ISTANBUL TECHNICAL UNIVERSITY ★ EURASIA INSTITUTE OF EARTH SCIENCES

SPATIO-TEMPORAL ANALYSIS OF PARTICULATE MATTER CONCENTRATIONS OF TURKEY

M.Sc. THESIS
Seden BALTACIBAŞI

Department of Climate and Marine Sciences

Earth System Sciences

Thesis Supervisor: Assoc. Prof. Tayfun KINDAP
Thesis Co-Supervisor: Assoc. Prof. Alper ÜNAL

January, 2014

ISTANBUL TECHNICAL UNIVERSITY ★ EURASIA INSTITUTE OF EARTH SCIENCES

SPATIO-TEMPORAL ANALYSIS OF PARTICULATE MATTER CONCENTRATIONS OF TURKEY

M.Sc. Thesis
Seden BALTACIBAŞI
601081012

Department of Climate and Marine Sciences

Earth System Sciences

Supervisor (Chairman) : **Assoc. Prof. Tayfun KINDAP (ITU)**
Co-Supervisor : **Assoc. Prof. Alper ÜNAL (ITU)**

January, 2014

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ AVRASYA YER BİLİMLERİ ENSTİTÜSÜ

TÜRKİYEDEKİ PARTİKÜL MADDE KONSANTRASYONUNUN ALANSAL VE MEKANSAL ANALİZİ

YÜKSEK LİSANS TEZİ

Seden BALTACIBAŞI
(601081012)

İklim ve Deniz Bilimleri Anabilim Dalı

Yer Sistem Bilimleri

Tez Danışmanı : Doç. Dr. Tayfun KINDAP
Tez Eş Danışmanı : Doç. Dr. Alper ÜNAL

Ocak, 2014

FOREWORD

I would like to express my deep appreciation and thanks for my advisor Assoc. Prof Tayfun Kındap and my co-advisor Assoc. Prof. Alper UNAL. I would like to thank to my family and for their support and help whenever I needed.

December 2013

Seden BALTACIBAŞI
Eurasia Institute of Earth Sciences

TABLE OF CONTENTS

Page

FOREWORD......v
TABLE OF CONTENTS......vii
ABBREVIATIONS......ix
LIST OF TABLES......xi
LIST OF FIGURES......xiii
SUMMARY......xv
ÖZET......xvii
1. INTRODUCTION......1
2. METHODOLOGY......7
2.1 Study Area......7
2.2 Air Quality Monitoring Stations......8
2.3 Air Quality Data Processing......9
2.3.1 Air quality monitoring stations data......10
2.3.2 Emissons data......11
2.3.3 Meteorology data......11
2.4 Analysis Tools......14
3. RESULTS......15
4. CONCLUSION......41
REFERENCES......43
APPENDICES......45
CURRICULUM VITAE......77

ABBREVIATIONS

PM_{10} :Particulate Matter (<10 µm diameter)
$PM_{2.5}$:Particulate Matter (<2.5 µm diameter)
AMI :Acute Myocardial Infarction
APHEA-2 :Air Pollution and Health: A European Approach 2
ABPRS :Address Based Population Registration System Results
CODP :Chorinic Obstructive Pulmonary Disease
EU :European Union
EPA :Environmental Protection Agency
VOCs :Volatile Organic Compounds
WRF :Weather Forecasting Model
CMAQ :Community Multiscale Air Quality Model
PCA :Principal Component Analysis
TUIK :Turkish Statistical Institute
PCs :Principal Components
NCEP :National Center of Environmental Prediction
NCAR :National Center of Atmospheric Research
NWP :Numerical Weather Prediction
NOAA :National Oceanic and Atmospheric Administration
SNAP :Standardized Nomenclature for Air Pollutants
TNO :Netherlands Organisation of Applied Scientific Research

LIST OF TABLES

Page

Table 1.1: PM10 standards 4
Table 2.1: Air quality monitoring stations list of each region and the cities have more than one stations 9
Table 2.2: TNO emssions inventory SNAP categories 11
Table 3.1: Summary statistics of PM10 levels in Turkey between 2008 -2010 15
Table 3.2: Summary of Principal Component Analysis 24
Table 3.3: Loadings of Principal Component Analysis of 4 parameters 25
Table 3.4: Euclidean distance 5 parameters and clusters 27
Table 3.5: Summary of Principal Component Analysis of 5 parameters 29
Table 3.6: Loadings of Principal Component Analysis of 5 parameters 29

LIST OF FIGURES

Page

Figure 1.1: Illustration of PM coarse and fine fractions (EPA, 2010) 2
Figure 2.1: 2012 population density of Turkey (TUİK) 7
Figure 2.2: Annual average growth rate (2007-2012) (TUİK) 8
Figure 2.3: Location of air quality monitoring stations in Turkey.............................. 9
Figure 2.4: Air quality data processing diagram... 10
Figure 2.5:NCEP-NCAR Reanalysis data grid cells.. 12
Figure 2.6:WRF model outer (d01) and inner (d02) domain 13
Figure 3.1: Summary of PM10 levels in Turkey of 3 years...................................... 16
Figure 3.2: Time series plot of three years data 17
Figure 3.3: Bayesian information criterian plot for three years 17
Figure 3.4:K-means clustering with Euclidean distance colored map of three years mean values in Turkey 19
Figure 3.5: Manhattan distance K-means clustering colored map of three years with only mean 19
Figure 3.6: Cluster differences between Euclidean distance and Manhattan distance K-means clustering colored map of three years with mean 20
Figure 3.7:a) Cumulative distribution functions (CDFs) of the Euclidean distance k-means clusters with only mean values. b) Zoom to the CDFs above 200 $\mu g/m^3$ 21
Figure 3.8:a) Cumulative distribution functions (CDFs) of the Manhattan distance k-means clusters with only mean values. b) Zoom to the CDFs above 200 $\mu g/m^3$ 22
Figure 3.9: Euclidean distance K-means clustering colored map of three years with 4 parameters: mean, 2.5 %, median, 97.5 %. 23
Figure 3.10:a) Cumulative distribution functions (CDFs) of the Euclidean distance k-means clusters with four parameters. b) Cumulative distribution functions (CDFs) of the Euclidean distance k-means clusters with mean 23
Figure 3.11:a) Cumulative distribution functions (CDFs) of the Euclidean distance k-means clusters with four parameters. b) Zoom to the CDFs above 200 µg/m3... 23
Figure 3.12: Euclidean distance K-means clustering colored map of three years with first two principal component.. 25
Figure 3.13: Cumulative distribution functions (CDFs) of the PCA K-means clusters with four parameter 26

Figure 3.14: Euclidean distance 5 parameter K-means clustering colored map of three years with first two principal component 30
Figure 3.15:TNO emissions map of averages of all sectors..................................... 30
Figure 3.16:NCEP-NCAR Reanalysis data 850 mb – 700 mb temperature differences in Afyon, Isparta and Antalya grid 31
Figure 3.17:NCEP-NCAR Reanalysis data 850 mb-700 mb temperature differences in Izmir, Aydın and Muğla grid... 32
Figure 3.18: Kahramanmaraş PM10 concentrations and PBL height in 2008........... 33
Figure 3.19:İstanbul PM10 concentrations and PBL height in 2008 33
Figure 3.20: Annual BL heights spatial plot 34
Figure 3.21:a) 2008 January and Februrary at 12 AM averages of PBL heights b) 2008 March, April, May at 12 AM averages of PBL heights c) 2008 June, July, August at 12 AM averages of PBL heights d) 2008 September, October, November at 12 AM averages of PBL heights......................... 35
Figure 3.22:a) 2008 January and Februrary at 18 PM averages of PBL heights b) 2008 March, April, May at 18 PM averages of PBL heights c) 2008 June, July, August at 18 PM averages of PBL heights d) 2008 September, October, November at 18 PM averages of PBL heights 35
Figure 3.23: Afyon 2008 PM10 concentrations and PBL heights (m) boxplot 36
Figure 3.24: Eskişehir 2008 PM10 concentrations and PBL heights (m) boxplot..... 37
Figure 3.25: Afyon PM10 concentrations distributions vs PBL heights levels (m) .. 38
Figure 3.26:İzmir PM10 concentrations distributions vs PBL heights levels (m).... 39

SPATIO-TEMPORAL ANALYSIS OF PM_{10} CONCENTRATIONS OF TURKEY

SUMMARY

Particulate matter pollution is one of the major concerns in the developing countries, due to its harmful effect on human health. In a study conducted by Atkinson *et al.* (2001), 10 µg increase of PM_{10} concentrations causes %0.6 increase in overall deaths.

PM_{10} measurements from 2008 to 2010 at 118 air quality monitoring stations of Ministry of Environment and Urbanization, were used to determine air pollution levels in Turkey. Out of 81 cities 12 of them have more than one station. The monitoring stations classified as rural, urban, suburban and in this study two urban cities are not used in analysis. Spatial and temporal analysis were conducted to identify clusters of high PM_{10} concentration and to identify possible trends in the data.

Annual average of PM_{10} over three years are 82.3, 76.5, 73.9 µg/m^3 for 2008, 2009 and 2010 respectively; all are above the EU air quality standard value, 40 µg/m^3. Temporal analysis showed that the decreasing of annual means is not explained by a significant decreasing trend. K-means clustering method, performed for spatial analysis, suggested that 118 stations in Turkey can be divided into five groups. The analysis conducted in five different steps. First three steps include the clustering with only one parameter (mean), four different parameters (mean, 2.5 %, median and 97.5 %) and five parameters (50%, 75%, 90%, 95%, 97.5%). Two distance method, Manhattan and Euclidean distances used in K-means algorithm as the last two steps. In addition to these analyses Principal Component Analysis computed with selected four parameters and five parameters. The eastern regions; belong to high polluted cluster with a range of 50 percentiles between 81.8 µg/m^3(Iğdır) to 126.9µg/m^3 (Van), which are more polluted than the industrial and populated western regions with cleanest cluster range of 50 percentiles between 32.6 µg/m^3 (Sinop) to 51.7µg/m^3 (Kırşehir).

In order to understand the basis for the variability in the clusters spatial distributions of TNO emissions inventory, NCEP-NCAR temperature differences at difference pressure level and Planetary Boundary Level (PBL) heights of the 2008 WRF model run outputs were analyzed. Both TNO emissions inventory and NCEP-NCAR based temperature differences estimate do not really explain the variation in the PM_{10} distribution. On the other hand, PBL heights time series plots, boxplots and spatial distribution plots shown that it is an effective parameter to explain the variation of PM_{10} distributions.

This study presented and eplained the findings of the spatio-temporal analysis of PM_{10} levels as well as possible reasons causing the spatial variance.

TÜRKİYEDEKİ PARTİKÜL MADDE KONSANTRASYONUNUN ALANSAL VE MEKANSAL ANALİZİ

ÖZET

Partikül madde kirliliği insan sağlığı üzerindeki zararlı etkisi nedeniyle, gelişmekte olan ülkelerde en önemli sorunlardan biri olmuştur. Atkinson ve diğerlerinin (2001) çalışmasına gore, PM_{10} konsantrasyonundaki 10 µg'lık artış genel ölümlerde %0.6'lık bir artışa sebep olmaktadır.

Türkiyedeki hava kirliliği seviyesini belirleyebilmek için, Çevre ve Şehircilik Bakanlığı'na ait 118 hava kalitesi izleme istasyonunun 2008-2010 yılları arasındaki PM_{10} ölçümleri kullanılmıştır. 81 il arasında 12 ilde birden fazla istasyon bulunmaktadır. Bu ölçüm istasyonları kırsal, kentsel ve yarıkentsel olarak sınıflandırılıp bu çalışmada 2 tane kırsal istasyon analizlerde kullanılmamıştır. Mekansal ve zamansal analizler ile yüksek PM_{10} konsantrasyonlarının sınıflandırılması ve olası eğilimler incelenmiştir.

Yıllık PM_{10} ortalamaları 2008, 2009 ve 2010 yılları için sırasıyla 82.3, 76.5, 73.9 µg/m^3 'tür ve bu ölçümlerin hepsi Avrupa Birliği yıllık PM_{10} standardı olan 40 µg/m^3 değerinin üstündedir. Zamansal analizler sonucuna göre yıllık ortalamalardaki düşüş belirgin bir düşüş eğilimi olarak açıklanmamaktadır. Mekansal analiz için kullanılan K-means sınıflandırma method, Türkiye'deki 118 istasyonun 5 gruba ayrılabileceğini göstermiştir. Analiz 5 aşamada gerçekleştirilmiştir. İlk aşamada tek değişken (ortalama değer), ikinci aşamada 4 değişken (ortalama, %2.5, medyanve %97.5), son olarak 5 değişken (50%, 75%, 90%, 95%, 97.5%) kullanılmıştır. Son iki aşama olarakta K-means sınıflandırma metodu uygulanırken iki farklı uzaklık hesaplama yöntemi kullanılmıştır; Manhattan ve Öklid uzaklıkları. Bu analizlere ek olarak, seçilen parametrelere Temel Bileşenler Analizi uygulanmış ve sonuçlar tekrar sınıflandırılmıştır. Doğu Anadolu Bölgesi, yüzde elli değeri 81.8 µg/m^3(Iğdır) ile 126.9µg/m^3 (Van) olan en yüksek PM_{10} konsantrasyonu sınıfında yer almıştır. Bunun yanı sıra daha kalabalık nüfusa ve daha geniş endüstriyel alana sahip olan batı bölgeleri yüzde elli değeri 32.6 µg/m^3 (Sinop) ile 51.7µg/m^3 (Kırşehir) olan en temiz PM_{10} konsantrasyonu sınıfında yer almıştır.

K-means sınıflandırma sonuçlarında ortaya çıkan mekansal dağılımdaki değişkenliği açıklamak için TNO emisyon envanteri, NCEP-NCAR verilerinden elde edilen farklı basınç seviyelerindeki sıcaklık farkı ve 2008 yılı için WRF modeli çalıştırılarak elde edilen PBL değerleri analiz edilmiştir. TNO emisyon envanteri ve NCEP-NCAR verilerinin analizi sonucunda PM_{10} dağılımındaki değişkenliğe kesin biraçıklama getirilememiştir. Bununla birlikte, PBL yükseklikleri zamansal ve mekansal analizi sonuçları PBL yüksekliklerinin PM_{10} dağılımında etkili bir parameter olduğunu göstermiştir.

Bu çalışma PM_{10} seviyelerinin mekansal ve zamansal analiz sonuçlarını belirtip, mekansal değişikliklerin sebeplerini açıklayacaktır.

1. INTRODUCTION

High air pollution level is the natural result of the rapid population growth, increase in energy requirements with industrial development, as well as increasing in vehicle usage and fossil fuel consumption. All these factors increase emissions of air pollutants to the atmosphere. Other external factors such as wind speed, atmospheric stability, landscape and topography provides the conditions where emissions turn into high air pollution concentrations. Among criteria air pollutants (i.e., carbon monoxide, ozone, nitrogen oxides, sulfur dioxide, particulate matter and lead) particulate matter pollution is one of the most critical ones.

Particulate matter (PM) is a mixture of particles, either solid or liquid, which are suspended in the air. They have a wide range of sizes. Generally, in terms of size particles are classified by their aerodynamic diameter. The aerodynamic diameter corresponds to the size of a unit-density sphere with the same aerodynamic characteristics as the particle of interest. Coarse particles' (PM_{10}) aerodynamic diameter range is 2.5 to 10 μm, for fine particles it is 0.1 to 2.5 μm. The illustration of different PM sizes are provided in Figure 1. As seen in Figure, even PM_{10} is much much smaller as compared to human hair.

The size of the particles defines the suspended time of the particles in the air. While $PM_{2.5}$ can stay on week or month based in the atmosphere, PM_{10} can be removed by precipitation in hours (WHO, 2005).

PM in urban and rural environments is a complex mixture with components having diverse chemical and physical characteristics. Research on PM exposure and risk are complicated by this heterogeneity, as well as the potential of particles to cause injury varies with size and other physical characteristics, chemical composition. By definition, PM_{10} includes $PM_{2.5}$ and thoracic coarse mass PM (the difference between PM_{10} and $PM_{2.5}$ is often referred to as "coarse" mass PM). PM_{10} includes those inhalable particles that are sufficiently small to penetrate to the thoracic region; the fine fraction of PM_{10} is cut off from the coarse fraction at 2.5 μm in aerodynamic diameter ($PM_{2.5}$), a size fraction with a high probability of deposition in the smaller conducting airways and alveoli.

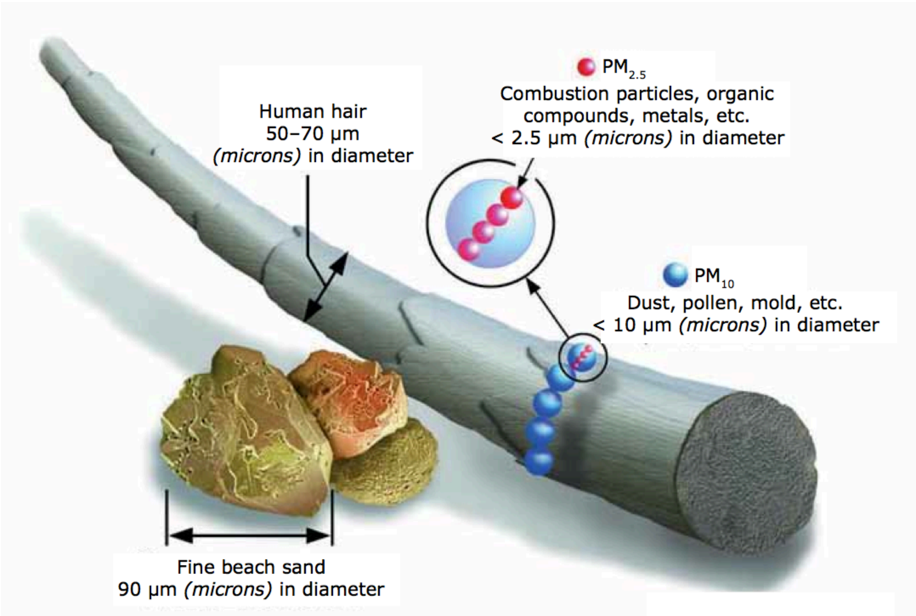


Figure 1.1: Illustration of PM coarse and fine fractions (EPA, 2010)

Many study results show that particulate matter pollution cause various diseases on short-term and long-term exposure. These studies are summarized in the following section.

Health Effects

As reviewed by Pope and Dockery (2006), Simkhovich *et al*. (2008) and Ren and Tong (2008), there are over 100 published articles that reports results on short-term exposure to air pollution and mortality. In a study conducted by Poloniecki *et al*. (1997), over 370,000 emergency cardiovascular admissions in London hospitals were analyzed between April 1987 and March 1994. They have found positive correlations between acute myocardial infarction (AMI) and black smoke and air pollutant gases (NO_2, CO and SO_2) and between angina and black smoke. The authors suggested that 1 in 50 heart attacks in London hospitals are triggered by air pollution. In another study, Ruidavets *et al*. (2005) found that short-term exposure to ozone (i.e., 1 to 2 days) is related to AMI events in middle-aged adults without heart disease. Nawrot and Nemery (2007), support these findings with their own study, which found that air pollution (especially pollution from traffic) ranks four in their list of environmental triggers.

APHEA-2 (Air Pollution and Health: a European Approach 2) (Atkinson *et al.*, 2001) study focused on the impact of increased particulate matter (PM) levels on daily mortality and hospital admissions for asthma and chronic obstructive pulmonary disease (COPD). APHEA-2 daily mortality studies were conducted in 29 European cities, covering over 43 million people for more than 5 years in the 1990s. The results showed that all-cause daily mortality increased by 0.6 percent for 10 $\mu g/m^3$ increase in PM_{10}. APHEA-2 hospital admission study was conducted in 8 European cities, covering 38 million people. Hospital admissions for asthma and COPD were observed to be increased by 1 percent per 10 $\mu g/m^3$ increase in PM_{10} among older people (65+) (Katsouyanni *et al.*, 2001). In other studies, the range for increase in all-cause daily mortality is between 0.6 and 1.2 percent per 10$\mu g/m^3$ increase in PM (Pope and Dockery, 2006).

Although long-term effect studies are not as numerous as the short-term effect studies, there are over 30 publications on this subject. As summarized by Pope and Dockery (2006), the range for all-cause mortality rates are between 1 and 17 percent per 10$\mu g/m^3$ increase in $PM_{2.5}$. For cardiopulmonary mortality rates this range is between 5 and 42 percent and for lung cancer it is between 0.8 and 81 percent.

In other studies, relation between air pollutants and reduced growth in children were analyzed. Guaderman *et al.* (2000) found that fourth graders who are exposed to PM, NO_2 and inorganic acid vapors, showed significant reduction in growth of lung function. Deficits were found to be higher for children spending more time outdoors. In a study conducted by Avol *et al.*, (2001), children who relocated to areas of lower PM_{10} showed increased growth in lung function whereas children who live in areas with high PM_{10} show decreased growth in lung function. The authors concluded that changes in air pollution exposure during growth years have a significant impact on lung function growth and performance. In another study, Perera *et al.* (2009), monitored children from birth till 5 years of age and showed that children in high exposure group had full-scale and verbal IQ scores that were 4.31 and 4.67 points lower, respectively, than those of less-exposed children.

Lim et al. (2012) conducted a research to quantify the disease burden and found that worldwide 3.1 million deaths and almost 3 percent of Disability Adjusted Life Years (DALYs) could be attributed to exposure to ambient $PM_{2.5}$. In western, central and eastern Europe 430 000 premature deaths and over 7 million years of DALYs were attributed to $PM_{2.5}$ pollution (EEA, 2013).

Regulations of Particulate Matter

For current regulatory purposes, PM has been classified by aerodynamic diameter, as size is a critical determinant of the likelihood and site of deposition within the respiratory tract and evidence has become available on the risk associated with specific size groups. Initially, regulations and guidelines were directed at very general measures of PM concentration, including total suspended particulate (TSP) matter in the United States and black smoke (BS) in Europe. In 1987, USEPA promulgated a standard for PM less than 10 μm in aerodynamic diameter (PM_{10}) and then, in 1997, a standard for PM less than 2.5 μm in aerodynamic diameter (PM2.5) was added. In WHO's 2000 air quality guidelines, guidance was given in relation to both of these PM indicators. Table 1.1 list the air quality standards set by European Union, World Health Organization and Turkish Ministry of Environment and Urbanization. As seen in the table, Turkish standards are in alignment with EU, both has daily standards of 50 μg/m^3 and annual value of 40 μg/m^3. Although WHO daily standard is similar, it has a lower annual value of 20 μg/m^3. It should be noted that for both EU and Turkish Ministry the daily standard is allowed to be exceed 35 times annually.

Table 1.1: PM_{10} standards

Company/Organization	Avg Period	Limit Value	Permitted exceedences each year
European Union	Daily	50 μg/m3	35 times
	Annual	40 μg/m3	
WHO	Daily	50 μg/m3	
	Annual	20 μg/m3	
Turkish Ministry of Environment and Urbanization	Daily	50 μg/m3	35 times
	Annual	40 μg/m3	

The largest particles, those in the coarse mode, are to a large extent mechanically produced by the break-up of larger solid particles. The amount of energy required to break these particles into smaller sizes increases as the size decreases. Biological sources may also contribute to this mode. Thus, in urban areas, the coarse particles typically contain resuspended dust from roads and industrial activities, and biological material such as pollen grains and bacterial fragments. The coarse particles also typically include the earth's crustal materials such as wind-blown dust from agricultural processes, uncovered soil, unpaved roads or mining operations. Traffic produces road dust and air turbulence that can re-entrain road dust near roadways. Near coasts, evaporation of sea spray can also produce large particles. Coarse particles may also be formed from the release of non-combustible materials in combustion processes, i.e. fly ash.

Smaller particles, those smaller than 2.5 μm in aerodynamic diameter and known as the fine mode are largely formed from gases, but combustion processes may also generate primary particles in this size range. Typically these particles are formed as ultrafine particles in two mode. The nucleation mode has particles with diameters up to about 0.1 μm and the Aitken mode has a size range from about 0.1μm to 1μm diameter. These particles produced by nucleation–condensation of low-vapour-pressure substances by product of high-temperature vaporization or chemical reactions in the atmosphere. Particles in this nucleation range (up to 0.1 μm) subsequently grow by coagulation (the combination of two or more particles to form a larger particle) or by condensation of gas or vapour molecules on the surface of existing particles. Coagulation is most efficient for large numbers of particles, while condensation is most efficient for large surface areas. Thus the efficiency of both coagulation and condensation decreases as particle size increases, and this decreasing efficiency effectively results in an upper limit of approximately 1 μm that cannot be exceeded by particle growth other than by hygroscopic growth in humid atmospheres.

In terms of their formation mechanism, particles divided into two groups; primary and secondary particles. Primary particles are directly emitted into the atmosphere by anthropogenic activity or natural sources, secondary particles are those formed in the air by chemical reactions like sulphur dioxide, nitrogen oxides and volatile organic compounds (VOCs) (Seinfeld and Pandis, 1998).

There are numerous sources of particles related to human activities as well as natural sources. By measuring the temporal and spatial patterns of the chemical composition of particles in the air and combining this information with meteorology, the particle mass can be apportioned to various sources. Methods for estimating the contributions of specific sources to PM concentrations have been reviewed in detail by Brook *et al*, (2004), and Watson *et al*, (2002). According to sources of particulate matter may be of natural or anthropogenic origin. The main natural resources are the sea sprays, terrestrial volcanoes and desert dust. Natural PM are in both coarse and fine mode in general. Anthropogenic sources can be listed as combustion processes, transport, agriculture and mining activities. Anthropogenic PM are generally in coarse mode.

Turkey is one of the largest and fastest developing countries in Europe and Middle East. Emission of industrial, traffic, agriculture activities as well as topographical conditions, meteorological factors and natural sources are the main causes of the particulate pollution in Turkey. There are several studies on air quality in Turkey that focuses on particulate matter pollution sources. In a study conducted by Kindap et al., (2006), particulate matter transport from Eastern European countries to Turkey was examined via meteorological and air quality model. High PM_{10} winter episodes over Istanbul in 2002 was found to be related to long-range transport of particles from Eastern Europe were investigated. Im *et al.,* (2010), analyzed PM_{10} level in Istanbul with Weather Research and Forcasting and Community Multiscale Air Quality models using a regional emissions inventory. In another study conducted by Tayanc (2000), increase in SO_2 concentrations during 1980s and 1990s were identified to be caused by fossil fuel heating and industrial emissions from 1980s to 1990s.

In this study, the main objectives are: i) to determine level of PM_{10} in Turkey using three years data (2008-2010), ii) to develop temporal analysis of PM_{10} levels, and iii) to develop a spatial analysis to understand if there are clusters of high PM_{10} levels. For this purpose, we have utilized the air quality data provided by the Turkish Ministry of Environment and Urbanization. Most widely used non-hierarchical clustering method, K-means, used for spatial analysis. Regression with time series has been used for temporal analysis. Emissions inventory data as well as meteorological parameters (i.e. PBL height) is used to explain spatial variability in th dataset.

2. METHODOLOGY

2.1 Study Area

Turkey is one of the largest countries in Europe and Middle East with its area that is covering approximately 780000 km^2. It is between latitudes 36 ° and 42° and longitudes 26 ° and 45° as a crossroad between Asia and Europe. There are three seas that encircle the country that are the Aegean Sea on the west, Black Sea on the north and the Mediterranean Sea on the south and it also contains Sea of Marmara is located on the northwest.

