

İSTANBUL TECHNICAL UNIVERSITY ★ INSTITUTE OF SCIENCE AND TECHNOLOGY

REAL TIME EARTHQUAKE RISK COMPUTATION

**M.Sc Thesis by
Serdar BAGIS**

Department: Computer Engineering

Programme: Computer Engineering

Thesis Supervisor: Asst. Prof. Dr. Burak Berk ÜSTÜNDAĞ

SEPTEMBER 2009

REAL TIME EARTHQUAKE RISK COMPUTATION

**M.Sc Thesis by
Serdar BAGIS**

Date of submission : 7 September 2009

Date of defence examination: 7 September 2009

Supervisor (Chairman) : Asst. Prof. Dr. Burak Berk ÜSTÜNDAĞ (İ.T.Ü.)

Members of the Examining Committee: Prof.Dr. Haluk EYİDOĞAN (İ.T.Ü.)

Asst. Prof. Dr.Mustafa KAMAŞAK (İ.T.Ü.)

SEPTEMBER 2009

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

GERCEK ZAMAN DEPREM RİSKİNİ HESAPLAM

YÜKSEK LİSANS TEZİ

Serdar BAGIS

Tezin Enstitüye Verildiği Tarih : 7 Eylül 2009

Tezin Savunulduğu Tarih : 7 Eylül 2009

Tez Danışmanı : Yrd. Doç.Dr. Burak Berk ÜSTÜNDAĞ (İ.T.Ü.)

Diğer Jüri Üyeleri : Prof.Dr. Haluk EYİDOĞAN (İ.T.Ü.)

Yrd. Doç. Dr.Mustafa KAMAŞAK (İ.T.Ü.)

EYLÜL 2009

FOREWORD

I would like to express my deep appreciation and thanks to my family that supported me on all matters. Also to my advisor who helped me understand and achieve the aim of the project.

September 2009

Serdar BAGIS
Computer and Mathematics Science

TABLE OF CONTENTS

	<u>Page</u>
ABBREVIATIONS	IX
LIST OF TABLES	XI
LIST OF FIGURES	XIII
LIST OF SYMBOLS	XV
SUMMARY	XVII
ÖZET	XIX
1. INTRODUCTION	1
1.1 Purpose of the Thesis.....	1
1.2 Background.....	2
2. EARTHQUAKE PREDICTION USING MONOPOLAR ELECTRIC FIELD PROBE	5
2.1 Introduction.....	5
2.2 Monopolar Electric Field Measurement	6
2.3 Possible Effect of Dilatancy to Permittivity and Precursory Surface Electric Field Variation	9
2.4 Network of Stations.....	12
3 REAL TIME, TIME SERIES FORECAST	15
3.1 Time Series	15
3.2 Forecasting Process	17
3.3 Time Series Forecasting	18
4. ANALYSIS OF COLLECTED DATA	21
4.1 Frequency Domain Approach To Time Series.....	21
4.1.1 Fourier transform.....	21
4.1.2 Discrete Fourier transform (DFT) and Fast Fourier transform (FFT)	23
4.2 Spectrum Analysis of Collected Data.....	24
4.2.1 Spectrum analysis.....	24
4.2.2 Recorded data.....	25
4.2.3 Periodogram analysis of data	26
4.2.4 Discussion on found periodic components	31
5. INVESTIGATION OF EARTHQUAKE PRECURSORY EVENTS, IN FREQUENCY DOMAIN	35
5.1 Data Description.....	35
5.2 Earthquakes and the Recorded Anomalies.....	37
5.3 Earthquake Precursory Anomaly Detection	40

5.3.1 Mathematical model.....	41
5.3.2 Results validation	42
6. CONCLUSION	45
REFERENCES	47

ABBREVIATIONS

GPS	: Global Positioning System
C_c	: spherical capacitor
MOS	: Metal Oxide Semiconductor
A/D converter	: analog-to-digital converter
NAF	: Northern Anatolian Fault
ARMA	: AutoRegressive Moving Average
MARS	: Multivariate Adaptive Regression Spline
NARMAX	: Nonlinear AutoRegressive Moving Average with eXogenous
DFT	: Discrete Fourier Transform
FFT	: Fast Fourier Transform
UTC	: Coordinated Universal Time
STFT	: Short Time Fourier Transform

LIST OF TABLES

	<u>Page</u>
Table 4.1 : Primary Tidal Components.....	32
Table 5.1 : A sample from earthquake list	37
Table 5.2 : Results.....	43

LIST OF FIGURES

	<u>Page</u>
Figure 2.1 : Electric field variation, with respect to distance ($r - r_1$) outside a charged sphere.....	6
Figure 2.2 : Monopolar electric charge measurement probe	8
Figure 2.3 : The elements of earthquake occurrence model having dielectric and mechanical parameters including liquid dilatancy.....	10
Figure 2.4 : Equivalent circuit model of dilatancy and piezoelectricity of a multilayer system.....	11
Figure 2.5 : Station locations in north-western Anatolia.....	13
Figure 3.1 : Forecast system architecture.....	17
Figure 4.1 : Data recorded at three different stations over a period of 7 days.....	26
Figure 4.2 : Data recorded at 3 stations over a period of 2200 days.....	27
Figure 4.3 : Periodogram on data recorded at 3 stations over a period of 2200 days.....	27
Figure 4.4 : Amplitude (upper picture) and the percent of the spectral energy (lower picture) of the frequency having a period of one year (366 days) on data recorded at 3 stations over a period of 2200 days.....	28
Figure 4.5 : Amplitude (upper picture) and the percent of the spectral energy (lower picture) of the frequency having a period of one day on data recorded at 3 stations over a period of 2200 days.....	29
Figure 4.6 : Amplitude (upper picture) and the percent of the spectral energy (lower picture) of the 12 hour period on data recorded at 3 stations over a period of 2200 days.....	30
Figure 4.7 : Amplitude (upper picture) and the percent of the spectral energy (lower picture) of the 8.5 days period corresponding to Sun's adiabatic expansion on data recorded at 3 stations over a period of 2200 days.....	30

Figure 4.8	: Amplitude (upper picture) and the percent of the spectral energy(lower picture) of 27.5 days period corresponding to Sun's adiabatic expansion on data recorded at 3 stations over a period of 2200 days.....	31
Figure 4.9	Fitting a signal having two harmonics first with a period of 1 year and the second with a period of 1 day to 2200 days data from Çanakkale Station.....	33
Figure 5.1	: 2200 days data filtered with a low-pass filter.	36
Figure 5.2	: Seven-days graph of Tekirdag and Beylikdüzü stations with M5.2 earthquake in Saros and its aftershock in June 2004....	38
Figure 5.3	: Graph of Canakkale station with M5.2 earthquake Mediterranean Sea.....	38
Figure 5.4	: One year data as mean of a window of 24 hours and sliding offset of 3 hours between 1.01.2007 – 31.12.2007.....	39
Figure 5.5	: 100 days data as mean of a window of 24 hours and sliding offset of 3 hours between 1.01.2007 – 22.09.2007 together with earthquakes of magnitude > 5 degree.....	39
Figure 5.6	: 100 day data as mean of a window of 24 hours and sliding offset of 3 hours between 1.01.2007 – 22.09.2007 together with earthquakes of magnitude > 3.5 degree.....	40
Figure 5.7:	: Prediction Scheme.....	42

LIST OF SYMBOLS

E	: Electric field strength
D	: electrical displacement vector
S	: surface
ϵ	: permittivity of the relevant medium
ϵ_a	: dielectric coefficient (permittivity) of the air
ϵ_1	: dielectric coefficient of the sedimentary layer
ϵ_2, ϵ_3	: dielectric coefficients of upper crustal granitic layers
ϵ_s	: dielectric coefficient of the leaking fluid
r	: radius of the spherical probe
s	: derivative equivalent operand of the Laplace transform
R_s	: equivalent reservoir output resistance of the leakages
C_s	: capacitor equivalent to simulate material change due to leakage charge-discharge phenomena
S_s	: switch is used in equivalent electrical circuit to simulate start of leakage from a reservoir
C_4	: couples the circuit model to the lower layers of the crust
U_p	: local stress dependent equivalent voltage source
qe	: equivalent charge flow supplied by coupled deeper layers
τ_f	: shear force
σ	: normal stress factor
a_n	: thickness of the nth layer
p	: pressure
d	: anisotropic mineral ratio
l	: average fault gouge thickness
ξ	: frequency

REAL TIME EARTHQUAKE RISK COMPUTATION

SUMMARY

Earthquakes are one of the most important natural disasters. They cause a lot of damage and because of them people get hurt or even killed. The need of some methods that are able to predict them was always a concern. The aim of this thesis is to analyze the data gathered at the Earthquake Prediction Project, try to model the recorded anomalies and to compute in real time the risk index of happening of earthquakes. The results show that the collected data are consistent with the happening of earthquakes and that the recorded anomalies precede major earthquakes. A system that makes real time computations of the earthquake risk was designed based on the fact that the static electrical energy at the surface of the earth changes little in short time, and thus any change could be triggered by the fact that an earthquake is ready to happen.

GERÇEK ZAMAN DEPREM RİSKİNİ HESAPLAMA

ÖZET

Deprem en önemli doğal afetlerden biridir. Birçok hasara neden olur ve depremlerin yüzünden insanlar yaralanır, hatta ölürlür. Depremleri tahmin etme yöntemlerine duyulan ihtiyaç her zaman ilgi konusu olmuştur. Bu tezin amacı, Deprem Tahmin Projesindeki toplanan verileri analiz etmek, kaydedilen anomalileri modellemeye çalışmak ve deprem olma risk indisini gerçek zamanlı hesaplamaktır. Sonuçlar gösteriyor ki, toplanan verilerdeki anomaliler ve kaydedilen depremler ilişkilidir. Dünya yüzeyindeki statik elektrik enerjinin değişimi kısa zaman içerisinde değişimi az olduğunu bildiğimiz için, herhangi bir ani değişim büyük bir depremin yolda olduğunu gösterir. Bunun için deprem riskini gerçek zamanda hesaplayan bir sistem hazırlanmıştır.

1. INTRODUCTION

Prediction of an earthquake refers to estimating the occurrence of an earthquake of a specific magnitude that will occur in a particular place at a particular time.

Despite considerable research efforts by seismologists, scientifically reproducible predictions cannot yet be made to a specific hour, day, or month but for well-understood faults, seismic hazard assessment maps can estimate the probability that an earthquake of a given size will affect a given location over a certain number of years.

Before an earthquake begins, early warning systems can provide a few seconds' or even a few hours before major shaking arrives at a given location. Technology takes advantage of the different speeds of propagation of the various types of vibrations produced and of different types of earthquake precursory events such as radon gas emission, gravity variation, static electric field variation, etc. Aftershocks are also likely after a major earthquake, and are commonly planned for, in earthquake disaster response protocols.

Experts do advise general earthquake preparedness, especially in areas known to experience frequent or large quakes, to prevent injury, death, and property damage if a quake occurs with or without warning. So an earthquake prediction system is a necessity in such areas.

1.1 Purpose of the Thesis

The aim of this project is to analyze data collected from the “*Earthquake Forecast by the method of regional stress measurement*” project conducted by Ass. Dr. Berk Ustundag at Istanbul Technical University and evaluate in realtime the risk of happening of an earthquake higher than a given magnitude. The data are measurements made at different stations around Marmara sea consisting of variations of electric charge in the air close to the surface of the earth. The recording area is a highly active seismic zone.

The data recordings show noticeable anomalies correlated with the occurrence of earthquakes within the surrounding area but not only. The purpose of the thesis is to exploit this fact and compute the risk of happening of an earthquake. The technique adopted is spectral analysis, using Fourier transform, which has been and still is one of the most important and most widely used tools in earthquake engineering for measuring the energy in signals.

1.2 Background

Earthquakes are a serious natural hazard in many parts of the world. They often seem to strike without warning, sometimes inflicting massive damage and casualties. Often they occur in predictable areas such as at plate margins where earthquakes have occurred in the past, so why can't scientists give adequate warning before an earthquake occurs?

Despite years of research, the use of advanced seismic recording equipment, and the efforts of many dedicated scientists, the task of predicting earthquakes is still, largely, a matter of informed guess work.

Before earthquakes happen, different types of precursory events start to occur and thus different techniques to try to predict the future earthquakes. Among these precursory events there are phenomena such as the piezoelectric effect, geomagnetic storm energy, solar storm activity, earthquake fault zone activity related electromagnetic energy fields, geomagnetic storms, phenomena such as lightning strikes, electromagnetic energy field fluctuations, gravitational pulls of the sun and the moon and also forces or phenomena related to geomagnetic storms etc.

