K., Wu, H. J. and Hou, J. J. (2012), "A fault diagnosis scheme of rolling element bearing based on near-field acoustic holography and gray level co-occurrence matrix," *Journal of Sound and Vibration*, **331**(15), 3663-3674.

Maafi, A. and Adane, A. (1989), "A two-state Markovian model of global irradiation suitable for photovoltaic conversion," *Solar & wind technology*, **6**(3), 247-252.

Makhoul, J. (1975), "Linear prediction: A tutorial review," *Proceedings of the IEEE*, **63**(4), 561-580.

Maragos, P., Schafer, R. W. and Mersereau, R. M. (1984), "Two-dimensional linear prediction and its application to adaptive predictive coding of images," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, **32**(6), 1213-1229.

Marzetta, T. (1980), "Two-dimensional linear prediction: Autocorrelation arrays, minimum-phase prediction error filters, and reflection coefficient arrays," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, **28**(6), 725-733.

Mellit, A., Benghanem, M., Arab, A. H. and Guessoum, A. (2005), "A simplified model for generating sequences of global solar radiation data for isolated sites: Using artificial neural network and a library of Markov transition matrices approach," *Solar Energy*, **79**(5), 469-482.

Mellit, A., Benghanem, M. and Kalogirou, S. A. (2006), "An adaptive wavelet-network model for forecasting daily total solar-radiation," *Applied Energy*, **83**(7), 705-722.

Moghaddam, A. A. and Seifi, A. R. (2011), "Study of forecasting renewable energies in smart grids using linear predictive filters and neural networks," *IET Renewable Power Generation*, **5**(6), 470-480.

Moser, B. K. (1996a), "5 - Least-Squares Regression," *Linear Models*, B. K. Moser, San Diego, Academic Press**,** 81-103.

Moser, B. K. (1996b), "6 - Maximum Likelihood Estimation and Related Topics," *Linear Models*, B. K. Moser, San Diego, Academic Press**,** 105-129.

Nandi, S., Toliyat, H. A. and Li, X. D. (2005), "Condition monitoring and fault diagnosis of electrical motors - A review," *IEEE TRANSACTIONS ON ENERGY CONVERSION*, **20**(4), 719-729.

Nunes, J., Guyot, S. and Delechelle, E. (2005), "Texture analysis based on local analysis of the bidimensional empirical mode decomposition," *Machine Vision and Applications*, **16**(3), 177-188.

Ocak, H. and Loparo, K. A. (2004), "Estimation of the running speed and bearing defect frequencies of an induction motor from vibration data," *Mechanical Systems and Signal Processing*, **18**(3), 515-533.

Pal, M., Madhusudana Rao, P. and Manimaran, P. (2014), "Multifractal detrended cross-correlation analysis on gold, crude oil and foreign exchange rate time series," *Physica A: Statistical Mechanics and its Applications*, **416**(0), 452-460.

Parsons, T. W. (1987), "*Voice and speech processing*," McGraw-Hill.

Pieczynski, W. (2010), "EM and ICE in hidden and triplet Markov models," *Proceedings of the Stochastic Modeling Techniques and Data Analysis international conference (SMTDA'10), Chania, Greece*.

Pinsky, M. and Karlin, S. (2010), "*An introduction to stochastic modeling*," Academic press.

Rabiner, L. R. and Juang, B. H. (1986), "An introduction to hidden markov models," *IEEE ASSP MAGAZINE*.

Rahoma, U. A. and Hassan, A. H. (2007), "Fourier transforms Investigation of Global Solar Radiation at True Noon: in the Desert climatology," *American Journal of Applied Sciences*, **4**(11), 902-907.

Rodgers, J. L. and Nicewander, W. A. (1988), "Thirteen Ways to Look at the Correlation Coefficient," *The American Statistician*, **42**(1), 59-66.

Russell, M., Brown, C., Skilling, A., Series, R., Wallace, J., Bonham, B. and Barker, P. (1996), "Applications of automatic speech recognition to speech and language development in young children," *Proceedings of the Spoken Language, 1996. ICSLP 96. Proceedings., Fourth International Conference on*, 176-179 vol.171.