Turkish Statistical Institute (TUIK), based on the population of Turkey is approximately 74 millions estimated that Address Based Population Registration System Results (ABPRS) of 2010 report. The distribution of the population is related to the topographic conditions as well as employment situation. The population increases from east to west and it is higher in the coastal areas due to mild climatic conditions and developed employment opportunities. The only exception is Ankara which is the second largest city and the capital of Turkey. Population is heavily concentrated in 16 metropolitan municipalities, as Istanbul Metropolitan Municipality population is 18% of the total population, Ankara Metropolitan Municipality is 6.5% and Izmır Metropolitan Municipality is 5.4% (ABPRS, 2011). All these can be seen in Figure 2.1 population density plot of Turkey.

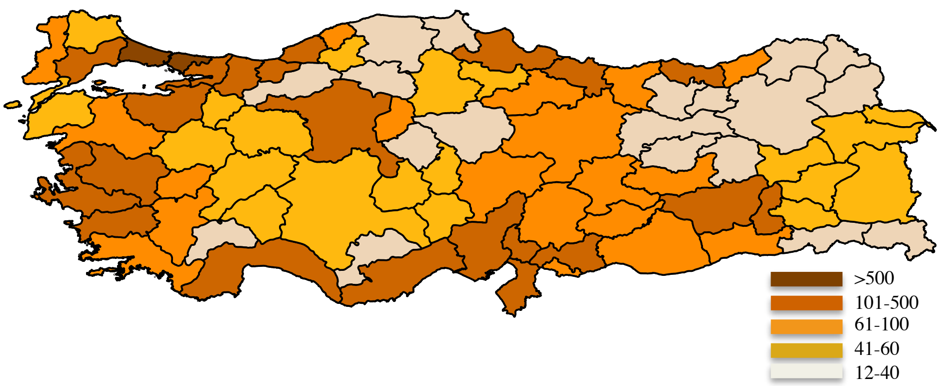


Figure 2.1: Population Density (person per km^2) of Turkey in 2012 (TUİK)

Figure 2.2 provided annual average growth rate of population for the period between 2007-2012. In general, the observations for population density are similar for average growth rate. However some cities, (Gaziantep, Şanlıurfa, Hakkari etc.) have higher growth rates although their 2012 population density is low.

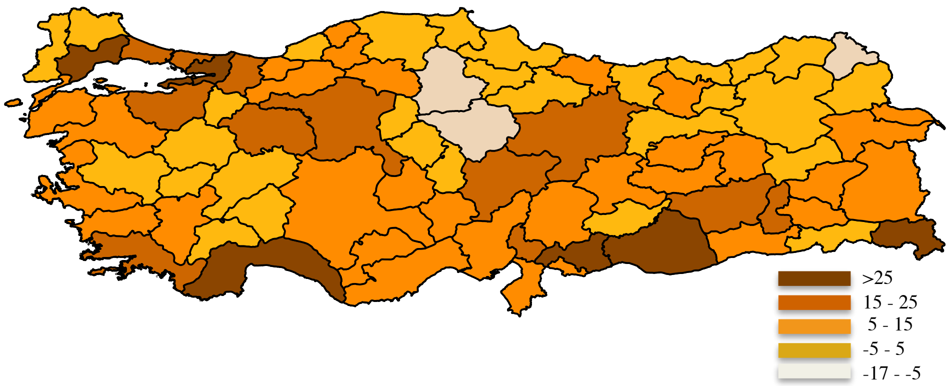


Figure 2.2 : Annual average growth rate (2007-2012) (TUİK)

2.2 Air Quality Monitoring Stations

The air quality network operated by the Turkish Ministry of Environment and Urbanization has over 118 sites throughout Turkey (since 2008), which are used to monitor air pollution. Out of 81 cities, 12 of them have more than one station. Marmara region (Northwestern part of Turkey) hosts 22 stations of which 10 of them are located in Istanbul and Kocaeli have 3 air quality monitoring stations. In the Aegean region (Western part of Turkey) there are 15 stations where 6 of them are located in İzmir and Denizli and Mugla have 2 stations each. Mediterrenean region (Southern part of Turkey) has 13 stations where only Adana and Kahramanmaras has multiple stations (4 and 2 respectively). Central Anatolia (Inner Anatolia) has 25 stations with Ankara having 10 stations and Kayseri and Konya have 3 and 2 stations. Black Sea region (Northern part of Turkey) has 20 stations where Trabzon and Samsun have 2 stations each. Eastern Anatolia has 14 and Southeastern Anatolia has 8 stations. In these regions every city has only one station. Table 2.1 shows that regional distribution of the stations. The location of the 118 monitoring stations are shown in Figure 2.3.

Table 2.1 : Air quality monitoring stations list of each region and the cities have more than one stations.

Regions	**City Name**	**Number of cities**	**Number of stations**
Marmara		**11**	**22**
	Istanbul		10
	Kocaeli		3
Agean		**8**	**16**
	Izmır		7
	Denizli		2
	Mugla		2
Black Sea		**18**	**20**
	Trabzon		2
	Samsun		2
Mediterrenean		**9**	**13**
	Adana		4
	Kahramanmaras		2
Central Anatolia		**13**	**25**
	Ankara		10
	Konya		2
	Kayseri		3
Eastern Anatolia		**14**	**14**
South Eastern Antolia		**8**	**8**
TURKEY		**81**	**118**

118 Air quality monitoring stations were classified by types, rural, urban and suburban. 2 stations in Adana, Catalan and Doğankent stations, are rural and suburban type, respectively. One station Ciğli in İzmir is a rural type station and 2 rural type stations Giresun and Artvin seperated from dataset.

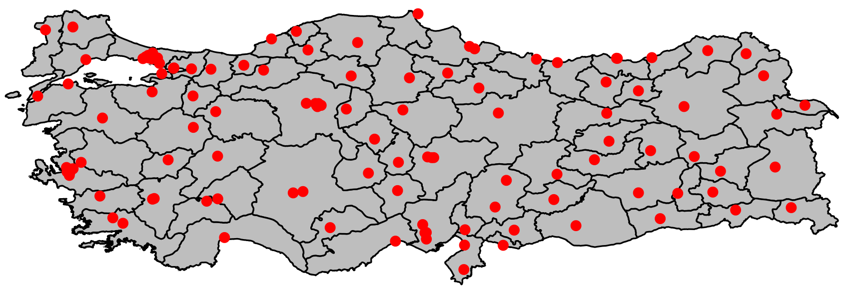


Figure 2.3: Location of air quality monitoring stations in Turkey

2.3 Air Quality Data Processing

2.3.1 Air quality monitoring stations data

The observations of station in cities 79 urban stations used in this study are hourly PM_{10} concentrations as in $\mu g/m^3$ between January 2008 and December 2010. Daily averages of PM_{10} concentration at 79 cities were calculated from hourly data over 3 years period by using scripts which are developed in R Programming. First step is to merge more than one station in the same city. For the cities that have multiple stations, average values were estimated. For example, there are 10 stations in Istanbul; the mean values of the 10 stations at each hour computed and then they were used for the data analysis as a single station. Data converted from 26304 rows by 118 columns matrix to 26304 rows 79 columns matrix. Each column of matrix refers the 79 cities of the Turkey. Second step is the quality check. If there are measurements at least 8 hours of a day (or % 30), data were used to estimate daily values. Otherwise the daily value is left as blank. The R programming used for this analysis is given in Appendix. Air quality monitoring stations process diagram is summurized in Figure 2.4.

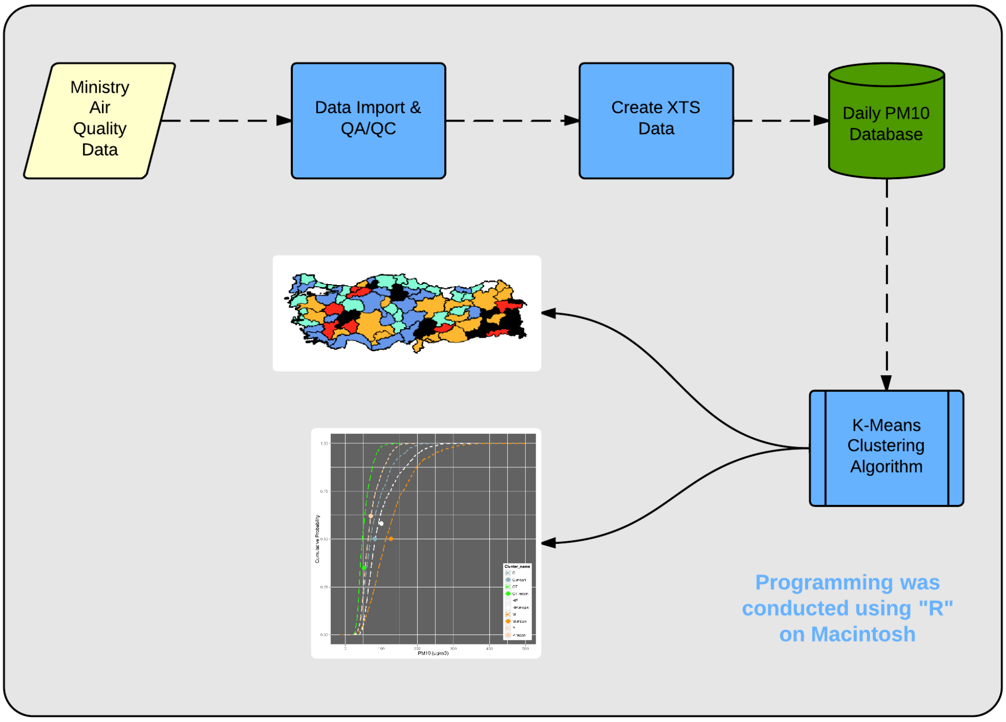


Figure 2.4: Air quality data processing diagram.

2.3.2 Emissons data

TNO/MACC-2005, is an emissions inventory with a resolution of 1/8x 1/16, has been prepared by TNO (http://www.tno.nl) to be used in the UBA PAREST project (http://www.parest.de/). The TNO/MACC-2005 inventory was built using the official reported emissions, downloaded from the European Environment Agency (EEA, 2008), but various consistency checks were included to spot gaps and/or errors in reported emissions by individual countries. Alternative data from the IIASA RAINS model (http://www.iiasa.ac.at/~rains) or TNO defaults were used to fill the gaps. The emissions in the TNO/MACC-2005 inventory are split between point sources (e.g., power plants, refineries, and major industries) and area sources (e.g. road transport and residential combustion). The complete list is given in Table 2.2. For the point sources a new highly detailed database was compiled whereas for the area sources new geographical distribution maps were compiled. This emission inventory covers Europe, including Turkey and partially Russia.

TNO inventory provide annual anthropogenic emissions of seven pollutants (CO, $NH_{3,}$ NMVOC, NOx, SOx, $PM_{2.5}$ and PM_{10} and for a number of anthropogenic activities aggregated into 10 SNAP97 source categories.

Table 2.2: TNO emssions inventory SNAP categories

SNAP	Description
S1	Combustion in energy and transformation industries (POW)
S2	Non-industrial combustion plant (RES)
S3	Combustion in manufacturing industry (IND)
S4	Production Processes (PRO)
S5	Extraction and distribution of fossil fuels and geothermal energy (FFE)
S6	Solvent and other product use (SOL)
S7	Road transport (ROAD)
S8	Other mobile sources and machinery (MOB)
S9	Waste treatment and disposal (WAS)
S10	Agriculture (AGR)

2.3.3 Meteorology data

NCEP/NCAR Data

The NCEP/NCAR Reanalysis data set is a contuniously updated gridded data set representing the state of the Earth's atmosphere, incorporating observations and numerical weather prediction (NWP) model output dating back to 1948. It is a joint product from the National Centers for Environmental Prediction (NCEP) and the National Center for Atmospheric Research (NCAR). The data is available for free download from the NOAA Earth System Research Laboratory and NCEP (NCAR, 2010). It is distributed in Netcdf and GRIB files. NCEP/NCAR Reanalysis 1 project is using advanced analysis/forecast system to perform data assimilation using past data from 1948 to the present. In this study, NCEP/NCAR Reanalysis 1 air temperature dataset is analyzed with respect to pressure level. 1000 mb, 925 mb, 850 mb, 700 mb temperatures choosen to determine the inversion days with 2.5^0 x 2.5^0 resolution grids. NCEP/NCAR Reanalysis 1 spatial coverage is 0.0 East to 357.5 East, 90.0 North to 90.0 South. Figure 2.5 showed the study domain, colors defines the each grid cell. (ESRL, 2013)

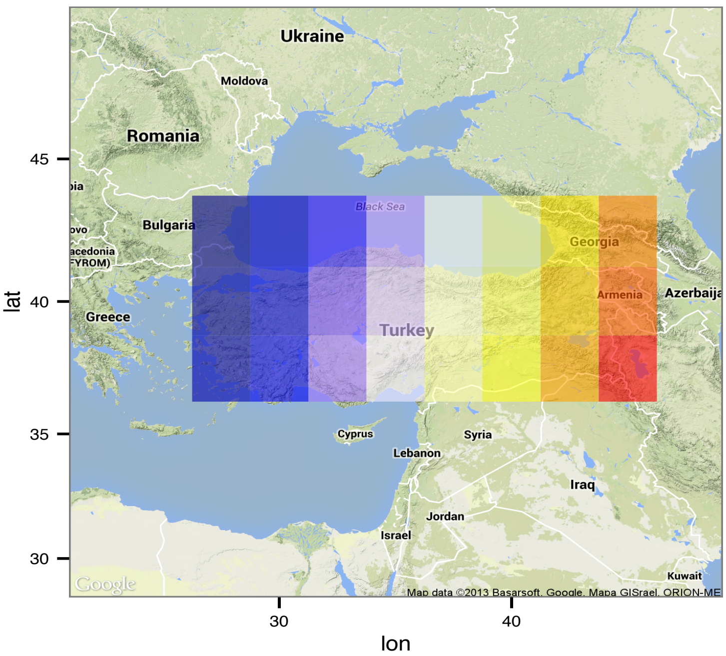


Figure 2.5: NCEP-NCAR Reanalysis data grid cells.

Weather Research and Forecasting Data

The Weather Research and Forecasting (WRF) Model is the next generation of the regional mesoscale model (MM5). WRF is a set of software, which is produced from National Center for Atmospheric Research (NCAR), the National Oceanic and Atmospheric Administration (represented by the National Centers for Environmental Prediction (NCEP) and the Forecast Systems Laboratory (FSL)), the Air Force Weather Agency (AFWA), the Naval Research Laboratory, the University of Oklahoma, and the Federal Aviation Administration (FAA) collabration, for numerical weather prediction method. It is open source, synoptic and creating climate projections. WRF involves two computational cores that are known as WRF-ARW (Advanced Research WRF) and WRF-NMM (Nonhydrostatic Mesoscale Model) for solving atmospheric differential equations. Model resolution changes meters to thousands of kilometers. Researchers may use real data (observations) or ideal case data to create simulations. The model uses 3rd order Runge Kutta time integration scheme and also offers one-way, two-way, and moving nest options (NCAR, 2010)

Microphysics parameters include schemes that contains various forms of water groups like vapor, cloud water, cloud ice, rain, snow and hail. The options suitable for inner domain (10 km) and outer domain (30km) are choosen.

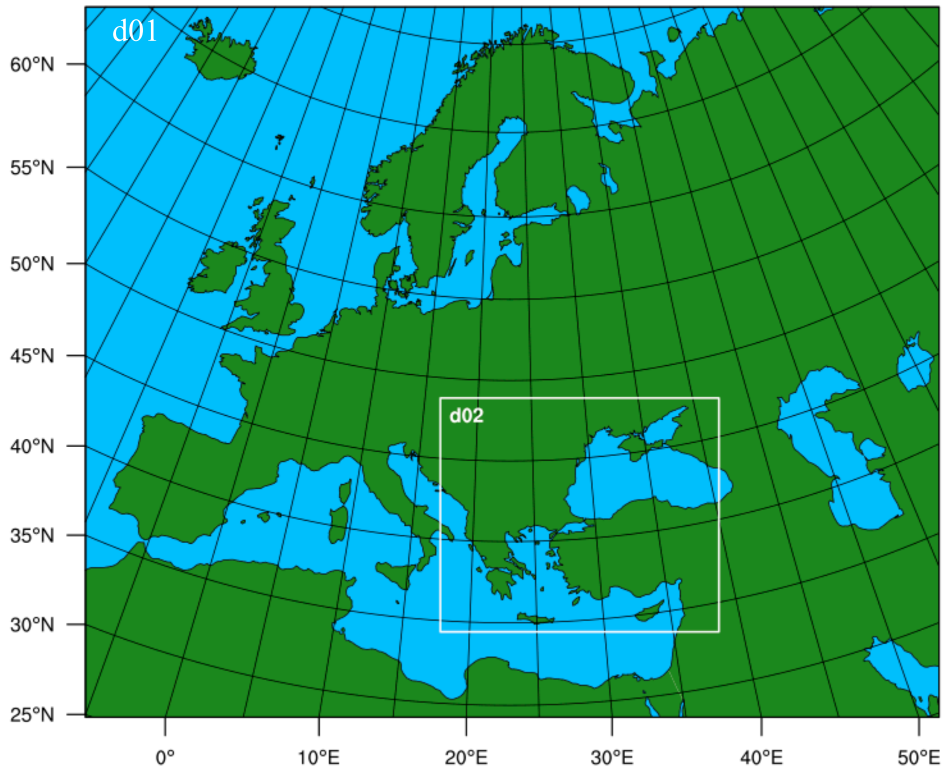


Figure 2.6:WRF model outer (d01) and inner (d02) domain

In this study, the chosen mother domain and nested domain are shown in Figure 2.6. The mother domain covers the North Africa, Europe and West Asia with a resolution of 30 km. The nested domain, which is the area under analysis, involves Mediterranean countries with resolution of 10 km. NCEP FNL (Final) Operational Global Analysis data was used as input for initial conditions to the model. These data are on 1.0x1.0 degree grids prepared operationally every six hours. The FNLs are made with the same model which NCEP uses in the Global Forecast System (GFS), but the FNLs are prepared about an hour or so after the GFS is initialized (NCEP, 2000). In this model, GRIB1 6 HOURLY FILES begin 1999.07.30, were used and the output has been generated from the model at every 60 minutes.The Planetary Boundary Layer height hourly outputs utilized as an indication of presence of inversions.

2.4 Analysis Tools

Clustering Analysis

Spatial analysis of the PM_{10} data were conducted using K-means method. K-means clustering method is the most widely used nonhierarchical clustering method to grouping the datasets. K-Means clustering method is computationally faster with a large number of variables if cluster numbers are small. The method divides the data in to a number of clusters, which is defined ahead of the analysis. Clusters must be determined iteratively with an initial guess (xi=2,3,….n). The centroids, mean values of the vector, computed of each cluster for the first guess. For each data point Euclidian distance (Eq 3.1) or Manhattan distance (Eq 3.2) is calculated from the data point to each cluster. If the data point is closest to its own cluster, the algorithm leaves it where it is. If the data point is not closest to its own cluster, it is moved into the closest cluster. The algorithm repeated until a complete pass through all the data points results in no data point moving from one cluster to another. At this point the clusters are stable and the clustering process ends.(Wilks, 2006).

$$\| \mathrm{x} - \mathrm{y} \| = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \qquad \text{Euclidean Distance} \qquad (2.1)$$

$$\| \mathrm{x} - \mathrm{y} \| = \sum_{i}^{n} |x_i - y_i| \qquad \text{Manhattan Distance} \qquad (2.2)$$

Principal Component Analysis (PCA) is one of the most frequently used statistical method in exploratory data analysis and for predictive linear model development. PCA reduces a data set containing a large number of variables to a data set containing fewer new compact variables. These new variables are linear combinations of the original ones, and these linear combinations are chosen to represent the maximum possible fraction of the variability contained in the original data. The new compact components are linear combinations of the input variables, where the first component maximizes the variance captured, and with each subsequent factor capturing as much of the residual variance as possible, while taking on an uncorrelated direction in space(Boslaugh & Watters, 2008).

3. RESULTS

Analysis were started with determination of PM_{10} levels in Turkey. Daily PM_{10} averages of 79 cities for each year used for the analysis. It is estimated that annual averages of PM_{10} in Turkey is 82.3, 76.5, 73.9 $\mu g/m^3$ for 2008, 2009 and 2010 respectively. It should be noted that all these values are above the EU limit value of 40 $\mu g/m^3$. Summary statistics (minimum, 1st quartile, median, mean, 3rd quartile and maximum) of PM_{10} levels in Turkey for three years period, 2008 to 2010 are given in Table 3.1.

Table 3.1: Summary statistics of PM_{10} levels in Turkey between 2008 -2010.

	2008($\mu g/m^3$)	**2009**($\mu g/m^3$)	**2010**($\mu g/m^3$)
Minimum	30.8	21.9	30.6
2.5th Percentile	57.8	51.9	50.4
Median	71.9	64.1	64.9
Mean	81.7	73.4	72.7
75th Percentile	99.9	89.5	86.9
Maximum	171.5	181.9	168.9

PM_{10} levels of Turkey shown in Figure 3.1 with a box-whisker plot. The Box-whisker plot indicates the mean (red points inside the box), 95 percent confidence bounds for the mean (short black lines on the red points inside the box), the median (the bold black line inside the box), the lower and upper quartiles of the data set (25th and 75th percentiles which is shown by the lower and upper ends of the box), and extreme values (top and bottom lines). The straight red line in the plot shows the EU PM_{10} annual limit value. Plot shows that more than 75 % of the data are above the PM_{10} standards.

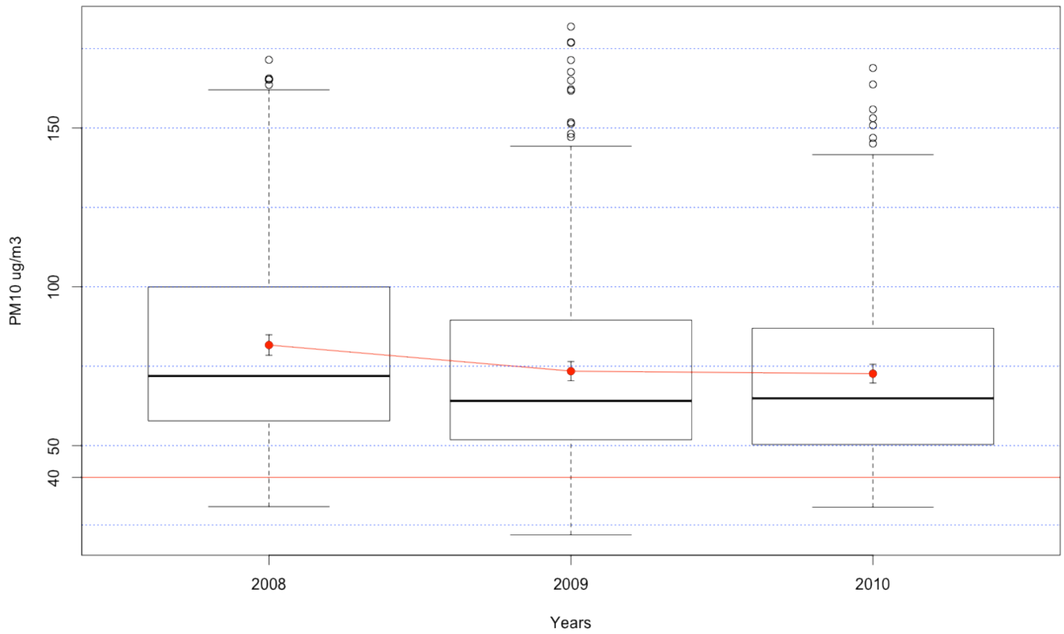


Figure 3.1: Summary of PM_{10} levels in Turkey of 3 years

Although the annual averages suggest a decreasing trend, in order to fully answer this question we have conducted a temporal analysis. For this purpose, we have utilized daily PM_{10} data averaged over 118 stations. Figure 3.2 presents these daily data. As seen in the Figure, daily averages ranges between 22 and and 182 µg/m^3. It should be noted that only 25 percent of the data have PM_{10} values lower than EU daily PM_{10} standard 50 µg/m^3, and 75 percent of the data are less than 90 µg/m^3. As expected, there is a strong seasonality in the data as highest values occur during winter period and the lowest values occur during summer. We have conducted a time series regression using time values as the independent variable. We have also included sine and cosine cycles to capture the seasonality (Blue line in Figure 3.2 presents the result of this time series regression). We should note that regression equation has a slope almost zero with a r-square value less than 20 percent. Based on this analysis it is concluded that there is no increasing or decreasing trend in the daily PM_{10} values.

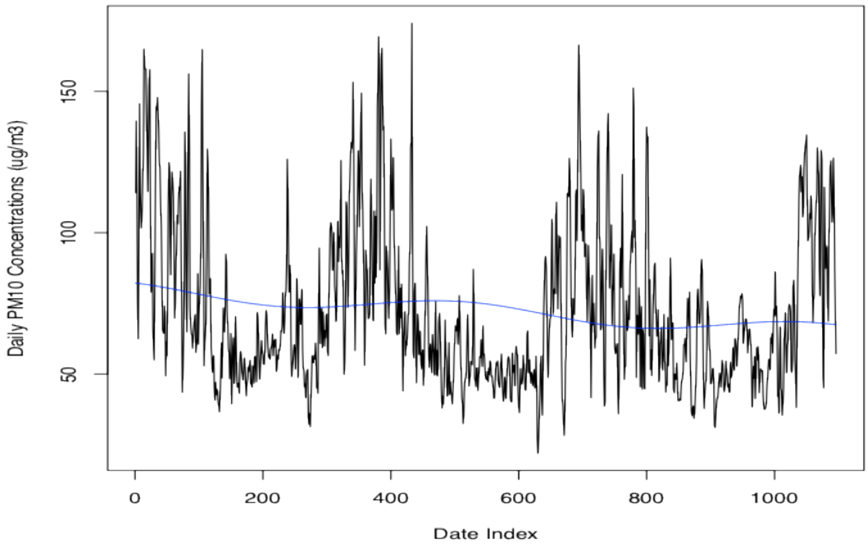


Figure 3.2 : Time series plot of three years data

Spatial analysis of the PM_{10} data were conducted using K-means method by two different distance, Manhattan and Euclidean distance, calculation. In this analysis, optimum number of cluster was identified by Bayesian Information Criterion (BIC) method, where within-cluster sum of squares is reduced for the given number of cluster. The result of the BIC suggested that 5 is the optimum number (Figure 3.3).

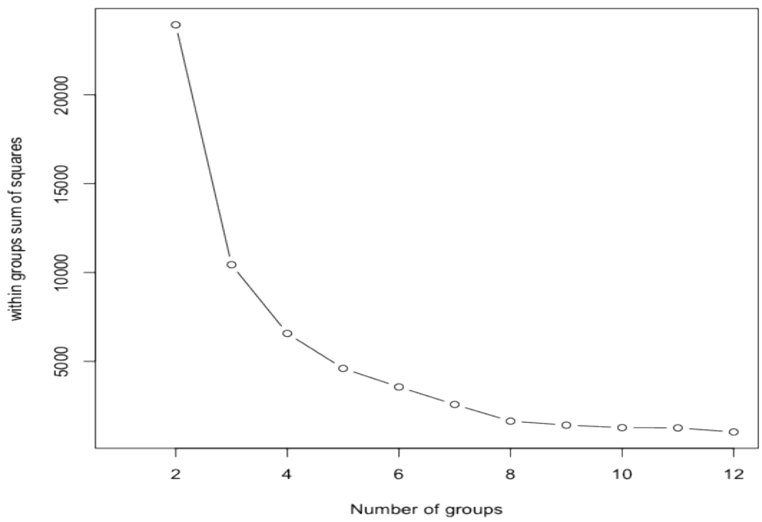


Figure 3.3 : Bayesian information criterian plot for three years.