Do we really want accurate predictions? The obvious answer may seem to be 'Yes, we do'. Sufficient warning would allow for the safe evacuation of the population, gas and water supplies could be cut before pipes were ruptured, and emergency services need never be caught by surprise again

Rigorous earthquake forecasts is challenging for several reasons:

- Earthquakes cluster strongly, and one event can change radically the probabilities of following ones.

- Overlapping shaking from multiple near-simultaneous earthquakes makes it impossible to identify individual events characterizing each .
- Earthquake forecasts come in many forms, including contour maps of earthquake probability, fault based forecasts, and forecasts that are only activated when prescribed geophysical anomalies are detected. Many forecasts prescribe probabilities of a finite set of scenario earthquakes without clear rules on how to associate future earthquakes to events in the set.

2. EARTHQUAKE PREDICTION USING MONOPOLAR ELECTRIC FIELD PROBE

2.1 Introduction

There have been many research activities on prediction of approximate time and the epicenters of probable earthquakes. Some of these researches depend on evaluation of physical changes on earth surface. Widely worked parameters are change in soil resistivity, spectral analysis of transmitted electromagnetic signals inside the ground, electric potential change between probe points, change in radon emission rate, temperature change in thermal spring waters, water level change inside the wells, measurement of long term movement of terrestrial points by GPS, temporal and spatial changes in background seismicity, observation of behavior of the biological beings and measurement of acoustic emission [1].

Anisotropic minerals, mainly the quartz highly common in nature and they have the property of piezoelectric effect. It is known that electric potential variation over a standard cubic test sample is directly proportional to the stress change ($d\delta/dt$) under time varying mechanical load. There had been some researches on earthquake prediction including VAN Method depending on this principle. Electrical potential is being measured by means of voltage difference between the electrodes inserted inside the soil at different locations. Electrical voltage does not only depend on regional stress change but also varies by changes in soil properties such as humidity, chemical reactions, impedance, and natural battery effects. This is a disadvantage of differential electric potential measurement inside the soil [1].

In this study, quasistatic waves and change in ultra low frequency component of electric fields are evaluated by using a specially designed monopolar probe. Variations of electric charge in the air close to the surface of the earth is measured and stored for evaluation of short and mid term earthquake prediction.

2.2 Monopolar Electric Field Measurement

In the proposed method, one part of the sensor mechanism is the Earth that is coupled to the monopolar electrode through air. Maximum electric field strength occurs at the surface of any sphere that is loaded by a voltage source (Figure 2.1). The electric field decreases inverse-square proportional to the distance from the surface of the source. This is also valid for the Earth as a globe since there are negative ions in the upper atmosphere. A high sensitive monopolar electric field probe has been developed and patented, which is to be installed close to the surface of the earth for this purpose.

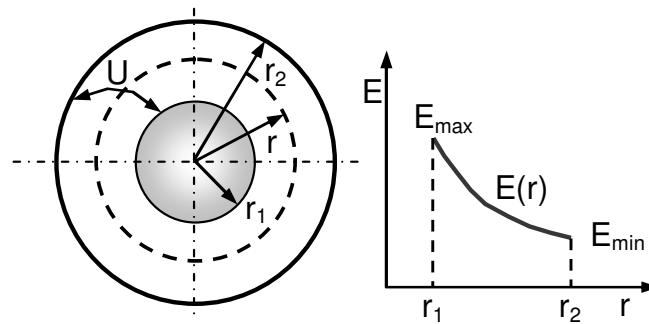


Figure 2.1 : Electric field variation, with respect to distance $(r - r_1)$ outside a charged sphere

A high sensitive monopolar electric charge probe has been developed which will be installed close to the surface of the earth for this reason. We assume that change in charge induction at the probe should be related to the change in regional resultant stress as an electric potential source. Although there also exist atmospherical electric field changes as noise, which can be filtered since the frequency range is much higher than what is assumed to be proportional to the change in regional stress. On the other hand superposition of the attenuated electric field changes from further regions should be considered. These two facts require vectoral measurement using a group of stations and adaptive filtering for the removal of the atmospheric noises.

Charge induction at the probe should possibly be related to a) the change in regional resultant stress, b) local electrochemical processes as electric potential sources, c) structural changes such as fractures and leakages and d) charge coupling of deeper Earth components. Although there also exists atmospherically electric field changes as noise, they can be filtered since the frequency range and signal patterns are different from those of the seismic components. On the other hand, superposition of

the attenuated electric field changes from farther regions should be considered. These two facts require vectoral measurement using a group of stations and adaptive filtering for the removal of the atmospheric noises.

The monopolar probe based measurement system consists of i. *a spherical capacitor* (C_c) on the probe as electric charge collector, ii. *MOS (Metal Oxide Semiconductor) circuit* as charge/bipolar voltage converter inside the probe, iii. *indicator device* for amplification, analog to digital conversion, signal processing and iv. *a server* for telemetric data acquisition and central data processing. Electrode of the first set of stationary probes were spherical and their diameter was just 40 mm. Collected charge is conducted to the charge/voltage converter via a high voltage cable, that is placed inside a dielectric pot as shown in Figure 2.2. The measurement sensitivity reaches down to 10^{-14} Coulomb level. U_{ext} is the stable reference voltage input of the probe. Cable denoted as U_{out} carries the output voltage of the sensor mechanism. This line is connected to an A/D converter over an instrumentation amplifier. It should be considered that high input impedance rate of the interconnection is important to preserve sensitivity of the MOS type charge/voltage converter. An external button S_b is used to provide offset charge via reference source. It is sometimes needed to speed up the first connection since output stabilizes in several hours due to high isolation rate of the pot and the equivalent internal capacity.

Following Gauss Law amount of induced charge Q on the electrode may be written as

$$Q = \oint_S \vec{D} \cdot d\vec{s} \quad (2.1)$$

where $D(C/m^2)$ is the electrical displacement vector and S is the total area of the Gaussian surface. Since the relation between the electric field strength, E (V/m) and electrical displacement, D is,

$$D = \epsilon E \quad (2.2)$$

where ϵ is the permittivity of the relevant medium.

The amount of charge collected by the probe's sphere is,

$$Q = 4\pi r^2 \epsilon E \quad (2.3)$$

where r is the radius of the spherical probe. Dynamic behavior of the charge/bipolar voltage converter is in terms of volt and can be written in the form:

$$U_{\text{out}}(t) = k_1 [dE(t)/dt] + k_2 \quad (2.4)$$

where $k_1 = 0.05$ and $k_2 = 0.186$ are constants. These constants are extracted to fit dynamic response to the applied known electric field change in isolated laboratory conditions. They

also verify the electrical circuit solution when material and MOS component characteristic parameters are used. Transfer function of the monopolar probe can be written as

$$T(s) = \frac{U_{\text{out}}(s)}{E(s)} = k_1 s + k_2 \quad (2.5)$$

where s denotes the derivative equivalent operand of the Laplace transform. This means steady state accuracy is 0.186 [counts/(V/m)]. It is clear that accuracy can be raised by enlarging the spherical electrode surface.

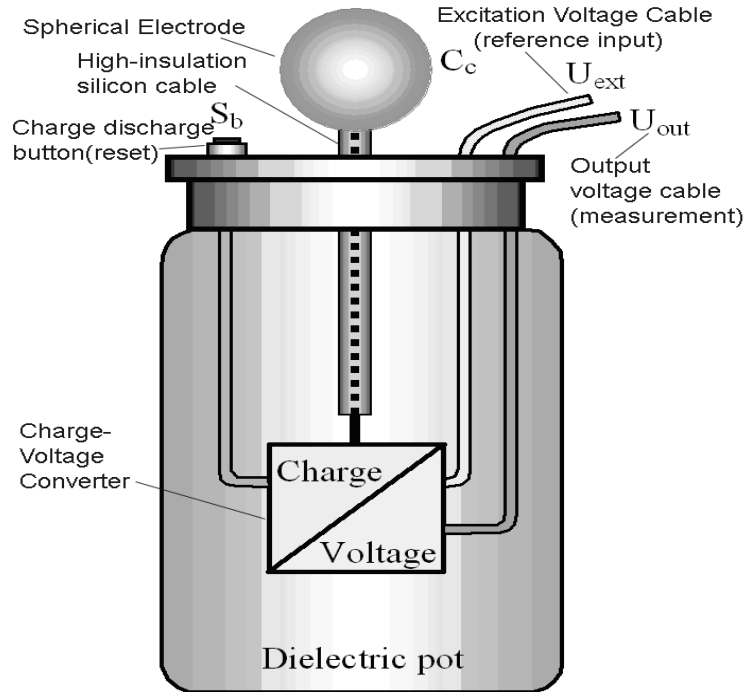


Figure 2.2 : Monopolar electric charge measurement probe.

2.3 Possible Effect of Dilatancy to Permittivity and Precursory Surface Electric Field Variation

Three different models of *liquid dilatancy*, *non-liquid dilatancy* and *long-term elasto-plastic instability* were proposed for the evaluation of real data of monopolar electric field measurements. These explanatory models are used for characteristic classification of the anomalies. On the other hand, variations and cooperative usage of these simplified models may take role in more realistic explanations.

Charged sphere approach for the Earth has been used here to explain behavior of the patterns measured by monopolar electric field probes. Earthquake process has precursory, instant as well as posterior effects on the fault structure. Since the occurrence time of the earthquake indicates a definite phase of this process, it provides experimental view to long term observations for structural analysis too. Let the depth of a probable earthquake be $d_{\text{hypocenter}}$. Since $d_{\text{hypocenter}}/r_{\text{earth}} \ll 1$, parallel plate equivalent circuit can be used for multilayer capacitor approach instead of spherical layers. This approach also gives the ability of adding regional parameters that can probably be used in seismo-tectonic analysis using the spatio-temporal data. The parameters in the models (Figures 2.3 and 2.4) are as follows,

ϵ_a is the dielectric coefficient (permittivity) of the air,

ϵ_1 is dielectric coefficient of the sedimentary layer,

ϵ_2 and ϵ_3 are dielectric coefficients of upper crustal granitic layers,

ϵ_s is dielectric coefficient of the leaking fluid

R_s represents equivalent reservoir output resistance of the leakages. It determines the time constant of liquid dilatancy,

C_s represents the capacitor equivalent to simulate material change due to leakage charge-discharge phenomena

S_s switch is used in equivalent electrical circuit to simulate start of leakage from a reservoir

C_4 couples the circuit model to the lower layers of the crust where piezoelectricity is negligible compared to the effects such as pyroelectricity,

U_p is the local stress dependent equivalent voltage source,

q_e is the equivalent charge flow supplied by coupled deeper layers

τ_f is the shear force.

The rectangular block in the model is exposed to the shear force τ_f that causes mechanical energy accumulation at either horizontal or the vertical directions. In earth sciences raised amount of the charged energy is generally expressed in terms of equivalent yearly displacement [centimeters]. σ is the normal stress that is a factor bonding the fault surface and it is also used in explanation of friction and instability models.

Equivalent electric circuit model of the upper crust which is used for explaining the effect of dilatancy on surface electric fields is shown in Figure 2.4. The proposed circuit model is a multilayer capacitor system having stress dependent voltage source. Upper crustal granitic layer is reduced into two different layers having pure piezoelectric material and non-piezoelectric material to simplify the simulations. Dielectric coefficient $\epsilon_{2,3}$ is replaced by ϵ_2 and ϵ_3 respectively. ϵ_3 represents the dielectric coefficient of the layer consisting of pure piezoelectric material.

Capacity of each layer in a n-layer capacitive system is,

$$C_1 = \frac{\epsilon_1 \cdot S}{a_1}, C_2 = \frac{\epsilon_2 \cdot S}{a_2} \dots C_n = \frac{\epsilon_n \cdot S}{a_n} \quad (2.6)$$

where S is the surface and a_n is the thickness of the n^{th} layer [Özkaya, 1996].

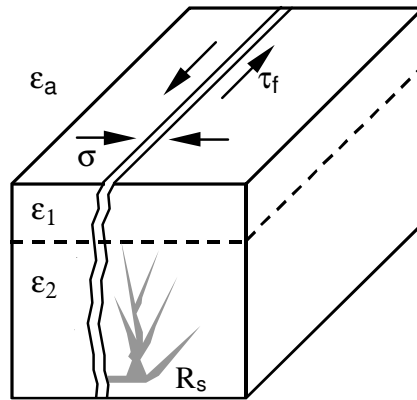


Figure 2.3. The elements of earthquake occurrence model having dielectric and mechanical parameters including liquid dilatancy.