Sahin, A. D. and Sen, Z. (2001), "First-order Markov chain approach to wind speed modelling," *Journal of Wind Engineering and Industrial Aerodynamics*, **89**(3-4), 263-269.

Schlapbach, A. and Bunke, H. (2004), "Off-line handwriting identification using HMM based recognizers," *Proceedings of the Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, 654-658 Vol.652.

Sebe, N. and Lew, M. S. (2000), "Wavelet based texture classification," *Proceedings of the Pattern Recognition, 2000. Proceedings. 15th International Conference on*, IEEE, 947-950.

Semler, L., Dettori, L. and Furst, J. (2005), "Wavelet-based texture classification of tissues in computed tomography," *Proceedings of the Computer-Based Medical Systems, 2005. Proceedings. 18th IEEE Symposium on*, IEEE, 265-270.

Shah, S. P., Xuan, X., DeLeeuw, R. J., Khojasteh, M., Lam, W. L., Ng, R. and Murphy, K. P. (2006), "Integrating copy number polymorphisms into array CGH analysis using a robust HMM," *Bioinformatics*, **22**(14), e431-439.

Shamshad, A., Bawadi, M. A., Hussin, W. M. A. W., Majid, T. A. and Sanusi, S. A. M. (2005), "First and second order Markov chain models for synthetic generation of wind speed time series," *Energy*, **30**(5), 693-708.

Shannon, W. D. (2007), "11 Cluster Analysis," *Handbook of Statistics*, J. P. M. C.R. Rao and D. C. Rao, Elsevier, **Volume 27,** 342-366.

Takiguchi, T., Nakamura, S. and Shikano, K. (2001), "HMM-separation-based speech recognition for a distant moving speaker," *IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING*, **9**(2), 127-140.

Torre, M., Poggi, P. and Louche, A. (2001), "Markovian model for studying wind speed time series in Corsica," *International Journal of Renewable Energy Engineering*, **3**(2), 311-319.

Ulgen, K. and Hepbasli, A. (2002), "Determination of Weibull parameters for wind energy analysis of Izmir, Turkey," *International Journal of Energy Research*, **26**(6), 495-506.

Unser, M. (1995), "Texture classification and segmentation using wavelet frames," *IEEE Trans Image Process*, **4**(11), 1549-1560.

Walker, G. (1931), "ON PERIODICITY IN SERIES OF RELATED TERMS," *Monthly Weather Review*, **59**(7), 277-278.

Weruaga, L. (2007), "All-Pole Estimation in Spectral Domain," *Signal Processing, IEEE Transactions on*, **55**(10), 4821-4830.

WU, Z., HUANG, N. E. and CHEN, X. (2009), "THE MULTI-DIMENSIONAL ENSEMBLE EMPIRICAL MODE DECOMPOSITION METHOD," *Advances in Adaptive Data Analysis*, **1**(3), 339-372.

Xiyun, Y., Feifei, J. and Huan, L. (2013), "Short-term solar radiation prediction based on SVM with similar data," *Proceedings of the Renewable Power Generation Conference (RPG 2013), 2nd IET*, 1-4.

Xu, H., Wei, C. L., Lin, F. and Sung, W. K. (2008), "An HMM approach to genome-wide identification of differential histone modification sites from ChIP-seq data," *Bioinformatics*, **24**(20), 2344-2349.

Yang, H.-D., Park, A. Y. and Lee, S.-W. (2007), "Gesture Spotting and Recognition for Human-Robot Interaction," *IEEE Transactions on Robotics*, **23**(2), 256-270.

Younus, A. M. D. and Yang, B. S. (2012), "Intelligent fault diagnosis of rotating machinery using infrared thermal image," *Expert Systems with Applications*, **39**(2), 2082-2091.

Yule, G. U. (1927), "On a method of investigating periodicities in disturbed series, with special reference to Wolfer's sunspot numbers," *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 267-298.

Zarzo, M. and Marti, P. (2011), "Modeling the variability of solar radiation data among weather stations by means of principal components analysis," *Applied Energy*, **88**(8), 2775-2784.