In the first step we have utilized mean values as the main parameter and the Euclidean distance is used for the distance calculation. Three years K- means clusters for the mean values is shown in Figure 3.4. The colors, black, red, orange, blue, and aqua respectively, represent the highest PM_{10} values of the cluster to lowest PM_{10} values. White color shows the two removed cities Giresun and Artvin. The highest centroid in 5 cluster is 117.2 μg/m^3 with cluster range 106.5 to 135.4 μg/m^3, represented by the black color and called as highly polluted cluster and the lowest centroid is 48.1 μg/m^{3}with cluster range 36.5 to 51.9 μg/m^3, represented by the aqua and called as cleanest cluster. Four of the clusters are above the EU PM_{10} limit value, only one city in the cleanest cluster; Sinop with 36.5 μg/m^3 in the Black Sea region shown with the aqua color below the limit. As it is clear from Figure 3.4, there is not a homogen distribution of the cities to the clusters. East and South East Anatolia region have the highest PM_{10} level dominantly, 7 (Hakkari, Van, Mardin, Siirt, Batman, Muş and Iğdır) of 11 cities in high polluted cluster are in these regions. The other 4 citieas are, Çorum and Karabük in the Black Sea region have 109.6 μg/m^3 and 108.6 μg/m^3 respectively, Kahramanmaraş in the Mediterranean region has 106.4 μg/m^3, Afyon in the Aegean Region has 108.3 μg/m^3 PM_{10} concentrations. Eight cities are clustered to the polluted cluster (red colored) around 80 μg/m^3 mean value with cluster range 88.4 μg/m^3 and 97.7 μg/m^3. Three of them, Adıyaman, Diyarbakır and Şırnak are in the Southeasth and East Anatolia Region, Denizli and Kütahya are in the Agean Region, Düzce and Bolu are in the Western Black Sea Region and Karaman is in the Central Anatolia Region in polluted cluster. Tunceli and Malatya are only in the cleanest cluster cities with 50.2 μg/m^3 and 52.7 μg/m^3 mean values respectively and Bingöl is the only city in the clean cluster, with 62.3 μg/m^3 mean value. Central Anatolia, Aegean and Mediteranean Region of the country are in the medium and clean clusters. The coastal area of the Black Sea Region is in the clean and cleanest cluster dominantly. The megacities, Istanbul with 54.6 μg/m^3 and Izmir with 52.5 μg/m^3 in the cleanest cluster and Ankara in the clean cluster with 67.3μg/m^3 PM_{10} concentrations.

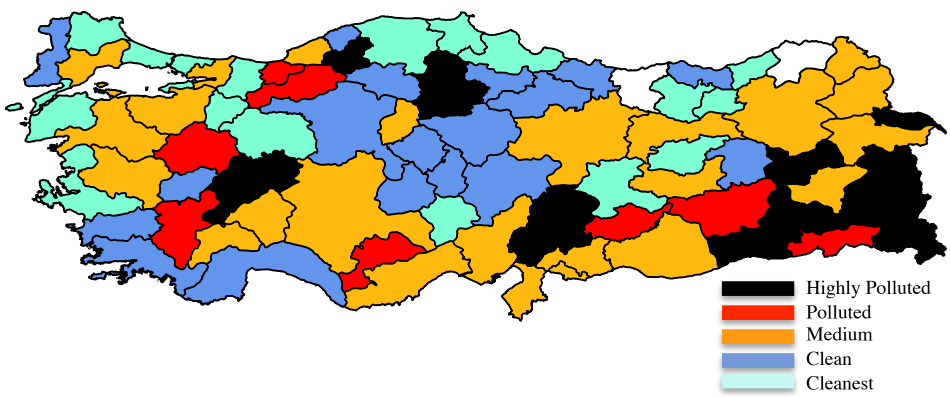


Figure 3.4 : K-means clustering with Euclidean distance colored map of three years mean values in Turkey.

K-means were utilized with Manhattan distance to understand the diffreneces between two distance calculation results. Three years K-means clusters for the mean values is shown in Figure 3.5. The high polluted cluster is 117.2 μg/m^3 with cluster range 106.5 to 135.4 μg/m^3 similar as the Euclidean distance result, represented by the black color and the lowest centroid is 48.1 μg/m^{3}with cluster range 36.5 to 54.6μg/m^3, represented by the aqua. All clusters are above the EU PM_{10} limit value, only one city in the fifth cluster; Sinop with 36.5 μg/m^3 in the Black Sea region shown below the limit. Homogenity were not be provided by euclidean distance calculations. East and South East Anatolia region have the highest PM_{10} level dominantly, Çorum and Karabük in the Black Sea region, Kahramanmaraş in the Mediterranean region, Afyon in the Aegean Region are in the highest clustered cities as Manhattan distance results.

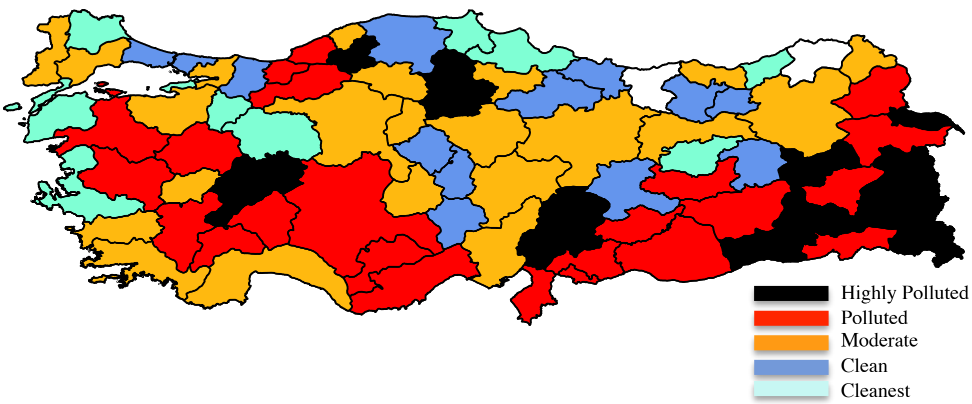


Figure 3.5 : Manhattan distance K-means clustering colored map of three years with only mean.

Difference plot of Euclidean distance and Manhattan distance K-means clustering presented in Figure 3.6 to show cluster changes of the cities. There are similarities and differences between two clustering results. Black color is represent the 2 cluster change, red color is 1 cluster change to high polluted cluster, orange is represent no change in clusters, blue and aqua represent the 2 and 1 cluster change to clean, respectively. As seen in the figure clusters increase or decrease only one cluster. İzmir is in the cleanest cluster as Euclidean distance K-means result with only mean, İstanbul is increased from cleanest cluster to clean cluster and Ankara is increased from clean cluster to medium cluster. The high polluted cluster cities did not change. Besides Hatay and Osmaniye in the Mediteranean Region cluster increased from medium cluster to polluted cluster, Aksaray, Kayseri and Çankırı in the middle of the

Central Anatolia Region cluster increased from clean cluster to medium cluster. Sinop, Trabzon and Rize in the Black Sea Region, Van, Hakkari and Iğdır in the East Anatolian Region Tekirdağ, Bursa and Kocaeli in the Marmara Region are some of cities whose clusters did not change. Malatya and Niğde in the Central Anatolia and Kastamonu in the Black Sea Region cluster changed from the cleanest cluster to clean cluster. Results shown that there are not distinct differences between Manhattan and Euclidean distance results. Therefore, analysis continue with Euclidean distance K-means.

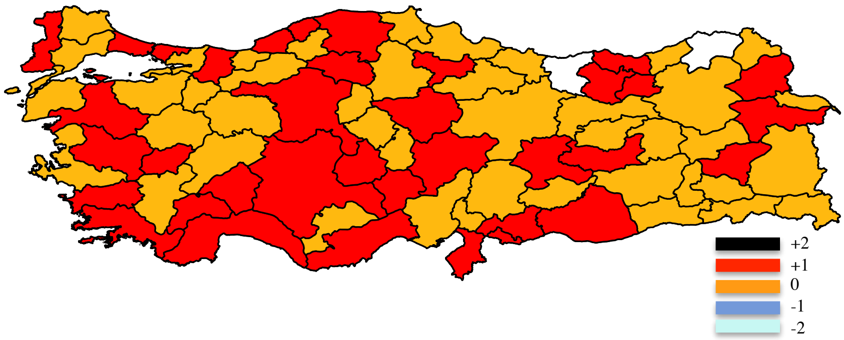


Figure 3.6 : Cluster differences between Euclidean distance and Manhattan distance K-means clustering colored map of three years with mean.

In order to understand whether using mean as the parameter is effective in defining the clusters, we have plotted Cumulative Distribution Functions (CDFs) for some of the selected cities in the clusters (Euclidean distance K-means presented in Figure 3.7). The colors used in the Figure 3.6 are white, light orange, orange, blue and green refers the highest PM_{10} level to the lowest, respectively. Two cities chosen for each cluster, one is the city which have minimum PM_{10} concentrations at the cluster and the other one is the maximum. Mean values of the cities are shown on the CDFs' with a circular shape. Van and Kahramanmaraş represent the Euclidean distance K-means high polluted cluster with white color in Figure 3.6. Van is the maximum of the cluster with 135.4 μg/m^3 and Kahramanmaraş is the minimum of the cluster with 106.5 μg/m^3 PM_{10} concentrations. Sinop and the Bayburt in the lowest cluster, with aqua color. Sinop is the minimum of the cluster with 36.5 μg/m^3 and Bayburt is the maximum of the cluster with 58.6 μg/m^3 PM_{10} concentrations. A zoom of the region towards the higher is also provided in Figure 3.6b. As seen in the Figure, although the clusters are distinct for the lower end (for example, green colored cleanest cluster is siginificantly lower as compared to white colored high polluted cluster), towards the higher end the difference reduces signficantly. For example, the difference between the green and white colors cluster is almost non-existent for the values over 200 µg/m^3.

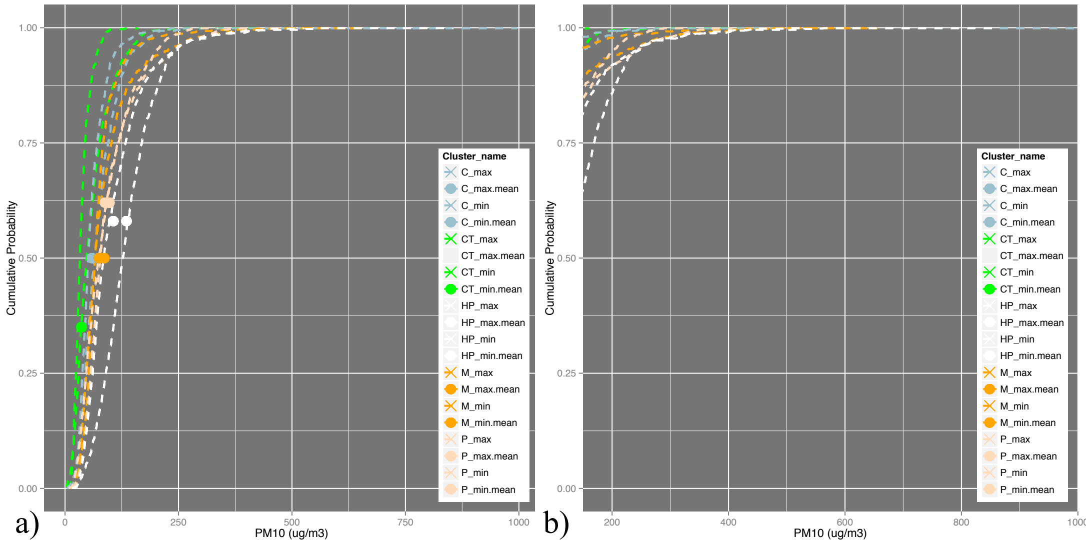
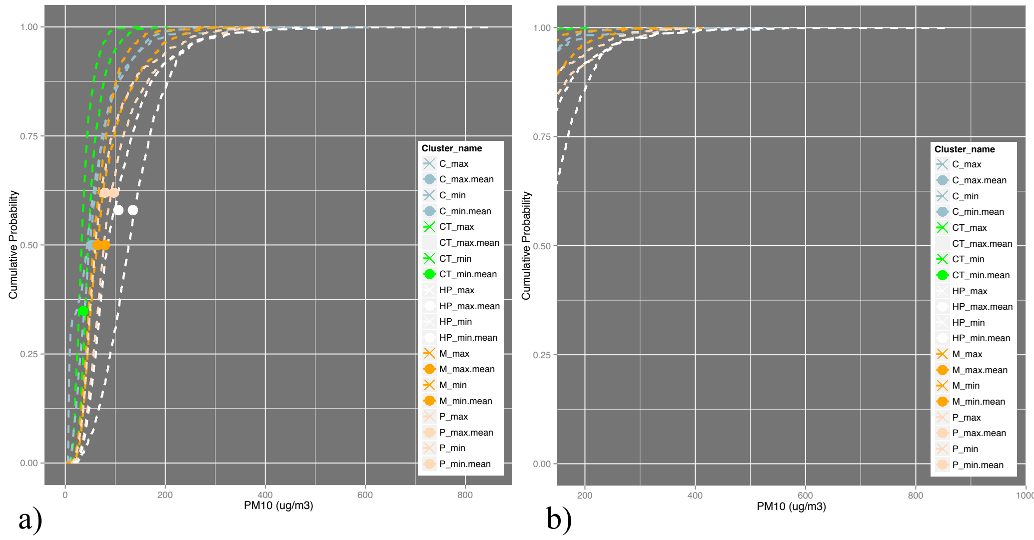


Figure 3.7 : a) Cumulative distribution functions (CDFs) of the Euclidean distance k-means clusters with only mean values. b) Zoom to the CDFs above 200 µg/m^3.

Cumulative Distribution Functions (CDFs) for Manhattan distance K-means with mean shown in Figure 3.8. Two cities choosen for each cluster, one is the city which have minimum PM_{10} concentrations at the cluster and the other one is the maximum as in Euclidean distance. The high polluted clusters cities are same as the Euclidean distance results. Van and Kahramanmaraşare represent the maximum and minimum cites of the highly polluted cities, respectively. Sinop and the Rize in the cleanest cluster, with green color. Sinop is the minimum of the cluster with 36.5 µg/m^3 and Rize is the maximum of the cluster with 50.6 µg/m^3 PM_{10} concentrations. As seen in the zoom of the region towards the higher in Figure 3.7b, although the clusters are distinct for the lower end. The difference towards the higher end, reduces signficantly as same in Euclidean distance CDFs. This might suggest that clustering method used is not effective to differentiate high PM_{10} values. In order to overcome this problem we have conducted K-Means clustering using 4 parameters (mean, median, 2.5th percentile and 97.5th percentile).

Three years Euclidean distance K-means result with four parameter is shown in Figure 3.8. The East and Southeast Anatolia regions also have the highly polluted and polluted clusters with more cities in the region. Central Anatolia in the medium



a) b)

Figure 3.8 : a) Cumulative distribution functions (CDFs) of the Manhattan distance k-means clusters with only mean values. b) Zoom to the CDFs above 200 µg/m^3.

and clean cluster. When we compare results in Figure 3.4 and Figure 3.9, there are similarities and differences. Three megacities Istanbul, Izmir and Ankara cluster not change, İstanbul and İzmir are in the cleanest PM_{10} level cluster and Ankara in the clean cluster. Antalya, Muğla in the Mediteranean Region, Yozgat, Kayseri and Nevşehir in the middle of the Central Anatolia Region, Sinop, Trabzon and Rize in the Black Sea Region, Tekirdağ, Bursa and Kocaeli are some of the other cities whose clusters did not change. Şanlıurfa and Gaziantep in the South East Anatolia region, Denizli and Kütahya in the Agean region, Elazığ, Erzincan and Erzurum in the East Anatolia regioncluster increased from the polluted cluster to highly polluted cluster.

.

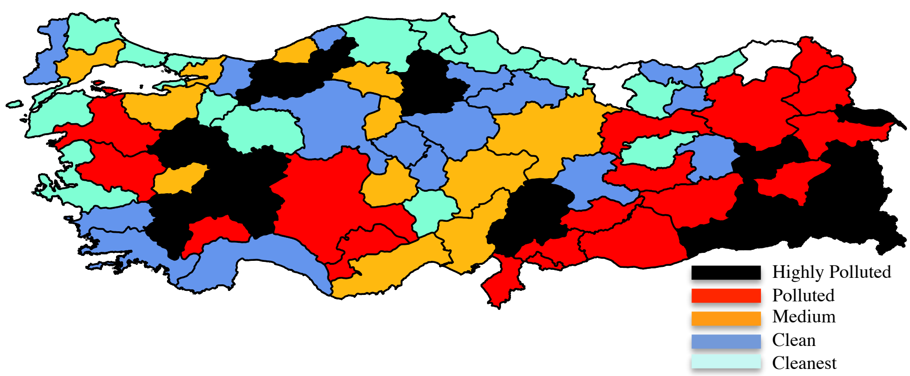
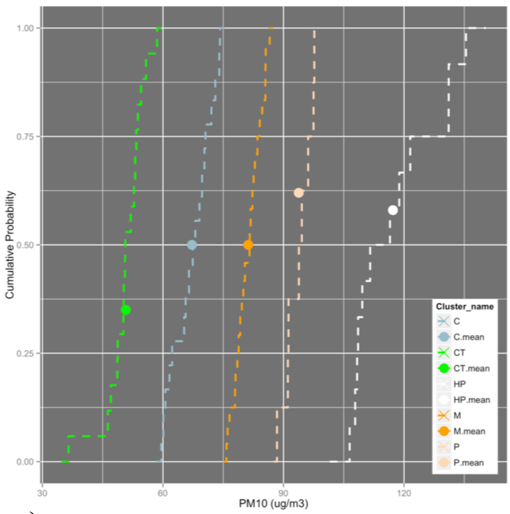
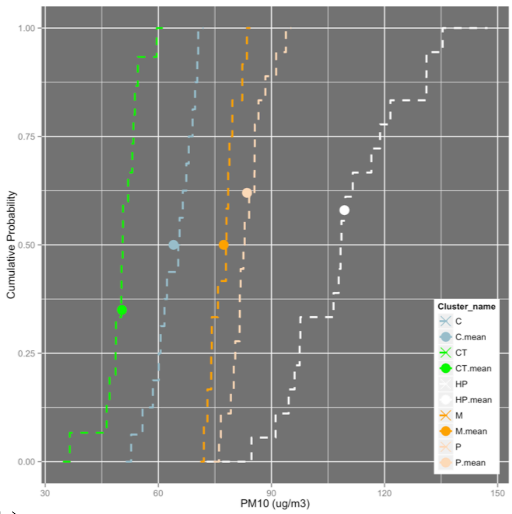


Figure 3.9 : Euclidean distance K-means clustering colored map of three years with 4 parameters : mean, 2.5 %, median, 97.5 %.



a)



b)

Figure 3.10 : a) Cumulative distribution functions (CDFs) of the Euclidean distance k-means clusters with mean b) Cumulative distribution functions (CDFs) of the Euclidean distance k-means clusters with four parameter

In order to understand clusters distributions of the mean and 4 parameter methods clustering, CDFs were plotted. As seen in the Figure 3.10.a Euclidean distance K-means with mean high polluted cluster (white color) is distinct from the other clusters. K-means with 4 parameter shown in Figure 3.10.b, medium and polluted cluster is close to each other and high polluted cluster is distinct from the other clusters.

Cumulative Distribution Functions (CDFs) for the 4-parameter case is shown in Figure 3.11.a Isparta and Van in the high polluted cluster with white color. Isparta is the minimum of the cluster with 84.7 µg/m^3 mean and 339.7 µg/m^3 for 97.5t^h percentile and Van is the maximum of the cluster with 135.4 µg/m^{3}mean and 294.9 µg/m^3 for 97.5th percentile Sinop and Ordu in cleanest cluster with green color. Sinop is the minimum of the cluster with 36.5 µg/m^3 mean and 82.8 µg/m^3 for 97.5th percentileand Ordu is the maximum of the cluster with 59.6 µg/m^3 mean and 140.3 µg/m^3 for 97.5th percentile PM_{10} concentrations. Figure 3.11.b show that the differences between clusters are also reduced at the towards the higher end. It should be noted that the difference is more distinctive in the 4 parameter K-means results for the values over 200 µg/m^3.

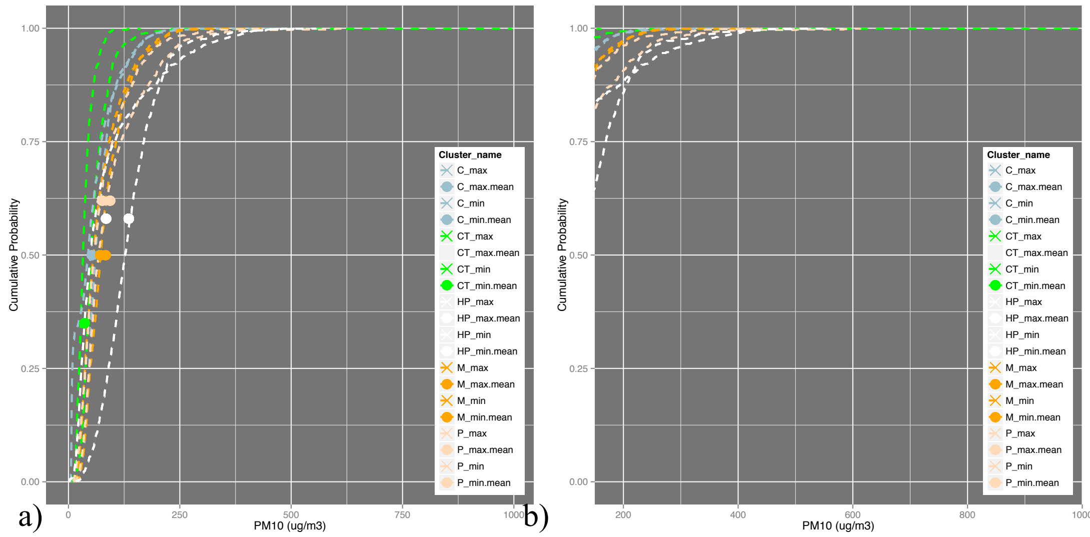


Figure 3.11 : a) Cumulative distribution functions (CDFs) of the Euclidean distance k-means clusters with four parameter. b) Zoom to the CDFs above 200 μg/m^3.

In order to understand the effective of the selected four parameters (mean, 2.5%, median and 97.5 %) on defining clusters, Principal Component Analysis (PCA) conducted. Summary of the PCA results showed (Table 3.2) that first two components cumulative proportion is represent the 96% of the data. PCA loadings of the four parameters given in Table 3.3, loadings are explain the weights of each standardized original variable on components.

Table 3.2: Summary of Principal Component Analysis

	PC1	**PC2**	**PC3**	**PC4**
Std.Dev.	1.76	0.85	0.40	0.08
Proportion of Var.	0.78	0.18	0.04	0.002
Cumulative Proportion	0.78	0.96	0.99	1

Table 3.3: Loadings of Principal Component Analysis of 4 parameter

	Comp.1	**Comp.2**	**Comp.3**	**Comp.4**
Mean	-0.560	-0.166	-0.116	0.803
2.5 %	-0.399	0.808	0.430	
Median	-0.482	-0.560	0.564	-0.370
97.50%	-0.482	-0.560	0.564	-0.370

In the step of the analysis Euclidean distance K-means method conducted with 2 PCs and given in the Figure 3.15. The result shown that K-means clustering with PCs is summarized all previous K-means analysis results. The highest PM_{10} levels dominate

East Anatolian region, Çorum Karabük and Kahramanmaraş are the additional cities in the high polluted. Afyon and Isparta in the Agean Region cluster change between high polluted cluster and polluted cluster depending on the previous results. In this result these two cities are in the polluted cluster with Iğdır and Hakkari in the East Anatolian Reagion and Mardin and Şırnak in the South East anatolian Region. Three megacities, İstanbul and İzmir in the cleanest cluster as Sinop, Çanakkale and Kırklareli.

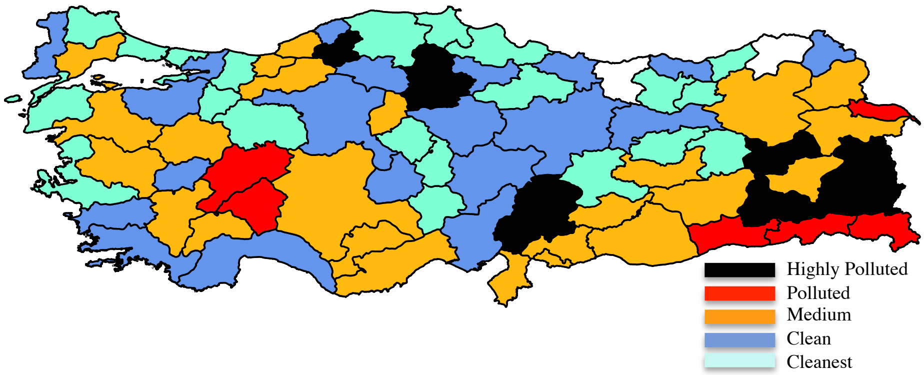


Figure 3.12 : Euclidean distance K-means clustering colored map of three years with first two principal component.

Cumulative Distribution Functions (CDFs) for the 4-parameter PCA case is shown in Figure 3.13. The differences between high polluted and polluted clusters are reduce at the towards the higher end. It should be noted that the difference is more distinctive in the 4 parameter K-means PCA results for the values over 200 $\mu g/m^3$.

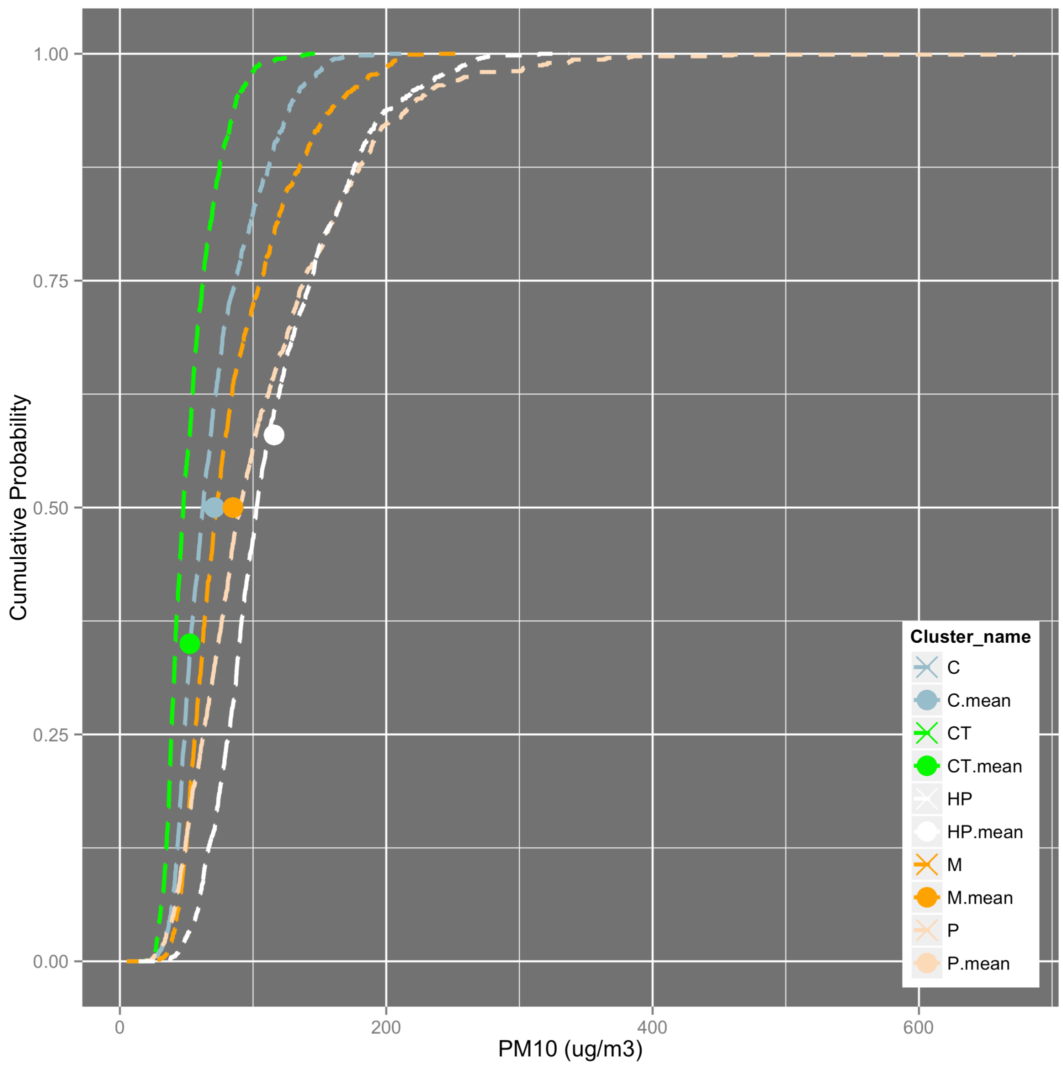


Figure 3.13 : Cumulative distribution functions (CDFs) of the PCA K-means clusters with four parameter.