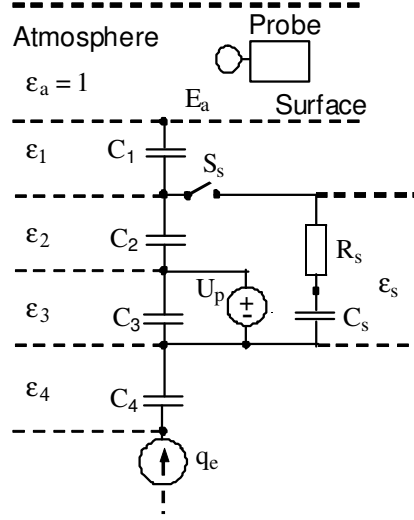


Figure 2.4 : Equivalent circuit model of dilatancy and piezoelectricity of a multi-layer system.

$$C = \frac{S}{\frac{a_1}{\epsilon_1} + \frac{a_2}{\epsilon_2} + \dots + \frac{a_n}{\epsilon_n}} \quad (2.7)$$

Since the charge of each layer is equivalent

$$Q = C \cdot U = C_1 U_1 = C_2 U_2 = \dots = C_n U_n \quad (2.8)$$

voltage drop over the layer can be expressed as,

$$U_k = \frac{a_k}{\epsilon_k} \frac{U}{A} \quad (k = 1, 2, \dots, n) \quad (2.9)$$

where

$$A = \sum_{i=1}^n \frac{a_i}{\epsilon_i} = \frac{a_1}{\epsilon_1} + \frac{a_2}{\epsilon_2} + \dots + \frac{a_n}{\epsilon_n} \quad (2.10)$$

and the electric field strength inside each layer is,

$$E_k = \frac{U}{\epsilon_k A} \quad (k = 1, 2, \dots, n) \quad (2.11)$$

which means that electric field strength is independent of the surface and varies with dielectric coefficient if we assume that the electric potential is constant.

$$E_a = E_3 \frac{\epsilon_3}{\epsilon_a} \quad (2.12)$$

stress dependent voltage source due to piezoelectricity can be expressed as,

$$U_p = 0.25 l p d \text{ [V]} \quad (2.13)$$

where p is σ oriented pressure [bar], d is the anisotropic mineral ratio and l is average fault gouge thickness (m). The factor 0.25 is valid only under the assumption that stress sensitivity of all piezoelectric minerals is the same as that of the quartz.

$$E_p = \frac{U_p}{l} = 0.25 p d \text{ [V/ m]} \quad (2.14)$$

Since pure piezoelectric portion is represented with a different capacitive layer, d ratio is 1 for C_4 . For example, if the change of pressure due to stress drop during the stress weakening and earthquake process is $P= 200$ bars, then the change in electric field strength will be

$$E_a = \frac{50 \cdot 5}{1} = 250 \text{ V/m} \quad (2.15)$$

without liquid dilatancy.

On the other hand, equation (16) shows that change in permittivity of the deeper layers affect the surface electric field as much as the stress dependent electric field variation. This can only be possible if there is a structural change or deformation.

In case of liquid dilatancy, change in E_a will also be a function of change in voltage U_p because of the new shunt capacitor representing the dielectric coefficient of the fluid. The dilatancy process begins with the switch S_s in the equivalent circuit and the time constant is determined by $\tau = R_s C_s$. The expected behavior of E_a will be

$$E_a = (\Delta E) e^{-(t/\tau)} \quad (2.16)$$

which is observed in many record examples before the earthquakes. Although the volume filled by the fluid inside the crack is relatively low, the change in E_a is still effective since $\epsilon_s = 81$ is much greater than $\epsilon_3 = 4$ and $\epsilon_4 = 5$.

2.4 Network of Stations

A network of fourteen on-line stations of monopolar electric field probes has been installed to detect earthquake precursory electrical anomalies around Northern Anatolian Fault (NAF) after the İzmit earthquake (1999). Electrodes of the probes were intended to measure electrical displacement between the upper lithosphere and the lower atmosphere. We propose a sort of passive electrostatic imaging in time to

detect beginning as well as rate of the structural change. In electrical engineering, “electrostatic imaging” requires a known electrical charge source and a coupling measurement probe. If the probe is mobile in space the induced electrical charges on the electrode will be a function of dielectric coefficient change of the material between the supply and the scan trajectory. The same feature holds for the lithosphere of the Earth since there is always equivalent internal charge distributed in layers. Non-stationary character of the background electric field is the main problem of passive electrostatic imaging. For this reason we collect long term data from fixed stations instead of mobile measurements. Hence we get the long term development of periodic components and compensate the patterns to be evaluated. Discontinuities and period deflections of the patterns have significant correlation to the seismic events as shown in the examples.

An internet based networking is used for data acquisition. Sixteen on-line stations, a test station in 1999, 4 stations in 2000, 8 stations in 2001, 12 stations in 2002 and 4 temporary stations, were installed in different locations around the North Anatolian Fault (NAF) since 1999. Fourteen of these stations are still active. Both the current and past data can be reached via internet site of the project: <http://www.deprem.cs.itu.edu.tr>. The station locations are shown in Figure 11 on the active fault map of North Western Anatolia.

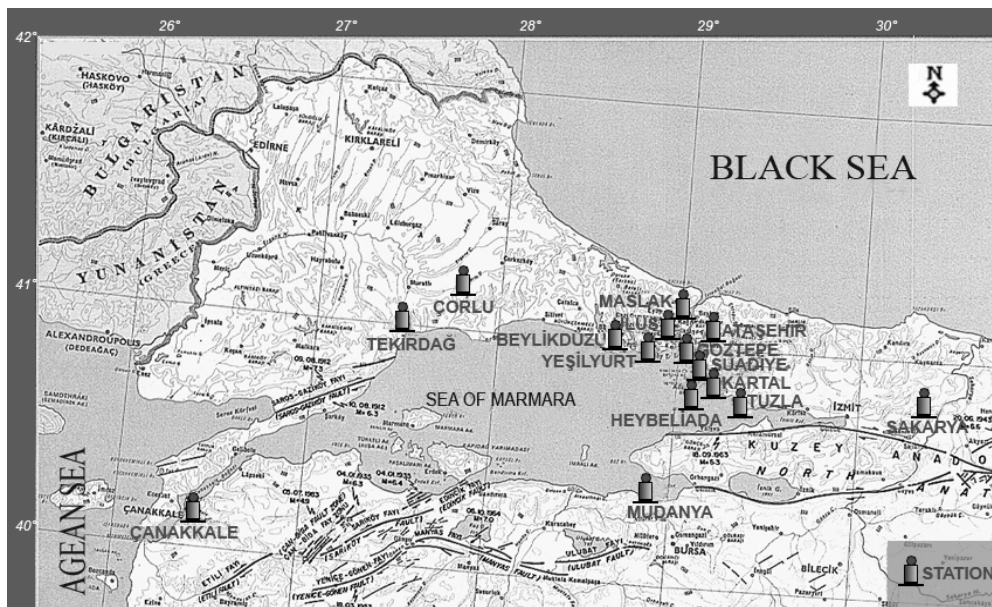


Figure 2.5 : Station locations in north-western Anatolia.

3 REAL TIME, TIME SERIES FORECAST

3.1 Time Series

A time series is sequence of observations to a variable recorded at equally spaced moments of time. The purpose of time series analysis is to determine some of the measured system's key properties by quantifying certain features of the time series. These properties can then help understand and predict the system's future behavior.

We merely notice that physical processes usually operate in continuous time. Most measurements, though, are done and recorded in discrete time. Thus the Earthquake Prediction Project's time series, as well as most climatic and other geophysical time series, are available in discrete time.

We concentrate here on time series in discrete time and consider therefore first the simple case of a scalar, linear ordinary difference equation with random forcing,

$$X_{t+1} = \sum_{j=1}^M a_j X_{t-M+j} + \sigma \zeta(t) \quad (3.1)$$

Its constant coefficients a_j determine the solutions $X(t)$ at discrete times $t = 0, 1, \dots, n, \dots$. In (Eq 3.1) the random forcing $\zeta(t)$ is assumed to be white in time, i.e., uncorrelated from t to $t + 1$, and Gaussian at each t , with constant variance equal to unity [4].

The investigation of time series becomes problematic due to seasonal fluctuations. Series are made up of four components.

The Seasonal Components of a series are often considered uninteresting and cause a series to be ambiguous.

The Trend Component (increasing or decreasing trend) and the *Cyclical Component* happen at the same magnitude during the same time period each year.

The Error, or irregular component, is the deviation of the sample from the (unobservable) *population* mean or actual function.

A *linear time series* is a time series whose graph lies on a straight line, and which can be described by giving its slope and its y intercept.

Seasonality is a characteristic of a time series in which the data experiences regular and predictable changes which recur every calendar year. Any predictable change or pattern in a time series that recurs or repeats over a one-year period can be said to be seasonal.

Stationarity is a common assumption in many time series techniques is that the data are stationary. A *stationary process* has the property that the mean, variance and autocorrelation structure do not change over time. Stationarity can be defined in precise mathematical terms, but we mean a flat looking series, without trend, constant variance over time, a constant autocorrelation structure over time and no periodic fluctuations

If the time series is not stationary, we can often transform it to stationarity with one of the following techniques.

1. We can difference the data. That is, given a time series, we create the new time series that is the difference of the data with itself.

The differenced data will contain one less point than the original data. Although you can difference the data more than once, one difference is usually sufficient.

2. If the data contain a trend, we can fit some type of curve to the data and then model the residuals from that fit. Since the purpose of the fit is to simply remove long term trend, a simple fit, such as a straight line, is typically used.
3. For non-constant variance, taking the logarithm or square root of the series may stabilize the variance. For negative data, you can add a suitable constant to make all the data positive before applying the transformation. This constant can then be subtracted from the model to obtain predicted (i.e., the fitted) values and forecasts for future points.

The above techniques are intended to generate series with constant location and scale. Although seasonality also violates stationarity, this is usually explicitly incorporated into the time series model.

3.2 Forecasting Process

A process is series of connected activities that transform one or more inputs into one or more outputs. All work activities are performed in processes and forecasting is no exception. In Figure 3.1 it is shown the way that the forecasting system works.

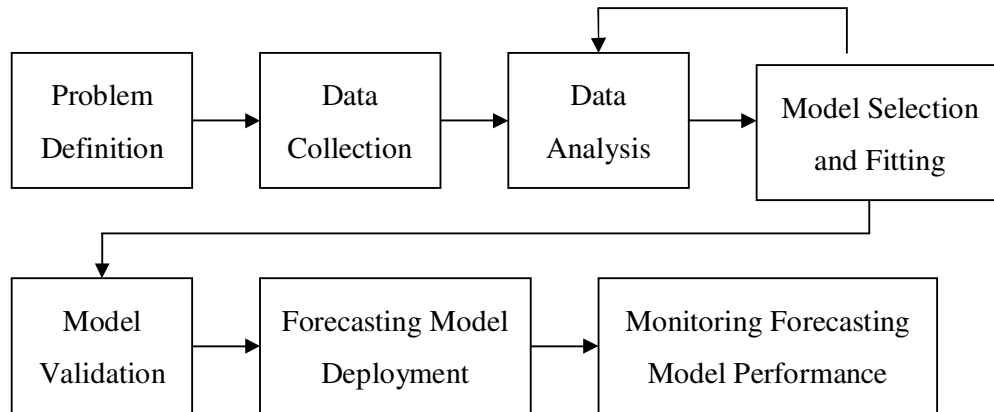


Figure 3.1: Forecast system architecture

The activities in the forecasting process are:

1. *Problem definition* involves developing understanding of how the forecast will be used. Often it is necessary to go deeply into the aspects of the monitored system that requires the forecast to properly define the forecasting component of the entire problem.
2. *Data collection* consists of obtaining the relevant history for the variables that are to be used for forecasting. Often it is needed to deal with missing values, potential outliers, or other data-related problems that have occurred in the past.
3. *Data analysis* is an important preliminary step to selection of the forecasting model to be used. Time series plots of the data often show trends and seasonal or other cyclical components. A trend is evolutionary movement, either upward or downward in the value of the variable. Trends can be long term or more dynamic and of relative short duration. Seasonality is the component of time series that repeats on a regular basis, such as year or month. Sometimes we shall smooth the data to make identification of summaries of the data, such as the sample mean standard deviation,

percentiles and autocorrelations. Unusual data points should be identified and flagged for further study. The purpose of this preliminary data analysis is to obtain some “feel” for the data. This information will usually suggest the initial types of quantitative forecasting methods and models to explore.