In order to overcome 4 parameter PCA case problem we have utilized PC analysis for K-Means clustering using five different parameters, 50^{th}, 75^{th}, 90^{th}, 95^{th} and 97.5^{th} percentile. The percentiles calculated over three yearsPM_{10} dataset given in Table 3.5. The metropolitan cities İstanbul and İzmir are in the cleanest cluster and Ankara is in the clean cluster in 5 parameter K-means results. Highly polluted cluster 50 % percentile range is 81.8 μg/m^3 and 126.9 μg/m^3. Cleanest cluster 50% percentile range is 32.6 μg/m^3 an 51.7 μg/m^3. There are 6 cities in the high polluted cluster; Batman, Hakkari, Siirt, Muş, Iğdır and Van and all of them are located in the East and South East Anatolian Region.

Table 3.4: Euclidean distance 5 parameters and clusters

Cities	50%	75%	90%	95%	97.50 %	Cluster name	Cluster
ADANA	66.2	84.0	118.4	145.6	184.3	C	3
ANKARA	53.3	85.7	129.3	153.7	175.2	C	3
KAYSERI	56.5	91.2	140.2	178.4	206.6	C	3
KOCAELI	62.8	93.1	141.8	169.6	185.1	C	3
MUGLA	60.7	82.0	105.9	129.4	145.5	C	3
TRABZON	54.7	82.6	122.3	155.4	190.4	C	3
AKSARAY	65.1	94.5	125.6	151.1	175.7	C	3
AMASYA	51.5	99.2	143.2	165.8	181.4	C	3
ANTALYA	58.8	81.6	126.7	156.7	181.0	C	3
AYDIN	60.6	84.9	114.8	149.0	172.9	C	3
BARTIN	58.9	90.0	128.0	151.5	175.8	C	3
BINGOL	48.8	77.0	115.7	153.1	192.5	C	3
BURSA	69.1	99.6	136.0	166.0	190.0	C	3
CANKIRI	60.8	98.0	134.5	162.6	188.4	C	3
EDIRNE	58.1	83.3	114.7	137.2	163.6	C	3
ICEL	68.0	98.3	140.5	171.8	194.8	C	3
TEKIRDAG	68.3	98.0	133.5	159.3	180.4	C	3
USAK	65.1	91.8	122.0	142.1	154.6	C	3
YOZGAT	56.7	81.9	107.5	129.1	155.1	C	3
ISTANBUL	47.1	66.1	91.9	108.4	122.4	CT	5
IZMIR	45.7	60.3	82.3	100.6	119.4	CT	5
SAMSUN	44.6	56.1	70.9	85.0	101.4	CT	5
BAYBURT	47.6	76.3	114.2	141.8	160.8	CT	5
BILECIK	45.0	62.8	81.4	96.3	103.6	CT	5
CANAKKALE	42.5	58.9	83.1	102.5	117.9	CT	5
ESKISEHIR	42.6	56.0	76.5	91.6	103.8	CT	5
GUMUSHANE	42.4	67.3	100.0	128.5	147.9	CT	5
KASTAMONU	43.6	67.0	95.9	121.1	151.4	CT	5
KIRKLARELI	42.3	57.8	82.2	102.9	123.0	CT	5
KIRSEHIR	51.7	75.3	103.8	128.5	159.8	CT	5
MALATYA	43.7	73.4	122.4	151.8	170.0	CT	5
NEVSEHIR	49.7	75.0	110.0	135.6	171.1	CT	5
NIGDE	46.8	69.1	92.7	115.7	145.6	CT	5
ORDU	50.8	71.0	94.7	111.8	140.3	CT	5
RIZE	37.4	62.9	100.8	123.9	142.5	CT	5
SAKARYA	44.5	67.3	104.3	146.4	190.0	CT	5
SINOP	32.6	43.7	60.1	71.9	82.8	CT	5
TOKAT	45.1	78.3	119.2	150.6	178.1	CT	5
TUNCELI	42.7	63.1	83.3	108.3	138.7	CT	5
YALOVA	45.2	60.8	81.0	96.3	117.7	CT	5

Table 3.4 (contd.) : Euclidean distance 5 parameters and clusters

Cities	50%	75%	90%	95%	97.50 %	Cluster name	Cluster
BATMAN	112.6	163.5	237.8	288.2	318.8	HP	2
HAKKARI	92.1	136.0	245.1	349.7	438.1	HP	2
IGDIR	81.8	161.2	320.3	409.4	474.5	HP	2
MUS	99.9	147.9	214.1	269.8	300.4	HP	2
SIIRT	107.2	148.7	198.5	238.9	290.4	HP	2
VAN	126.9	169.2	213.3	244.0	294.9	HP	2
KONYA	60.5	105.2	163.5	199.0	244.1	M	1
AGRI	61.4	106.0	164.4	233.3	283.7	M	1
ARDAHAN	56.7	99.7	163.7	206.3	241.1	M	1
BALIKESIR	60.5	93.0	151.4	207.2	249.9	M	1
BITLIS	69.2	104.2	152.6	193.5	270.9	M	1
BURDUR	67.1	102.1	145.8	185.8	221.7	M	1
DIYARBAKIR	75.6	108.0	152.2	188.8	244.6	M	1
ELAZIG	70.4	108.2	153.3	181.8	235.4	M	1
ERZINCAN	57.4	99.2	152.5	188.4	223.4	M	1
ERZURUM	60.8	95.1	148.6	214.1	279.8	M	1
GAZIANTEP	64.4	111.2	165.4	205.3	251.6	M	1
HATAY	64.8	102.8	162.5	194.1	221.1	M	1
KARAMAN	74.5	117.7	170.9	199.8	222.5	M	1
KARS	64.0	104.6	159.7	210.8	237.6	M	1
KILIS	66.2	91.4	149.4	196.1	247.1	M	1
KIRIKKALE	64.2	98.0	152.4	187.4	216.8	M	1
MANISA	65.9	95.3	146.1	189.6	222.5	M	1
SIVAS	61.3	96.0	144.0	177.9	208.8	M	1
URFA	67.0	101.9	155.0	223.5	274.2	M	1
ZONGULDAK	72.1	107.3	146.7	174.2	199.2	M	1
DENIZLI	77.3	112.8	174.3	236.5	282.9	P	4
K.MARAS	85.0	133.2	187.0	239.7	283.6	P	4
ADIYAMAN	68.0	117.3	197.6	233.2	263.1	P	4
AFYON	84.7	141.4	212.9	263.3	305.0	P	4
BOLU	65.1	111.0	206.9	278.1	329.3	P	4
CORUM	85.4	135.6	209.5	269.7	299.9	P	4
DUZCE	60.5	100.9	203.9	278.0	334.6	P	4
ISPARTA	49.3	104.5	213.1	282.4	339.7	P	4
KARABUK	86.9	131.9	193.1	242.9	289.5	P	4
KUTAHYA	77.3	116.1	179.7	241.3	279.9	P	4
MARDIN	81.5	125.3	200.8	277.1	358.1	P	4
OSMANIYE	89.2	129.5	190.6	238.1	291.7	P	4
SIRNAK	76.2	109.4	178.8	238.2	329.3	P	4

PCA results showed (Table 3.6) that first two components cumulative proportion is represent the 98% of the data. PCA analysis results shown in Table 3.4, first 2 components represent the 98 percentage of the data. First two component utilized to K-means analysis .

Table 3.5: Summary of Principal Component Analysis of 5 parameter

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5
Standard deviation	2.126	0.645	0.221	0.099	0.060
Proportion of Variance	0.904	0.083	0.010	0.002	0.001
Cumulative Proportion	0.904	0.988	0.997	0.999	1

Table 3.6: Loadings of Principal Component Analysis of 5 parameter

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5
50%	-0.412	0.732	-0.42	0.338	
75%	-0.456	0.332	0.481	-0.638	-0.211
90%	-0.462	-0.212	0.495	0.402	0.579
95%	-0.456	-0.372	0.355	-0.726	
97.50%	-0.449	-0.413	-0.589	-0.438	0.298

Euclidean distance K-means method with 5 parameter conducted with 2 PCs given in the Figure 3.16. The result shown that K-means clustering with PCs is summarized all previous K-means analysis results. Hakkari, Van, Iğdır, Siirt and Muş and Batman in the East and Southeast Anatolian Region are clustered to highly polluted cluster. Çorum, Karabük, Bolu and Düzce in the Black Sea, Kahramanmaraş, Osmaniye in the Mediterranean, Adıyaman, Mardin and Şırnak in the East Anatolian, Afyon, Isparta, Denizli and Kütahya in the Agean Region are the cities in the polluted cluster. Three megacities, İstanbul and İzmir in the cleanest cluster as Sinop, Çanakkale and Kırklareli.

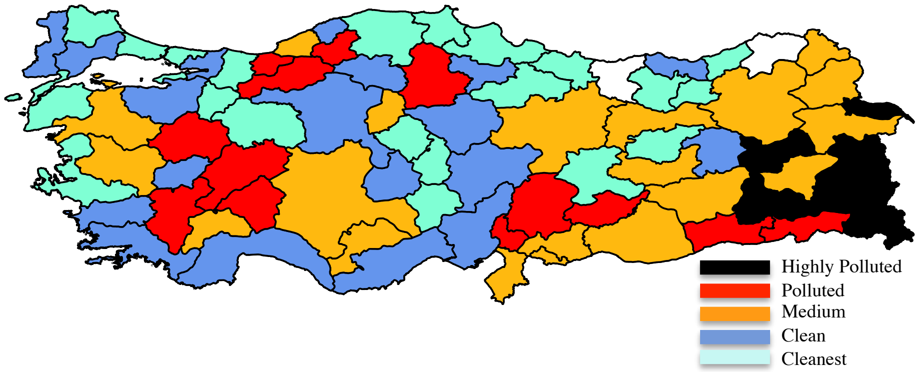


Figure 3.14 : Euclidean distance 5 parameter K-means clustering colored map of three years with first two principal component.

In order to understand the basis for the variability in the clusters, spatial distributions of emissions inventory was checked. For this purpose, TNO emissions inventory data were plotted (Figure 3.15). As seen in the figure, for PM10, north-western Turkey (Istanbul and its vicinity), along with western Turkey, south-western Turkey, Inner Anatolia have high emission values. On the other hand, South-West Turkey, Eastern Turkey and North-Eastern Turkey has the lowest emission values. High polluted cluster cities as Hakkari, Van, Şırnak in the Eastern Anatolia Region have low emission values besides Cleanest cluster cities as Istanbul, Izmir and Çanakkale have high emission values.

Therefore, it is clear that the variation in the emissions distribution do not really explain the variation in the PM_{10} distribution. This is probably due to the fact that there is high uncertainty in the emissions database. It should be remembered that TNO emissions for Turkey is based mostly on the expert data as the required database are not available. The distribution for each sector is provided in Appendix but it should be noted that they do not have the explanatory power either.

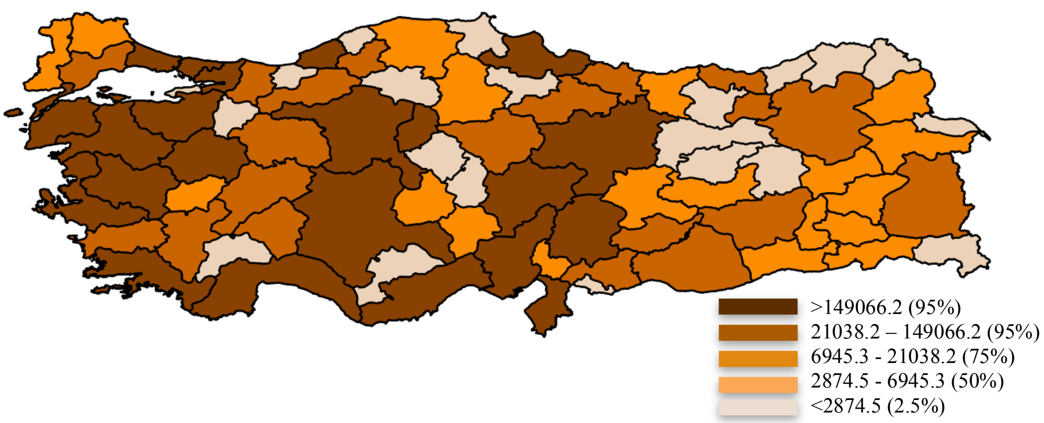


Figure 3.15 : TNO emissions map of averages of all sectors.

Another factor that effect the air pollution levels is the meteorological conditions. In order to quantfiy such effect one method is to develop indicators for presence of inversion. As widely known, under inversion conditions air pollution levels significantly increase. In this study we first utilized NCEP-NCAR data to identify inversion via estimating temperature at different pressure levels. NCEP-NCAR provides temperature data at 1000 mb, 925 mb, 850 mb and 700 mb. Temperature differences were estimated between different levels (e.g. temperature at 850 mb – temperature at 700 mb). Negative values suggests a presence of inversion. Figure 3.16 presents the difference of temperature at 850 mb and temperature at 700 mb for NCEP-NCAR cells which cover cities Afyon, Isparta and Antalya. As seen in the figure negative values occure in two times in a year. Similar in Figure 3.17 for ncep-ncar cells which cover cities Izmir, Aydın and Muğla negative values occure in two times in a year.

For both grid cells NCEP-NCAR based temperature differences estimate do not really provide any indication of inversion conditions. It migth be do the fact that NCEP-NCAR dataset is at very course level (with 2.5^0 x 2.5^0) and the data used to compile do not have the capability to understand the atmosferic temperature profile (the difference plot for the other presuure levels provides similar results and are given in Appendix).

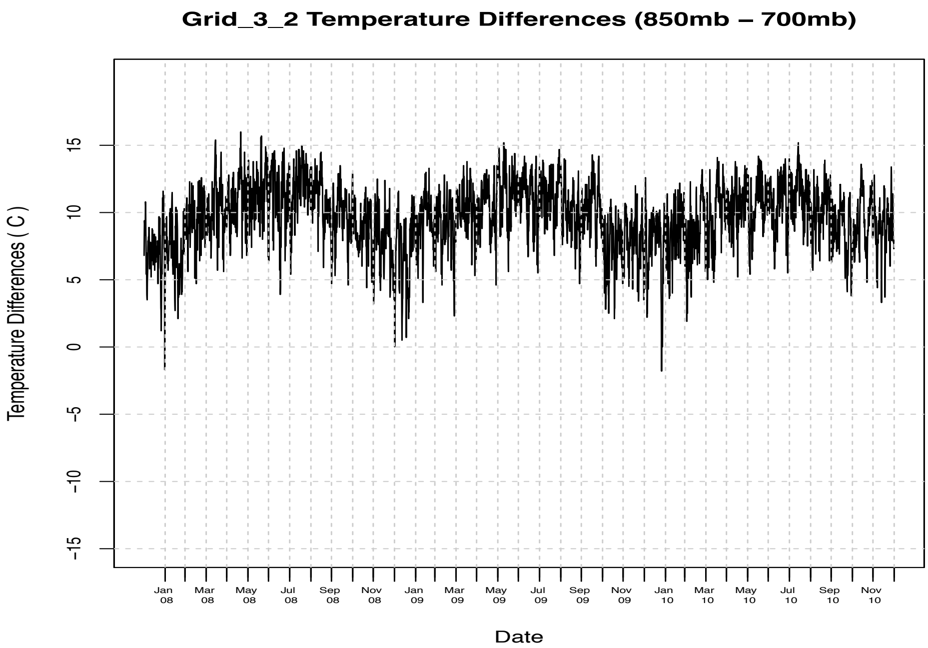


Figure 3.16 : NCEP-NCAR Reanalysis data 850mb-700 mb temperature differences in Afyon, Isparta and Antalya grid.

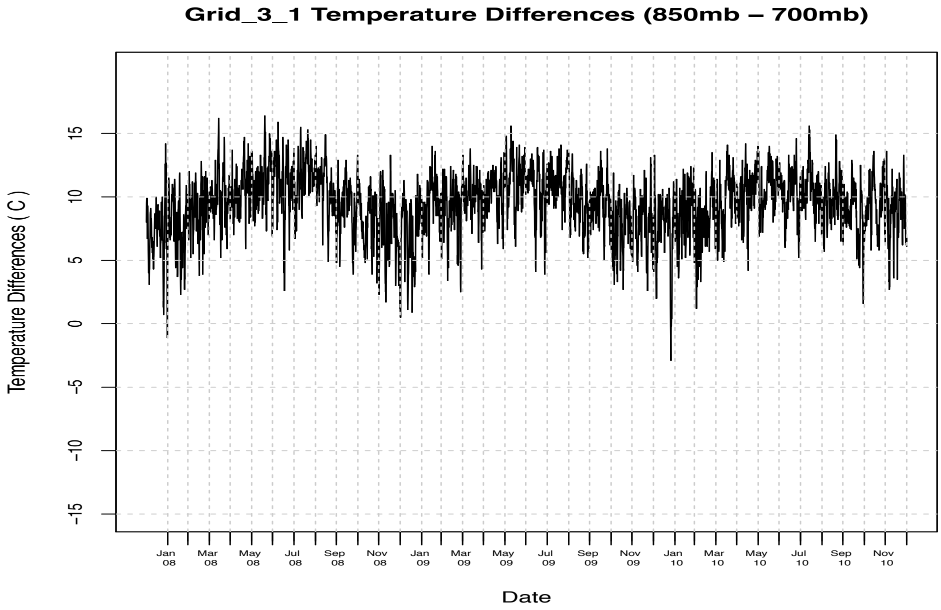


Figure 3.17 : NCEP-NCAR Reanalysis data 850mb-700 mb temperature differences in Izmir, Aydın and Muğla grid.

In the next step we decided to use WRF output and analyse PBL height as an indication of presence of inversions. We have utilized PBL height outputs of 2008 WRF run. PM_{10} concentrations vs PBL heights of the selected 13 cities (Afyon, Isparta, Iğdır, Antalya, Eskişehir, İzmir, Kahramanmaraş, Muş, Düzce, Van, Bolu, Ankara, İstanbul) were plotted. Kahramanmaraş is in the highly polluted cluster, 2008 PM_{10} concentrations vs PBL heights shown in Figure 3.18. As seen in the figure, high PBL heights (above 400 m) occure in the spring and summer months (April to August) and PM_{10} concentrations decreasing from 200 $\mu g/m^3$ to 100 $\mu g/m^3$. In winter and autumn months (September to March) lowest PBL heights occurred besides highest PM_{10} concentrations observed. 2008 PM_{10} concentrations vs PBL height of İstanbul, cleanest cluster city, shown in Figure 3.19. PM_{10} concentrations are below 100 $\mu g/m^3$ during the year but concentration peaks was occured under low PBL height. (Antalya, Afyon, Iğdır and Muş plots given in Appendix E)

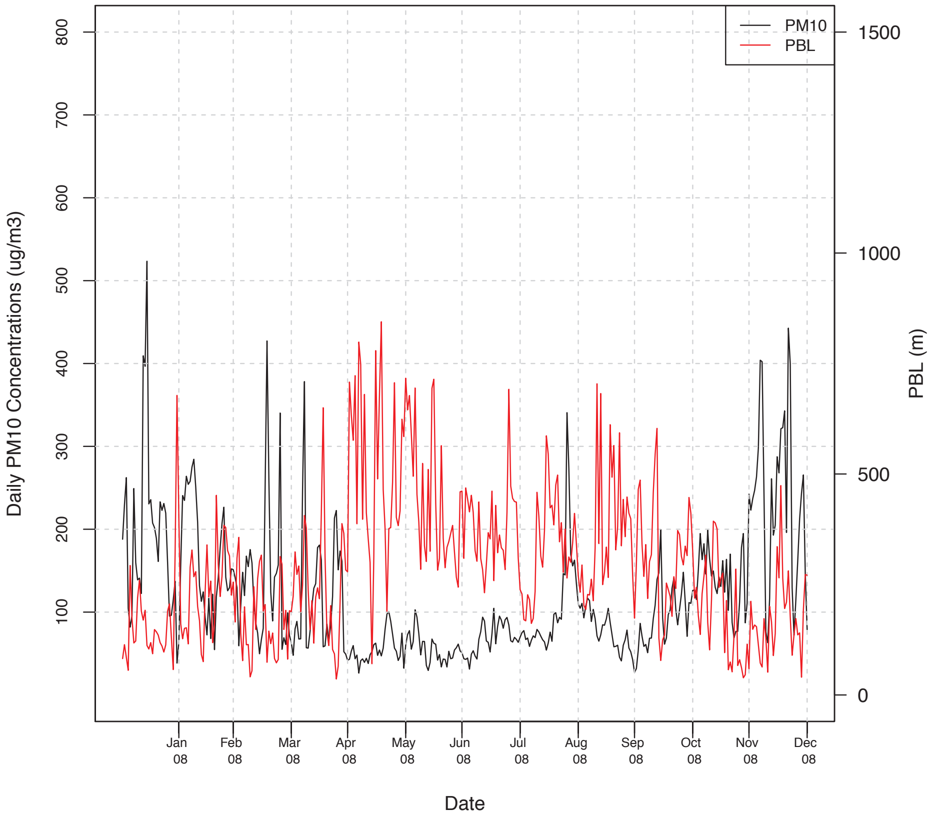


Figure 3.18 : Kahramanmaraş PM_{10} concentrations and PBL height in 2008.

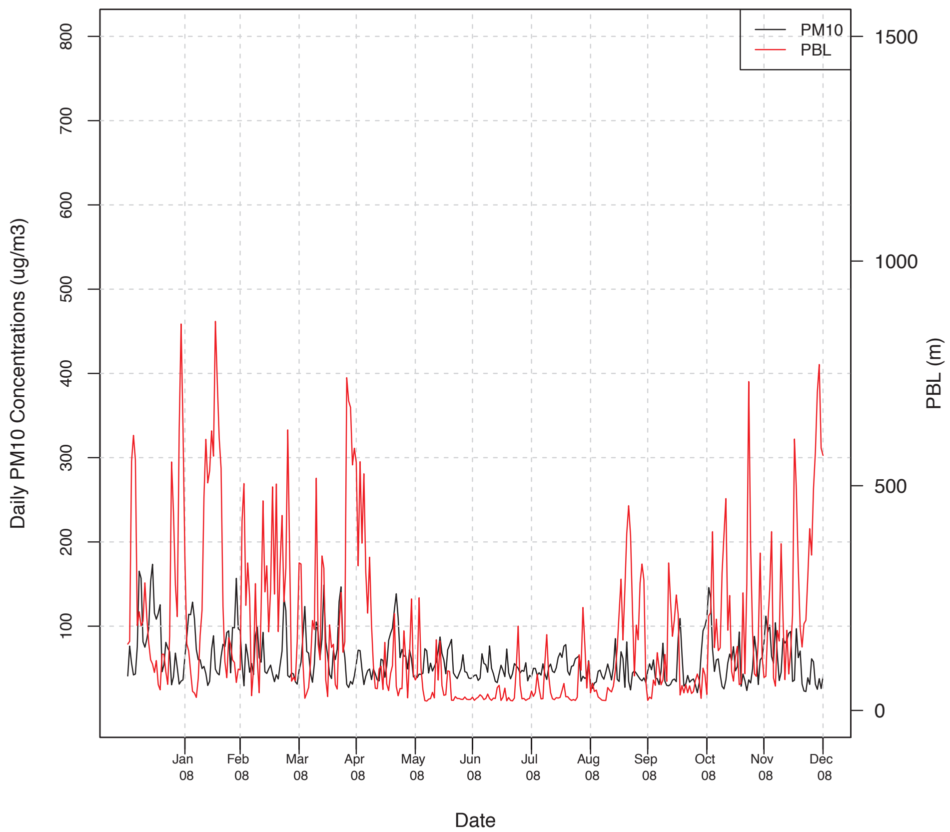


Figure 3.19 : İstanbul PM_{10} concentrations and PBL height in 2008.

In order to understand the effect of the PBL heights on cluster results, we have plotted the spatial distribution of PBL height. Annual PBL heights spatial plot is shown in Figure 3.20. As seen in the figure, there is a wide range from western part of Turkey to Eastern part, low PBL height occurred on the coastal parts and high PBL heights occurred on the interior parts. Eastern Turkey and North-Eastern Turkey and North-East Mediterranean region have the lowest PBL height. Hakkari, Van and Iğdır in the East Anatolia region, under the lowest PBL height annually with Kahramanmaraş in the North-East part of Mediterranean Region. Seasons and hours are the effective parameters PBL heights changes. Dataset utilized in daily averages, at 06:00 AM, at 12:00 AM and 18:00 PM for seasons. In Figure 3.21, 12:00 averages of PBL heights for each season are shown. Quarter 1 represent January and February averages, Quarter 2 March, April and May averages, Quarter 3 June, July and August averages and Quarter 4 September, October and November averages (the daily and 06:00 AM plots given in Appendix).

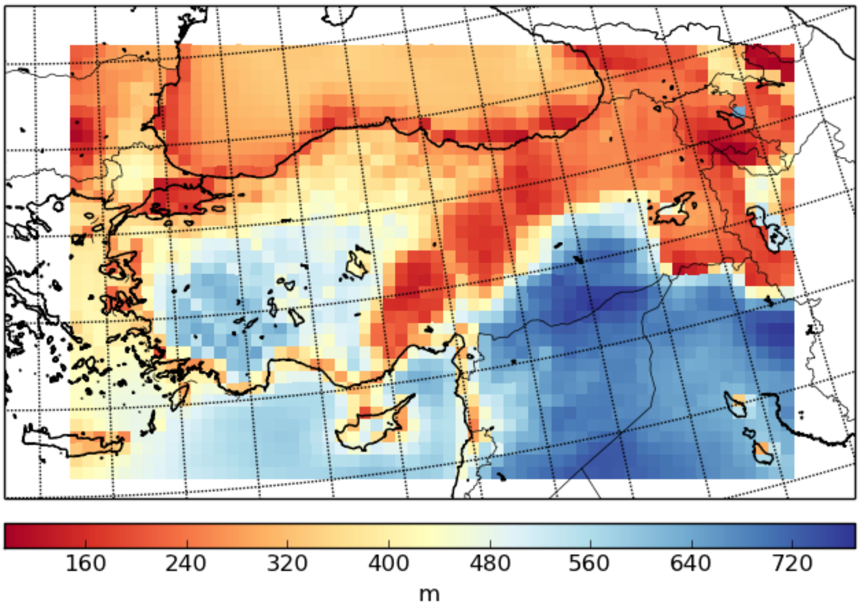


Figure 3.20 : Annual PBL heights spatial plot

As seen in the figure 3.21, low PBL heights at 12:00 AM occurred on East and North-East Turkey, North-East part of Mediterranean Region which the Taurus Mountains extended and on the sea surface. The high polluted cities in the Aegean Region, Afyon, Isparta, Denizli and Kütahya have high PBL heights in Quarter 1,2 and 3 besides the high polluted cities in the South East and East Anatolian Region Hakkari, Iğdır, Van, Şırnak and Mardin.