4. *Model selection and fitting* consists of choosing one or more forecasting models and fitting the model to the data, i.e., estimating a unknown model parameters, usually by the method of least squares.
5. *Model validation* consists of an evaluation of the forecasting model to determine how it is likely to perform in the intended application. This must go beyond just evaluating the “fit” of the model to the historical data and must examine what magnitude of forecast errors will be experienced when the model is used to forecast “fresh” or new data.
6. *Forecasting model deployment* involves getting the model and the resulting forecasts in use.
7. Monitoring forecasting model performance should be an ongoing activity after the model has been deployed to ensure that it is still performing satisfactorily. It is the nature of forecasting that the conditions change over time, and a model that performed well in the past may deteriorate in performance. Therefore monitoring of forecast errors is an essential part of good forecasting system design.

3.3 Time Series Forecasting

An intrinsic feature of a time series is that typically adjacent observations are dependent. The nature of this dependence among observation of a time series is of considerable practical interest. In time series prediction the goal is to construct a model that can reveal the underlying process and can predict the future of the measured process under interest. In order to construct a model for a process, data is gathered by measuring value of certain variables sequentially in time. Usually data is incomplete and includes noise. Stochastic and dynamic models for time series data have use in important application areas, such as forecasting of future value of a time series from current and past value, determination of the transfer function of a system subject to inertia, use of indicator input variable in transfer function models to

represent and assess the effect of unusual intervention events on the behavior of a time series, design of simple control schemes by means of which potential deviations of the system output from a desired target may, so far as possible be compensated by adjustment of the input series value [7].

Several computation techniques have been proposed to gain more insight into process and phenomena that include temporal information. Statistical methods based on linear (eg. AR and ARMA) and non-linear processes (eg. NARMAX and MARS) have been effectively used in many applications [8]. Recently neural networks have gained a lot of interest in time series prediction or to forecast certain outcomes by extrapolation due to their ability to learn effectively nonlinear dependencies from large volume of possibly noisy data with a learning algorithm. Many types of neural networks have been used in time series prediction. Traditional way of using neural networks in time series prediction is to convert the temporal sequence into concatenated vector via a tapped delay line, and to feed the resulting vector as an input to a network. This time delay neural network approach however has its well-known drawbacks, one of the most serious ones of being the difficulty to determine the proper length for the delay line [9].

4. ANALYSIS OF COLLECTED DATA

Two basic approaches to time series analysis are associated with the time domain or the spectral domain. The time domain approach analysis mainly by modeling and fitting an Auto Regressive Moving Average (ARMA(p;q))-process to the sequences of observations.

The spectral domain approach is motivated by the observation that the most regular, and hence predictable, behavior of a time series is to be periodic. This approach then proceeds to determine the periodic components embedded in the time series by computing the associated periods, amplitudes, and phases, in this order

4.1 Frequency Domain Approach To Time Series

4.1.1 Fourier transform

Frequency domain is used to the analysis of mathematical functions or signals with respect to frequency, rather than time. Time-domain graph shows how a signal changes over time, whereas a frequency-domain graph shows how much of the signal lies within each given frequency band, over a range of frequencies. A given function or signal can be converted between the time and frequency domains with a pair of mathematical operators called a transform.

The *Fourier transform* decomposes a function into the sum of a (potentially infinite) number of sine and cosine wave frequency components. The 'spectrum' of frequency components is the frequency domain representation of the signal. The inverse Fourier transform converts the frequency domain function back to a time function.

There are several common conventions for defining the Fourier transform of an integrable function $f : \mathbf{R} \rightarrow \mathbf{C}$ (*Kaiser 1994*).:

$$\hat{f}(\xi) := \int_{-\infty}^{\infty} f(x) e^{-2\pi i x \xi} dx, \quad (4.1)$$

for every real number ξ .

When the independent variable x represents *time*, the transform variable ζ represents ordinary frequency (in hertz). Under suitable conditions, f can be reconstructed from \hat{f} by the *inverse transform*:

$$f(x) = \int_{-\infty}^{\infty} \hat{f}(\xi) e^{2\pi i x \xi} d\xi, \quad (4.2)$$

for every real number x .

In the study of Fourier series, complicated periodic functions are written as the sum of simple waves mathematically represented by sines and cosines. Due to the properties of sine and cosine it is possible to recover the amount of each wave in the sum by an integral. In many cases it is desirable to use Euler's formula, which states that

$$e^{2\pi i \theta} = \cos 2\pi \theta + i \sin 2\pi \theta \quad (4.3)$$

to write Fourier series in terms of the basic waves $e^{2\pi i \theta}$. This has the advantage of simplifying many of the formulas involved and providing a formulation for Fourier series. This passage from sines and cosines to complex exponentials makes it necessary for the Fourier coefficients to be complex valued. The usual interpretation of this complex number is that it gives you the amplitude (or size) of the wave present in the function and the phase (or the initial angle) of the wave. This passage also introduces the need for negative "frequencies". If θ were measured in seconds then the waves $e^{2\pi i \theta}$ and $e^{-2\pi i \theta}$ would both complete one cycle per second, but they represent different frequencies in the Fourier transform. Hence, frequency no longer measures the number of cycles per unit time, but is closely related.

We may use Fourier series to motivate the Fourier transform as follows. Suppose that f is a function which is zero outside of some interval $[-L/2, L/2]$. Then for any $T \geq L$ we may expand f in a Fourier series on the interval $[-T/2, T/2]$, where the "amount" (denoted by c_n) of the wave $e^{2\pi i n x / T}$ in the Fourier series of f is given by

$$\hat{f}(n/T) = c_n = \int_{-T/2}^{T/2} e^{-2\pi i n x / T} f(x) dx \quad (4.4)$$

and f should be given by the formula

$$f(x) = \frac{1}{T} \sum_{n=-\infty}^{\infty} \hat{f}(n/T) e^{2\pi i n x / T}. \quad (4.5)$$

If we let $\xi_n = n/T$, and we let $\Delta\xi = (n+1)/T - n/T = 1/T$, then this last sum becomes the Riemann sum

$$f(x) = \sum_{n=-\infty}^{\infty} \hat{f}(n/T) e^{2\pi i x \xi_n} \Delta\xi. \quad (4.6)$$

By letting $T \rightarrow \infty$ this Riemann sum converges to the integral for the inverse Fourier transform. Under suitable conditions this argument may be made precise. Hence, the Fourier transform can be thought of as a function that measures how much of each individual frequency is present in our function, and we can recombine these waves by using an integral (or "continuous sum") to reproduce the original function.

4.1.2 Discrete Fourier transform (DFT) and Fast Fourier transform (FFT)

Discret Fourier transform is a specific kind of Fourier transform, used in Fourier analysis. It requires an input function that is *discrete* and whose non-zero values have a limited (*finite*) duration. Such inputs are often created by sampling a continuous function, like a person's voice. And unlike the discrete-time Fourier transform (DTFT), it only evaluates enough frequency components to reconstruct the finite segment that was analyzed. Its inverse transform cannot reproduce the entire time domain, unless the input happens to be periodic (forever). Therefore it is often said that the DFT is a transform for Fourier analysis of finite-domain discrete-time functions. The sinusoidal basis functions of the decomposition have the same properties.

The DFT is widely employed in signal processing and related fields to analyze the frequencies contained in a sampled signal, to solve partial differential equations, and to perform other operations such as convolutions. The DFT can be computed efficiently in practice using a fast Fourier transform (FFT) algorithm.

A *fast Fourier transform (FFT)* is an efficient algorithm to compute the discrete Fourier transform (DFT) and its inverse. There are many distinct FFT algorithms involving a wide range of mathematics, from simple complex-number arithmetic to group theory and number theory;

A DFT decomposes a sequence of values into components of different frequencies. This operation is useful in many fields, but computing it directly from the definition is often too slow to be practical. An FFT is a way to compute the same result more quickly: computing a DFT of N points in the obvious way, using the definition, takes $O(N^2)$ arithmetical operations, while an FFT can compute the same result in only $O(N \log N)$ operations.

The most well known FFT algorithms depend upon the factorization of N , but (contrary to popular misconception) there are FFTs with $O(N \log N)$ complexity for all N , even for prime N . Many FFT algorithms only depend on the fact that $e^{-\frac{2\pi i}{N}}$ is an N th primitive root of unity, and thus can be applied to analogous transforms over any finite field, such as number-theoretic transforms.

Since the inverse DFT is the same as the DFT, but with the opposite sign in the exponent and a $1/N$ factor, any FFT algorithm can easily be adapted for it.

4.2 Spectrum Analysis of Collected Data

4.2.1 Spectrum analysis

Spectral analysis decomposes the series evolution in periodic contributions, thus allowing a more insightful view of its structure and on its cyclical behavior at different time scales [10].

The frequency spectrum is a complex, describing the magnitude and phase of a signal, or of the response of a system, as a function of frequency. In many applications, phase information is not important. By discarding the phase information it is possible to simplify the information in a frequency domain representation to generate a frequency spectrum or spectral density.

"Power Spectra" answers the question "which frequencies contain the signal's power?" The answer is in the form of a distribution of power values as a function of frequency, where "power" is considered to be the average of the signal. In the frequency domain, this is the square of FFT's magnitude.

Power spectra can be computed for the entire signal at once (a "periodogram") or periodogram of segments of the time signal can be averaged together to form the "power spectral density".

The power spectral density is a frequency-domain description that can be applied to a large class of signals that are neither periodic nor square-integrable; to have a power spectral density a signal needs only to be the output of a wide-sense stationary random process.

The periodogram computes the power spectra for the entire input signal:

$$\text{Periodogram} = \frac{|\text{abs}(F(\text{signal}))|^2}{N} \quad (4.6)$$

where $F(\text{signal})$ is the Fourier transform of the signal, and N is the normalization factor.

The *periodogram* is based on the definition of the power spectral density (PSD). Let $x_\omega(n) = \omega(n)x(n)$ denote a windowed segment of samples from a random process $x(n)$, where the window function ω (classically the rectangular window) contains M nonzero samples. Then the periodogram is defined as the squared-magnitude DFT of x_ω divided by M . In the limit as M goes to infinity, the expected value of the periodogram equals the true power spectral density of the process $x(n)$ [13].

4.2.2 Recorded data

Data is recorded every 5 seconds from a network of 14 stations. Figure 4.1 is the graph of the variations of the electric field over a period of seven days, recorded during December 2007. These are the raw data, without any kind of preprocessing or filtering.

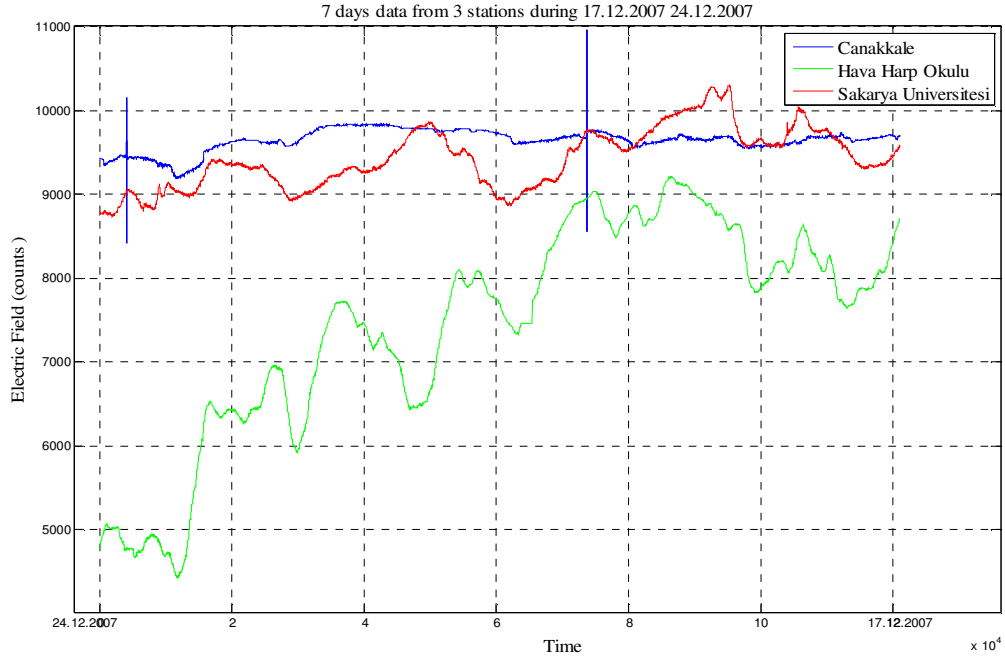


Figure 4.1 Data recorded at three different stations over a period of 7 days

The blue curve is data recorded at Canakkale station, the green curve is data recorded at Hava Harp Okulu in Istanbul, and the red curve is the data recorded at Sakarya University. The horizontal axis depicts UTC time while the vertical axis presents Electric Field.