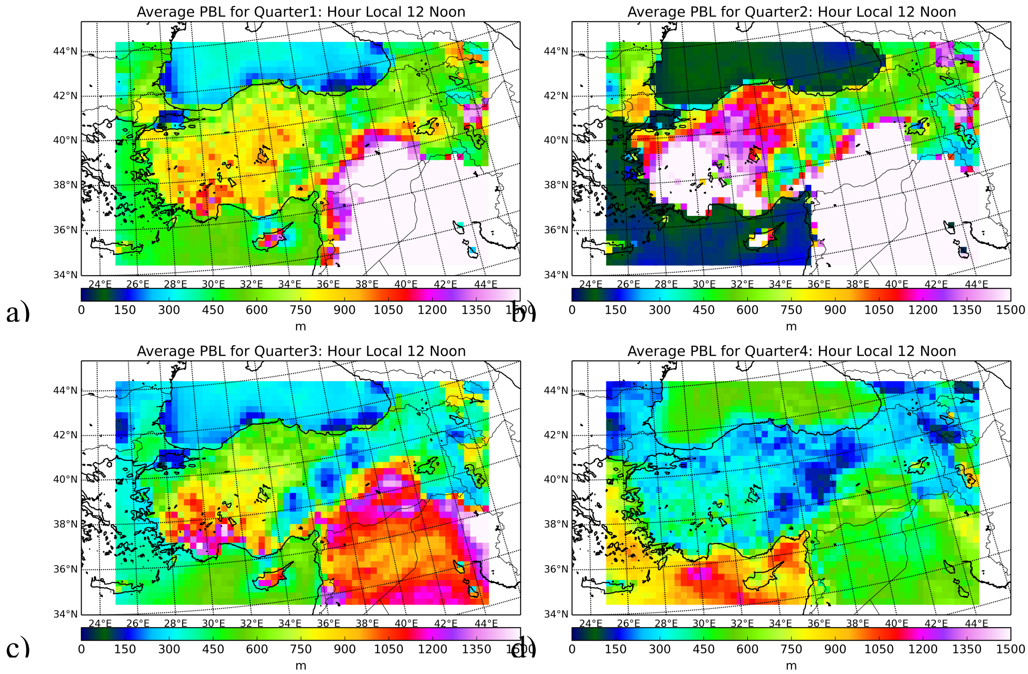


Figure 3.21 : a) 2008 January and Februrary at 12 AM averages of PBL heights b) 2008 March, April, May at 12 AM averages of PBL heights c) 2008 June, July, August at 12 AM averages of PBL heights d) 2008 September, October, November at 12 AM averages of PBL heights.

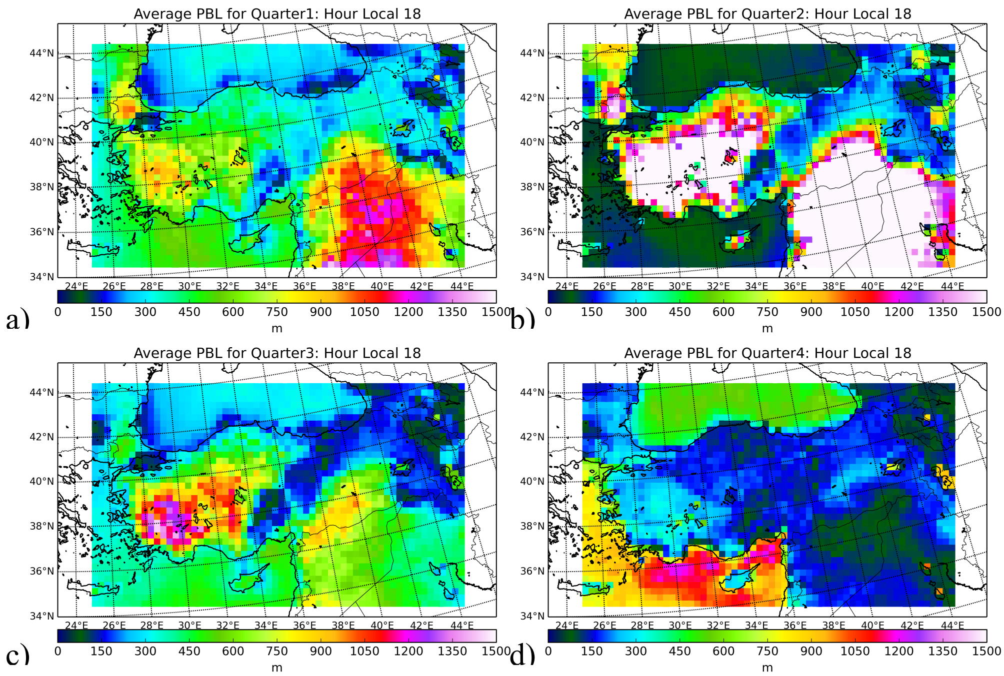


Figure 3.22 : a) 2008 January and Februrary at 18 PM averages of PBL heights b) 2008 March, April, May at 18 PM averages of PBL heights c) 2008 June, July, August at 18 PM averages of PBL heights d) 2008 September, October, November at 18 PM averages of PBL heights.

Similar in Figure 3.22, low PBL heights at 18:00 AM occurred on East and Northeast Turkey, Northeast part of Mediterranean Region. The PBL heights level is significantly decreasing with hour differences. While the lowest PBL height is approximately 100-150 m at 12:00 am, it decreases to below 50m. It is clear that the spatial distributions of PBL height can explain the variation in the PM_{10} distribution in East and South-East Anatolia Region. However, Aegean Region spatial distribution plots (Figure 3.21 and 3.22) show that the region has high PBL height, Afyon, Isparta, Denizli and Kütahya (the polluted cluster cities in the Aegean region) have high PM_{10} concentrations. It should be remembered that, spatial distribution maps plotted based on the grid cells. PM_{10} concentrations and PBL heights boxplot shown in the Figure 3.23 to explain high PM_{10} levels in the selected city Afyon.

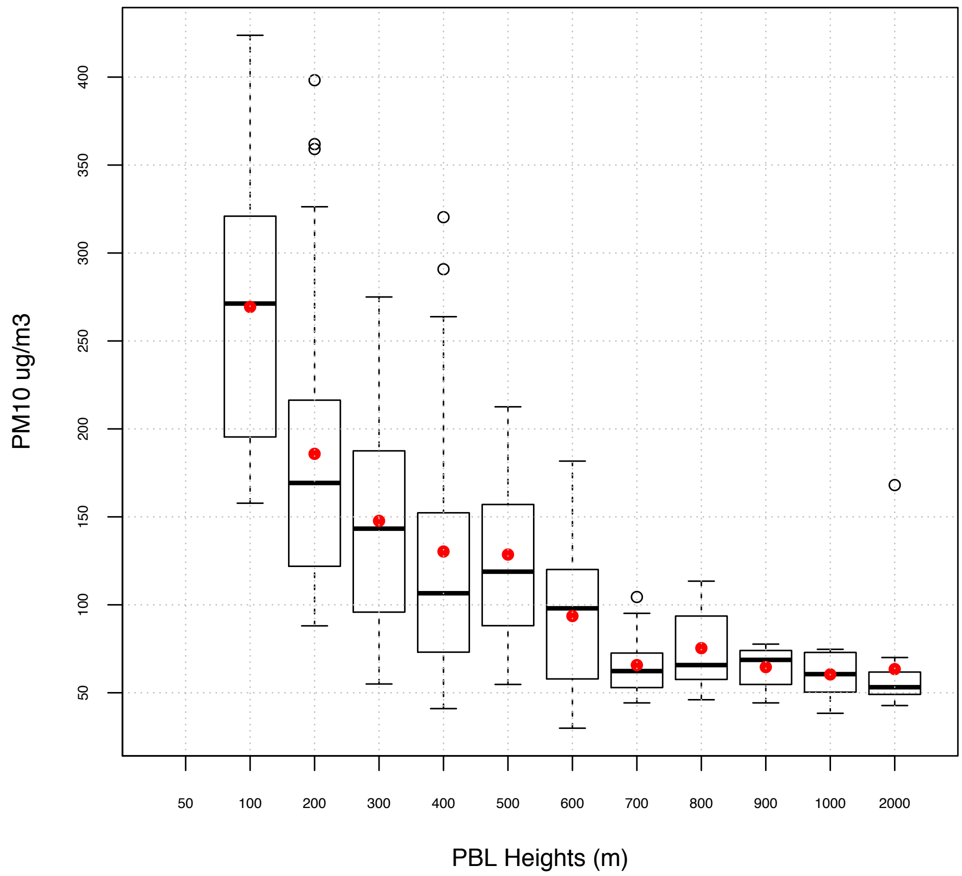


Figure 3.23 : Afyon 2008 PM_{10} concentrations and PBL heights (m) boxplot. The Box-whisker plot indicates the mean (red points inside the box), 95 percent confidence bounds for the mean (short black lines on the red points inside the box), the median (the bold black line inside the box), the lower and upper quartiles of the data set (25th and 75th percentiles which is shown by the lower and upper ends of the box), and extreme values (top and bottom lines).

As seen in the figure 3.23, there is an exponential trend between PM_{10} concentrations and PBL heights. Contrary to spatial distributions map, low PBL heights occurred in the city. Topography is the effective parameter of low PBL heights in Afyon. It is located in mountainous countryside inland from the Aegean coast. In order understand the topography effect, Eskişehir is selected city that is a plain city in Central Anatolia Region (Figure 3.24). PBL heights range is 100 m to 2000 m as Afyon. Low PBL heights occurred in the city but PM_{10} concentrations were observed below 100 μg/m^3. The reason is that PM_{10} concentrations can be spread around in the plain city.

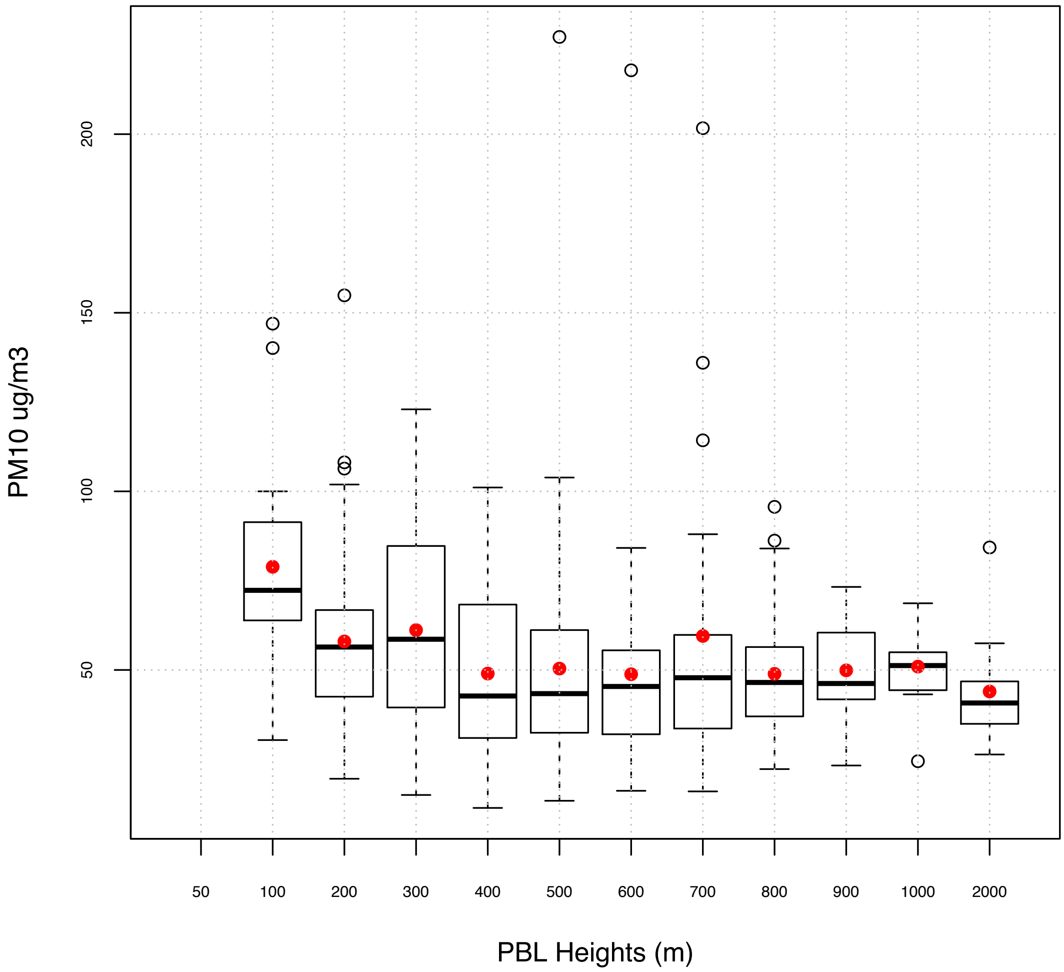


Figure 3.24 : Eskişehir 2008 PM_{10} concentrations and PBL heights (m) boxplot. The Box-whisker plot indicates the mean (red points inside the box), 95 percent confidence bounds for the mean (short black lines on the red points inside the box), the median (the bold black line inside the box), the lower and upper quartiles of the data set (25th and 75th percentiles which is shown by the lower and upper ends of the box), and extreme values (top and bottom lines).

In order to show PM_{10} concentrations distribution on PBL heights levels represented in histogram figures. The count of PM_{10} concentrations shown in y axis, PM_{10} concentrations in x axis and PBL height level are noted as the upper label of the blocks in meters. Afyon and İzmir are the selected cities in the high polluted and cleanest cluster cities, respectively.

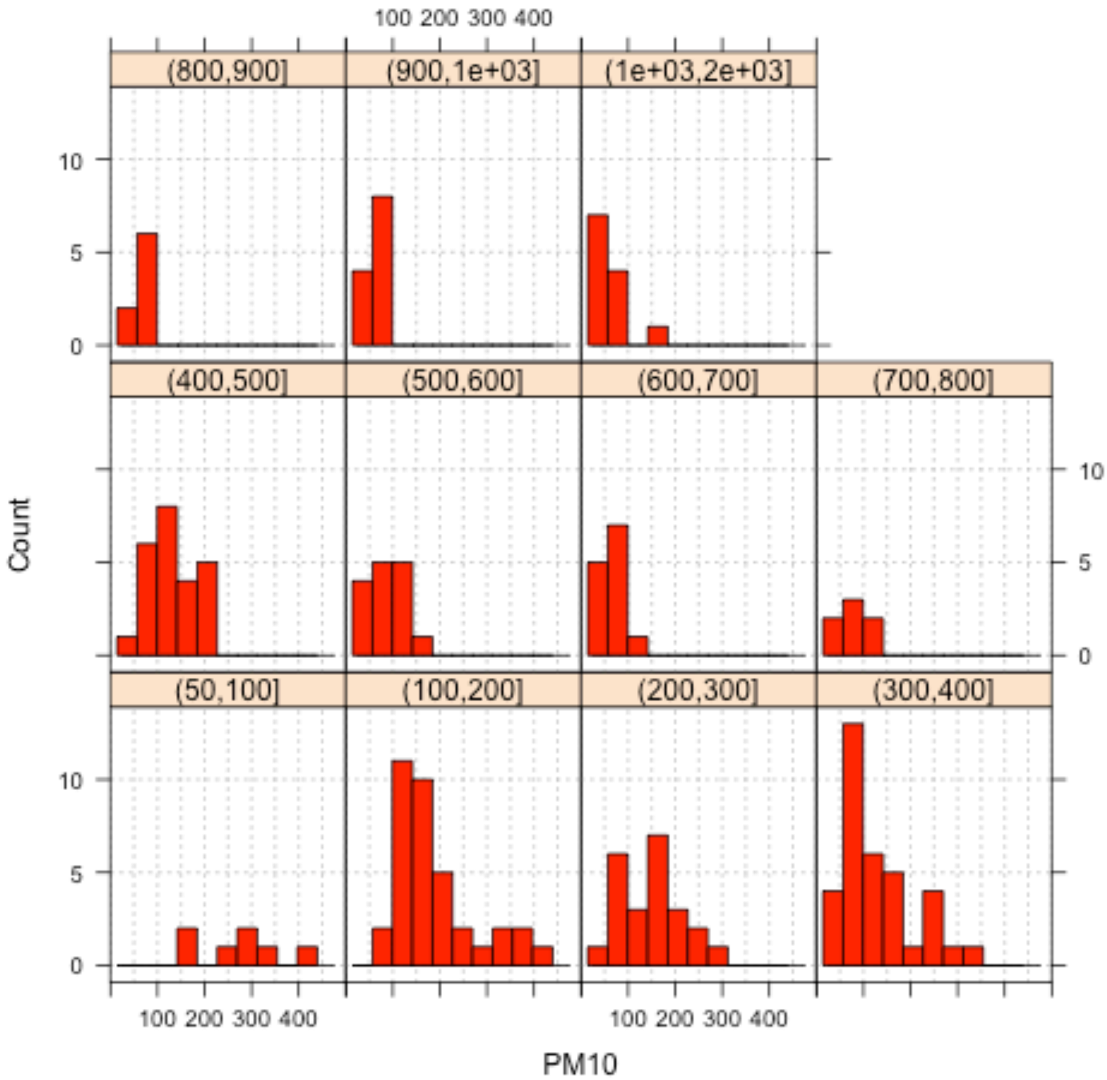


Figure 3.25 : Afyon PM_{10} concentrations distributions vs PBL heights levels (m).

Afyon PM_{10} concentrations distribution shown in Figure 3.25. As seen in the figure, PBL heights range is 0 to 2000 meters and it mostly between 100 meters to 500 meters. High PM_{10} values ocuured below the 500 m PBL heights. PM_{10} concentrations between 100 and 200 μg/m^3 observed in 22 times in 100 meters to 200 meters PBL heights. Above 400 meters PM_{10} concentrations below the 200μg/m^3. On the other hand, in Figure 3.26 İzmir PBL heights range is 0 to 2000 meters and it mostly between 100 meters to 700 meters. This wide range of PBL height in Izmir causes low PM_{10} concentrations.

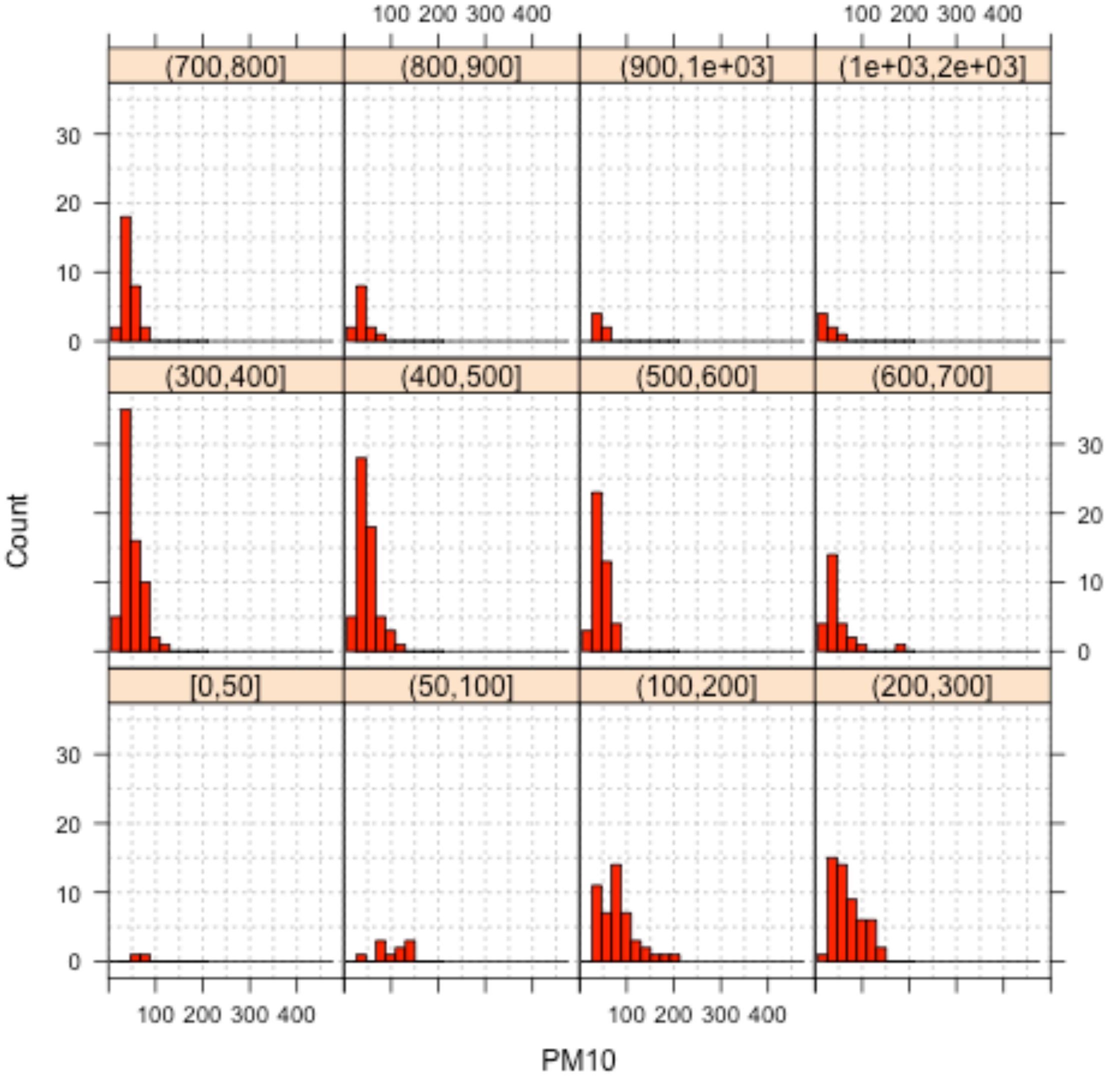


Figure 3.26 : Izmir PM_{10} concentrations distributions on PBL heights levels (m).

It is clear from the analysis, PBL height is an effective parameter to understand the variation in the PM_{10} distribution. However, it is not the only parameter to explain distributions of the PM_{10} concentrations, complex meteorological conditions, topography and emissions are important for distribution.

4. CONCLUSION

In this study PM_{10}levels in Turkey was evaluated via spatio-temporal analysis. Determination of PM_{10} level is critical as paticulate pollution have significant health effects. PM_{10} measurements of 118 monitoring stations of the Turkish Ministry of Environment and Urbaniztion over three years (2008-2010) were used for this study. Temporal analysis revealed that daily average PM_{10} values do not have a decreasing or increasing trend between 2008 and 2010 when average over all the monitors are considered. For spatial analysis, K-means analysis was used to identify clusters in the data. Bayesian Information Criterian (BIC) method result suggested that data can be divided into five different clusters. K-means analysis conducted using two different distance calculations, Manhattan and Euclidean distance. The analysis first utilized three years average values. The results suggested the existance of clusters in the data. The East and Southeast Anatolian Region have the highest PM_{10} levels. Cumulative Distribution Functions (CDFs) of the selected cities in the clusters were prepare to examine whether the mean is the correct parameter to identify the clusters. The differentiation of the clusters above the 200 µg/m^{3}is not clear with this method.In the next step four different parameters (mean, 2.5%, median, 97.5%) were used to overcome this reduction. However, similarities between the clusters were still there. Results have not change even PCA analysis conducted on the four parameter dataset. In the final stage (%50, %75, %90, %95 and %97.5) with PCA, were found to succesfully differentiate the clusters. Therefore, in this analysis five parameters PCA utilized. The eastern part of the country, (East and South East Region) have high PM_{10} levels. The high polluted cluster include 6 cities; Hakkari, Van Siirt, Muş, Iğdır,Batman. All high polluted cluster cities are in the East and South East Anatolian Region. The polluted cluster include 13 cities, six of them are Kahramanmaraş, Denizli, Bolu, Afyon, Düzce, Isparta. There are 20 cities in medium cluster, six of them are; Konya, Sivas, Ağrı, Balıkesir, Hatay and Diyarbakır. The clean cluster include 19 cities, six of them are Ankara, Bursa, Adana, Antalya, Edirne and Bartın. The cleanest cluster include 21 cities, six of them are İstanbul, İzmir, Eskişehir, Sinop, Rize and Çanakkale.

In order to understand the basis for the variability in the clusters spatial distributions of emissions inventory and meteorological conditions were analized. For this TNO emissions inventory was utilized. However, spatial distribution of emissionsdo not conform with the spatial distributions of the clusters. For example, emissions inventory have the highest values in western part of Turkey, the cities of the regions distributed to cleanest cluster. The eastern part in the high polluted cluster, emissions inventory have lowest values.

The temperature and PBL heights analyzed to understand the meteorological conditions effects on PM_{10} distributions. In the first step NCEP-NCAR data utilized to identify inversion via estimating temperature at different pressure levels. The difference of temperature at 850 mb and temperature at 700 mb for two different NCEP-NCAR cells. One of them cover cities Afyon, Isparta and Antalya and the other one cover cities İzmir, Aydın and Muğla negative values occure in two times in a year. For both grid cells NCEP-NCAR based temperature differences estimate do not really provide any indication of inversion conditions.

In the next step we decided to use WRF output and analyse PBL height as an indication of presence of inversions. We have utilized PBL height outputs of 2008 WRF run. PM_{10} concentrations vs PBL heights time series plots, boxplots and spatial distribution plots shown that higher PM_{10} concentrations observed under low PBL heights.

Overall, the results of this study show that the western part of the country, more populated and industrialized region, (Marmara, Aegean and also Black Sea region) have lower PM_{10} levels than the eastern part. This can be explain partly due to the differences in PBL variation. For example, in Afyon PBL and PM_{10} concentration have strong correlations. However, in other cities (e.g. Istanbul, Eskişehir) PBL can not explain the results. This suggests that complex topographycal/meteorological conditions have nonlinear impact of pollutions. In order to explain such conditions 3D atmospheric models such as CMAQ should be employed along with good quality emissions inventory data.

REFERENCES

Abdalmogith S.S, Harrison R.M., 2005, The use of trajectory cluster analysis to examine the lon-range transport of secondary inorganic aerosol in the UK. *Atmospheric Environment*, 39, **35,** 6686-6695.

Austin E., Coull B., Thomas D., Koutrakis P., 2012, A framework for identifying distinct multipollutant profiles in air pollution data, *Environment International,* **45**, 112-121.

Atkinson R., Anderson H., Sunyer J., Ayres, J. Baccini M., Vonk J., Boumghar A., Forastiere F., Forsbert B., Touloumi G., Schwartz J., Katsouyanni K., 2001, Acute effects of particulate air pollution on respiratory admissions: results from APHEA 2 project,*Am J Respir Crit Care Med*, 164:1860-1866

Avol EL., Gauderman WJ., Tan SM., London SJ., Peters JM., 2001, Respiratory Effects of Relocating to Areas of Differing Air Pollution Levels,*Am. J. Respir. Crit. Care Med.* Vol. 164, No 11, 2067-2072.

Boslaugh S., Watters P. A., 2008, *Statistics in Nutshell* (First Edition ed.).,Sebastopol,, CA, USA: O'Reilly Media, Inc.

Brook RD et al., 2004, Air pollution and cardiovascular disease: a statement for healthcare professionals from the Expert Panel on Population and Prevention Science of the American Heart Association. *Circulation,* 109:2655–2671.

EEA , (2013), Air Quality in Europe 2013, *European Environment Agency*, 978-92-9213-406-8

ESRL, (2013), Earth System Research Laboratory, Physical Science Division, *Retrieved from Website:* http://www.esrl.noaa.gov

Im U., Markakis K., Unal A., Kindap T., Poupkou A., Incecik S., Yenigun O., Melas D., Theodosi C., Mihalopoulos N., 2010, Study of a winter PM episode in Istanbul using the high resolution WRF/CMAQ modeling system, *Atmospheric Environment*, 44: 3085-3094.

Katsouyanni K., Touloumi G., Samol E., Gryparis A., Tertre A.L., Monopolis Y., Rossi G., Zmirou D., Ballester F., Boumghar A., Anderson H.R., Wojtyniak B., Paldy A., Braunstein R., Pekkanen J., Schindler C., Schwartzet J., 2001, Confounding and effect modification in the short-term effects of ambient particles on total mortality: results from 29 European cities within the APHEA-2 project,*Epidemiology*, 12:521-531.

Kindap T., Unal A., Chen S.H., Hu Y., Odman M.T., Karaca M., 2006, Long-range aerosol transport from Europe to Istanbul, Turkey,*Atmospheric Environment* 40, pp. 3536–3547.

Lua H., Chang C., Hsieh J., 2005, Classification of PM_{10} distributions in Taiwa, *Atmospheric Environment,* 40 1452–1463.

Rodŕiguez S., Querol X., Alastuey A., Viana M., Alarcón M., Mantilla E., Ruiz C.R., 2003, *Science of the Total Environment,* 328,1-3, 95-113

Tayanc M., 2000, An assessment of spatial and temporal variation of sulfur dioxide levels over Istanbul, Turkey. *Environmental Pollution* 107, 61-69.

Wilks Daniel S., 2006, Statistical Methods In The Atmospheric Sciences, *Second Edition*, pg:559-561

Perera F.P., Li Z., Whyatt R., Hoepner L., Wang S., Camann D., Rauh V., 2009, Prenatal Airborne Polycyclic Aromatic Hydrocarbon Exposure and Child IQ at Age 5 Years,"*Pediatrics* (doi:10.1542/10.1542/peds.2008-3506).

Guaderman W., McConnell R., Gilliland F., London S., Thomas D., Avol E., Vora H., Berhane K., Rappaport E.B., Lurmann F., Margolis H.G., Peters J., 2000, Association between air pollution and lung function growth in southern California children,*Am J Respir Crit Care Med*, 162:1383-1390.

Nawrot T., Nemery B., 2007, How Relevant is Air Pollution as a Trigger for Myocardial Infarction,*Epidemiology*, 18(5):S123.

NCAR, 2010, National Centre for Atmospheric Research, *ARW Version 3 Modeling System's User's Guide*, NCAR, Boulder, Colo, USA, 2010.

NCEP, 2000, National Centers for Environmental Prediction, National Weather Service, NOAA, U.S. Department of Commerce. 2000, updated daily. *NCEP FNL Operational Model*

Poloniecki J.D., Atkinson R.W., de Leon A.P., Anderson R.H., 1997, Daily time series for cardiovascular hospital admissions and previous day's air pollution in London, UK, *Occup Environ Med*, 54:535– 40.

Pope C.A., Dockery D.W., 2006, Health Effects of Fine Particulate Air Pollution: Lines That Connect, *J. Air and Waste Management*, 56:709-742.

Ren C., Tong S., 2008, Health effects of ambient air pollution – recent research development and contemporary methodological challenges,*Environmental Health*, 7:56 doi:10.1186/1476-069X-7-56.

Ruidavets J.B., Cournot M., Cassadou S., Giroux M., Meybeck M., Ferrieres J., 2005, Ozone air pollution is associated with acute myocardial infarction,*Circulation*, 111:563–9.

Seinfeld J.H., Pandis S.N., 2006, *Atmospheric chemistry and physics: from air pollution to climate change*, 2nd edition, ISBN 13 9780471720171, John Wiley & Sons Inc.

Simkhovich B.Z., Kleinman M.T., Kloner R.A., 2008, Air Pollution and Cardiovascular Injury: Epidemiology, Toxicology, and Mechanisms, *J. American College of Cardiology*, Vol 52 No 9.

Watson J.G. et al., 2002, Receptor modeling application framework for particle source apportionment. *Chemosphere,* 49:1093–1136.

WHO, 2006, Air Quality Guidelines Global Update 2005, *World Health Organization*, ISBN 92 890 2192 6

TUIK, 2013, February 28, *Turkish Statistical Institute*. *Retrieved from TUIK Website:* www.tuik.gov.tr

APPENDICES

APPENDIX A: R Programing Codes

APPENDIX B: K-Means Plots

APPENDIX C: TNO Emissions Distribution Maps

APPENDIX D: NCEP-NCAR Temperature Differences Plots

APPENDIX E: PM_{10} concentrations vs PBL heights Plots

APPENDIX F: Spatial Distribution Plots of PBL Heights

APPENDIX A: R Programming Codes

Data Processing Codes

Convert to Time Series

```
# This code was written by Seden on 03-may-2011 to create a matrix in time series data format.
# data taken from the ministry of environment and forestry, it is from 01-10-2007 to 31-12-2010 for PM10 and from 02-10-2007 to 31-12-2010 for SO2
rm(list=ls())
# Need to load Matlab library to read .mat files, XTS library to create timeseries data, and GDATA library to read Excel data.
library(R.matlab)
library(xts)
library(gdata)
#First create parameters for pollutants and time period
parameters<-c("PM10","SO2")
#Start Date of the data is 2007-10-01 for PM10, 2007-10-02 for SO2
start_date<-c("2007-10-01","2007-10-02")
start.time <- '00:00:00'
# Directory for data output
dir.create("/Users/seden/Desktop/AirQD/RData")
# Directory where the original data resides
orig_dir="/Users/seden/Desktop/AirQD/OriginalData"
# Directory where the output data (.RData) resides
Rdata_dir="/Users/seden/Desktop/AirQD/RData"
# Column Names Files in the directory of original data
xlCol=c("TurkeyPM10KeyTable2.xls","TurkeySO2KeyTable2.xls")
#Loop over Pollutants
for ( o in 1:length(parameters)){
p=parameters[o]#p=PM10
comm1=paste("data<-readMat('",orig_dir,"/",p,"_Allv6.mat')",sep="")
eval(parse(text=comm1))
comm2=paste(p,"<-data$",p,".Allv6[,]",sep="")#PM10<- data$PM10.Allv6[,]
eval(parse(text=comm2))
rm(data)
comm3=paste("ll=length(",p,"[,1])",sep="")#ll=length(PM10[,1])
```

```
eval(parse(text=comm3))
#create date and hour index
comm4=paste("start.date<-'",start_date[o],"'",sep="")#start.date <- '2007-10-01'
eval(parse(text=comm4))
interval <- 60
increment.mins <- interval * 60
x <- paste(start.date, start.time)
se2<-print(strptime(x, "%Y-%m-%d %H:%M:%S")  + c(1:ll)*increment.mins)
comm5=paste("deneme<-xts(",p,",se2)",sep="") # deneme in xts format
eval(parse(text=comm5))
comm6=paste("tt<-read.xls('",orig_dir,"/",xlCol[o],"')",sep="")#tt<-
read.xls('TurkeyPM10KeyTable2.xls')
eval(parse(text=comm6))
colnames(deneme)<-tt[,1]
maximum=10000  #max(deneme)=13945
minimum=0
deneme[deneme>maximum]=NaN #Set everything higher than 10000 to NaN
deneme[deneme<minimum]=NaN #Set everything lower than 0 to NaN
comm7=paste(p,"_xts_hourly<-deneme",sep="")
eval(parse(text=comm7))
rm(deneme)
comm8=paste("save(",p,"_xts_hourly,","file='",Rdata_dir,"/",p,"_Hourly_xts.RData')
",sep="")
eval(parse(text=comm8))
# export data to a text file==> 'write.table()'
comm9=paste("write.table(",p,"_xts_hourly,","file='",orig_dir,"/",p,"_Hourly.txt')",se
p="")
eval(parse(text=comm9))}
# ".mat" data format convert to time series for each parameter
```

Convert to Daily

```
###This code was written by Seden on 04-may-2011 to convert hourly time series
data of pollutants to daily.
rm(list=ls())
# Need to load XTS library to compute the xts data
library(xts)
```

```
#First create parameters for pollutants and time period
parameters<-c("PM10","SO2")
#to seperate data for each year, need to year range
year_min=2007
year_max=2010
# Directory where the original data resides
orig_dir="/Users/seden/Desktop/AirQD/OriginalData"
# Directory where the output data (.RData) resides
Rdata_dir="/Users/seden/Desktop/AirQD/RData"
# Loop over Pollutants
for (o in 1:length(parameters)){
        p=parameters[o]
#load hourly data from RData directory (PM10_xts_hourly)
temp1=paste("load('",Rdata_dir,"/",p,"_Hourly_xts.RData')",sep="")
eval(parse(text=temp1))
temp2=paste("colname<-colnames(",p,"_xts_hourly)",sep="")
eval(parse(text=temp2))
#create a Na matrix to write the daily data
temp3=paste(p,"_daily_xts=c()",sep="")
eval(parse(text=temp3))
#computing daily means from xts_hourly
#loop over stations
for(i in 1:length(colname)){
a=colname[i]#station name
#"endpoints" ==> specify the time index in a period (days,months,years....)
temp4=paste("ep=endpoints(",p,"_xts_hourly$",a,",'days')",sep="")
eval(parse(text=temp4))
#period.apply==>apply a specified function in a time interval which specified with
endpoints function
temp5=paste("day.Mean=period.apply(",p,"_xts_hourly$",a,",INDEX=ep,FUN=mea
n,na.rm=TRUE)",sep="")
eval(parse(text=temp5))
temp6=paste(p,"_daily_xts=cbind(",p,"_daily_xts,day.Mean)",sep="")#
eval(parse(text=temp6))
rm(day.Mean)
```

```
}#daily mean station loop
temp7=paste("colnames(",p,"_daily_xts)<-colname",sep="")
eval(parse(text=temp7))
temp8=paste("save(",p,"_daily_xts,","file='",Rdata_dir,"/",p,"_Daily_xts.RData')",sep="")
eval(parse(text=temp8))
#write data to a text file
temp9=paste("write.table(",p,"_daily_xts,","file='",orig_dir,"/",p,"_Daily.txt')",sep="")
eval(parse(text=temp9))
##to create daily data for each year
for (k in year_min:year_max){
temp33=paste(p,"_daily_",k,"<-",p,"_daily_xts['",k,"']",sep="")
eval(parse(text=temp33))
temp88=paste("save(",p,"_daily_",k,",file='",Rdata_dir,"/",p,"_Daily_",k,".RData')",sep="")
eval(parse(text=temp88))
#write data to a text file
temp99=paste("write.table(",p,"_daily_",k,",file='",orig_dir,"/",p,"_Daily_",k,".txt')", sep="")
eval(parse(text=temp99))}}#parameters loop
# hourly time series data convert to daily time series for each parameter
```

Merge Stations

```
###This code was written by Seden on 05-may-2011 to merged data of the cities which have more than one station
rm(list=ls())
# Need to load XTS library to compute the xts data and  GDATA library to read xls files
library(gdata)
library(xts)
#First create parameters for pollutants and time period
parameters<-c("PM10","SO2")
# station list files in the directory of original data
station_list=c("stationsPM10.csv","stationsSO2.csv")
#create a matrix of cities which have more than one station, cities listed from the stations list csv files.
```

```
city<-c("ADANA","ANKARA","DENIZLI","ISTANBUL","IZMIR","KAHRAMANMARAS","KAYSERI","KOCAELI","KONYA","MUGLA","SAMSUN","TRABZON")
# Need to maximum station number of the city to create matrix
#i.e. ISTANBUL has the max station number(10) for the PM10 data.
colnum=10
# Directory where the original data resides
orig_dir="/Users/Macbookpro/Desktop/AirQualityData/OriginalData"
# Directory where the output data (.RData) resides
Rdata_dir="/Users/Macbookpro/Desktop/AirQualityData/RData"
for (o in 1:length(parameters)){
p=parameters[o]
temp1=paste("load('",Rdata_dir,"/",p,"_Daily_xts.RData')",sep="")
eval(parse(text=temp1))
#stationlist.csv files include all station names(118) in first column and numbers given to the more than one stations city in the second column.
 #Need to this .csv file to find the stations location and then merge them.
temp2=paste("stations<-read.csv('",orig_dir,"/",station_list[o],"')",sep="")
eval(parse(text=temp2))
#convert the data from xts format to a matrix
temp3=paste("dataset<-as.matrix(",p,"_daily_xts)",sep="")
eval(parse(text=temp3))
#create a NA matrix to write the new merged data
kolip=paste(p,"_daily<-c()",sep="")
eval(parse(text=kolip))
##need to create station location matrix to take the stations column numbers in the original data, this stations will be removed from the original data
temp=paste("station.location<-matrix(data=NA,nrow=length(city),ncol=",colnum,")",sep="")# en çok istasyonu olan il istanbul==> ncol=10
eval(parse(text=temp))
for (i in 1:length(city)){
a<-city[i]#station name
# find the stations column numbers in the data
d=which(stations[,2]==i)
for (j in 1:length(d)){
tem=paste("station.location[",i,",",j,"]<-d[",j,"]",sep="")
```

```
eval(parse(text=tem))}#(j) loop
#create matrix for each city
temp4=paste(a,"_daily_column<-dataset[,d]",sep="")
eval(parse(text=temp4))
#create city matrix which has 1 column
temp5=paste(p,"_",a,"_daily<-matrix(",a,"_daily_column,ncol=1)",sep="")
eval(parse(text=temp5))
temp28=paste("save(",p,"_",a,"_daily",",file='",Rdata_dir,"/",p,"_",a,".RData')",sep=
"")
eval(parse(text=temp28))
tempo=paste("zo=length(",a,"_daily_column[,1])",sep="")
eval(parse(text=tempo))
##create an one column '.._daily_mean' matrix to write the means of the each stations
of a city
temp6=paste(a,"_daily_mean<-matrix(data=NA,nrow=",zo,",ncol=1)",sep="")
eval(parse(text=temp6))
## take means of the each row of the stations of a city ('..._daily_column' matrix
used) than write it to the '..._daily_mean' matrix
for (h in 1:zo){
temp7=paste("wd<-mean(",a,"_daily_column[",h,",],na.rm=TRUE)",sep="")
eval(parse(text=temp7))
temp8=paste(a,"_daily_mean[",h,",]<-wd",sep="")
eval(parse(text=temp8))}#(h) loop
 ## merged the matrix of each city
kipo=paste(p,"_daily<-cbind(",p,"_daily,",a,"_daily_mean)",sep="")
eval(parse(text=kipo))}# (i) city loop
temp9=paste("colnames(",p,"_daily)<-city",sep="")
eval(parse(text=temp9))
## the cities which have more than one station merged and written to the '..._daily'
matrix => str(PM10_daily)==>nrow=1188,ncol=12 and numeric matrix
## to remove the stations of the city from the original data
num=c("12","11","10","9","8","7","6","5","4","3","2","1")
#'num' created to take stations numbers from the station location matrix reversely.
#need to take numbers reversely to control the column number changes in the
original data
num=as.numeric(num)
for (y in 1:length(num)){
```

```
e=num[y]
#find staions column number in the original data
temp10=paste("t=which(is.na(station.location[",e,",])==0)",sep="")
eval(parse(text=temp10))
temp11=paste("rem.stat=station.location[",e,",t]",sep="")
eval(parse(text=temp11))
#remove them from original data
dataset<-dataset[,-rem.stat]}#(y) loop
# only cities which have only one station remind in the original data (dataset)
temp12=paste(p,"_single_stats<-dataset",sep="")
eval(parse(text=temp12))
#column numbers of single_stats matrix it used for the station loop to save the data
temp122=paste("sq<-length(colnames(",p,"_single_stats))",sep="")
eval(parse(text=temp122))
temp222=paste("colnm<-colnames(",p,"_single_stats)",sep="")
eval(parse(text=temp222))
#single stations cities loop to save each one
for (w in 1:sq){
a<-colnm[w]#station name
temp55=paste(p,"_",a,"_daily<-",p,"_single_stats[,",w,"]",sep="")
eval(parse(text=temp55))
temp28=paste("save(",p,"_",a,"_daily",",file='",Rdata_dir,"/",p,"_",a,".RData')",sep=
"")
eval(parse(text=temp28))}
# merge all single station cities
temp13=paste(p,"_daily_merged<-cbind(",p,"_daily,dataset)",sep="")
eval(parse(text=temp13))
comm28=paste("save(",p,"_daily_merged,",p,"_single_stats",",file='",Rdata_dir,"/",p
,"_Merged.RData')",sep="")
eval(parse(text=comm28))}#parameter loop
### '..._Merged.RData' include '.._daily_merged'(81 city daily data) and
'.._single_stats' (69 single station cities daily data)
```

K-means Analysis

```
rm(list=ls())
```

```
#"amap" is the Kmeans library which include both Euclidean and Manhattan distance choices
library(amap)
library(xts)
orig_dir="/Users/seden/Desktop/AirQD/UrbanAnalysis/OriginalData/"
RData_dir="/Users/seden/Desktop/AirQD/UrbanAnalysis/RData/"
plots_dir="/Users/seden/Desktop/AirQD/UrbanAnalysis/Plots/"
file_to_process="PM10_Urban_Mean.rds"
#kmeans_method="euclidean" #or "manhattan"
kmeans_method="euclidean"
cluster_number=5#must be determine before kmeans analysis
# this variable name will be used in saving results (clusters and index) of kmeans analysis.
processed_variableName="Kmeans_Euclidean_Mean"
color_data="kmeans_color.csv"
cluster_colors=eval(parse(text=paste("read.csv('",orig_dir,color_data,"',sep=',',header=TRUE)",sep="")))
dataset=eval(parse(text=paste("readRDS('",RData_dir,file_to_process,"')",sep="")))# (2008-2010)
Clusters=Kmeans(dataset,cluster_number,method="euclidean")
index=cbind(dataset,Clusters$cluster)
rownames(index)[74]="SANLIURFA"
temp1=paste("save(Clusters,index,file='",RData_dir,processed_variableName,".RData')",sep="")
eval(parse(text=temp1))
color_index=matrix(data=NA,nrow=length(rownames(index)),ncol=4)
color_index[,1:3]=index[,1:3]
rownames(color_index)=rownames(index)
colnames(color_index)=c("Comp1","Comp2","Clus","Color")
color_index=as.data.frame(color_index)

for (j in 1:nrow(index)){
  d=index[j,1]
  if( d <= cluster_colors$Value[1]){
    color_index[j,3]="aquamarine1"
  }else{if(d > cluster_colors$Value[1] && d<=cluster_colors$Value[2]){
    color_index[j,3]="cornflowerblue"
```

```
  }else{if(d>cluster_colors$Value[2] && d <= cluster_colors$Value[3]){
    color_index[j,3]="darkgoldenrod1"
  }else{if(d > cluster_colors$Value[3] && d <=cluster_colors$Value[4]){
    color_index[j,3]="red"
  }else{
    color_index[j,3]="black"
  }
  }}}}

for (j in 1:81){
  d=index[j,3]
  if(d==1){
    color_index[j,4]= "cornflowerblue"
  }else{if(d==2){
    color_index[j,4]="black"
  }else{if(d==3){
    color_index[j,4]="red"
  }else{if(d==4){
    color_index[j,4]="darkgoldenrod1"
  }else{
    color_index_euc[j,4]= "aquamarine1" }
  }}}}
##K-means map
library(maptools)
library(maps)
eval(parse(text=paste("png('",plots_dir,processed_variableName,".png')",sep="")))
deneme=readShapePoly("/Users/seden/Desktop/AirQualityData/shape/Turkiye.shp")
plot(deneme)
for(c in 1:81){
  b=deneme$AD[c]
  n=which(rownames(color_index)==b)#row number
  u=color_index[n,3]#color
  plot(deneme[which(deneme$AD==b),], col=u, add=T)}
dev.off()
```

Cumulative Distribution Function Plots

```
rm(list=ls())
library(ggplot2)
setwd("/Users/seden/Desktop/AirQD/UrbanAnalysis/RData/")
#load("ManMean_Man4prm_PCA_Clusters.RData")
load("Euc_mean_4prm_Clusters.RData")
load("PM10_Merged_Urban.RData")
colnames(PM10_Merged_Urban)[74]="SANLIURFA"
newdata=Euc_4prm_Cluster
cities=rownames(newdata)
newdata$Cluster_name=as.factor(newdata$Cluster_name)
Clean_clus=newdata[which(newdata$Cluster_name=="C"),]
b_min=which(newdata$Mean==min(Clean_clus$Mean))
b_max=which(newdata$Mean==max(Clean_clus$Mean))
Clean_min_max=cities[c(b_min,b_max)]
rm(b_min,b_max)
Cleanest_clus=newdata[which(newdata$Cluster_name=="CT"),]
b_min=which(newdata$Mean==min(Cleanest_clus$Mean))
b_max=which(newdata$Mean==max(Cleanest_clus$Mean))
Cleanest_min_max=cities[c(b_min,b_max)]
rm(b_min,b_max)
Medium_clus=newdata[which(newdata$Cluster_name=="M"),]
b_min=which(newdata$Mean==min(Medium_clus$Mean))
b_max=which(newdata$Mean==max(Medium_clus$Mean))
Medium_min_max=cities[c(b_min,b_max)]
rm(b_min,b_max)
Polluted_clus=newdata[which(newdata$Cluster_name=="P"),]
b_min=which(newdata$Mean==min(Polluted_clus$Mean))
b_max=which(newdata$Mean==max(Polluted_clus$Mean))
Polluted_min_max=cities[c(b_min,b_max)]
rm(b_min,b_max)
HPolluted_clus=newdata[which(newdata$Cluster_name=="HP"),]
b_min=which(newdata$Mean==min(HPolluted_clus$Mean))
b_max=which(newdata$Mean==max(HPolluted_clus$Mean))
HPolluted_min_max=cities[c(b_min,b_max)]
rm(b_min,b_max)
```

```
#Merge data for these cities
a=PM10_Merged_Urban[,Clean_min_max]
#Cleans=rowMeans(a,na.rm=TRUE)
Cleans=as.data.frame(a)
rm(a)
a=PM10_Merged_Urban[,Cleanest_min_max]
#MostCleans=rowMeans(a,na.rm=TRUE)
MostCleans=as.data.frame(a)
rm(a)
a=PM10_Merged_Urban[,Medium_min_max]
#Mediums=rowMeans(a,na.rm=TRUE)
Mediums=as.data.frame(a)
rm(a)
a=PM10_Merged_Urban[,Polluted_min_max]
#Polluteds=rowMeans(a,na.rm=TRUE)
Polluteds=as.data.frame(a) #sanlıurfa needs to be changed to urfa
rm(a)
a=PM10_Merged_Urban[,HPolluted_min_max]
#HighlyPolluteds=rowMeans(a,na.rm=TRUE)
HighlyPolluteds=as.data.frame(a) #sanlıurfa needs to be changed to urfa
OverallMeans=cbind(MostCleans,Cleans,Mediums,Polluteds,HighlyPolluteds)
#legend_labels=colnames(OverallMeans)
colnames(OverallMeans)=c("CT_min","CT_max","C_min","C_max","M_min","M_
max","P_min","P_max","HP_min","HP_max")
library(ggplot2)
require(reshape2)
GroupsData=melt(OverallMeans)
colnames(GroupsData)=c("Cluster_name","PM10")

cols <- c("CT_min" = "green","CT_max" = "green",
        "C_min" = "lightblue3","C_max" = "lightblue3",
        "M_min" = "orange", "M_max" = "orange",
        "P_min" = "peachpuff1", "P_max" = "peachpuff1",
        "HP_min" = "white","HP_max" = "white",
"CT_min.mean"="green","CT_min.mean"="green","C_min.mean"="lightblue3","C_
max.mean"="lightblue3",
```

```
"M_min.mean"="orange","M_max.mean"="orange","P_min.mean"="peachpuff1","P
_max.mean"="peachpuff1","HP_min.mean"="white","HP_max.mean"="white")

tip<-
c("CT_min"=2,"CT_max"=2,"C_min"=2,"C_max"=2,"M_min"=2,"M_max"=2,"P_
min"=2,"P_max"=2,"HP_min"=2,"HP_max"=2,
"CT_min.mean"=1,"CT_max.mean"=1,"C_min.mean"=1,"C_max.mean"=1,"M_min.
mean"=1,"M_max.mean"=1,"P_min.mean"=1,"P_max.mean"=1,"HP_min.mean"=1,
"HP_max.mean"=1)

shp<-
c("CT_min"=4,"CT_max"=4,"C_min"=4,"C_max"=4,"M_min"=4,"M_max"=4,"P_
min"=4,"P_max"=4,"HP_min"=4,"HP_max"=4,
"CT_min.mean"=16,"CT_max.mean"=16,"C_min.mean"=16,"C_max.mean"=16,"M
_min.mean"=16,"M_max.mean"=16,"P_min.mean"=16,"P_max.mean"=16,"HP_min
.mean"=16,"HP_max.mean"=16)

pp2=ggplot(GroupsData, aes(x=PM10,
color=Cluster_name,linetype=Cluster_name,shape=Cluster_name))+

  stat_ecdf(size=1)+
geom_point(aes(x=mean(GroupsData$PM10[GroupsData$Cluster_name=="CT_min
"],na.rm=TRUE),

 color="CT_min.mean",linetype="CT_min.mean",shape="CT_min.mean",
y=.35),size = 5)+

  coord_cartesian(xlim=c(150,1000))+
geom_point(aes(x=mean(GroupsData$PM10[GroupsData$Cluster_name=="CT_ma
x"],na.rm=TRUE),

 color="CT_max.mean",linetype="CT_max.mean",shape="CT_max.mean",

  y=.35),size = 5) +

  coord_cartesian(xlim=c(150,1000))+

geom_point(aes(x=mean(GroupsData$PM10[GroupsData$Cluster_name=="C_min"
],na.rm=TRUE),

color="C_min.mean",linetype="C_min.mean",shape="C_min.mean",

   y=.5),size = 5) +

  coord_cartesian(xlim=c(150,1000))+

geom_point(aes(x=mean(GroupsData$PM10[GroupsData$Cluster_name=="C_max"
],na.rm=TRUE),

color="C_max.mean",linetype="C_max.mean",shape="C_max.mean",

 y=.5),size = 5) +

 coord_cartesian(xlim=c(150,1000))+

geom_point(aes(x=mean(GroupsData$PM10[GroupsData$Cluster_name=="M_min"
],na.rm=TRUE),

color="M_min.mean",linetype="M_min.mean",shape="M_min.mean",

y=.5),size = 5) +
```

```
 coord_cartesian(xlim=c(150,1000))+
geom_point(aes(x=mean(GroupsData$PM10[GroupsData$Cluster_name=="M_max
"],na.rm=TRUE),
color="M_max.mean",linetype="M_max.mean",shape="M_max.mean",
y=.5),size = 5) +
coord_cartesian(xlim=c(150,1000))+
geom_point(aes(x=mean(GroupsData$PM10[GroupsData$Cluster_name=="P_min"]
,na.rm=TRUE),
 color="P_min.mean",linetype="P_min.mean",shape="P_min.mean",
 y=.62),size = 5) +
coord_cartesian(ylim=c(150,1000))+
geom_point(aes(x=mean(GroupsData$PM10[GroupsData$Cluster_name=="P_max"
],na.rm=TRUE),
 color="P_max.mean",linetype="P_max.mean",shape="P_max.mean",
  y=.62),size = 5) +
  coord_cartesian(xlim=c(150,1000))+
geom_point(aes(x=mean(GroupsData$PM10[GroupsData$Cluster_name=="HP_min
"],na.rm=TRUE),
color="HP_min.mean",linetype="HP_min.mean",shape="HP_min.mean",
 y=.58),size = 5) +
 coord_cartesian(xlim=c(150,1000))+
geom_point(aes(x=mean(GroupsData$PM10[GroupsData$Cluster_name=="HP_ma
x"],na.rm=TRUE),
color="HP_max.mean",linetype="HP_max.mean",shape="HP_max.mean",
y=.58),size = 5) +
  coord_cartesian(xlim=c(150,1000))+
  scale_color_manual(values=cols)+
  scale_linetype_manual(values=tip) +
  scale_shape_manual(values=shp)+
  ylab("Cumulative Probability")+
  xlab("PM10 (ug/m3)")+
theme(panel.background = element_rect(fill = "gray46"))+
theme(legend.justification=c(1,0), legend.position=c(1,0))
setwd("/Users/seden/Desktop/AirQD/UrbanAnalysis/Plots/CDFs/")
ggsave(pp2, file="Euc_mean _CDF.pdf", width=8,height=8)
```