4.2.3 Periodogram analysis of data

In the analysis of the periodogram, choosing the size of the window is a very important. A window that captures less than a full period of a frequency leads problems such as frequency leakage.

Looking at the data recorded over several years, several periods are clearly noticeable. The main period seems to be of one year long. Figure 4.2 shows the data recorded at three different stations, every 5 seconds over a period of 2200 days, i.e. 6 years, period between 31 December 2007 and 15 December 2002. So the chosen window length for analysis is of 2200days \times 17280 values.

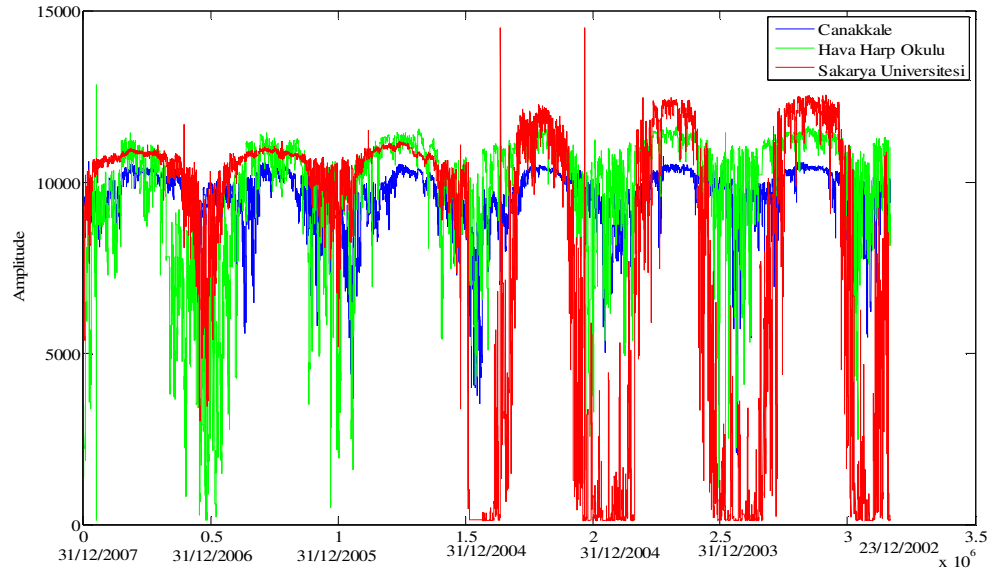


Figure 4.2 Data recorded at 3 stations over a period of 2200 days

In order to see if there are any periodic frequencies, the periodogram was used. In Figure 4.3 we can see the periodogram of the data plotted at Figure 4.2.

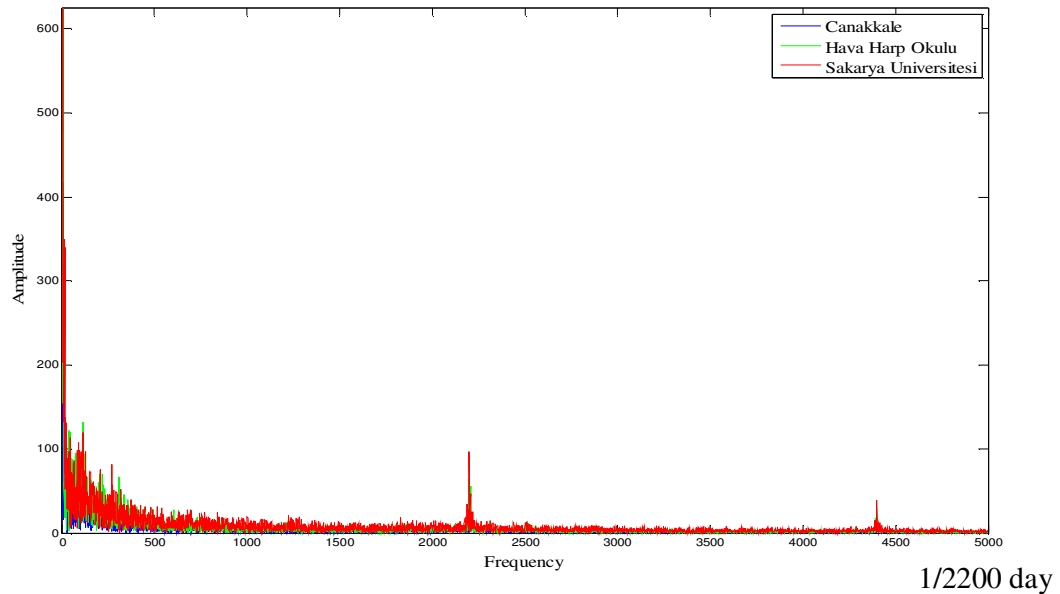


Figure 4.3: Periodogram on data recorded at 3 stations over a period of 2200 days

Analyzing the periodogram from low frequency to high frequency, we can observe several clear peaks. The first one is the frequency having the value 6. This frequency corresponds to a period of about 366 days, i.e. yearly period. Figure, 4.4 shows the

amplitude and the percent of spectral energy on that frequency. So the variation of electric charge in the air shows a cyclical period of one year.

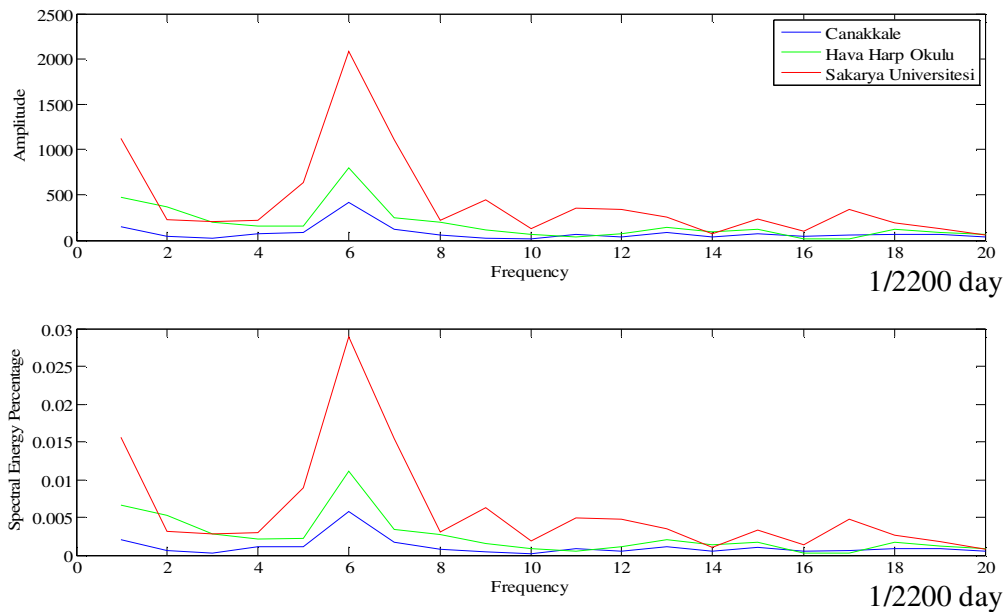


Figure 4.4 Amplitude (upper picture) and the percent of the spectral energy (lower picture) of the frequency having a period of one year (366 days) on data recorded at 3 stations over a period of 2200 days

Another period that is clearly noticeable is the one at frequency 2200 out of 2200 days data. This corresponds to a frequency with a period of around 1 day, i.e. 24 hours. Beside the big peak in the middle, there are two other smaller peaks, which correspond to a period of 2194 and 2206, i.e. 23.93 hours and 24.07 hours respectively. Figure 4.5 shows the amplitude and the percent of spectral energy of the frequencies. So the variation of electric charge in the air shows additional cyclical periods of around 24 hours.

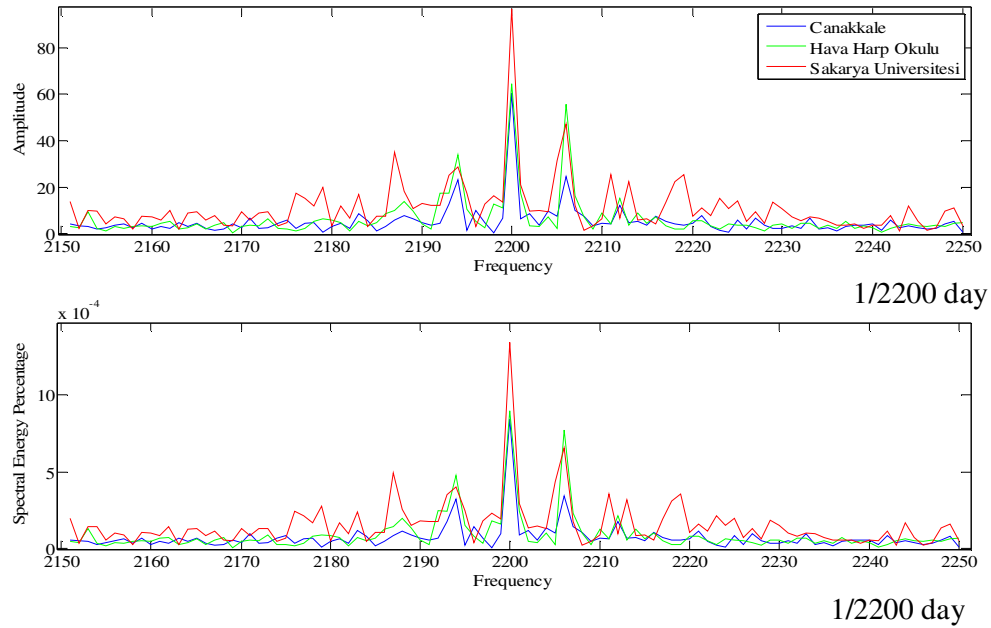


Figure 4.5: Amplitude (upper picture) and the percent of the spectral energy (lower picture) of the frequency having a period of one day on data recorded at 3 stations over a period of 2200 days

The characteristic period of 24 hours is located in the components of the mechanic lithosphere oscillation (well-known to the geophysics) which is caused by the tidal gravity influence.

Other available measurements of the components of the Earth's magnetic field (included in the measurements carried out by special scientific projects in Greece and around the world) show the existence of a permanent periodic variation with a constant period of about 24 hours. Taking into consideration the above statement, we may infer that the periodic variation of the measured electric signal is the electric component of an electromagnetic wave. It appears that this electromagnetic wave is caused by the tidal deformity of the Earth.

Other periods are the ones around frequency 4400 out of 2200 days data, corresponding to about 12 hours. Figure 4.6 shows the amplitude and spectral energy of the frequency. The electric charge variation has a period of 12 hours.