Cumulative Distriburion Function Plots of Clusters

```
OverallMeans2=OverallMeans-OverallMeans$MostCleans
```

```
y2=rep(0,1096)
x=1:1096
yot=paste("ManMean_TimeSeries.ps",sep="")
eval(postscript(file=yot,paper="letter"))
op <- par(mar = par("mar")/2)
par(mar=c(2.55,3.0,1.0,1.05))
par(mai=c(0.4,0.7,0.1,0.1))
plot.new()
par(mfrow=c(5,1))
plot(x,OverallMeans$MostCleans,type='n',lty=1,lwd=1.5,axes=FALSE,
ann=FALSE,col="green",ylim=c(0,200))
lines(x,OverallMeans$MostCleans,lty=1,lwd=1.5,col="green")
lines(x,y2,lty=1,lwd=1.5,col="white")
polygon(c(x, rev(x)), c(OverallMeans$MostCleans, rev(y2)),
        col = "gray97", border = NA)
#axis(1,at=c(90,180,270,360,450,540,630,720,810,900,990,1096),lab=c("1-3","1-
6","1-9","1-12","2-3","2-6","2-9","2-12","3-3","3-6","3-9","3-12"))
axis(2, las=1, at=c(0,50,100,150,200))
box()
abline(v=c(90,180,270,360,450,540,630,720,810,900,990,1096), col="black",
lty=1,lwd=0.5)
abline(h=c(0,50,100,150,200), col="gray", lty="dotted")
plot(x,OverallMeans2$Cleans,type='n',lty=1,lwd=1.5,axes=FALSE,
ann=FALSE,col="dodgerblue2",ylim=c(0,200))
lines(x,OverallMeans$Cleans,lty=1,lwd=1.5,col="dodgerblue2")
lines(x,y2,lty=1,lwd=1.5,col="white")
polygon(c(x, rev(x)), c(OverallMeans$Cleans, rev(y2)),
        col = "gray97", border = NA)
#axis(1,at=c(90,180,270,360,450,540,630,720,810,900,990,1096),lab=c("1-3","1-
6","1-9","1-12","2-3","2-6","2-9","2-12","3-3","3-6","3-9","3-12"))
axis(2, las=1, at=c(0,50,100,150,200))
box()
abline(v=c(90,180,270,360,450,540,630,720,810,900,990,1096), col="black",
lty=1,lwd=0.5)
abline(h=c(0,50,100,150,200), col="gray", lty="dotted")
#abline(h=c(0), col="black", lty=1)
```

```
plot(x,OverallMeans2$Mediums,type='n',lty=1,lwd=1.5,axes=FALSE,
ann=FALSE,col="darkorange2",ylim=c(0,200))
lines(x,OverallMeans$Mediums,lty=1,lwd=1.5,col="darkorange2")
lines(x,y2,lty=1,lwd=1.5,col="white")
polygon(c(x, rev(x)), c(OverallMeans$Mediums, rev(y2)),
        col = "gray97", border = NA)
#axis(1,at=c(90,180,270,360,450,540,630,720,810,900,990,1096),lab=c("1-3","1-
6","1-9","1-12","2-3","2-6","2-9","2-12","3-3","3-6","3-9","3-12"))
axis(2, las=1, at=c(0,50,100,150,200))
box()
abline(v=c(90,180,270,360,450,540,630,720,810,900,990,1096), col="black",
lty=1,lwd=0.5)
abline(h=c(0,50,100,150,200), col="gray", lty="dotted")
#abline(h=c(0), col="black", lty=1)
plot(x,OverallMeans2$Polluteds,type='n',lty=1,lwd=1.5,axes=FALSE,
ann=FALSE,col="firebrick3",ylim=c(0,300))
lines(x,OverallMeans$Polluteds,lty=1,lwd=1.5,col="firebrick3")
lines(x,y2,lty=1,lwd=1.5,col="white")
polygon(c(x, rev(x)), c(OverallMeans$Polluteds, rev(y2)),
        col = "gray97", border = NA)
#axis(1,at=c(90,180,270,360,450,540,630,720,810,900,990,1096),lab=c("1-3","1-
6","1-9","1-12","2-3","2-6","2-9","2-12","3-3","3-6","3-9","3-12"))
axis(2, las=1, at=c(0,100,200,300))
box()
abline(v=c(90,180,270,360,450,540,630,720,810,900,990,1096), col="black",
lty=1,lwd=0.5)
abline(h=c(0,100,200,300), col="gray", lty="dotted")
#abline(h=c(0), col="black", lty=1)
plot(x,OverallMeans2$HighlyPolluteds,type='n',lty=1,lwd=1.5,axes=FALSE,
ann=FALSE,col="gray12",ylim=c(0,300))
lines(x,OverallMeans$HighlyPolluteds,lty=1,lwd=1.5,col="gray12")
lines(x,y2,lty=1,lwd=1.5,col="white")
polygon(c(x, rev(x)), c(OverallMeans$HighlyPolluteds, rev(y2)),
        col = "gray97", border = NA)
axis(1,at=c(90,180,270,360,450,540,630,720,810,900,990,1096),lab=c("1-3","1-
6","1-9","1-12","2-3","2-6","2-9","2-12","3-3","3-6","3-9","3-12"))
axis(2, las=1, at=c(0,100,200,300))
box()
```

```
abline(v=c(90,180,270,360,450,540,630,720,810,900,990,1096), col="black", lty=1,lwd=0.5)
abline(h=c(0,100,200,300), col="gray", lty="dotted")
#abline(h=c(0), col="black", lty=1)
par(op)
dev.off()
```

Emossions Data Plots

```
rm(list=ls())
library(R.matlab)
# %S1 POW Combustion in energy and transformation industries
# %S2 RES Non-industrial combustion plants
# %S3 IND Combustion in manufacturing industry
# %S4 PRO Production processes
# %S5 FFE Extraction and distribution of fossil fuels and geothermal energy
# %S6 SOL Solvent and other product use
# %S7 ROAD Road transport
# %S8 MOB Other mobile sources and machinery
# %S9 WAS Waste treatment and disposal
# %S10 AGR Agriculture
# %Species: CO, NH3, NMVOC, NOX, PMcoarse, PM25, SOx
data=readMat("/Users/seden/Desktop/AirQD/UrbanAnalysis/TNO_emissions/emi_turkey_city_tno_orig2005_v6.mat")
Emissions=data$tot.emi.city
PM_only=Emissions[,5,]+Emissions[,6,]
station_names=as.character(unlist(data$Cnames)) #attr(x=Sec3_PM,"names")
station_names[c(8,13,17,21,27,35,42,45,59,80)]=c("KARABUK","CORUM","DUZCE","CANKIRI","GUMUSHANE","CANAKKALE","KUTAHYA","BINGOL","KAHRAMANMARAS","ICEL")
Sector_PM=c()
for(i in 1:10){
  temp=PM_only[,i]
  Sector_PM=cbind(Sector_PM,temp)
  rm(temp)}
Sector_PM=as.data.frame(Sector_PM)
rownames(Sector_PM)=station_names
rm(i)
```

```
Sector_all_PM=cbind(Sector_PM,rowSums(Sector_PM))
colnames(Sector_all_PM)=c("Sec_1","Sec_2","Sec_3","Sec_4","Sec_5","Sec_6","Sec_7","Sec_8","Sec_9","Sec_10","Sec_all")
quantiles=apply(Sector_all_PM,2,quantile,probs=c(25,50,75,95)/100,na.rm=TRUE)
Sec_all_quantiles=t(quantiles)
color_index=matrix(data=NA,nrow=nrow(PM_only),ncol=10)
color_index=as.data.frame(color_index)
for(j in 1:length(Sector_all_PM)){
 for(k in 1:81){
 if(Sector_all_PM[k,j]<= Sec_all_quantiles[j,1]){
   color_index[k,j]="bisque2"
 }else if(Sector_all_PM[k,j]< Sec_all_quantiles[j,1] & Sector_all_PM[k,j]<=
Sec_all_quantiles[j,2]){
   color_index[k,j]="darkgoldenrod1"
 }else if(Sector_all_PM[k,j]< Sec_all_quantiles[j,2] & Sector_all_PM[k,j]<=
Sec_all_quantiles[j,3]){
   color_index[k,j]="darkorange"
 }else if(Sector_all_PM[k,j]< Sec_all_quantiles[j,3] & Sector_all_PM[k,j]<=
Sec_all_quantiles[j,4]){
   color_index[k,j]="darkorange3"
 }else{
   color_index[k,j]="darkorange4" }
   }#k loop/ station}#j loop/ Sector
rownames(color_index)=station_names
deneme=readShapePoly("/Users/seden/Desktop/AirQualityData/shape/Turkiye.shp")
for(g in 1:11){
eval(parse(text=paste("pdf('/Users/seden/Desktop/AirQD/UrbanAnalysis/Plots/TNO_emi/Sec_",g,".pdf')",sep="")))
 plot(deneme)
 for(c in 1:81){
 b=deneme$AD[c]
 n=which(rownames(color_index)==b)#row number
 u=color_index[n,g]#color
plot(deneme[which(deneme$AD==b),], col=u, add=T)  }#c loop/station
if(g==11){
par(mai=c(1.0,1.5,4.75,1.0))
title(main=list('TNO Sectors Sum Emission Map'),cex=0.9)
```

dev.off() }else{par(mai=c(1.0,1.5,4.75,1.0))

eval(parse(text=paste("title(main=list('TNO Sector ",g," Emission Map'),cex=0.9)",sep="")))

dev.off() }#if}# g loop/ sector

NCEP-NCAR Analysis

```
rm(list=ls())
## Not run:
#library(RNCEP)
setwd("/Users/seden/Desktop/AirQD/UrbanAnalysis/November_2013/")
## Retrieve the temperature from a particular pressure level for
## a specified spatial and temporal extent
#pressure level, sigma levels and single level files.17 pressure levels (hPa): 1000, 925, 850, 700, 600, 500, 400, 300, 250, 200, 150, 100, 70, 50, 30, 20, 10
load("NCEP.RData")
# temp700<- NCEP.gather(variable='air', level=700,
#                    months.minmax=c(1,12), years.minmax=c(2008,2010),
#                    lat.southnorth=c(30,50), lon.westeast=c(20,50),
#                    reanalysis2 = TRUE, return.units = TRUE)
# temp850 <- NCEP.gather(variable='air', level=850,
#                      months.minmax=c(1,12), years.minmax=c(2008,2010),
#                      lat.southnorth=c(30,50), lon.westeast=c(20,50),
#                      reanalysis2 = TRUE, return.units = TRUE)
# temp1000 <- NCEP.gather(variable='air', level=1000,
#                     months.minmax=c(1,12), years.minmax=c(2008,2010),
#                     lat.southnorth=c(30,50), lon.westeast=c(20,50),
#                     reanalysis2 = TRUE, return.units = TRUE)
# temp925 <- NCEP.gather(variable='air', level=925,
#                     months.minmax=c(1,12), years.minmax=c(2008,2010),
#                     lat.southnorth=c(30,50), lon.westeast=c(20,50),
#                     reanalysis2 = TRUE, return.units = TRUE)
save(temp1000, temp925, temp850,temp700,file = "NCEP.RData")
#p = 101325 (1 - (2.25577x10^-5)xh)^5.25588            (1)
#z=44330.8 - 4946.54xP^0.1902632
#where
#p = air pressure (Pa)
#h = altitude above sea level (m)
```

```
p=c(1000,925,850,700)
p=p*100#convert from hectopascals to pascals
z=44330.8-(4946.54*(p^0.1902632))
# 1000hPa=100kPa==0 meters
# 925hPa=92.5kPa==762 meters
# 850hPa=85kPa==1400 meters
# 700hPa=70kPa==3000meters
#data<-read.csv(file = "NY subsample.csv")
lat=dimnames(temp1000)[[1]]
lon=dimnames(temp1000)[[2]]
lat=as.numeric(lat)
lon=as.numeric(lon)
require(reshape2)
lat_tur=lat[4:6]
lon_tur=lon[4:11]
temp1000_tur=temp1000[4:6,4:11,]
temp925_tur=temp925[4:6,4:11,]
temp850_tur=temp850[4:6,4:11,]
temp700_tur=temp700[4:6,4:11,]
tempDiff850_700=temp850_tur-temp700_tur
a=melt(tempDiff850_700)
rm(tempDiff850_700)
tempDiff850_700=a
Hakkari_NCEP_850=temp850_tur[3,8,]
Hakkari_NCEP_700=temp700_tur[3,8,]
HakkDiff850_700=Hakkari_NCEP_850 - Hakkari_NCEP_700
HakkDiff850_700=as.matrix(HakkDiff850_700)
x=1:4384
plot(x,HakkDiff850_700[,1],type="l")
colnames(tempDiff850_700)=c("lat","lon","tempdiff")
#temp$tempdiff=temp$temp1000-273.15
require('ggmap')
map.in <- get_map(location = c(20, 30, 50, 50),
          maptype="terrain",source = "google")
theme_set(theme_bw(base_size = 8))
```

```
colormap <- c("Darkblue","Blue","White","Yellow","Red")
alper=ggmap(map.in) %+% tempDiff850_700 +
  aes(x = lon, y = lat, z = tempdiff) +
  geom_tile(aes(lon, lat, fill=tempdiff),alpha=0.6, data=tempDiff850_700) +
  scale_fill_gradientn(name = "Max",
colours = colormap,limits=c(-20,20))
ggsave(filename = "tempDiff850_700.png",
       plot = alper,
       scale = 1,
       width = 6, height = 4,
       dpi = 600)
tempo1000=temp1000_tur[3,8,]
tempo850=temp850_tur[3,8,]
tempo925=temp925_tur[3,8,]
tempo700=temp700_tur[3,8,]
inversion_data=as.data.frame(readRDS("/Users/seden/Desktop/AirQD/UrbanAnalysi
s/OZAN_DATA/Iversion_HP_P.rds"))
Hakkari_temp=inversion_data$HAKKARI
min_temp=Hakkari_temp #Add for env. lapse rate
min_temp=as.data.frame(min_temp)
min_Temp_NCEP=temp700_tur[3,8,seq(1,4384,4)]-273.15
tempodiff=min_temp-min_Temp_NCEP
dd=which(tempodiff<=-10)
x=1:1096
png("/Users/seden/Desktop/AirQD/UrbanAnalysis/OZAN_DATA/inversion_hakkari
-700mb.png")
plot(x,tempodiff$min_temp,type="l",xlab="index",ylab="Temp
Diff",main="Hakkari-700 mb")
dev.off()
PM10_merged=readRDS("/Users/seden/Desktop/AirQD/UrbanAnalysis/RData/PM1
0_Merged_Urban.rds")
hakkari_pm10=PM10_merged[,"HAKKARI"]
png("/Users/seden/Desktop/AirQD/UrbanAnalysis/OZAN_DATA/Hakkari_PM&Te
mpDiff700.png")
plot(as.matrix(hakkari_pm10[,1]),as.matrix(tempodiff),xlab="PM10",ylab="Temp
Diff",main="Hakkari PM10 & Temp Diff")
dev.off()
```

```
png("/Users/seden/Desktop/AirQD/UrbanAnalysis/OZAN_DATA/Hakkari_PM_timeSeries.png")
plot(hakkari_pm10,type="l")
dev.off()
Mardin_temp=inversion_data$MARDIN
min_temp_mardin=Mardin_temp #Add for env. lapse rate
min_temp_mardin=as.data.frame(min_temp_mardin)
min_Temp_NCEP=temp850_tur[3,8,seq(1,4384,4)]-273.15
tempodiff_mardin=min_temp_mardin-min_Temp_NCEP
x=1:1096
png("/Users/seden/Desktop/AirQD/UrbanAnalysis/OZAN_DATA/inversion_mardin.png")
plot(x,tempodiff_mardin$min_temp,type="l")
dev.off()
mardin_pm10=PM10_merged[,"MARDIN"]
plot(as.matrix(mardin_pm10[,1]),as.matrix(tempodiff_mardin))
```

PBL heights Boxplot

```
rm(list=ls())
library(ncdf4)
library("xts")
stations_nc=c("izmir","hakkari","afyon","van","kmaras","bolu","istanbul","duzce","eskisehir")
stations_pm=c("IZMIR","HAKKARI","AFYON","VAN","KAHRAMANMARAS","BOLU","ISTANBUL","DUZCE","ESKISEHIR")
for(i in 1:length(stations_nc)){
city=stations_nc[i]
city_x=stations_pm[i]
setwd("/Users/AlperUnal/Alper/Projects/Students/Seden/November_2013/Afyon/")
fil=paste(city,"_pblh_Correct.nc",sep='')
 nc <- nc_open(fil)
 pblh <- ncvar_get( nc,"PBLH" )
 nc_close(nc)
start.date=c("2007-12-31")
start.time=c("23:00:00")
interval=60
increment.mins=interval*60
```

```
x=paste(start.date,start.time)
se2=print(strptime(x,"%Y-%m-%d %H:%M:%S") + c(1:8785)*increment.mins)
deneme=xts(pblh,se2)
ep=endpoints(deneme,"days")
daily_pblh=period.apply(deneme,INDEX=ep,FUN=mean,na.rm=TRUE)
ep2=endpoints(deneme,"months")
monthly_pblh=period.apply(deneme,INDEX=ep2,FUN=mean,na.rm=TRUE)
setwd("/Users/AlperUnal/Alper/Projects/Students/Seden/November_2013/")
load("pm10dataclusters//PM10_Merged_Urban.RData")
colnames(PM10_Merged_Urban)[74]="SANLIURFA"
d=paste("a=PM10_Merged_Urban$",city_x,sep="")
eval(parse(text=d))
daily_PM=a["2008-"]
daily_pbl=daily_pblh[1:366]
rm(monthly_pblh,daily_pblh,PM10_Merged_Urban,a,deneme,ep,ep2,fil,increment.m
ins,interval,nc,pblh,se2,start.date,start.time,x)
#eval(parse(text=paste("pdf('",city_x,"_PBLvsPM.pdf')",sep="")))
PBL=as.data.frame(daily_pbl)
colnames(PBL)=c('pbl')
PM10=as.data.frame(daily_PM)
colnames(PM10)=c('PM10')
#plot(PBL$pbl,PM10$PM10)
#dev.off()
setwd("/Users/AlperUnal/Alper/Projects/Students/Seden/November_2013/Boxplots_
PM_PBL/")
#library(psych)
dene=cbind(PM10$PM10,PBL$pbl)
colnames(dene)=c('PM10','PBL')
heights=c(0,50,100,200,300,400,500,600,700,800,900,1000,2000)
fiti=which(max(dene[,"PBL"])> heights)
lim=c(heights[fiti],heights[max(fiti)+1])
dene2=data.frame(dene, bin=cut(dene[,2], lim, include.lowest=TRUE))
summary(dene2$bin)
eval(parse(text=paste("pdf('",city_x,"_Boxplot_PM&PBL.pdf')",sep="")))
boxplot(dene2$PM10~dene2$bin,axes=FALSE,ylab="PM10 ug/m3",xlab="PBL
Heights (m)")
```

```
men=tapply( dene2$PM10, dene2$bin, mean, na.rm=T )
points(men,pch=19,col='red')
abline(v=seq(0,length(lim),1),col="gray",lty="dotted")
abline(h=seq(0,max(lim),50),col="gray",lty="dotted")
axis(1,seq(0,length(lim)-1,1),lim[1:length(lim)],cex.axis=0.6)
axis(2,seq(0,max(dene2$PM10,na.rm=TRUE),50),cex.axis=0.6)
box()
dev.off()}#nc loop
```

PBL Heights Histogram

```
rm(list=ls())
library(ncdf4)
library(xts)
stations_nc=c("izmir","hakkari","afyon","van","kmaras","bolu","istanbul","duzce","eskisehir")
stations_pm=c("IZMIR","HAKKARI","AFYON","VAN","KAHRAMANMARAS","BOLU","ISTANBUL","DUZCE","ESKISEHIR")
for(i in 1:length(stations_nc)){
city=stations_nc[i]
city_x=stations_pm[i]
setwd("/Users/seden/Desktop/AirQD/UrbanAnalysis/Plots/PBL_09_12_AU/Afyon/)
fil=paste(city,"_pblh_Correct.nc",sep=")
 nc <- nc_open(fil)
 pblh <- ncvar_get( nc,"PBLH" )
 nc_close(nc)
start.date=c("2007-12-31")
start.time=c("23:00:00")
interval=60
increment.mins=interval*60
x=paste(start.date,start.time)
se2=print(strptime(x,"%Y-%m-%d %H:%M:%S") + c(1:8785)*increment.mins)
deneme=xts(pblh,se2)
ep=endpoints(deneme,"days")
daily_pblh=period.apply(deneme,INDEX=ep,FUN=mean,na.rm=TRUE)
ep2=endpoints(deneme,"months")
monthly_pblh=period.apply(deneme,INDEX=ep2,FUN=mean,na.rm=TRUE)
# setwd("/Users/AlperUnal/Alper/Projects/Students/Seden/November_2013/")
```

```
# load("pm10dataclusters//PM10_Merged_Urban.RData")
load("/Users/seden/Desktop/AirQD/UrbanAnalysis/RData/PM10_Merged_Urban.R
Data")
colnames(PM10_Merged_Urban)[74]="SANLIURFA"
d=paste("a=PM10_Merged_Urban$",city_x,sep="")
eval(parse(text=d))
daily_PM=a["2008-"]
daily_pbl=daily_pblh[1:366]
rm(monthly_pblh,daily_pblh,PM10_Merged_Urban,a,deneme,ep,ep2,fil,increment.m
ins,interval,nc,pblh,se2,start.date,start.time,x)
#eval(parse(text=paste("pdf('",city_x,"_PBLvsPM.pdf')",sep="")))
PBL=as.data.frame(daily_pbl)
colnames(PBL)=c('pbl')
PM10=as.data.frame(daily_PM)
colnames(PM10)=c('PM10')
setwd("/Users/seden/Desktop/AirQD/UrbanAnalysis/Plots/PBL_09_12_AU/Histogra
ms_PM_PBL/")
dene=cbind(PM10$PM10,PBL$pbl)
colnames(dene)=c('PM10','PBL')
heights=c(0,50,100,200,300,400,500,600,700,800,900,1000,2000)
fiti=which(max(dene[,"PBL"])> heights)
lim=c(heights[fiti],heights[max(fiti)+1])
dene2=data.frame(dene, bin=cut(dene[,2], lim, include.lowest=TRUE))
pp=paste("png('",city_x,"_Histogram_PM&PBL.png')",sep="")
eval(parse(text=pp))
library(lattice)
histogram(~ PM10 | factor(bin), type='count',data = dene2,xlim=seq(0,500,50),
  panel=function(x,params,...){
  panel.abline(v=seq(0,500,50),h=seq(0,40,5),col="grey",lty="dotted")
  panel.histogram(x,...,col="red")})
dev.off()}#nc loop
```

APPENDIX B: K-Means Plots

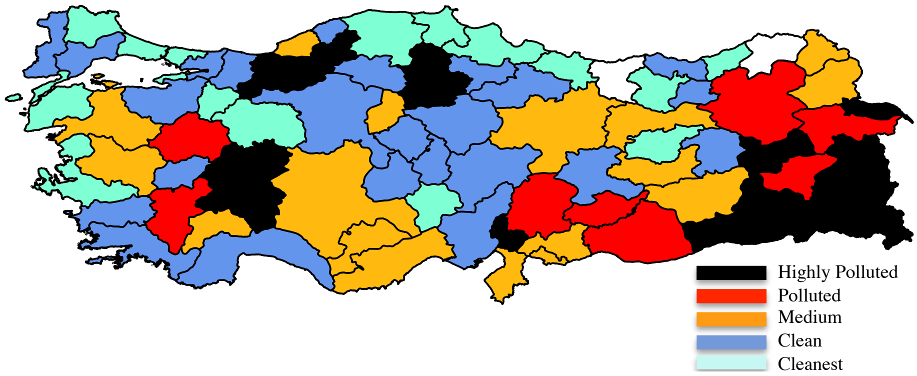


Figure B.1: Mahattan distance K-means clustering colored map of three years with 4 parameters : mean, 2.5 %, median, 97.5 %.

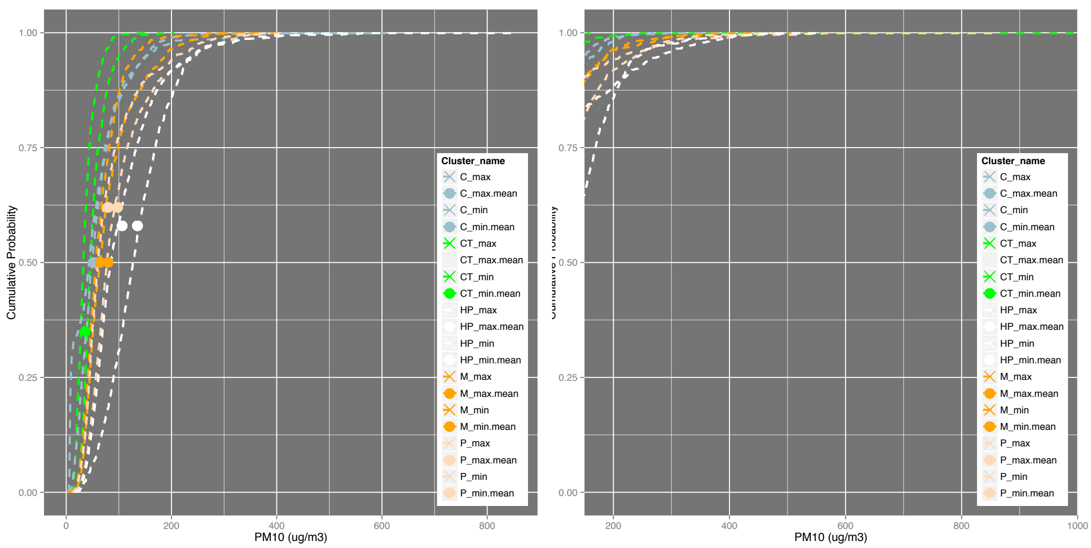


Figure B.2: a) Cumulative distribution functions (CDFs) of the Manhattan distance k-means clusters with four parameter. b) Zoom to the CDFs above 200 μg/m^3.