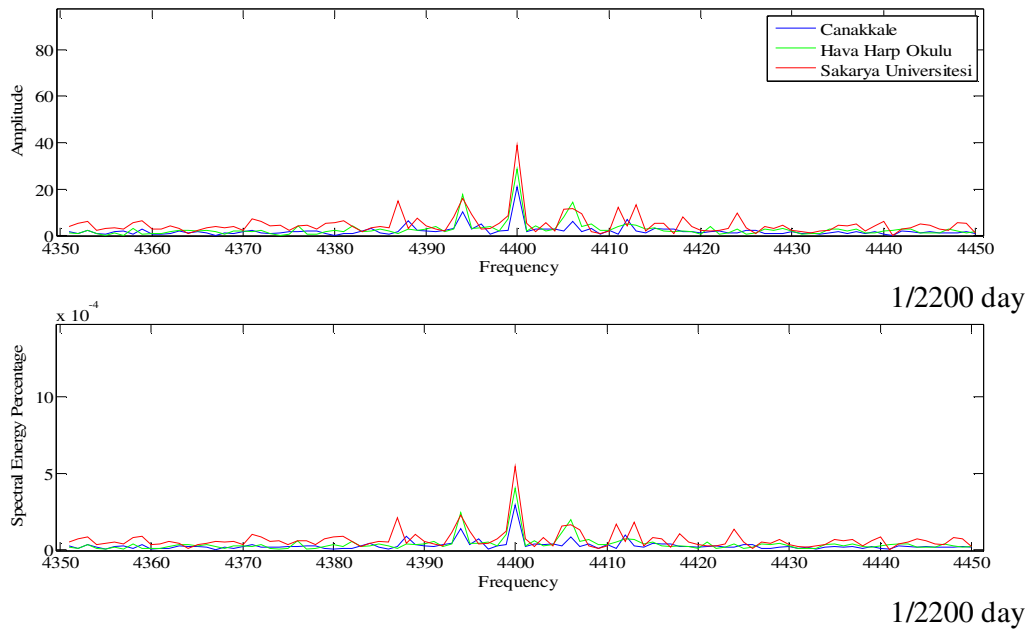


Figure 4.6 Amplitude (upper picture) and the percent of the spectral energy (lower picture) of the frequency having a 12 hour period on data recorded at 3 stations over a period of 2200 days

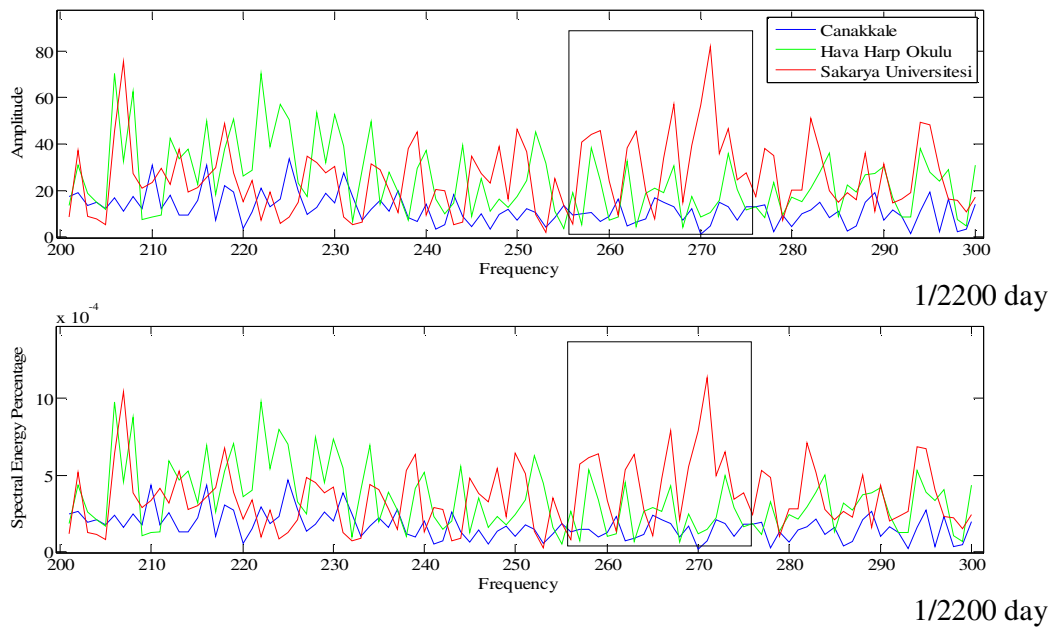


Figure 4.7 Amplitude (upper picture) and the percent of the spectral energy (lower picture) of the 8.5 days period corresponding to Sun's adiabatic expansion on data recorded at 3 stations over a period of 2200 days

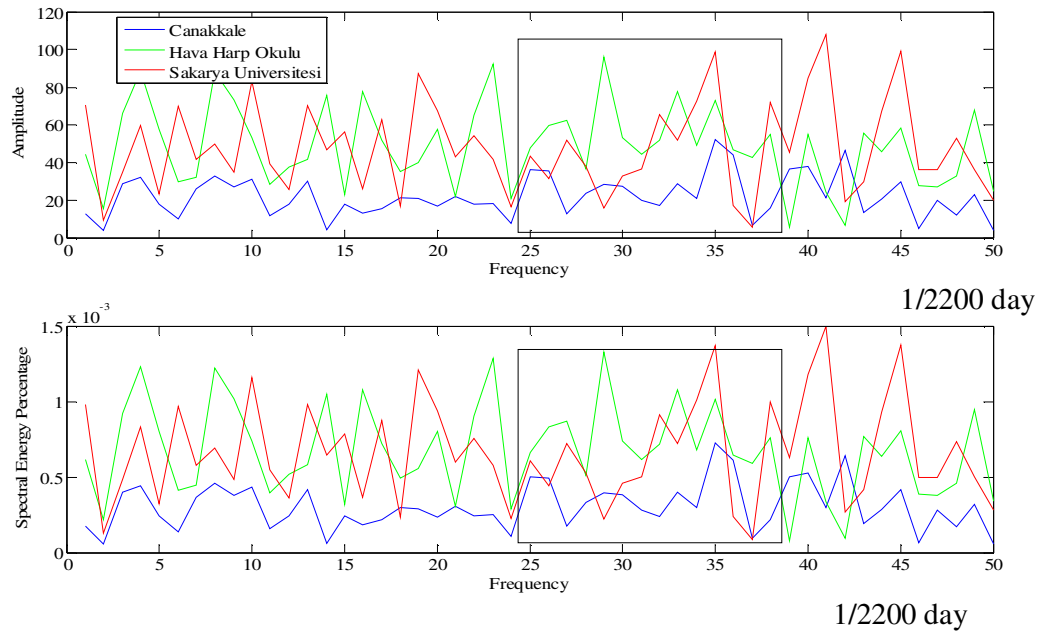


Figure 4.8 Amplitude (upper picture) and the percent of the spectral energy (lower picture) of 27.5 days period corresponding to Sun's adiabatic expansion on data recorded at 3 stations over a period of 2200 days

4.2.4 Discussion on found periodic components

Gravitational attraction of the sun and the moon are responsible for the tidal motions on the globe, in all three media, the atmosphere, the oceans and the solid Earth.

The rhythmic rise and fall of sea level along the world's coastlines are the most apparent manifestation of tides in the global oceans. In some places, the sea level rises and falls with a period of about half a day (these are called semi-diurnal tides), whereas in other places the period is more like a day (called diurnal tides). Yet again, in some locations the tides are mixed.

Tidal forcing varies quite rapidly with time. Resonances in the oceanic response push tides in certain localities to be above the values predicted by the equilibrium theory. While the equilibrium theory predicts two bulges to form, one underneath the moon and the other on the opposite side of the globe, in reality the high water may significantly precede or lag the transit of the moon. These differences are due to the dynamical response of the oceans to tidal forcing. Oceanic tides are considered to be the response of the fluid medium to the astronomical forcing by the sun and moon's gravitational attractions. The tides result from dynamical balances principally

between the tractive forces due to gravitation and the retarding forces due to friction in the water column.

The sun and the moon are the only important celestial bodies in producing terrestrial tides. While the moon is much smaller than the sun, it is nevertheless more important for tidal processes, because of its proximity to the Earth. It is the small imbalance between the centrifugal force and gravitational attraction of the moon on the water column that give rise to horizontal tractive forces, causing water motions that tend to cause two bulges in the sea surface, one immediately underneath and the other on the other side of the globe. These bulges tend to rotate around the globe along with the moon resulting in semi-diurnal tides with a period of half a lunar day (12.4 hours), even though the Earth's rotation is diurnal (period of 24 hours).

To model tidal motions then, it is necessary to prescribe the astronomical forcing. The tidal potential term in the equations of motion can be expressed as a Fourier series, with each term representing a tidal constituent. While there are tens of such constituents in a general expansion, usually in the deep ocean, only 11 so-called primary components are important. Of these, four are semi-diurnal, M2, N2, S2 and K2; four are diurnal, K1, O1, P1 and Q1; and the rest, Mf, Mm and Ss are long-term.

Table 4.1 Primary Tidal Components.

Tidal component	Description	Period (hours)	Equivalent frequency on 2200 day data
M2	Principal Lunar	12.43	4247
S2	Principal Solar	12.00	4400
N2	Larger Lunar Elliptic	12.66	4170
K2	Luni-Solar	11.97	4411(4406)
K1	Lunar-Solar Declinational	23.93	2206
O1	Principal Lunar Diurnal	25.82	2045
P1	Principal Solar Declinational	24.07	2194
Q1	Large Lunar Elliptic	26.87	1965
MF	Lunar Fortnightly	327.90	
MM	Lunar Monthly	661.30	79
SSA	Solar Semi Annual	4383.00	13
	Solar Annual	8766.00	6

The effect of the gravitational attraction of the sun and the moon is also observed on the recorded data. The previously mentioned periods in Table 4.1, are seen in the periodogram from Figure 4.3. Marked components are seen in the Earthquake Prediction Project's recorded data

The yearly period shown in Figure 4.4 corresponds to the Solar Annular tidal component of the Earth. The Solar Semi Annual period is also noticeable.

The peaks presented in Figures 4.5 correspond to a period of 24 hours, i.e one day. The first smaller peak coincides with the period of the Lunar-Solar declinational tidal component K1 and the second smaller peak corresponds to Solar declinational component P1.

The peaks presented in Figures 4.6 correspond to a period of 12 hours, i.e. half day.

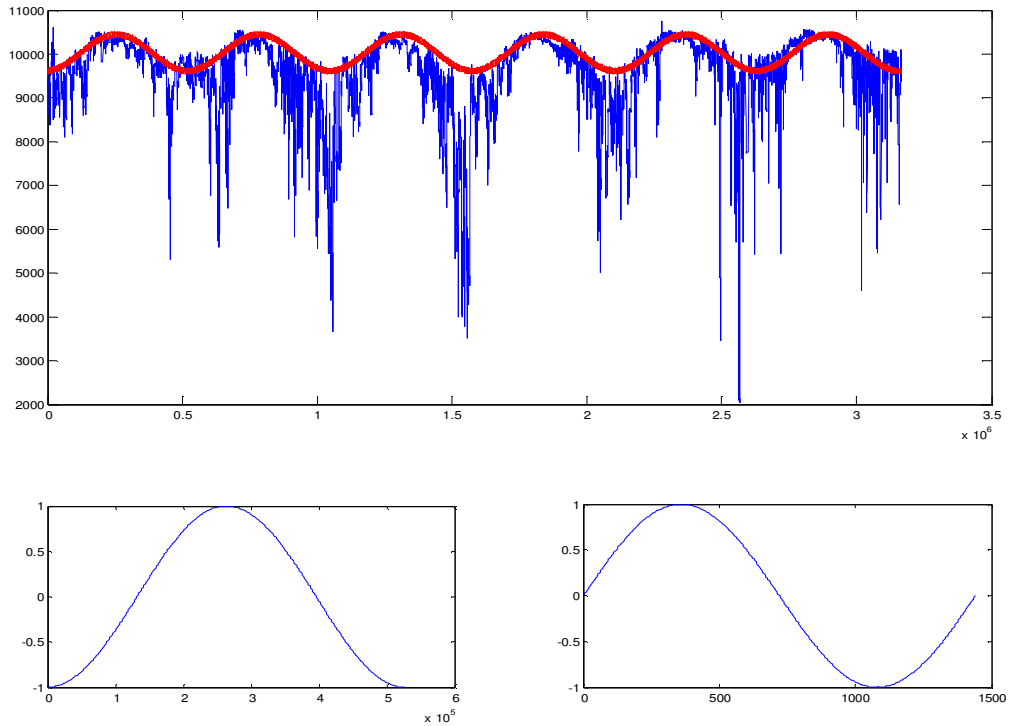


Figure 4.9 Fitting a signal having two harmonics first with a period of 1 year and the second with a period of 1 day to 2200 days data from Çanakkale Station.

In Figure 4.9 we fitted a function having two harmonics one with a daily period and one with yearly period. We can see that it fits very nicely, except for the high frequency components.

5. INVESTIGATION OF EARTHQUAKE PRECURSORY EVENTS, IN FREQUENCY DOMAIN

5.1 Data Description

The data to be analyzed has been recorded since 2001, at an interval of 5 seconds.

The study is mainly conducted on data coming from 3 stations, Çanakkale, Hava Harp Okulu and Sakarya Üniversitesi. The reason for this fact is that the data is much more continuous, has less missing values, than the other stations.

Analyzing the data, we can observe some interesting behaviors, as mentioned in the previous chapter. The data has major low frequency signals ranging from one year period to one day.

After removing these low frequency signals with a high pass filter, then we could see the variance of the data increases during winter and decreases during summer. This is shown in Figure 5.1. The reason for this should be analyzed. One reason could be the effect of temperature on the monopolar probe, another reason could be the effect of the sun ionizing the atmosphere or the probe could get saturated and could not detect small changes. But as previously mentioned these assumptions are not yet verified. So the algorithm to be used for analyzing this data should be dynamic, to be able to make good predictions knowing these facts.

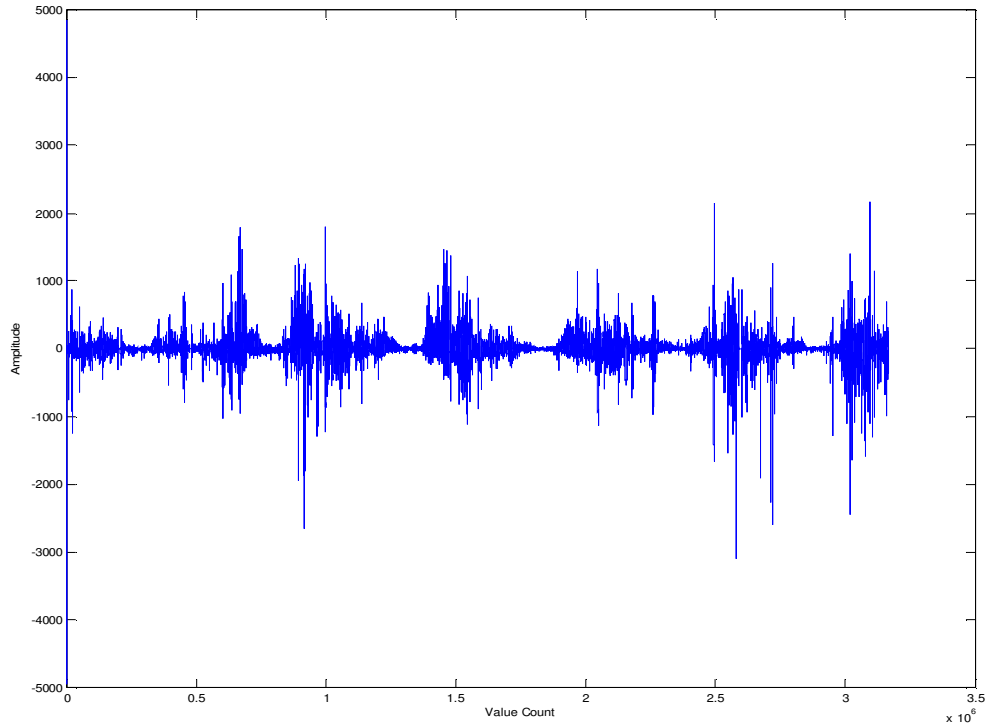


Figure 5.1 : 2200 days data filtered with a low-pass filter.

Additional to the data recorded by the project, the list of all the earthquakes with magnitude over 2 on Richter scale that happened between 25-45 E and 35 – 42 N, was taken from Kandilli Rasathanesi ve Deprem Arařtırma Enstitüsü. This list of earthquake was used to measure the performance of our predictions.

The earthquakes were not taken one by one, but by the energy released during a defined period of time and over a defined zone. Equation 5.1 shows the way earthquake energy was calculated.

$$E(l, L, t, \Delta l, \Delta L, \Delta t) = \sum_{m=t}^{t+\Delta t} e^{q(m)} \quad (5.1)$$

l = latitude, L = Longitude, t = time, and the corresponding intervals, q being the magnitude of the earthquake.

Table 5.1: A sample from earthquake list.

Date	Time	Latitude	Longitude	Depth(Km)	Magnitude (Richter)	Locations
02.01.2002	12:02:35.03	39.4000	38.4200	06.6	3.9	Kemaliye-ERZINCAN
04.01.2002	00:22:26.58	37.0500	36.3600	12.5	2.4	Osmaniye
04.01.2002	02:48:10.47	39.2500	27.8300	10.6	2.3	Savastepe-Balikesir
04.01.2002	17:36:27.45	40.2100	34.2800	02.2	3.1	Sungurlu-ÇORUM
04.01.2002	20:44:14.05	40.9500	30.9300	04.6	3.4	Cumayeri-BOLU
04.01.2002	22:45:55.53	40.8200	30.5600	04.5	2.4	Hendek-ADAPAZARI
08.01.2002	14:17:57.41	36.2900	29.5700	23.8	3.3	Kas-ANTALYA
10.01.2002	14:26:29.06	36.9800	29.1500	24.0	2.8	Çameli-DENIZLI
10.01.2002	21:01:16.45	36.5400	29.3600	04.9	3.0	Fethiye-MUGLA
10.01.2002	21:21:50.29	40.7900	30.9700	05.8	3.2	Gölyaka-DÜZCE

5.2 Earthquakes and the Recorded Anomalies

The anomalies were studied in the higher frequency domain, having a period of at most 1 day. But this does not make it a rule, since studying the data, days before the earthquake happen, there are noticeable increases and decreases in the amplitude of the data. Some patterns in 0.1-10 hours' time intervals are repetitively recorded before seismic events. These can be classified as: step function, impulse function, impulsive increment followed by negative exponential decrement and exponential increment followed by negative exponential decrement that begins with a discontinuity.

In Figure 5.2 an example of earthquake with magnitude greater than M4 is shown. The blue arrow shows the time when earthquakes happen. As it can be observed, before a major earthquake happens, there are major changes in the path followed by data.

In Figure 5.3 is another example showing the way the path changes before a major earthquake happens. Something interesting is noticed. Analyzing the data in time domain, with a window as long as one day, one could easily say that the peaks that arise at the middle of the window can be classified as anomalies. But we observe this kind of behavior every day within that period of time. So, the technique to distinguish the anomalies is based not only on the presence of sudden changes in the path, but changes relative to the values recorded at similar past time intervals.

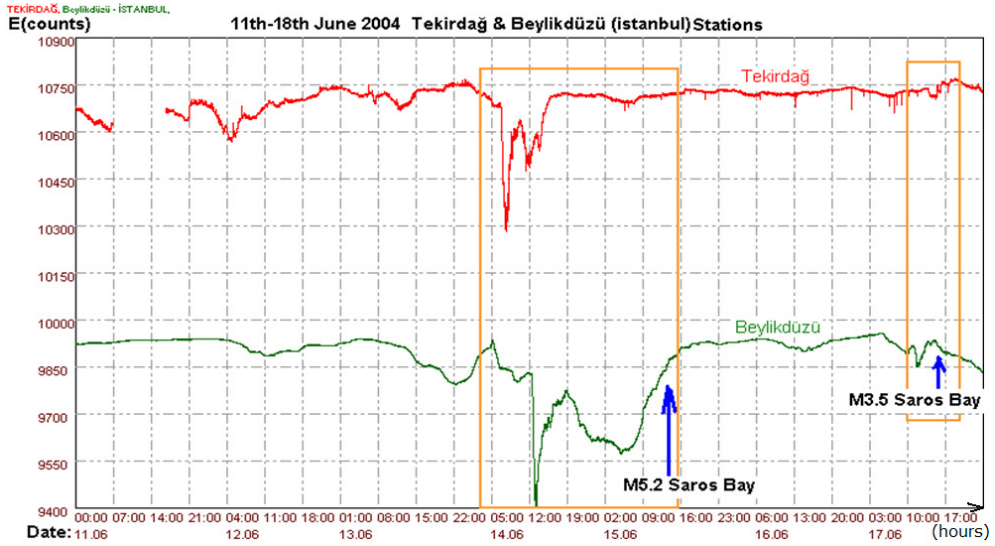


Figure 5.2: Seven-days graph of Tekirdağ and Beylikdüzü stations with M5.2 earthquake in Saros and its aftershock in June 2004.

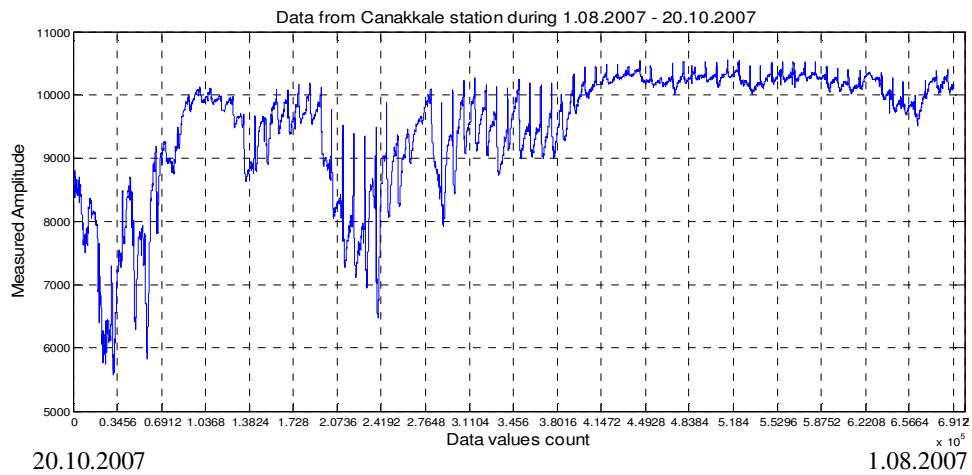


Figure 5.3: Graph of Canakkale station with M5.2 earthquake Mediterranean Sea

Data shows high seasonal variability. This is clearly visible if we remove the high frequency components of the data. Figure 5.4 was obtained by computing the mean of 24 hours data window sliding it by 3 hours. Some major drops of energy are seen. The reason for their happening is not clear. But looking at Figure 5.5 we can see a high correlation between the ups and downs from the path and the happening of earthquakes.

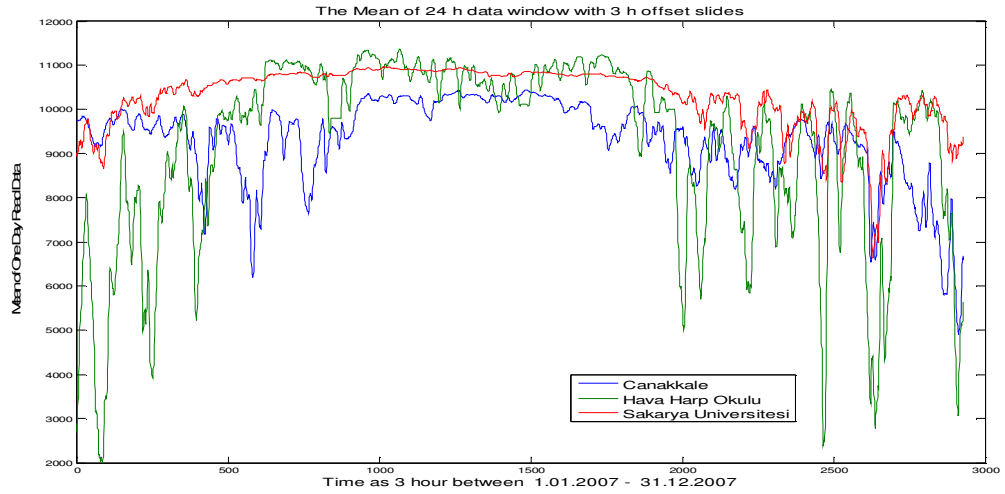


Figure 5.4 : One year data as mean of a window of 24 hours and sliding offset of 3 hours between 1.01.2007 – 31.12.2007

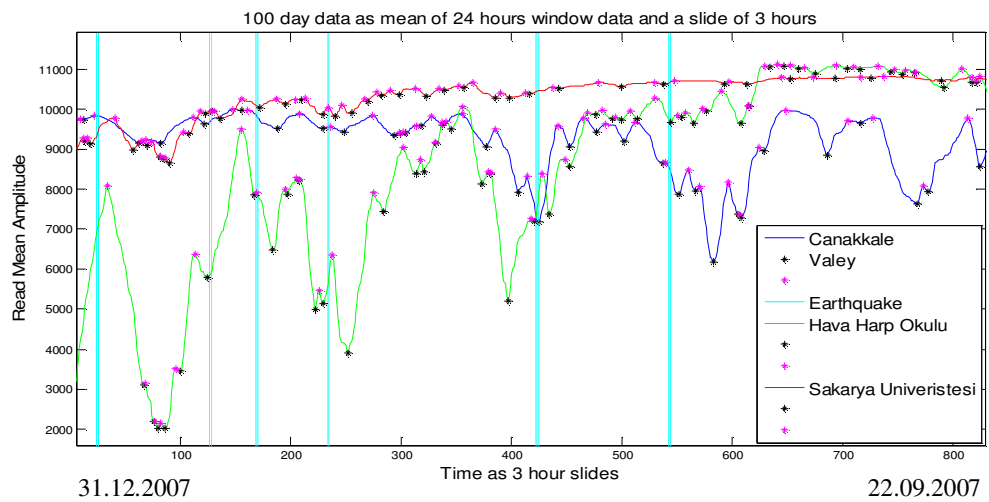


Figure 5.5: 100 days data as mean of a window of 24 hours and sliding offset of 3 hours between 1.01.2007 – 22.09.2007 together with earthquakes of magnitude > 5 degree