APPENDIX C: TNO Emissions Distribution Maps

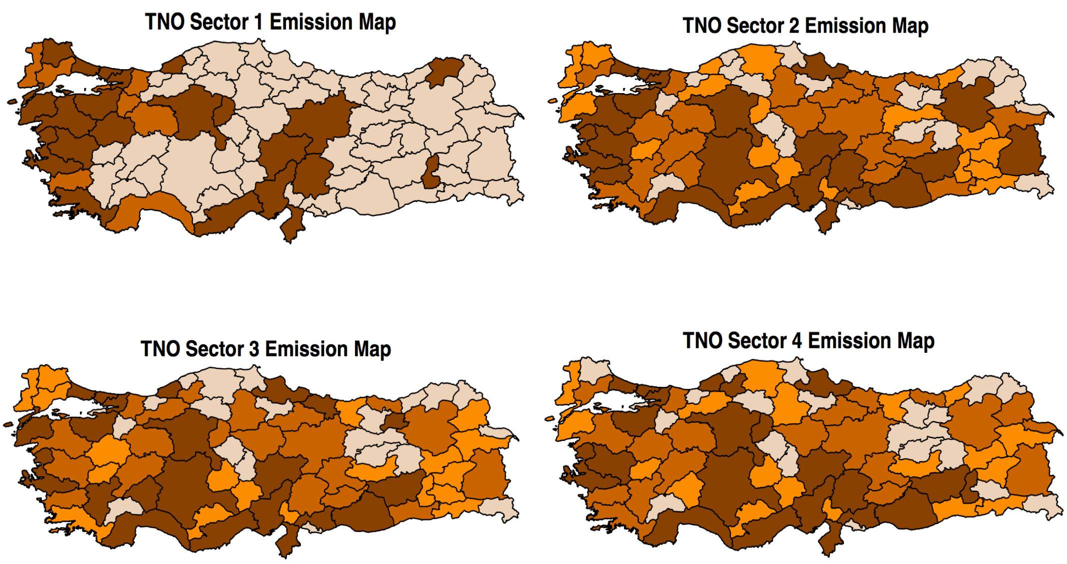


Figure C.1 : TNO emissions map of SNAP 1 (Combustion in energy transformation industries (POW)), SNAP 2 (Non-industrial combustion plant (RES)), SNAP 3 (Combustion in manufacturing industry (IND)) and SNAP 4 (Production Processes (PRO)).

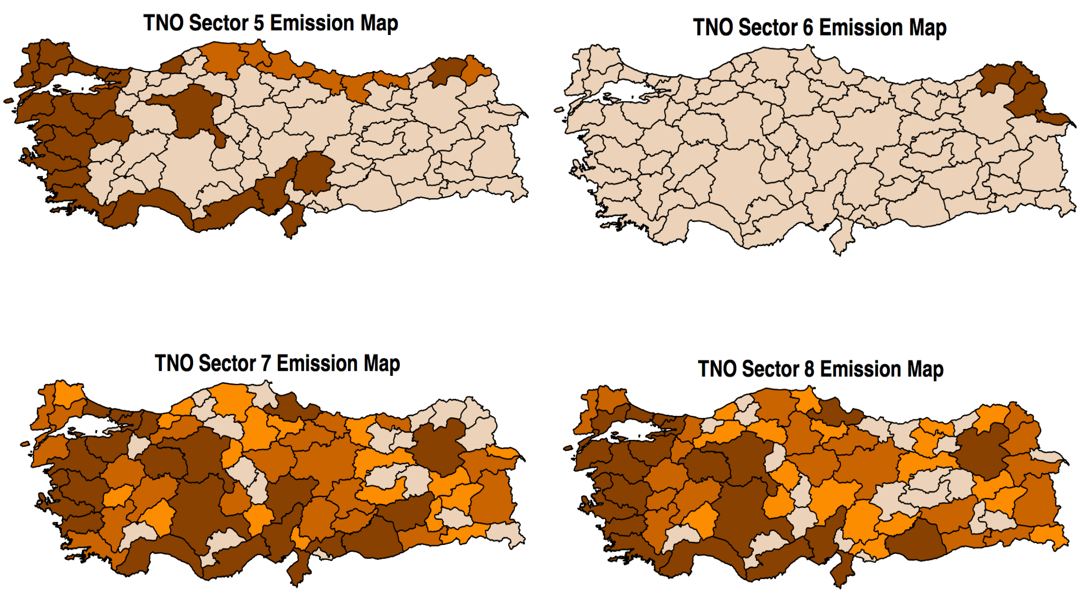


Figure C.2 : TNO emissions map of SNAP 5 (Extraction and distribution of fossil fuels and geothermal energy (FFE)), SNAP 6 (Solvent and other product use (SOL)), SNAP 7 (Road transport (ROAD)) and SNAP 8 (Other mobile sources and machinery (MOB)).

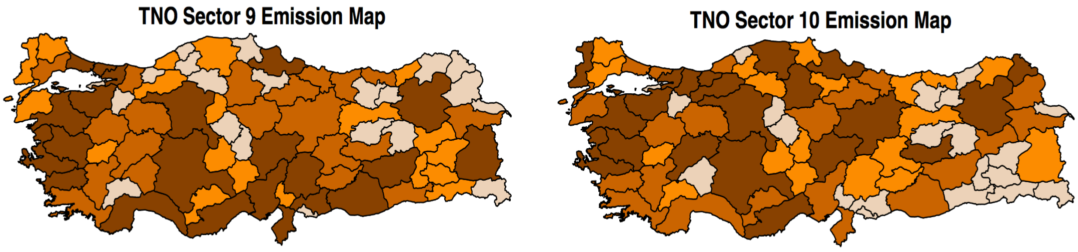


Figure C.3 : TNO emissions map of SNAP 9 (Waste treatment and disposal (WAS)) and SNAP 10 (Agriculture (AGR)).

APPENDIX D : NCEP-NCAR Temperature Differences Plots

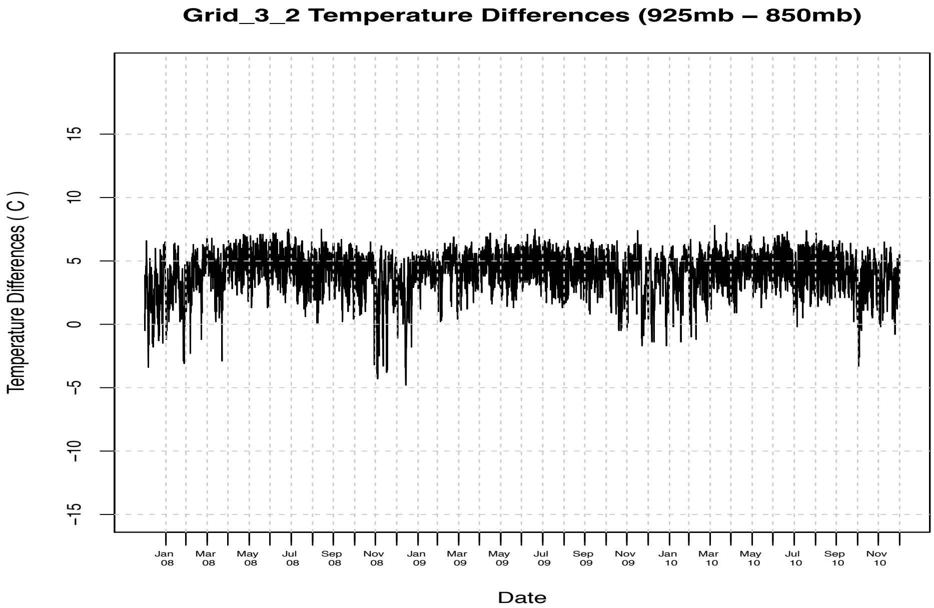


Figure D.1 : 925 mb-850 mb temperature differences in Afyon, Isparta and Antalya grid

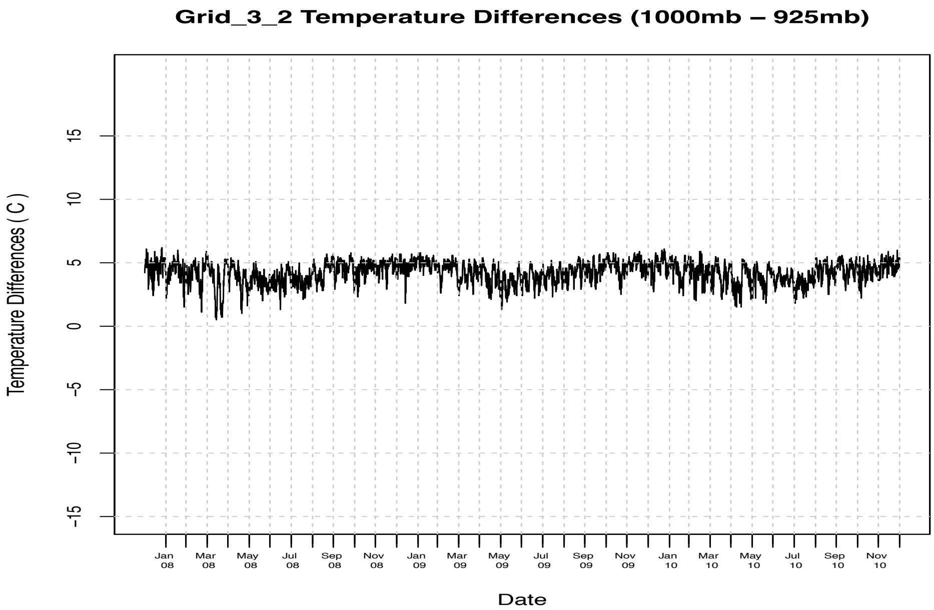


Figure D.2 : 1000 mb-925 mb temperature differences in Afyon, Isparta and Antalya grid

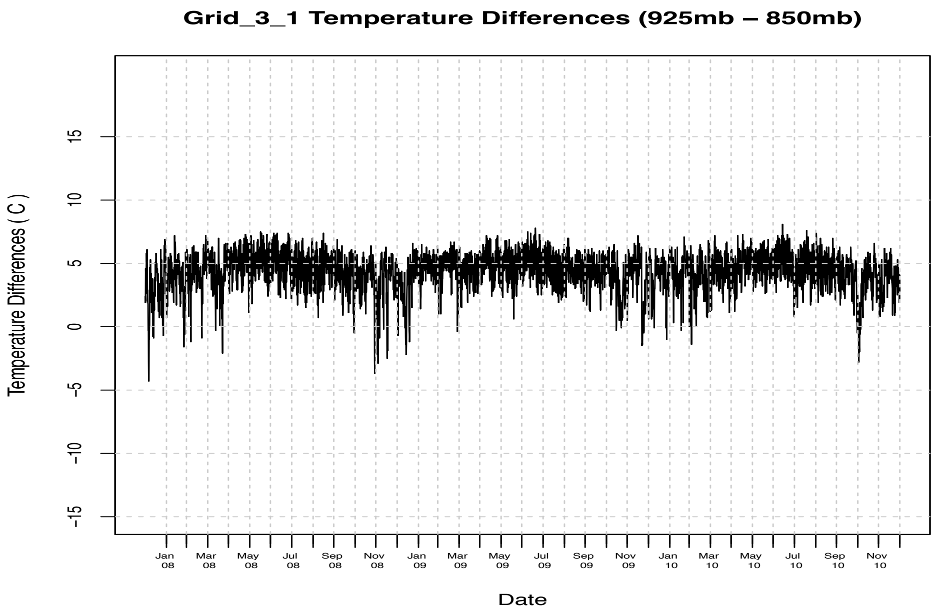


Figure D.3 : 925 mb-850 mb temperature differences in İzmir, Aydın and Muğla grid

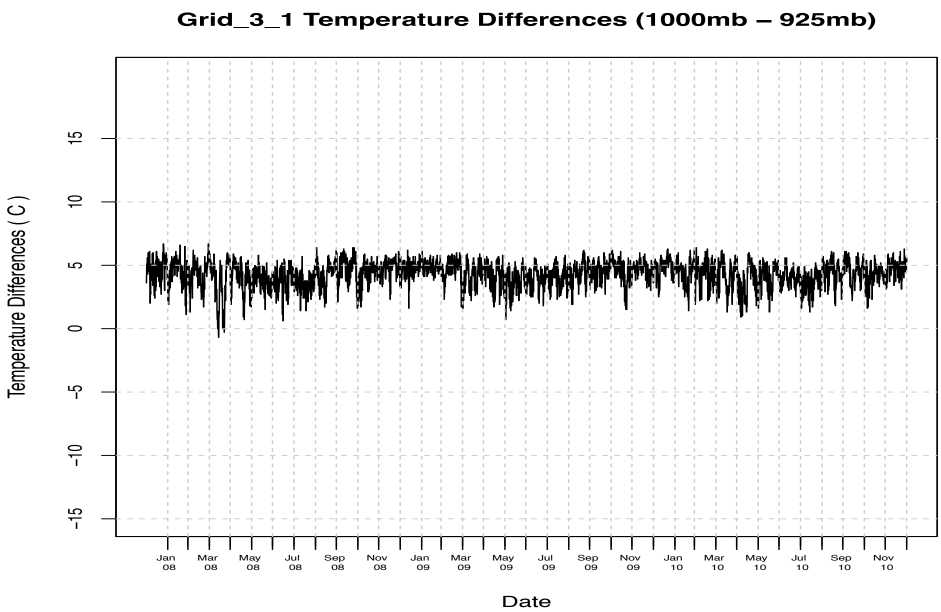


Figure D.4 : 1000 mb-925 mb temperature differences in İzmir, Aydın and Muğla grid

APPENDIX E : PM$_{10}$ concentrations vs PBL heights Plots.

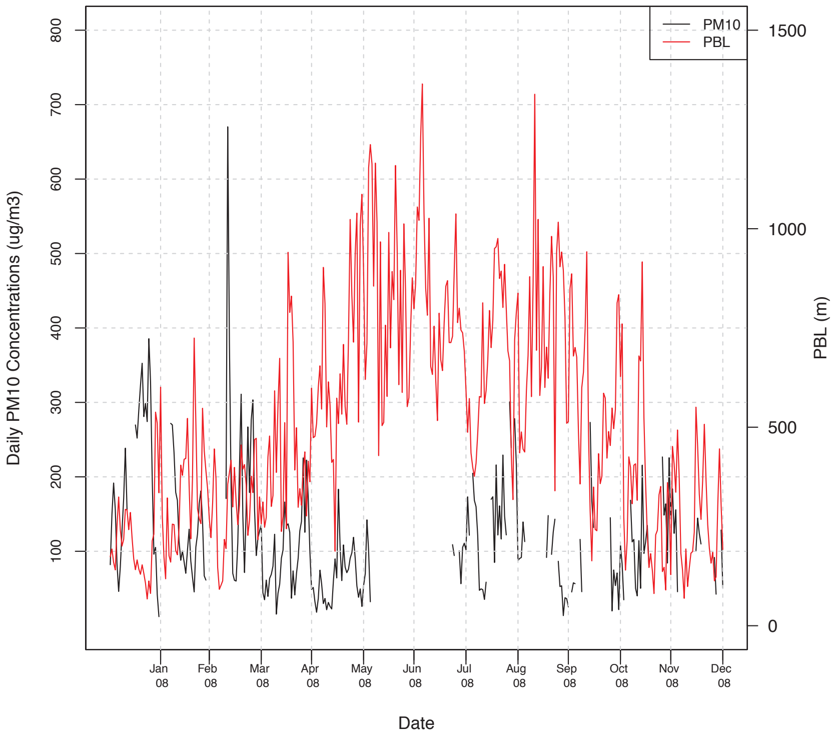


Figure E.1 : Muş PM$_{10}$ concentrations and PBL height in 2008.

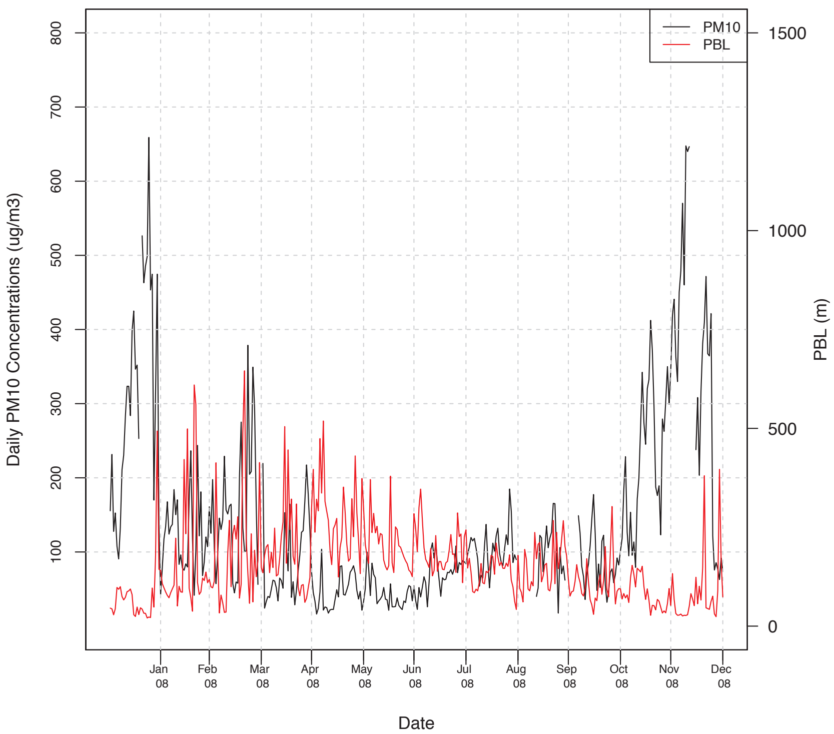


Figure E.2 : Iğdır PM$_{10}$ concentrations and PBL height in 2008.

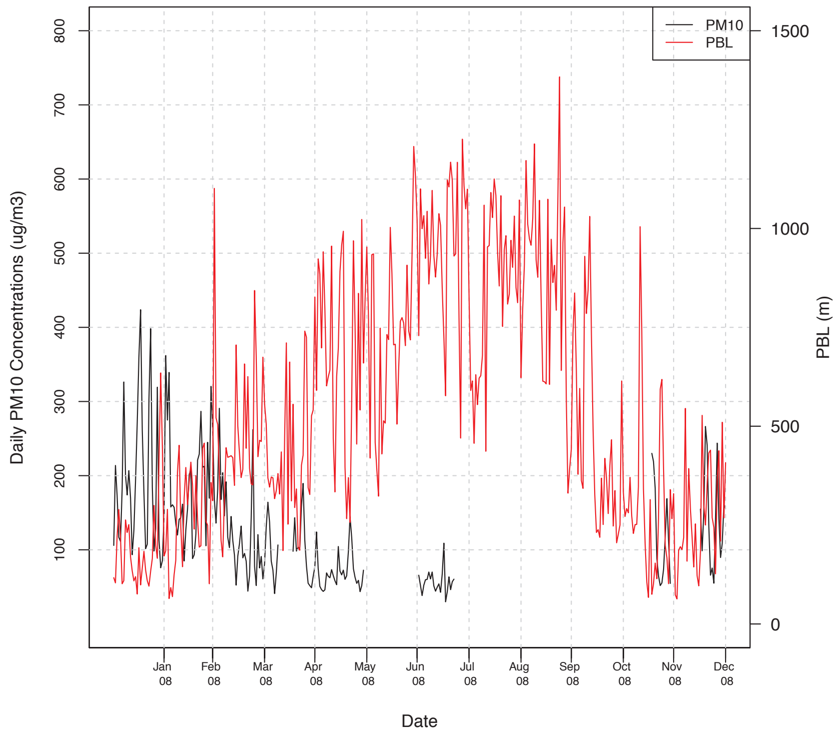


Figure E.3 : Afyon PM_{10} concentrations and PBL height in 2008.

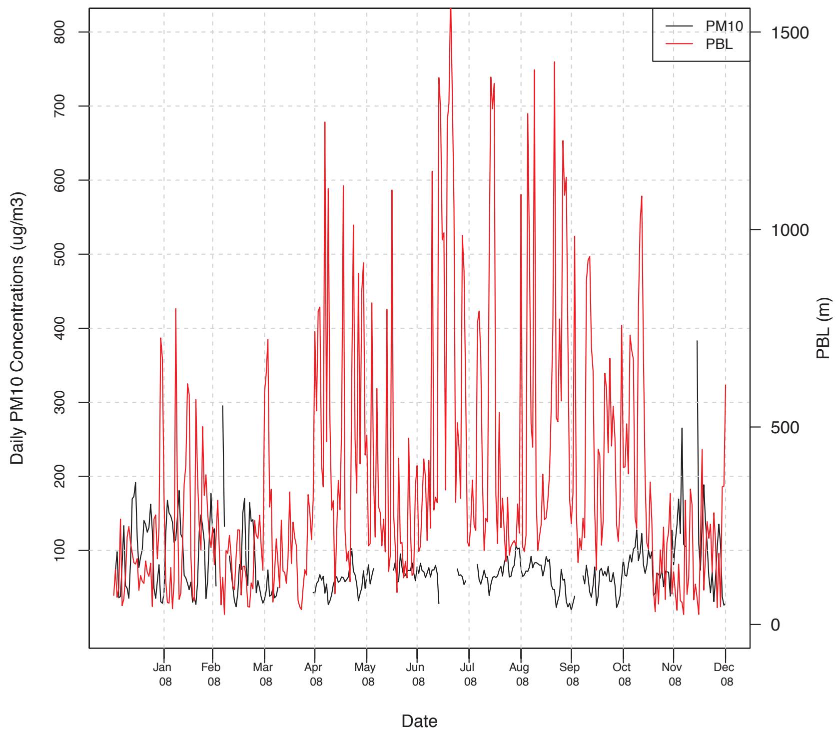


Figure E.4 : Antalya PM_{10} concentrations and PBL height in 2008.

APPENDIX F : Spatial Distribution Plots of PBL Heights

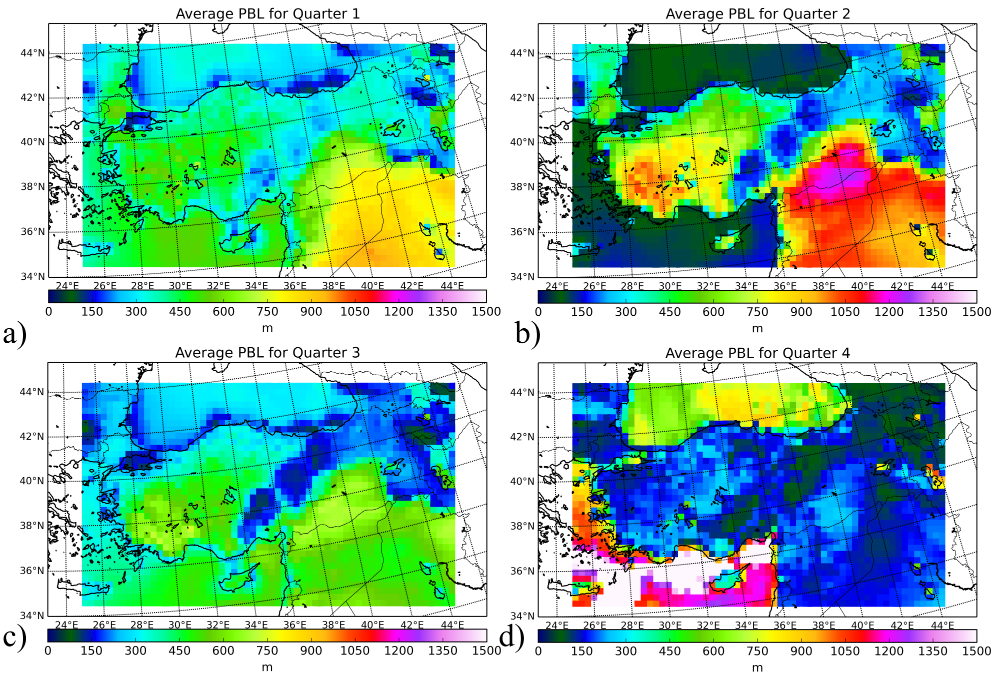


Figure F.1 : a)2008 January and February daily averages of PBL Heights b)2008 March, April, May daily averages of PBL Heights c) 2008 June, July, August daily averages of PBL Heights d) 2008 September, October, November daily averages of PBL Heights.

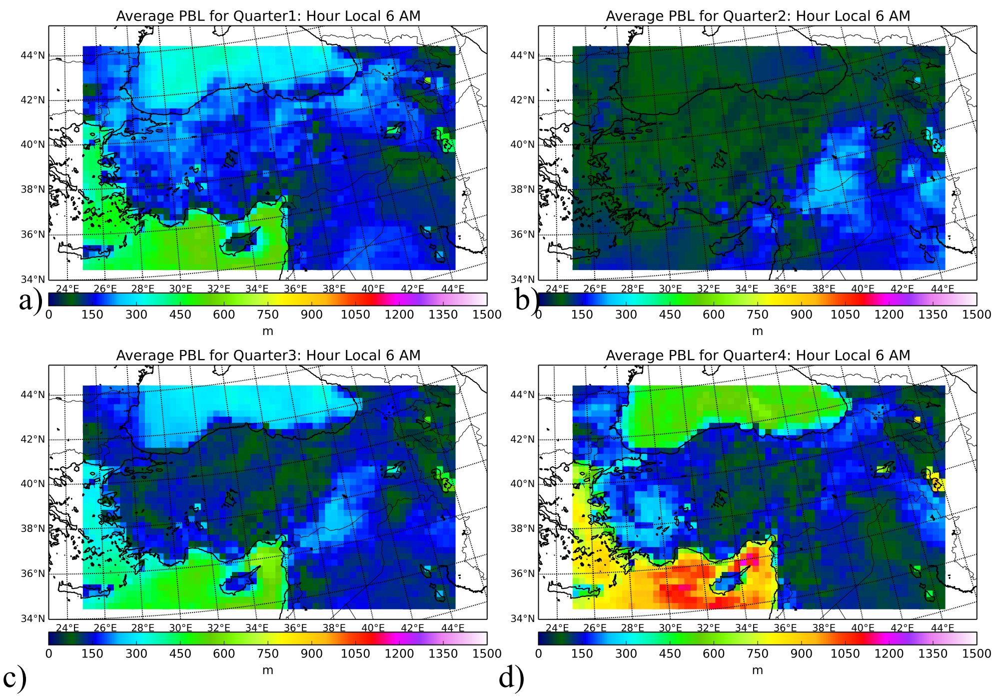


Figure F.2 : a)2008 January and February at 6 AM averages of PBL Heights b) 2008 March, April, May at 6 AM averages of PBL heights c) 2008 June, July, August at 6 AM averages of PBL heights d) 2008 September, October, November at 6 AM averages of PBL heights.

CURRICULUM VITAE

Name Surname: Seden Baltacıbaşı

Place and Date of Birth: İskenderun, 02-11-1984

Address: FeneryoluMah. FeneryoluSok. No:4 Daire:8 Kadıköy/İstanbul

E-Mail: sedenbaltacibasi@gmail.com

B.Sc.: Marmara University, Environmental Engineering Department

Professional Experience and Rewards:

Projects :

Impact of European Emissions on Air Quality on Marmara Region, Turkey: Development Modelling Framework (Meteorology, Emissions, and Air Quality Models) and Climatological Evaluation, Funded by: Scientific and Technologic Research Council of Turkey (TUBITAK), 2009

"Quantification of the effect of Sahara Dust on Air Pollution in Turkey", supported by The Scientific and Technological Research Council of Turkey, 110Y078, 2010

Computer Skills:

Mac OS X, Windows, Matlab, R Programming and ArcGIS.