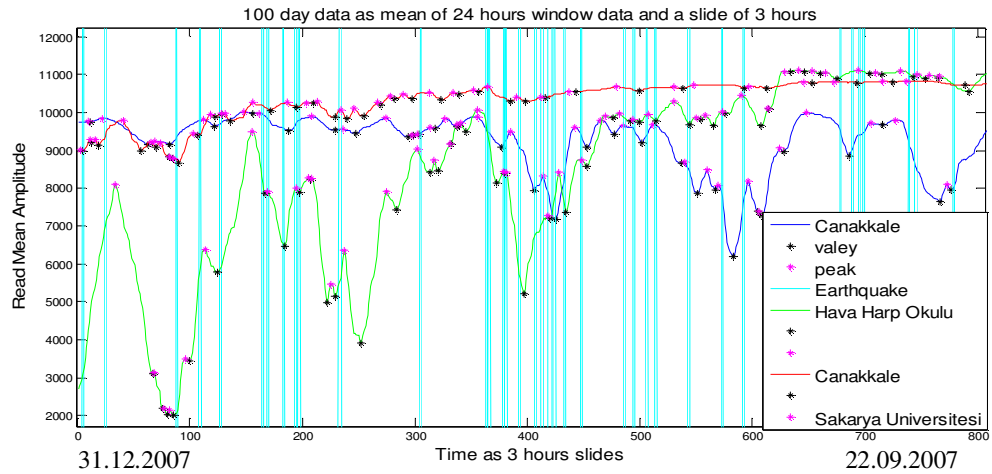


Figure 5.6: 100 day data as mean of a window of 24 hours and sliding offset of 3 hours between 1.01.2007 – 22.09.2007 together with earthquakes of magnitude > 3.5 degree

5.3 Earthquake Precursory Anomaly Detection

The anomalies previously mentioned leave different marks on the spectral decomposition of a portion of the data from a specific station.

Before we start working with the data we run a high pass filter over the data so that we can remove abnormal read data. The filter is actually a mean of the previous data recorded over 5 minutes.

Each station has its own unique behavior. This is seen in the oscillations in time domain. Earthquake precursory anomalies are mostly the same and should be identified regardless of the station.

The filtered data of each station is used to compute short time Fourier transforms (STFT), i.e. we compute the spectral density of the data. This is done by taking a fix time length window and slide it by a specific amount of time.

It is quite difficult to choose the size of the window that is going to be use for STFT. This is caused by the fact that in spectral analysis one of the biggest problems is the spectral leakage. So to avoid this, we chose the window to be of one day, and the size of the sliding offset to be 3 hours data.

Since the characteristics of the signal normally changes little over time, though it has seasonal variations as we can see in Figure 5.1, we compute a mean spectral window of the several previous windows at every step. The number of windows to be

accounted for the mean is chosen by taking in consideration the rate of seasonal change and the rate of success of the method.

All these methods were implemented in Matlab (model.m), having as data source the data recorded from stations and as validation data the list of earthquakes taken from Kandilli Rasathanesi ve Deprem Arařtırma Enstitüsü. The earthquakes taken in consideration were higher than 3.5 degrees Richter and on a region between latitudes $36^{\circ} - 41^{\circ}$ N and between longitudes $26^{\circ} - 31^{\circ}$ E, i.e. the region around Marmara Sea.

5.3.1 Mathematical model

The time series is made of recorded static energy measured from the surface of the earth at intervals of 5 seconds.

$$x = \{x_1, x_2, x_3 \dots\}$$

A low-pass filter is used to remove the abnormal read values and the noise.

$$x_i = \left(\sum_{j=0}^{60} x_{i+j} \right) / 60 \quad (5.2)$$

The data is then transformed from time domain to frequency domain by taking the Fourier transform of a sliding window as long as one day data, i.e. $w=24h * 60' * (60'')/5 = 17280$ data values. The window is slid through the whole data by an offset of 3 hours, i.e. $\text{offset} = 3h * 60' * (60'')/5 = 2160$ data values.

Thus we obtain a new time series of vectors.

$$Y_t = \text{FFT}(x_t, x_{t+1}, \dots, x_{t+w}) = [y_{t,1}, y_{t,2}, \dots, y_{t,w}] \quad (5.3)$$

The first 10 lower frequency components of the obtained transformation are summed, i.e. the frequencies having a period starting from 12 hours up to 2 hours.

And so we obtain another time series that keeps track of the spectral change.

$$z_t = \left(\sum_{i=3}^{12} y_{t,i}^2 \right) / w \quad (5.4)$$

The time series thus obtained was used to compute two different means

The first mean is the mean of the previous one week of obtained z_t .

$$\text{Mean1}_t = \sum_{i=0}^{56} z_{t+i} \quad (5.5)$$

The second mean is the mean of the previous two weeks of obtained z_t

$$\text{Mean2}_t = \sum_{i=0}^{112} z_{t+i} \quad (5.6)$$

This two means keep the trend of the spectral change, since data is dependent on the near past.

All the z_t that passes over the two mean or are neighbors with ones that pass are denoted to be anomalies.

$$v_t = \sum_{i=0}^7 x_{i+j} \{ z_{t+i} \mid z_{t+i} > \text{Mean1}_{t+i} \parallel z_{t+i} > \text{Mean2}_{t+i} \} \quad (5.7)$$

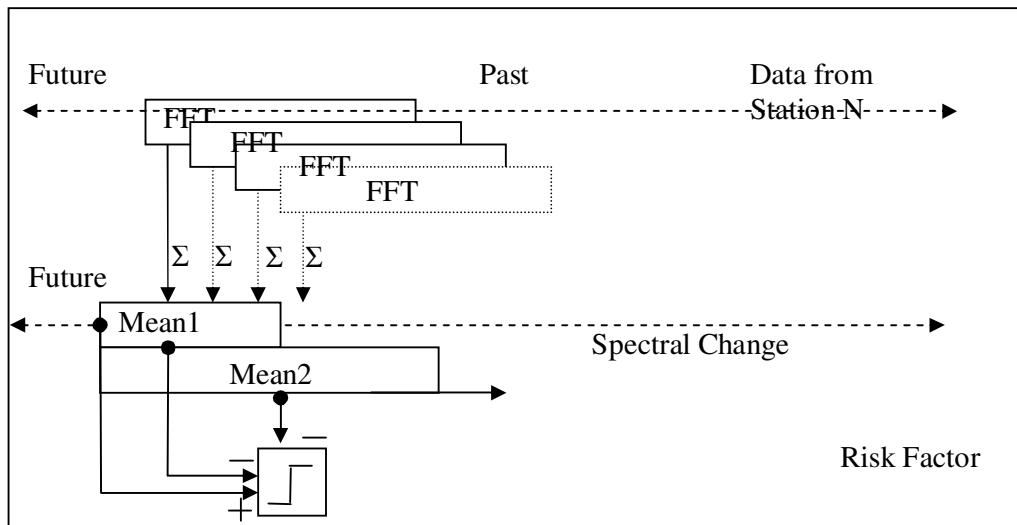


Figure 5.7: Prediction Scheme

5.3.2 Results validation

The test was done on 3 years data, between 1.01.2005 until 31.12.2007. The validation data was taken from Kandilli Rasathanesi ve Deprem Araştırma Enstitüsü. The earthquakes taken in consideration were higher than 3.5 degrees Richter and on a region between latitudes $36^\circ - 41^\circ$ N and between longitudes $26^\circ - 31^\circ$ E, i.e. the region around Marmara Sea.

Table 5.2 Results

size of the prediction window		6 h	9 h	12 h	15 h	18 h	21 h	Total # of Quakes
Çanakkale	> 3.5	230	256	275	230	243	254	414
	> 3.5 %	0,56	0,62	0,66	0,56	0,59	0,61	
	> 4	42	50	51	44	48	52	75
	> 4 %	0,56	0,67	0,68	0,59	0,64	0,69	
	> 5	6	7	7	7	8	8	8
	> 5 %	0,75	0,88	0,88	0,88	1,00	1,00	
	quake prediction	4102	4407	4650	4876	5090	5270	8664
	quake prediction %	0,47	0,51	0,54	0,56	0,59	0,61	
	not quake prediction	4562	4257	4014	3788	3574	3394	
	total predictions	8664	8664	8664	8664	8664	8664	
total days	1083	1083	1083	1083	1083	1083		

Havaharp	>3.5	242	263	275	284	294	305	509
	>3.5 %	0,48	0,52	0,54	0,56	0,58	0,60	
	>4	46	52	56	57	59	60	97
	>3.5 %	0,47	0,54	0,58	0,59	0,61	0,62	
	>5	5	5	6	6	6	6	11
	>3.5 %	0,45	0,45	0,55	0,55	0,55	0,55	
	quake prediction	3916	4258	4554	4803	5013	5206	
	quake prediction %	0,45	0,49	0,53	0,55	0,58	0,60	
	not quake prediction	4748	4406	4110	3861	3651	3458	
	total predictions	8664	8664	8664	8664	8664	8664	
	total days	1083	1083	1083	1083	1083	1083	

Different sizes of the prediction window were taken into account, each of them generating different results for different earthquake magnitudes. Decreasing the window size decrease the number of earthquakes predicted, while increasing also the number of predictions we made.

6. CONCLUSION

Many different techniques were adopted to predict the earthquake. All of them measure different physical properties of the planet. The collected data hides interesting things about the phenomena they encounter. Data processing helps uncover these phenomena and derive some mathematical models about them. The models thus obtained can be used to predict the happening of future such phenomena.

In this thesis I tried two different techniques to predict the risk of happening of an earthquake on data collected from the “*Earthquake Forecast by the method of regional stress measurement*” project conducted by Ass. Dr. Berk Ustundag.

The data shows interesting frequencies generated by sun and moon that are not related to earthquakes, so data had to be firstly filtered to remove them.

After preprocessing task, spectral analysis was used to to compress the available data and still preserve it's characteristics. The spectrum was obtained by a taking the Fourier transform of a moving window over the collected data and over the difference between two different stations.

REFERENCES

- [1] **Üstündağ, B.**, 2008, Earthquake Precursory Electric Field Signatures, Nova Science Publishers
- [2] **Wencai, Y., Robinson, P., Rongshan, F. and Ying, W.**, 2001. Geodynamic Processes and Our Living Environment, Geological Publishing House, Institute of Geology, CAGS, China
- [3] <<http://www.phy.ornl.gov/csep/CSEP/OM/NODE30.html>>, accessed at 2.09.2009
- [4] **Ghil, M., Allen, M. R., Dettinger, M. D., Ide, K., Kondrashov, D., Mann, M. E., Robertson, A. W., Saunders, A., Tian, Y., Varadi, F. and Yiou, P.**, 2002: Advanced spectral methods for climatic time series, 40, 1 / Reviews of Geophysics
- [5] **Iacobucci, A.**, November 2003. Spectral Analysis for Economic Time Series
- [6] **Tsatsaragos, J. M.**, 2008. Measurements of periodically electric field from the interior of the earth, Journal of Engineering Science and Technology Review
- [7] **Box, G., Jenkins, G. M., Reinsel, G.**, 1994. Time Series Analysis: Forecasting & Control, Prentice Hall; 3rd edition
- [8] **Weigend, A.S. and Gershenfeld, N.A.**, editors, 1993. Time Series Prediction: Forecasting the Future and Understanding the Past. Addison Wesley
- [9] **Koskela, T., Varsta, M., Heikkonen, J. and Kaski, K.**, 1998. Time series prediction using RSOM with local linear models. International Journal of Knowledge-Based Intelligent Engineering Systems
- [10] **Falk, M., Marohn, F., Michel, R., Hofmann, D., Macke, M.**, 2006. A First Course on Time Series Analysis Examples with SAS, by Chair of Statistics, University of Wurzburg
- [11] <http://ccrma.stanford.edu/~jos/sasp/Matlab_Periodogram.html>, accessed at 19.08.2009

CURRICULUM VITAE

Candidate's full name: Serdar BAGIS

Place and date of birth: Constanta / Romania

Permanent Address: Constanta / Romania

Universities and

Colleges attended: International Computer High school Constanta
Romania (1996 - 2000)

Ovidius University, Mathematics-Informatics
Faculty Constanta Romania (2000 – 2004)